2002

Site-Specific Incorporation of Biochemical and Biophysical Probes into Proteins Using Expressed Protein Ligation

Mande Holford

Site-Specific Incorporation of Biochemical and Biophysical Probes into Proteins Using Expressed Protein Ligation

by

Mande Holford

Submitted in Partial Fulfillment

of the Requirements for the

degree of Doctor of Philosophy

The Rockefeller University

August, 2002

THE ROCKEFELLER UNIVERSITY

1230 YORK AVENUE • NEW YORK, NEW YORK 10021-6399

Thomas P. Sakmar, M.D.
Acting President (RU)
Richard M. & Isabel P. Furlaud Professor
  & Head of Laboratory of Molecular Biology
  & Biochemistry (RU)
Associate Investigator (HHMI)
Senior Physician (RU Hospital)

Telephone: *(212) 327-8288*
           (212) 327-8676
Facsimile: *(212) 327-7904*
E mail: *sakmar@mail.rockefeller.edu*

Dr. Sidney Strickland
Dean and Vice President for Educational Programs
Office of Graduate Studies
The Rockefeller University

September 10, 2002

Dear Dr. Strickland:

As members of the committee appointed to conduct the final examination of Mande Holford, we have read her thesis and have found it to be acceptable (with/without) minor revision in partial fulfillment of the University's requirements for the granting of the doctoral degree. Therefore, we the undersigned, are satisfied that Ms. Holford has met the requirements for the Ph.D. degree with respect to her written thesis, thesis presentation and examination by this committee.

_____
Dr. Thomas P. Sakmar, Chairman
The Rockefeller University

_____
Dr. Tom Muir, Advisor
The Rockefeller University

_____
Dr. Seth Darst
The Rockefeller University

_____
Dr. Robert Landick
University of Wisconsin-Madison

"When I look back at my old notebooks, I am amazed at how inefficient the early developmental work was. I seemed always to choose the wrong reaction to do first and was not able to identify the most important parameters as the work was progressing. At the end of the first two years the results were so poor, I wonder what made me think that this approach would ever succeed. But from the outset I had a strong conviction that this was a good idea, and I am glad that I stayed with it long enough."

--Bruce Merrifield *(Merrifield, 1993)*

Acknowledgements

Primarily, I would like to thank my ambassadors of science Drs. Lawrence Johnson and Tom Muir. Dr. Johnson introduced me to the laboratory as an undergraduate and Dr. Muir has skillfully guided me throughout my graduate years. Tom has a contagious enthusiasm for science that is particularly visible when he's lecturing to you in his office. He has the ingenious ability of building bridges between the different branches of science. As a result, he has achieved a lot of acclaim in his short tenure as professor, and he has done this with out insisting his people work crazy hours. He has shown me that a successful career in science does not require dead-bolting yourself to the lab bench. The occasional trip to the beach or to a Yankee game is not only allowed, it's encouraged. I am grateful to have conducted my graduate study with such a truly talented individual.

I would also like to thank ALL past and current members of the Muir lab. When I joined the lab in 1997 there were only four other people in the lab and as a spoof on the British pop band 'Spice Girls' we called ourselves 'The Splice Girls and Boys' after the protein splicing technique we were all beginning to use. (I was 'scary splice', Tom, as senior member of the lab, was 'old splice'.) As I write my thesis the lab number hovers somewhere around 13-15 strong, and we're now referred to as "Tom's Gang." The transition from teen pop to street thug in the group description is in no way a criticism of the personalities in the laboratory. My lab colleagues have always made the lab someplace I referred to as my second home. I've made a lot of deep friendships that will continue after I leave the lab. It would be a disservice to name only a few lab members because all of them, Ginger Splice included, have contributed to the work presented in this document. So if you'll indulge me, I'd like to list all their names (this isn't so much for sentimental reasons, it's mostly because I have a my memory like an Alzheimer's patient): Brenda Ayers, Graham Cotton, Julio Camarero, Patricia Mayville, Ulrich Blaschke, Michelle Trester-Zedlistz, Izabela Giriat, John Silberstein, Gholson Lyon, Roseanne Hofmann, Robert Flavell, Morgan Huse, Francis Valiyaveetil, Henning Mootz,

Jennifer Ottesen, Alessandra Romanelli, Vasant Muralidharan, Mike Hahn, and Matt Sekedat.

. To my committee members Drs. Seth Darst and TomSakmar, I say thank you for always taking time out of your busy schedules to accommodate me. Scheduling a FAC meeting was never something I dreaded. Don't get me wrong, FAC meetings did make me sweat, but this wasn't because they were difficult to organize, it was more the thought that I would be presenting before a group of talented individuals with a proven record for practicing science at the highest levels. Thank You. I would also like to give special thanks to my outside member Robert Landick for venturing back to the urban jungle of New York City a little sooner than he planned.

Words can't express how much my family and friends have supported me throughout this effort. If I'm forced to try, I can only say a child could not ask for more in a parent than my mother and father, Janice and Winston Holford. To my friends, the bond cannot be broken, even with all the visa and green card issues that try to separate us.

Lastly, I would like to thank the ENTIRE Rockefeller community for giving me the experience of being in an oasis. Specifically, I'd like to thank Dr. Bruce Merrifield for being a shining example of scientific innovation, integrity, and longevity. As I wrote my thesis, I found it inspirational to reread Dr. Merrifield's autobiography <u>Life During a Golden Age of Peptide Chemistry</u>. Consequently, I have littered his words like fairy dust throughout my thesis with the hopes of creating a little magic.

Table of Contents

List of figures

List of Tables

Abbreviations

Abl-SH3, Abelson tyrosine kinase-Src homology 3

AsiA, bacteriophage T4 antisigma

Boc, tertiary butyloxycarbonyl

CBD, chitin binding domain

CD, circular diochroism

CNBr, cyanogen bromide

DAPA, L-diaminopropionic acid

DIEA, diisopropyl-ethylamine

DMF, dimethylformamide

DMSO, dimethylsulfoxide

EPL, expressed protein ligation

ESMS, electrospray mass spectrometry

Fl, fluorescein

Fmoc, 9-fluorenylmethoxycarbonyl

FRET, fluorescence resonance energy transfer

GdmCl, guanidinium chloride

GFP, green fluorescent protein

GSΔTβR-I, transforming growth factor β receptor I minus the GS region

GS-0Bio, unphosphorylated GS thioester peptide

GS-0BioTβR-I, unphosphorylated transforming growth factor β receptor I ligation
product

GS-4, tetra-phosphorylated GS thioester peptide

GS-4TβR-I, tetra-phosphorylated transforming growth factor β receptor I ligation product

HBTU, 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate

HOBt, 1-hydroxybenzotriazole

HTH, Helix-Turn-Helix

IAc-3BP2, N-iodoacetylated 3BP2 peptide

ICH₂CN, iodoacetonitrile

MALDI-TOF MS, Matrix-Assisted Laser Desorption-Ionization Time of Flight Mass Spectroscopy

MBHA, 4-methylbenzylhydrylamin

MESNA, mercaptoethanesulfonic acid

Mxe GyrA, *Mycobacterium xenopi* DNA gyrase A

NMR, nuclear magnetic resonance spectroscopy

Pam, phenylacetamidomethyl

PyBOP, benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate

Rh, tetramethylrhodamine

RNAP, RNA polymerase

RP-HPLC, reverse-phase high performance liquid chromatography

Sce VMA, *Saccharomyces cerevisiae* vascuolar ATPase

SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis

SH3, Src Homology 3

SPPS, solid phase peptide synthesis

TGFβ, transforming growth factor β

TβR-I, transforming growth factor β receptor I

THF, tetrahydofuran

TSQ-MS, Triple Stage Quadrupole Mass Spectroscopy

Z, benzyloxycarbonyl

Standard IUPAC single- and triple letter codes for amino acids used throughout

Abstract

Protein engineering can be made far more powerful if a protein is not only expressed recombinantly but also altered covalently using synthetic chemistry. These two methods are brought together in the protein semi-synthesis technique Expressed Protein Ligation (EPL). In EPL, recombinant and synthetic polypeptides are joined together via a chemoselective ligation reaction. EPL was originally used to attach synthetic constructs to the C-terminus of recombinant proteins, but is now used to attach recombinant or synthetic polypeptides either at the N- or C-terminus of a protein or into the core of a protein. This thesis illustrates, with three distinct applications, the development of EPL from its original definition to its current understanding. In the first application, a general strategy was developed for the site-specific incorporation of fluorophores into proteins using Abl-SH3 as a model system. In the second application, chemistries were developed that allowed the site-specific introduction of phospho-amino acids into proteins, in this case using the transforming growth factor β receptor I as the model system. In the final application, EPL was used to synthesize several modified versions of the *E. coli* sigma factor σ^{70}, demonstrating that this method can be used to probe extremely large macromolecules. These studies revealed that EPL works under a variety of reaction conditions and provided paradigms for using this technique to site-specifically insert fluorophores and phosphate groups into proteins. The chemical manipulation of proteins by EPL will be an important tool as researchers strive to characterize the proteomes of organisms.

Chapter 1: Introduction

*"The name protein which I propose to you for the organic oxide of fibrin and of albumin,
I wanted to derive it from πρωτειος [in the lead] because it seems to be the original
or principal substance of animal nutrition."*

--G. J. Mulder *(Soderbaum, 1916).*

Importance of proteins

Proteins have achieved what most rock bands struggle to obtain: to go from a small unknown to world-renowned recognition that spans the test of time. When G. J. Mulder first coined the phrase 'protein' in 1838, it was used to represent a poorly defined conglomeration of atoms. At the time, organic chemists busied themselves trying to analyze the components of proteins, from the point of view that they were derived from plants and passed onto human beings through consumption of animal and plant products *(Tanford, 2001)*. Since that time it has been well established that proteins in the body are not derived from nutrition, but rather are manufactured as gene products. The completion of the human genome, in addition to the genomes of several other organisms, has heralded the protein molecule to super star status.

With the genetic blueprint of several organisms now available, a new challenge has emerged. Namely, the characterization of many of tens of thousands of novel gene products at the chemical and biological levels. Often referred to as the 'proteome project,' this is an enormously daunting undertaking, particularly given the emerging picture of complexity in biological processes, where proteins are involved in the everyday "molecular business of biology" *(Jones, 1994)*. In order to fully comprehend a biological process, one must understand at a fundamental level the role proteins play. Consequently, this requires access to proteins, and the ability to manipulate them. In this instance, access refers to the capability of obtaining workable amounts of protein, and manipulation is the ability to design and construct proteins endowed with the probes appropriate for the process under investigation. As a direct response to this need, the field of protein engineering has developed several methods for manufacturing peptides and proteins through the use of protein engineering techniques that encompass the purely synthetic *(Jackson et al., 1994)* and the purely biosynthetic *(Chang et al., 1994) (Offer, 2000)*. However, before continuing the discussion on proteins, it is necessary to take a step back and discuss the proteins' junior counterpart peptides. For it is from the field of peptide synthesis that initial stabs at protein engineering were made.

Evolution of peptide synthesis

The field of peptide synthesis began with the contributions of Theodore Curtis and Emil Fischer. In 1881 Curtis was the first to synthesize a protected peptide *(Curtius, 1881)*, then, in 1901, Fischer was the first to synthesize an unprotected dipeptide *(Fischer, 1901)*. In the hundred years since these founding endeavors the field of peptide synthesis has achieved astounding progress (Figure 1.1). Each mark in the timeline shown in Figure 1.1 represents a hurdle peptide chemists had to overcome in order to attempt to reproduce nature's accuracy and efficiency in a synthetic form.
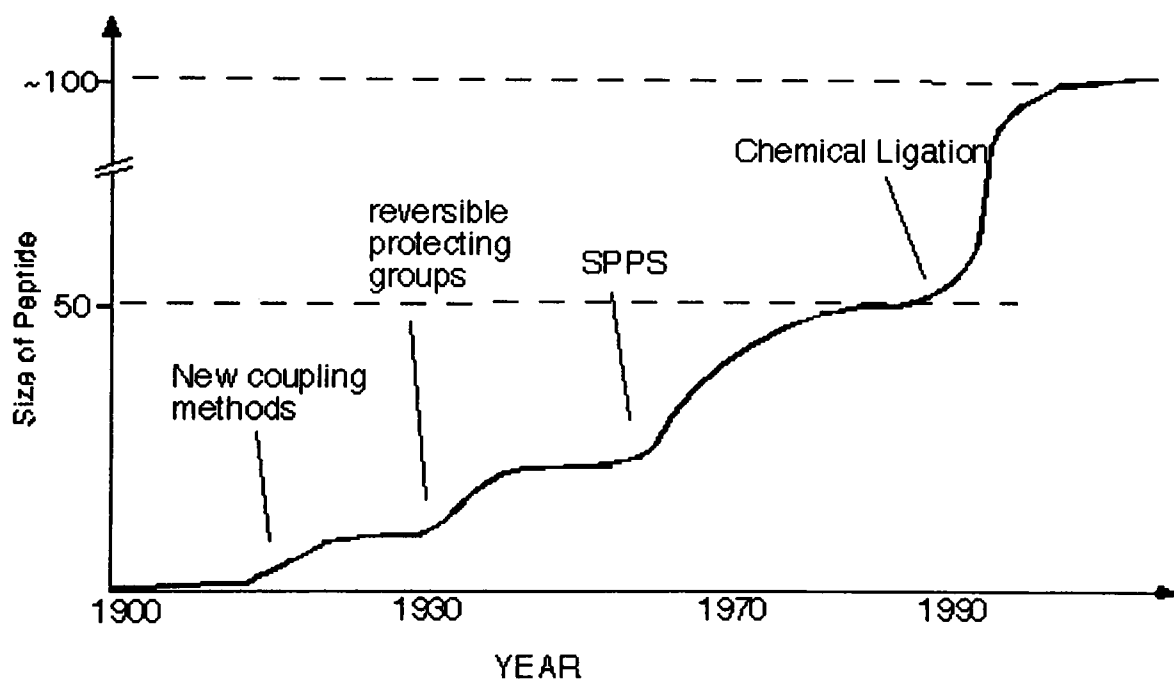
Figure 1-1. 100 years of chemical peptide synthesis. A plot depicting the achievements of the field of peptide synthesis over the past 100 years.

The first of these hurdles was to devise a method for reacting amino acids. Due to the zwitteronic nature of amino acids, two amino acids will not react on their own to form a peptide bond. To initiate the formation of a peptide bond an activating agent, called a "coupling agent," has to be added to the mixture. The initial coupling agents used were devised by the founding fathers and involved activating the carboxyl group of an amino acid to nucleophilic attack. Activation is achieved by replacing the hydroxyl of the

carboxylic acid with an electron withdrawing group, thus increasing the electrophilicity of the carboxyl group. Curtius utilized an acid azide coupling method, where he activated the carboxyl group of an amino acid by converting it to an acid azide *(Curtius, 1881)*. Fischer devised the acid chloride method, activating the carboxyl group by replacing the hydroxyl function with a chloride atom *(Fischer, 1901)*.

The knowledge of how to react two amino acids did not immediately result in the efficient synthesis of peptides. Both Curtius's and Fischer's methods had significant drawbacks which made the task of engineering a peptide extremely laborious. Specifically, Curtius's method is limited by the tendency of the acid azide to rearrange to the isocyanate, this is known as the Curtius rearrangement. With Fischer's method the acid chloride rearranges to a highly reactive amino acid derivative, N-carboxyanhydride, which is unsuitable for peptide synthesis due to instability and loss of chiral integrity. Such difficulties were due in part to the fact that there was not an efficient method for reversibly protecting the $^\alpha$amino group of an amino acid. Bergman and Zervas solved the problem in 1932 when they introduced the benzyloxycarbonyl (Z) group *(Bergmann, 1932)*. The Z group is a urethane protecting group that is readily removed by hydrogenolysis or acidolysis (Figure 1.2A). Bergman and Zervas's innovation was three fold. First, it presented an easy solution to the problem of reversibly protecting the $^\alpha$amino group, as hydrogenolysis and acidolysis were techniques that were well established at the time. Second, it served as the basis for which all modern day $^\alpha$amino protecting groups are designed, namely, urethane protecting groups. Thirdly, it prevented racemization. Up to this point racemization of amino acids was a persistent obstacle, thus it greatly advanced the field to discover that N^α protection of amino acids by urethane derivatives essentially eliminated the problem. The Z group was the first of a series of pivotal developments that provided the infrastructure for peptide synthesis as it is known today.

What followed were new, more efficient coupling agents to replace Curtius and Fischer's methods, namely amino acid anhydrides *(Wieland, 1951)* and active esters *(Bodanszky, 1955; Wieland, 1951)*. Also, Bergman and Zervas's urethane protecting group strategy was greatly improved with the introduction of the tertiary butyloxycarbonyl (Boc) and 9-fluorenylmethoxycarbonyl (Fmoc) groups (Carpino,

5

1957), which could be removed by milder acidic conditions than the Z group (Figure 1.2 B &C).

Removal of the Z group

Removal of the Boc group

Removal of the Fmoc group
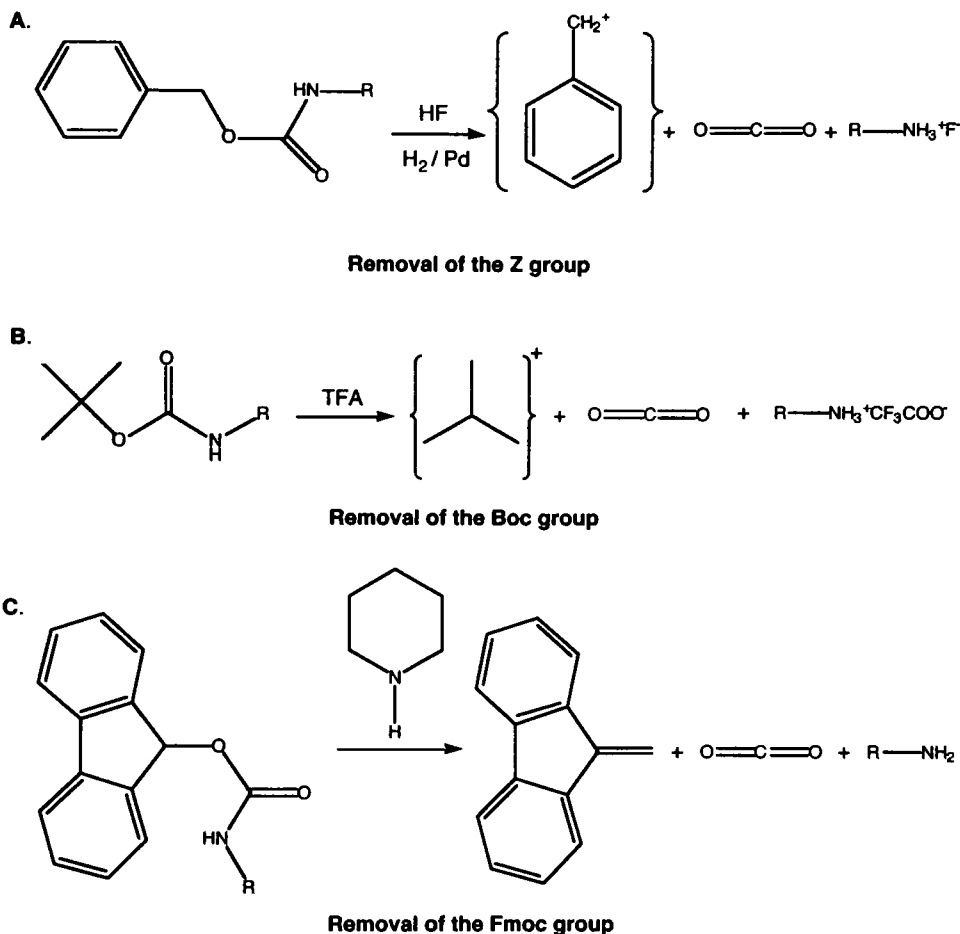
Figure 1-2. Removal of urethane N$^\alpha$ protecting groups. (A) Removal of Z-group, (B) Removal of Boc group, (C) Removal of the Fmoc group. Both the Boc and Fmoc groups are removed under milder conditions as compared to the hydrogenolysis or strong acid treatments of the Z group.

When Merrifield introduced the Solid Phase Peptide Synthesis (SPPS) method in 1963, he harvested the previous advances and introduced an innovation that changed the field forever *(Merrifield, 1963)*. Up to that point peptide synthesis was done in the solution phase, with the most memorable contribution made by Du Vigneaud in 1953 when he synthesized the octapeptide Oxytocin *(du Vigneaud, 1953)*. One major drawback to solution phase synthesis is that it is very labor intensive. This is due to the

fact that each step in the synthesis has to be purified to get rid of excess activated amino acid required to optimize each coupling. Merrifield simplified the process by using an insoluble polymer support to anchor the peptide as it was being synthesized (Figure 1.3).
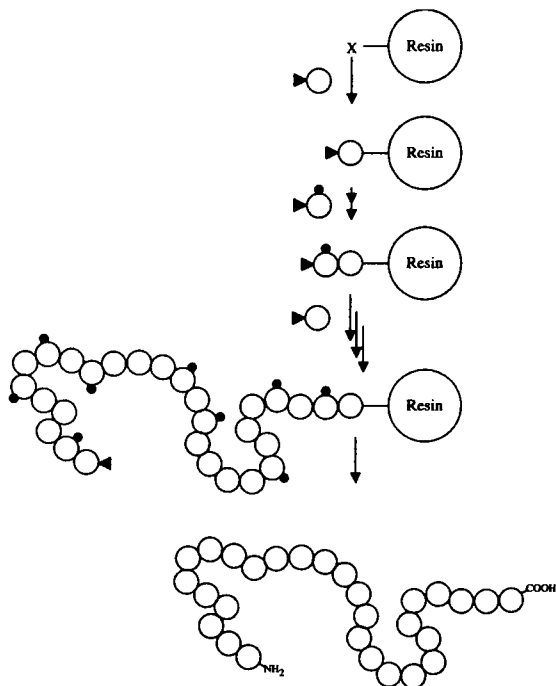
Figure 1-3. Schematic of Merrifield's Solid Phase Synthesis Strategy (SPPS). Dark circles and triangles depict side chain and N^{α}- protecting groups, respectively. At each step a protected amino acid is attached to the growing peptide chain that is connected to a resin support. The final step is removal of the peptide chain from the resin support to yield the free peptide.

The anchoring of the amino acids residues as they are sequentially elongated made it extremely easy to wash away excess material, purifying the reaction in a manner of minutes versus hours with the solution phase approach. In his book, Life during the Golden Age of Peptide Synthesis, Merrifield describes his approach and the advantages it would have over solution phase methods *(Merrifield, 1993)*:

> One day in 1959 I had an idea about how we might
>
> synthesize peptides in a simpler, faster, and more efficient
>
> way then could be done by the classical methods. The idea
>
> was to make use of an insoluble solid support to hold the

7

peptide chain while it was being assembled. The plan was to begin with some stable insoluble polymeric material that could be derivatized with a functional group through which the terminal amino acid residue of the peptide chain could be attached. The next residue would be suitably protected and then activated and coupled to the first one to form the peptide bond. A repetition of the process would yield the tripeptide, and so on until the desired peptide sequence was assembled. Removal of protecting groups and cleavage from the insoluble support would then yield the final peptide. The expected advantage of this approach was the ability to filter the mixture after each reaction and remove the excess soluble reagents and by products. The solid support and the growing peptide that was covalently attached to it would remain on the filter. In principle, all the reactions could be performed in a single reaction vessel, and higher yields and fewer losses of material could be expected... Because the reagents were in solution but the peptide chain was in the solid phase, I gave the technique the name 'solid –phase peptide synthesis.'

Merrifield's accomplishments, coupled with those of Sheehan (carbodiimide activation) *(Sheehan, 1955)*, Carprino (Boc, Fmoc protecting groups) *(Carpino, 1957)*, Konig and Geiger (HOBt activation) *(Konig, 1970)*, Sheppard (Fmoc SPPS) *(Atherton, 1978)*, and other's paved the way for the efficient synthesis of longer and longer peptide chains. Several proteins were prepared by SPPS *(Akaji, 1985; Sieber, 1978; Zawadzke, 1993)*, the most memorable being the HIV-1 protease that led to the crystal structure of the protein *(Miller et al., 1989; Schneider and Kent, 1988)*. But this success was limited, usually requiring large teams and taking years to complete. By the 1990's modern improvements in the SPPS method made the synthesis of peptide fragments up to 50 amino acids in length a routine and reliable process *(Kiso, 1990; Schnolzer, 1992; Tam,*

1983), but there are very few proteins that are 50 amino acids long. The logical next step was to find a way to link the peptide chains together, in essence, to synthesize big polypeptides or proteins.

Evolution of chemical protein synthesis

When Emil Fischer synthesized the first free dipeptide in 1902, it was with the intention of one day being able to synthesize an enzyme. He wrote:

> My entire yearning is directed toward the first synthetic
> enzyme. If its preparation falls within my lap with the
> synthesis of a natural protein material, I will consider my
> mission fulfilled *(Fruton, 1985)*.

For eight years, 1902-1910, Fischer tried to synthesis an enzyme. In 1910, frustrated that his mission would not be fulfilled, Fischer terminated all peptide synthesis work in his laboratory. He cited the limitations of his synthetic methods as the reason. Many years later peptide chemists would take up the protein synthesis torch again. This time equipped with far superior synthesis methods than those available to Fischer. However, due to the sheer scale of the problem (proteins being significantly larger than peptides) modern day peptide chemists would find the challenge just as daunting as Fischer did.

Merrifield and others had succeeded in making peptide synthesis extremely successful and efficient, to the point where eventually it was automated. The next obvious step would be to apply the techniques learned in peptide synthesis to chemical protein synthesis. Unfortunately, the two fields are not so readily integrated. Not only are proteins much larger than peptides, they also have higher dimensions of organization that have to be accounted for. As polypeptide chains increase in size they start to acquire secondary and eventually tertiary structure. The formation of these structures make SPPS synthesize of large proteins almost impossible, as they decrease the efficiency of the coupling reactions that lead to chain elongation. A logical solution to this problem was to make smaller peptide fragments and link them together in a convergent approach.

The initial fragment condensation approach to synthesizing proteins called for small protected peptide segments of the protein of interest to be first constructed using SPPS, then chemically linked together in solution to produce the target sequence. Fragment condensation of peptide chains exploits the advantages of SPPS, as peptide fragments up to 50 amino acids were becoming readily available. However the disadvantage of this method is similar to that of classic solution phase peptide synthesis, namely, in order to prevent branching of the peptide or undesired chain assembly, it is necessary to protect the reactive side-chain functionalities of each fragment. Protected peptide fragments are difficult to purify, characterize, and tend to have solubility problems. Despite these problems the total chemical synthesis of proteins via fragment condensation has enjoyed significant successes. The syntheses of insulin *(Sieber, 1978)*, Rnase A *(Akaji, 1985)*, human PTH *(Tam, 1986)*, and human EGF *(Heath, 1988)* serve to illustrate this point (reviewed in *(Lloyd-Williams et al., 1993)*).

Several influential contributions have been made in an effort to solve the problems associated with the classic fragment condensation approach. These include the thiol-capture strategy of Kemp *(Kemp and Carey, 1993)*, which relies on a proximity-based entropic activation step, and the minimal protected fragment condensation approaches of Blake and Yamashiro *(Blake and C.H., 1981; Yamashiro and Li, 1988)*, which utilizes peptide °thioacid derivatives. However, it was not until the emergence of the chemical ligation approach that the problems of classic fragment condensation were significantly addressed. The conceptual breakthrough at the heart of the chemical ligation approach is the realization that a synthetic protein need not only contain amide linkages within its polypeptide backbone. Accordingly, chemical ligation strategies utilize alternative organic reactions all of which, initially, generated non-amide linkages between the peptides being joined together. Thioester *(Schnolzer, 1992; Williams, 1994)* and hydrazone *(Gaertner, 1992)* forming reactions were among the first to be described, but have since been joined by oxime *(Rose, 1994)*, thioether *(Muir, 1994; Nefzi et al., 1995)* and thiozolidine (Liu and Tam, 1994) (Figure 1.4). All chemical ligation strategies rely on highly selective reactions between synthetic peptide fragments, each bearing unique and mutually reactive groups. The chemoselective nature of the chemical ligation method means that fully unprotected peptide segments can be used.
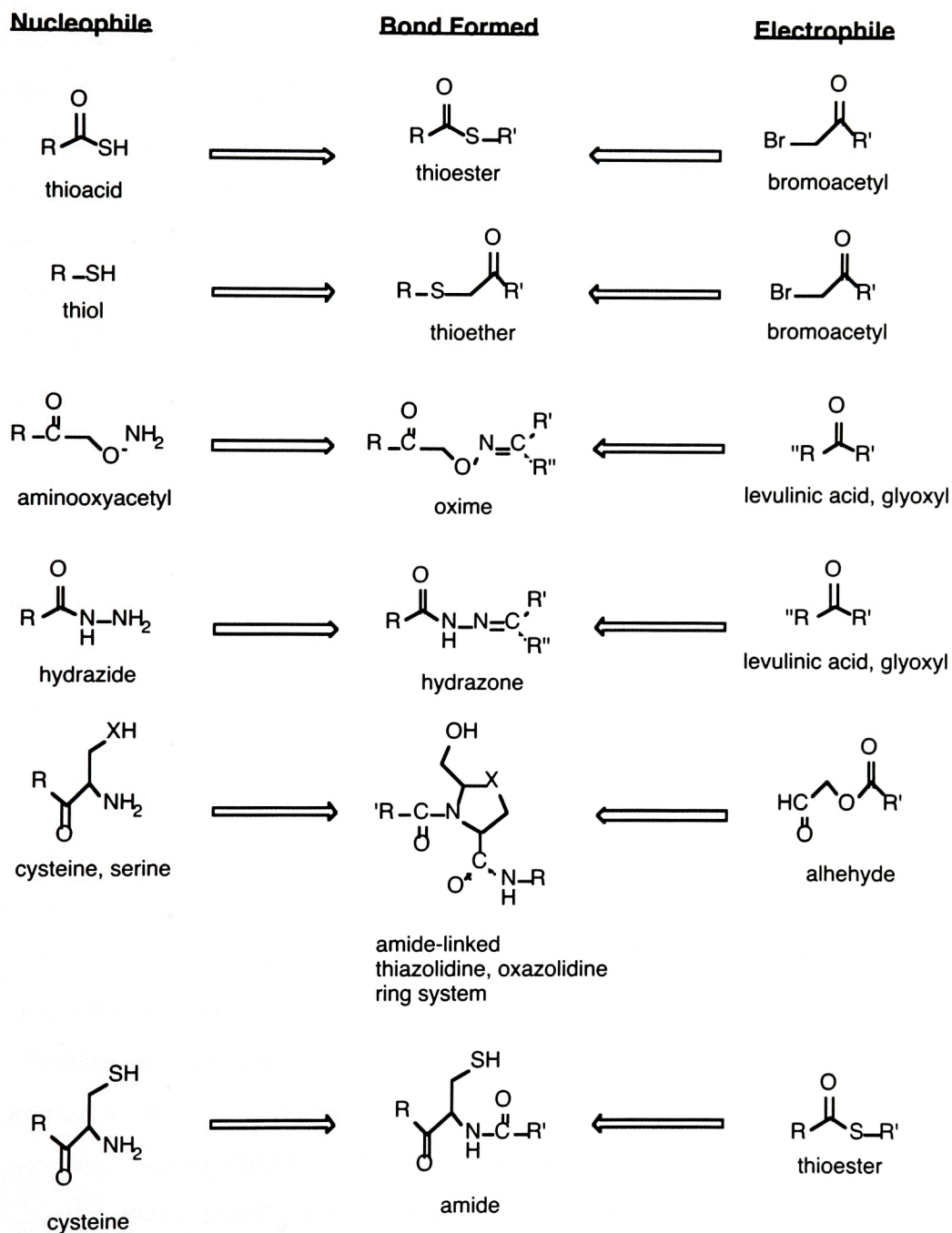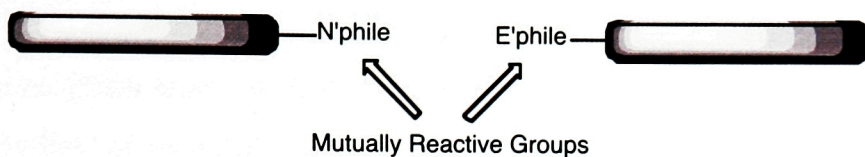
Figure 1-4. Chemical ligation strategies. Different methods for chemoselectively combining peptide fragments.

11

These chemical ligation strategies have allowed synthetic and, in some instances, recombinant polypeptides to be joined together through non-native linkages (for reviews see *(Dawson, 2000; Muir, 1995; Wallace, 1995)*).

The field of chemical ligation was revolutionized by the efforts of Steve Kent and colleagues with the advent of Native Chemical Ligation *(Dawson et al., 1994)*. In native chemical ligation, as with other chemical ligation strategies, two fully unprotected synthetic peptide fragments are chemically ligated under neutral aqueous conditions, but unlike other ligation strategies, native chemical ligation results in the formation of a native peptide bond at the ligation site. As illustrated in Figure 1.5A, native chemical ligation involves the chemoselective reaction between a peptide fragment containing a ᵅthioester and another peptide fragment containing an N-terminal cysteine residue. The initial transthioesterification reaction is followed by a rapid intramolecular S->N acyl shift to generate an amide bond at the ligation junction. Accordingly, the target full-length polypeptide is obtained in the desired final form without further manipulation. Native chemical ligation can be performed in the presence of all the functionalities commonly found in proteins, including free cysteine sulfhydryls. This is because cysteine residues that are not at the N-terminus cannot spontaneously rearrange to give an amide bond, and due to the reversible nature of the initial transthioesterification step, are simply converted back to starting materials. This advantage has made native chemical ligation the most successful and commonly used chemical ligation approach to be utilized in the synthesis of proteins with an average size of 100 amino acid residues *(Baca et al., 1995; Dawson et al., 1997; Muir et al., 1997; Tam, 1995)* (for review *(Dawson, 2000)*). Native chemical ligation increases by two fold the size of peptides that can be made by SPPS. However, in order to break the size barrier and synthesize larger proteins, it is necessary to broaden the scope of chemical protein synthesis beyond peptide chemistry.

Protein semi-synthesis is an elegant chemistry driven approach to engineering proteins that involves a multidisciplinary cast of characters from biologists to organic chemists. With semi-synthetic approaches scientists meet nature halfway and effectively eliminate the "size issue" in chemical protein synthesis. If scientists surrender completely to nature, heterologous expression of recombinantly cloned genes is by far the most efficient route to engineered proteins. Site-directed mutagenesis is facilitated by the

"replicative fidelity and efficiency of living organisms *(Woods et al., 1996)*," a property which also constrains the approach to the naturally occurring amino acids. Protein semi-synthesis surmounts the limitation of recombinant expression by allowing for the incorporation of unnatural amino acids into proteins.
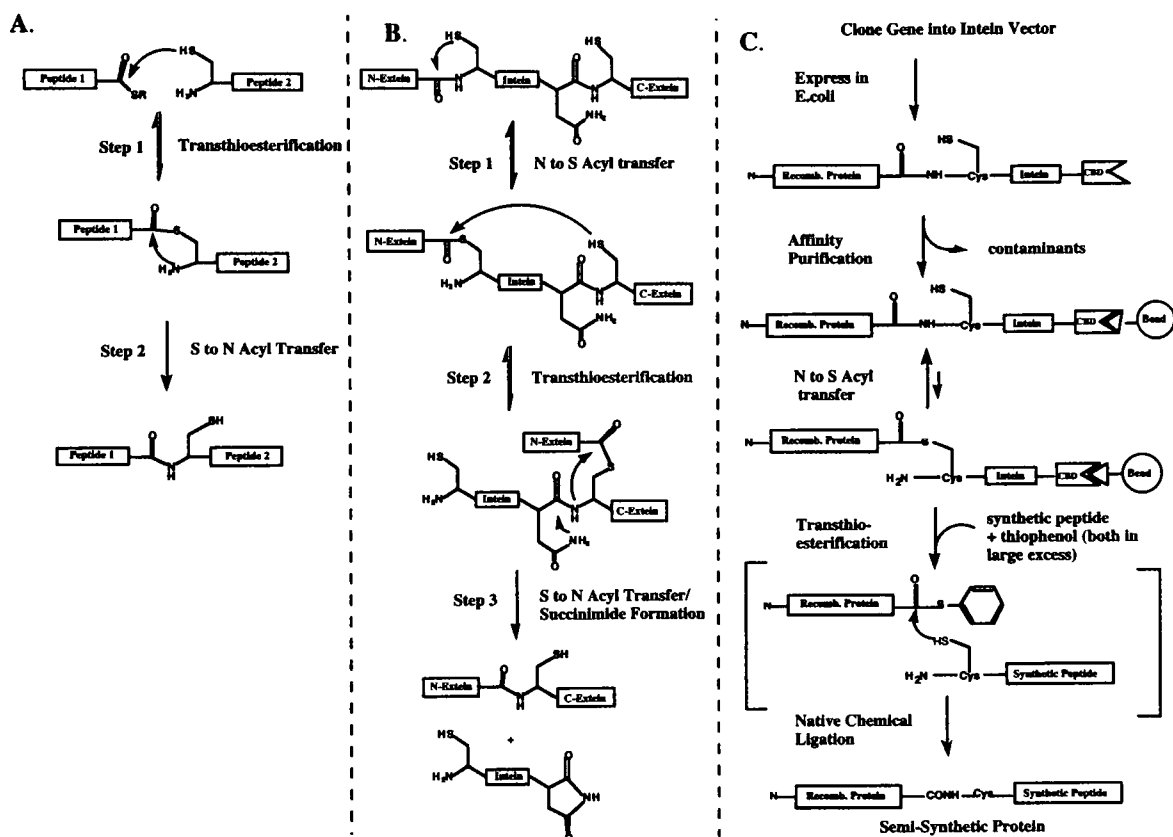
Figure 1-5. Mechanism of (A) Native Chemical Ligation, (B) Protein Splicing, (C) Expressed Protein Ligation.

The term "protein semi-synthesis" originally referred to processes in which proteolytic or chemical cleavage fragments of natural proteins were used as the building blocks for the resynthesis of the protein (reviewed in *(Wallace, 1995)*). For example, it was noted nearly 30 years ago that CNBr fragments of pancreatic trypsin inhibitor spontaneously condense to reform the native peptide bond between them, an effect that is also observed with cytochrome c *(Wallace and Clark-Lewis, 1992)*. In both cases, the two fragments produced after CNBr cleavage cooperatively refold, bringing the homoserine lactone at the carboxy-terminus of one fragment in close proximity to the

13

free amino terminus of the other. This autocatalytic fragment religation approach, termed confirmationally assisted protein synthesis, has been used to incorporate natural and unnatural amino acids into cytochrome c *(Wallace and Clark-Lewis, 1992)*.

Another general approach to protein semi-synthesis involves the use of proteolytic enzymes to facilitate the regioselective ligation of peptide fragments. Classically, this "enzyme-assisted reverse proteolysis" strategy involved altering the reaction conditions such that aminolysis of an acyl-enzyme intermediate is favored over hydrolysis. This is typically achieved by including high concentrations of organic solvents such as glycerol, DMF, or acetonitrile in the reaction medium *(Homandberg, 1978)*. Significant progress in the area of enzyme-assisted reverse proteolysis has been achieved by genetically engineering the active site of the proteolytic enzyme. This approach was pioneered in the mid 1980's by Kaiser and co-workers who chemically incorporated a thiol group into the active site of the serine protease subtilisin. The resulting thiol subtilisin analog possessed improved acylation activity relative to the wild-type enzyme *(Nakatsuka et al., 1987)*. In an extremely elegant extension of this work, Wells and co-workers have further engineered subtilisin affording an enzyme capable of efficiently catalyzing the ligation of peptide fragments *(Jackson et al., 1994)*. Their double mutant form of the enzyme, termed subtiligase, functions as an effective acyl transferase and importantly has vastly reduced proteolytic activity compared to the wild-type enzyme. Mutation of the active site serine to a cysteine allows the enzyme to be acylated with peptides esterified at their C-terminus with a glycolate phenylalanyl amide ester group. A second active site mutation (proline to alanine) relieves steric crowding, allowing efficient aminolysis of the enzyme-peptide thioester intermediate by the amino-terminus of a second peptide fragment. Subtiligase has been used to prepare totally synthetic proteins *(Jackson et al., 1994)* and semi-synthetic proteins *(Chang et al., 1994)*.

Semi-synthesis techniques like conformationally assisted protein synthesis *(Wallace and Clark-Lewis, 1992)*, enzyme-assisted reverse proteolysis *(Jackson et al., 1994; Nakatsuka et al., 1987)*, and unnatural amino acid mutagenesis *(Cornish et al., 1995)* have been successfully applied to the construction of proteins with unnatural probes. The methods listed above have provided researchers with a chemical toolbox for manipulating proteins in order to investigate biological processes. However, there is a

dilemma in that not too many researchers are equipped to access the toolbox. For example, in Schultz's unnatural amino acid mutagenesis, the ribosomal machinery is essentially tricked into incorporating the desired unnatural amino acid into the protein by mutating the codon of interest with an amber codon. The mutant gene is then added to an *in vitro* translation mix containing semi-synthetic suppressor tRNA charged with the unnatural residue. Although highly innovative (having been used to introduce a range of unnatural amino acids into proteins such as lysozyme and ras p21), unnatural amino acid mutagenesis is labor intensive, technically demanding, and often results in a low yield of the desired protein. If we were to construct a plot of the evolution of protein semi-synthesis similar to that constructed for peptide synthesis in Figure 1.1, where instead of size on the Y-axis we plotted utility or accessibility of approach, one would find that not many researchers are using the protein semi-synthesis techniques that are currently available. Expressed protein ligation (EPL), a novel protein semi-synthesis tool developed in the Muir lab in 1997, combines recombinant biology and synthetic chemistry in a way that promises to make proteins containing unnatural building blocks accessible to most researchers *(Evans et al., 1998; Muir et al., 1998; Severinov and Muir, 1998)*(review *(Holford, 1998)*).

Development of Expressed Protein Ligation

Expressed protein ligation, is a protein semi-synthesis method that developed from the unexpected convergence of two, at first glance, unrelated areas of research; protein total synthesis *(Dawson, 2000; Kent, 1988)* and protein self-splicing *(Perler et al., 1997)*. EPL was developed to address some of the limitations of earlier chemical protein synthesis approaches, such as the size of proteins synthesized, and accessibility of the techniques. At the core of the EPL technique is the chemical synthesis method native chemical ligation. As described above, native chemical ligation is a technique where the chemoselective reaction of a C-terminal thioester and an N-terminal cysteine yields a native peptide bond at the site of ligation (Figure 1.5A.). The two reactive moieties

15

necessary for native chemical ligation to occur, an N-terminal cysteine and a C-terminal αthioester, are also necessary for EPL to occur.

As originally noted by Kent and co-workers *(Dawson et al., 1994)* and later demonstrated with chimeric heterodimers of the transcription factors c-Jun and c-Fos *(Erlandson et al., 1996)*, the N-terminal Cysteine moiety may be generated via recombinant DNA methods. Most approaches currently in use for preparing N-terminal cystein proteins involve the in vitro use of exogenous proteases. In the approach developed by Verdine and co-workers, a factor Xa recognition sequence is appended immediately in front of the cryptic N-terminal cysteine in the protein of interest *(Erlandson et al., 1996)*. Treatment of this recombinant fusion protein with the protease gives the requisite N-terminal cysteine protein directly, which can then be used in subsequent ligation reactions *(Erlandson et al., 1996; Huse, 2000; Xu et al., 1999)* (demonstrated in section 3.2). Tolbert and Wong recently showed that the cysteine protease from tobacco etch virus (TEV) can also be used to release N-terminal cysteine proteins from suitable precursors *(Tolbert, 2002)*. The ability to produce N-terminal cysteine moieties via recombinant technology has a two fold effect: first, it eliminates the size constraints of C-terminal fragments used in native chemical ligation, and subsequently EPL thus, expanding the chemical protein engineering tool box. Second, recombinant DNA technology is a technique that is readily attainable by most researchers, and therefore it broadens the appeal of the approach.

Though N-terminal cysteines could be obtained recombinantly, αthioesters were not as readily accessible. The solution to this problem was obtained with the discovery of the protein splicing phenomenon.

Protein splicing is a post-translational process by which an intervening sequence, termed an intein, is removed from a host protein, termed an extein *(Perler et al., 1994)*. The first intein was discovered in 1990 as a 'spacer' within the *Saccharomyces cerevisiae* vascuolar ATPase (*Sce* VMA) gene that was absent in other VMA homologues *(Hirata, 1990; Kane et al., 1990)*. Since then, over 100 putative inteins have been identified and catalogued in InBase, the online intein registry and database (http://www.neb.com/neb/inteins.html) *(Perler, 1999)*. Intein-mediated protein splicing is the protein equivalent of RNA splicing. However, unlike RNA splicing, which

16

requires the help of an elaborate ribonuclear protein complex, protein splicing is an intrinsic process, and can take place *in vitro* using purified pro-proteins *(Xu and Perler, 1996)*. This has allowed the mechanism of intein-mediated splicing to be studied in some detail using both mutagenesis and protein chemistry based techniques *(Chong et al., 1996; Shao et al., 1996; Xu and Perler, 1996)* (for review see *(Perler, 2000)*).

The conventional mechanism of protein splicing (alternative mechanisms have been identified *(Southworth, 2000)* is summarized in Figure 1-5B. The initial step in the mechanism is an N->S or N->O acyl shift, where the amino terminal peptide bond of Ser1 or Cys1 is replaced with a (thio)ester bond. This initial step in the protein splicing mechanism is essentially the reverse of the final step in native chemical ligation, (compare Figure 1.5 A and B). The N->S or N->O acyl shift appears to be energetically unfavorable and the driving force(s) behind this event is still the subject of study. Although, crystal structures of several inteins have been solved *(Duan et al., 1997; Hall, 1997; Hashimoto, 2000; Ichiyanagi, 2000; Klabunde et al., 1998; Poland, 2000)*, there is still a question as to why the first step in the mechanism occurs. In the crystal structure determined by Sacchettini and co-workers of the GyrA intein, the amide bond between the final residue in the N-extein and the first residue in the intein is observed to be in the higher energy *cis*-conformation. It is possible that this unusual stereochemical constraint helps push the equilibrium of the system towards the (thio)ester side.

After the initial N->(S,O) rearrangement, a trans-esterification reaction occurs, moving the N-extein from the side chain of the first residue in the intein to the side chain of the first residue in the C-extein, which is always a Ser, Thr or Cys. This step results in ligation of the exteins and formation of a branched protein intermediate. In the next step, the amide bond at the intein-C-extein junction is broken as a result of succinimide formation involving the conserved C-terminal asparagine residue within the intein. In the final step of the process, a peptide bond is formed between the N-extein and C-extein following an (S,O)->N acyl shift; identical to the final step in native chemical ligation.

The uncanny chemical similarities between native chemical ligation (Figure 1-5A) and protein splicing (Figure 1-5B) raised the question: is it possible to use protein splicing as a route to recombinant protein αthioesters for use in native chemical ligation? The answer to this question was demonstrated by Xu and co-workers with the discovery

that protein splicing can be halted at the first step of splicing, presumably at the (thio) ester intermediate stage by mutation of the key asparagine residue in the intein to an alanine *(Chong et al., 1997)*. Proteins expressed as in-frame amino-terminal fusions to a mutated intein can be cleaved by thiols via an intermolecular trans-thioesterification reaction to generate a recombinant protein αthioester derivative. The recombinant αthioester can then be reacted with a polypeptide containing an amino-terminal cysteine via native chemical ligation. *E .coli* expression vectors are now commercially available which allow the generation of protein fusions to two different engineered inteins, the *Saccharomyces cerevisiae* vacuolar ATPase (*Sce* VMA) intein and the *Mycobacterium xenopi* DNA gyrase A (*Mxe* GyrA) intein.

The stage was thus set for the genesis of a new protein semi-synthesis approach that would pull together protein splicing and native chemical ligation. This new method, termed Expressed Protein Ligation, initially, allowed synthetic peptides to be chemically ligated to the C-terminus of recombinant proteins through a native peptide bond *(Evans et al., 1998; Muir et al., 1998; Severinov and Muir, 1998; Holford, 1998;)*. The mechanism of EPL is shown in Figure 1-5C. In the first step, the protein fragment of interest is expressed as an intein-affinity tag fusion, where the affinity tag allows the protein of interest to be purified. By using an engineered intein, protein splicing is unable to proceed to completion and is stalled at the stage involving the thioester-linked intermediate (Figure 1-5B, step 1). Consequently, exposure of the immobilized fusion protein to an aqueous solution containing an N-terminal cysteine peptide and a catalytic thiol reagent result in native chemical ligation of the peptide to the protein. The final product is an engineered protein that includes a synthetic fragment via which unnatural probes may be introduced.

The three initial works of EPL each used the mechanism shown in Figure 1-5C. Namely, EPL was used to attach synthetic peptides to the C-terminus of recombinant proteins. However, since its inception in 1997 (the year I joined the Muir lab), the goal of EPL has been to expand to include the addition of not only synthetic fragments, but also recombinant fragments, to either the N or C-termini, or to insert fragments into the core of a protein, see Figure1-6. Such attributes would make EPL a useful tool for the researchers involved in the proteome project. As mention before, the proteome project is

18

daunting and will require a collective effort from diverse branches of science. Everyone from applied physicists to zebra fish biologists will have an input deciphering the vast amount of proteins. This varied congregation will need tools that can extend across traditional barrier lines. Expressed protein ligation has the ability to be that kind of tool. At its origin EPL encompasses scientific collaboration (synthetic chemistry and protein splicing). In this thesis alone, EPL has facilitated the joining of peptide/protein chemists with structural and molecular biologists. The objectives of this thesis helped to characterize many of the early underlying breakthroughs that have significantly aided in advancing EPL to a technique that is each day gaining ground with different investigators and being used in innovating ways to address questions from protein folding to cell signaling (for reviews *(Hofmann, 2002; Muir, 2001)*).

A. N-terminal Ligation.
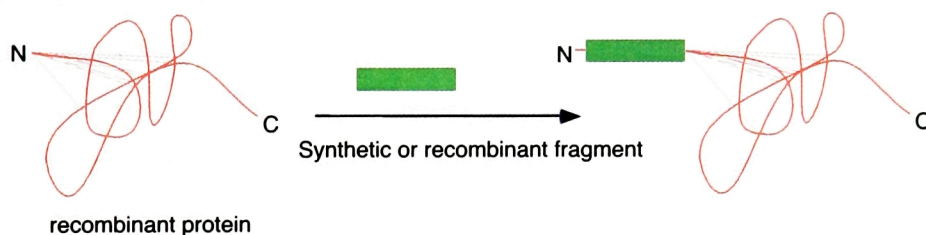
B. C-terminal Ligation.
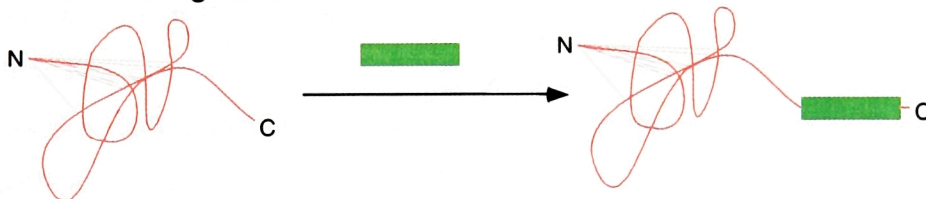
C. Protein Insertion.
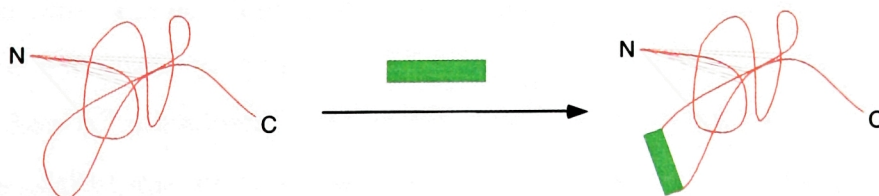
Figure1-6. Envisioned EPL protein synthesis strategies. Three different possibilities for combining recombinant proteins with synthetic or recombinant fragments: **(A)** C-terminal ligation. The fragment is attached to the C-terminus of the recombinant protein. **(B)** N-terminal ligation. The fragment is attached to the N-terminus of the recombinant protein. **(C)** Protein insertion. Insertion of the fragment into the core of the protein.

Objectives of thesis

The overall goal of this thesis is to illustrate the development and application of the protein semi-synthesis technique expressed protein ligation as a readily accessible and effective tool for the incorporation of unnatural amino acids and physical probes into proteins. Specifically, EPL was applied to the incorporation of fluorophores, post-translational modifications, and UV photochemical cross-linkers into proteins of biological interest.

The scope and versatility of the expressed protein ligation method was demonstrated using three distinct applications. In the first application, a strategy was developed for the site-specific incorporation of fluorophores into large proteins. Using the Src homology 3 (SH3) domain of the Abelson protein tyrosine kinase (Abl) as a model system, a method was be developed for the incorporation of fluorophores into so called contiguous receptor-ligands, designed to act as biosensors for macromolecular interactions. With the second application, a general strategy was developed to introduce site-specific post-translational modifications into proteins for analysis of signal transduction. This strategy was applied to the selective phosphorylation of the transforming growth factor β receptor I (TβR-I) GS region to investigate the role phosphorylation plays in the receptor's signaling activities. To this end, an approach was developed for the synthesis of phosphorylated thioester peptides for use in EPL reactions. The final application calls the for design and synthesis of several modified versions of the *E. coli* sigma factor, σ^{70}, in order to investigate inhibition of the protein. Using EPL, a general technique was developed for incorporating UV photochemical cross-linkers and fluorophores into modified versions of σ^{70}, which was used in a variety of experiments to investigate how σ^{70} is inhibited by the bacteriophage T4 anti-sigma (AsiA) protein.

The goal of this thesis was achieved through a synergy of the existing methods (native chemical ligation, protein splicing, recombinant biology) and novel techniques (expressed protein ligation) that, when used in conjunction, provide unique opportunities for expanding the boundaries of protein engineering.

Chapter 2: Materials and Methods

"Great things are not done by impulse, but by a series of small things brought together."

--Van Gogh (1853-1890)

2.1 General Reagents and Methods

Solid-phase peptide synthesis.

Boc-amino acids were obtained from Novabiochem (San Diego, CA), Bachem (Torrance, CA) or Neosystems (France). HBTU, and 4-methylbenzylhydrylamine (MBHA) resin were obtained from Richelieu Biotechnologies (Montreal, Canada), and Peninsula Laboratories (Belmont, CA), respectively. *N,N*-Dimethylformamide and HPLC-grade acetonitrile was purchased from Fischer. Trifluoroacetic acid was purchased from Halocarbon (River Edge, NJ). HF was purchased from Matheson Gas. All other reagents were obtained from Aldrich Chemical Co. All polypeptides were manually synthesized according to the *in situ* neutralization/HBTU activation protocols for Boc SPPS (Schnolzer, 1992) on a MBHA resin. Global deprotection and cleavage from the support was achieved by the treatment with hydrogen fluoride (HF) containing 5% (v/v) *p*-cresol, for one hour at 0^0 C. Following the removal of HF, crude peptide products were precipitated and washed with anhydrous cold Et_2O before being dissolved in degassed aqueous acetonitrile (50%) and lyophilized. Polypeptides were purified by preparative HPLC and in all cases, polypeptide composition and purity were confirmed by ESMS and analytical HPLC.

Reverse phase HPLC.

Analytical HPLC were preformed on a Hewlett-Packard 1100 series instrument with diode array detection. Analytical runs were performed on a reversed-phase Vydac C18 column (5 micron, 4.6 mm x 150 mm) at a flow rate 1 mL/minute. Preparative HPLC was performed on a Waters DeltaPrep 4000 system fitted with a Waters 486 tunable absorbance detector using a Vydac C18 column (15-20 micron, 50 mm x 250 mm) at a flow rate of 30 mL/minute. All runs used linear gradients of 0.1% (w/v) aqueous TFA (solvent A) *versus* 90% (v/v) acetonitrile plus 0.1% TFA (solvent B).

Mass spectrometry.

Electrospray mass spectrometry (ESMS) was routinely applied to all synthetic peptides and components of reaction mixtures. ESMS was performed on a Sciex API-100 single quadrupole electrospray mass spectrometer. Matrix-assisted laser desorption ionization (MALDI) mass spectrometry was performed on a Perseptive Biosystems Voyager-DE STR Biospectrometry Workstation and analyzed using the m/z program. Calculated masses were obtained using the program MacProMass (Sunil Vemuri & Terry Lee, City of Hope, Duarte, CA).

2.2 Materials and Methods used in Section 3.1

Synthesis of H_2N-Lys[Rh]-(Gly)$_{11}$-(3BP2)-Ahx-Arg-Ala-Lys[Fl]-CONH$_2$ (1).

The peptide was synthesized (0.1 mmole scale) on a MBHA resin using a Boc/Bzl protection strategy. Orthogonal protection of the lysine ε-NH_2 groups with fluorenylmethoxycarbonyl allowed directed attachment of the tetramethylrhodamine and fluorescein groups (activated as N-hydroxysuccinimide esters) before the final cleavage step. Following cleavage, the peptide (140 mg) was purified by preparative RP-HPLC using a linear gradient of 25-40% B over 60 min to give 15 mg (0.04 mmole, 8% yield) of material. The purified peptide was characterized as the desired product (1) by ESMS [observed mass = 3024.32 ± 0.8 Da; expected mass (average isotopic mass) = 3025.0 Da].

Synthesis of Cys-Abl-SH3 (2).

Synthesis of the 57 amino acid peptide (residues Leu^{65}- Val^{119}of native sequence) [NH$_2$-CLFVALYDFVASGDNTLSITKGEKLRVLGYNHNGEWBEAQTKNGQGWVP SNYITPVN-COOH] was preformed on a 0.5 mmole scale on an Asn-Pam resin. The naturally occurring Cys^{101} in wild-type Abl-SH3 was mutated to α-aminobutyric acid (B) to avoid side reactions during thioether ligation. Following HF cleavage (400 mg), the peptide was purified by preparative RP-HPLC using a linear gradient of 35-45%B over

23

60 min. The purified peptide was characterized by ESMS [observed mass = 6302.58 ± 0.75 Da; expected mass average isotopic mass) = 6301.8 Da].

Synthesis of N-iodoacetylated 3BP2 peptide (IAc-3BP2) (3).

In a 1 mL eppendorf, 4 mg of purified (1) was solubilized in 800 mL of 8M urea, 100 mM 2-[N-Morpholino]ethanesulfonic acid (MES) pH= 6.3. Iodoacetic anhydride solution was prepared by dissolving 5.28 mg of iodoacetic anhydride in 200 ml of tertahydrofuran (THF). 21 mL (1.5 mmol) of iodoacetic anhydride solution was then added to the eppendorf containing the solubilized peptide (1). The reaction was allowed to proceed for 30 min then the desired N-iodoacetylated 3BP2 peptide (IAc-3BP2) (3) was purified by semi-preparative RP-HPLC using a 25-45% B linear gradient over 60 min. The purified product was characterized by ESMS [observed mass = 3191.6 ± 0.1 Da; expected mass (average isotopic mass) = 3192.0 Da].

Synthesis of Abl-SH3-3BP2 FRET biosensor using thioether approach (4).

2.5 mg (0.8 μmol) of IAc-3BP2 (3) and 12.5 mg (1.9 μmol) of Cys-Abl-SH3 (2) were solubilized in 1 mL of degassed 8M urea, 0.1M sodium phosphate buffer pH 7.4. The reaction proceeded for 1 hour then the reaction was purified by semi-preparative RP-HPLC using a 30-55% B linear gradient over 60 min. The purified product (4) was characterized by EMS [observed mass = 9370.0 ± 0.8 Da; expected (average isoptopic mass = 9368.1 Da].

Preparation of Recombinant Abl-SH3-Intein-CBD (5).

E. coli BL21 cells transformed with pTYB2Abl-SH3 *(Xu et al., 1999)* were grown to mid-log phase in Luria Bertani (LB) medium and induced with 0.1 mM IPTG at room temperature for two hours to yield protein in the soluble fraction. Following centrifugation, cells were re-suspended in 60 ml of lysis buffer (25 mM HEPES, pH 8.0, 0.1 mM EDTA, 250 mM NaCl, 5% glycerol, 1.0 mM PMSF) and lysed using a French press. The lysate was clarified by centrifugation and then loaded onto a 15 ml chitin column pre-equilibrated in column buffer (20 mM HEPES, pH 7.0, 250 mM NaCl, 1mM EDTA, 0.1% Triton X-100), the column was extensively washed using the same buffer

24

and then stored at 4°C until further use. The loading of the column was determined by cleavage of 100 μL of beads in ligation buffer (0.2 M phosphate, 0.2 M NaCl, pH 7.2) containing 100 mM DTT. The amount of cleaved protein in solution was quantified by analytical RP-HPLC by comparison to a stock solution of known concentration. This afforded a resin loading of 0.4 mg/mL.

Synthesis of Cys-3BP2 (6).

The peptide was synthesized as described for peptide (1) with chain elongation continued to include an N-terminal cysteine residue. The purified peptide was characterized by ESMS [observed mass = 3128.1 ± 0.8 Da; expected mass (average isotopic mass) = 3129.5 Da].

Semi-synthesis of Abl-SH3-3BP2 FRET biosensor by EPL (7).

Purified peptide (6) (1.0 mg) was dissolved in 2.0 ml of ligation buffer containing 10% (v/v) DMSO. The solubilized peptide was added to 1.5 ml of Abl-SH3-Intein-CBD (5) chitin beads in the presence of 1.5 % (v/v) thiophenol. The ligation reaction was allowed to proceed for 44 hours at room temperature with gentle mixing, at which point the desired ligation product (7) was purified by semi-preparative RP-HPLC using a linear gradient of 25-50% B over 1 hour to yield 0.7 mg (74 nmole, 77% yield) of material. The purified product was characterized by ESMS [observed mass = 9371.2 ± 0.8 Da; expected mass (average isotopic mass) = 9370.5 Da].

Synthesis of NH$_2$-Cys-Gly-Arg-Gly-Arg-Gly-Arg-Lys[Fl]-CONH$_2$ (8).

The peptide was synthesized (0.5 mmole scale) on an MBHA resin using a Boc/Bzl protection strategy. Orthogonal protection of the lysine ε-NH$_2$ group with fluorenylmethoxycarbonyl allowed specific attachment of the fluorescein moiety (activated as *N*-hydroxysuccinimide ester) before the final HF cleavage step. The cleaved peptide was purified by preparative RP-HPLC using a linear gradient of 0-73% B over 60 min to give 25 mg (0.02 mmole, 52%) of material. The purified peptide was characterized by ESMS [observed mass = 1245.9 ± 0.1 Da; expected mass (average isotopic mass) = 1246.5 Da].

25

Model EPL Studies Between Abl-SH3-Intein-CBD (5) and peptide (8).

Peptide (8) was dissolved (1 mg/mL) in a ligation buffer containing either 0, 2, 3, 4 or 5.5 M GdmCl (final concentration after dilution effect of the beads). This solution (100-200 μL) was then added to typically 50-100 μL of Abl-SH3-intein-CBD beads (5) equilibrated in ligation buffer and either 1% thiophenol (v/v) or 1% 2-mercaptoethanesulfonic acid (w/v) added to the mixture. The ligation reaction was allowed to proceed at room temperature overnight with gentle agitation. The amount of ligation product Abl-SH3-Cys-Gly-Arg-Gly-Arg-Gly-Arg-Lys[Fl]-CONH$_2$ (9), present at different time-points was determined by analyzing aliquots of the reaction supernatant by analytical RP-HPLC and ESMS: [observed mass 7488.0 ± 1.5 Da; expected for (9) (average isotope composition) = 7488.5 Da].

Synthesis of NH$_2$-CNTDYLEEVGKQFDVTRERIRQIEAKALRKLRH-CONH$_2$ (10).

The peptide was synthesized (0.33 mmole scale) on a *p*-methylbenzhydrylamine resin using a Boc/Bzl protection strategy. Following cleavage (355 mg) and global deprotection with HF, the peptide was purified by preparative HPLC using a linear gradient of 30-60% B over 60 min to give 44 mg (0.011 mmole, 16% yield) of material. The purified peptide was characterized as the desired product (10) by ESMS [observed mass = 4115.9 ± 0.5 Da; expected mass (average isotopic mass) = 4117.7 Da].

Semi-synthesis of Abl-SH3-[CNTDYTLEEVGKQFDVTRERIRQIEAKALR KLRH-CONH2] using EPL (11).

Purified peptide (10) (0.6 mg) was dissolved in 0.1 ml of ligation buffer containing 2 M GdmCl. The solubilized peptide was added to 100 μl of Abl-SH3-intein-CBD chitin beads (5) in the presence of 1.0 % (v/v) MESNA. The ligation reaction was allowed to proceed for 24 hours at room temperature with gentle mixing, at which point the desired

ligation product was isolated by analytical HPLC and characterized by ESMS [observed mass = 10357.5 ± 2.0 Da; expected mass (average isotope comp. = 10358.3 Da).

Synthesis of P40 (12) and NS (13) ligands.

P40 (12) (APTYSPPPPP) was synthesized (0.5 mmole scale) on an MBHA resin using standard Boc SPPS. Following HF cleavage the peptide was purified by preparative reverse-phase HPLC using a linear gradient of 7-27% B over 60 min. The purified peptide was characterized as the desired product by ESMS [observed mass = 1022.9 + 0.7 Da; expected mass (average isotopic mass) = 1023.2 Da]. NS (13) (YSPWTNF) was obtained from a co-worker in the lab.

Fluorescence spectroscopy.

Experiments were conducted at 18^0C in a stirred 0.5 cm-path length cuvette using a SPEX FL3-11C fluorimeter. Excitation was at 490 nm with a 1.5 nm slit width and the fluorescence emission was monitored at 520 nm and 580 nm through a 5 nm slit. The protein was titrated with increasing concentrations of each ligand; P40 (12) or NS (13), and the change in the emission ratio of fluorescein to rhodamine (580:520) was monitored as a function of ligand concentration. In all cases, the protein concentration was kept at 50 nM in a buffer containing 150 mM NaCl, 100mM sodium phosphate buffer pH 7.2.

Circular dichroism spectroscopy.

CD measurements were preformed on an Aviv 62DS spectropolarimeter at 25^0 C using a 0.02cm path length cuvette. Protein concentrations were 50μM in 150mM NaCl, 100mM sodium phosphate buffer pH 7.2. Spectra were recorded from 250 to 190 nm using a step of 0.5 nm and an averaging time of four seconds.

2.3 Materials and Methods used in Section 3.2

Synthesis of N$^\alpha$-Biotin-GTTLKDLIYDMTTSGSGSGLPL-COS-propion-amide (GS-0Bio (14)).

The peptide was synthesized (0.03 mmole scale) on an (MBHA) resin *(Camarero, 1998)* according to the *in-situ* neutralization/HBTU activation protocol for Boc SPPS, and the N-terminus of the peptide was biotinylated using biotin N-hydroxysuccinimide. Following chain assembly, global deprotection and cleavage from the support was achieved by treatment with HF containing 4% (v/v) *p*-cresol, for 1 h at 0°C. Following removal of the HF, the crude peptide product was precipitated and washed with anhydrous cold Et$_2$O before being dissolved in degassed aqueous acetonitrile (50% B) and lyophilized. The crude peptide (20 mg, 7.87 μmole) was purified by semi-preparative HPLC using a linear gradient of 35-45% B over 60 min to give 4 mg (1.6 μmole, 20% yield) of material. The purified peptide was characterized as the desired product (GS-0Bio) (14) by ESMS [observed mass = 2542.2 ± 0.1 Da; expected mass (average isotopic mass) = 2540.7 Da].

Synthesis of H$_2$N-TTLKDLIYDMpTTpSGpSGpSGLPL-COS-Bzl (GS-4 (15)).

The peptide was synthesized (0.3 mmole scale) on a sulfamylbutyryl AM resin using the Fmoc N$^\alpha$ SPPS strategy. The C-terminal leucine residue was loaded onto the resin using the method of Backes et al. *(Backes, 1999)*, which afforded an initial loading of 0.3 mmole/g. All amino acids (2.2 mmole) were coupled using HBTU activation with the exception of Fmoc-Ser(PO(OBzl)OH)-OH and Fmoc-Thr(PO(OBzl)OH)-OH which were coupled using PyBOP as the activating agent. Coupling times ranged from 1-2 hours, and all residues from Met[184]-pSer[191] of GS[175-195] were double coupled. The N-terminal Thr was attached as a Boc-Thr(O^{t}Bu)-OH derivative. Following chain assembly, the peptide resin was activated by an overnight reaction with ICH$_2$CN (9.5 mmole) in DMF (3.2 mL) in the presence of DIEA (220 mL). Iodoacetonitrile used for N-alkylation was prefiltered through an basic alumina plug. After extensive washing with DMF and DCM,

a portion (600 mg, 0.075 mmole) of this resin was treated with BzlSH:DIEA:DCM (2.7 mL : 1 mL : 17.8 mL) for 6 hours at room temperature with gentle stirring at which point the slurry was concentrated *in vacuo*. Global deprotection was then achieved by treating the dried resin with 25 mL TFA containing 5% H_2O (v/v) and 5% p-cresol (v/v) as scavengers. Following removal of the TFA *in vacuo*, the crude peptide product was precipitated and washed with anhydrous cold Et_2O before being dissolved in degassed aqueous acetonitrile (50% B) and lyophilized. The crude peptide (117 mg, 0.045 mmole) was purified by semi-preparative HPLC at 50°C using a linear gradient of 35-45% B over 60 min to give 6 mg of purified GS-4 (15) (2.3 mmole, 5.1% yield) of material. We attributed the modest yield of GS-4 (15) to 1) to incomplete thiolysis of the peptide from the resin, confirmed by amino acid analysis of the resin after cleavage, and 2) incomplete deprotection of the phosphate groups, confirmed by ESMS of the primary contaminant, which displayed an additional 90 Da. The purified peptide was characterized as the desired product (15) by ESMS [observed mass = 2597.0 ± 0.5 Da; expected mass (average isotopic mass) = 2596.5 Da].

GSΔTβR-I (16) Expression and Purification (performed by M. Huse).

Expression

In general, followed the protocols found in the Bac-To-Bac™ baculovirus expression system manual (Gibco BRL) to produce recombinant TβR-I protein. The Leu[196]-Cys mutation and the coding sequence for Factor Xa cleavage (IEGR) were encoded in the N-terminal primer used to amplify the GSΔTbR-I (16) DNA fragment. The DNA was cloned into the pFastBacHTa expression vector (Gibco BRL) using the SpeI and SphI restriction endonuclease sites. This strategy placed the Factor Xa cleavage site and GSΔTβR-I (16) coding sequence downstream of a hexahistidine tag. For expression, 4-6 liters of confluent SF9 cells in suspension culture (~2 X 10^6 cells/mL) were infected with GSΔTβR-I (16) baculovirus at a multiplicity of infection of 1-10 virus particles per cell. Cells were harvested after 50 hours and washed in PBS before being frozen at –78°C.

Purification

4-6 liters of frozen, GSΔTβR-I (16) expressing SF9 cells were thawed and resuspended in 70 mL of 20 mM Tris pH 8.5, 10 mM β-mercaptoethanol (β-ME), 8% glycerol (v/v), and

1 mM phenyl-methyl sulphonyl fluoride (PMSF). The cells were lysed by sonication and the resultant lysate subjected to centrifugation at 16,500 rpm in a midspeed Beckman J2-MI centrifuge for 1 hour. The soluble fraction of the lysate was filtered through a syringe filter with a 5 μm pore size (Nalgene). The sample was then applied to a large Q sepharose column (~50 mL bed volume; Pharmacia) equilibrated in a buffer containing 20 mM Tris pH 8.5, 5 mM β-ME, 8% glycerol (v/v). Under these conditions, the His-tagged GSΔTβR-I (16) protein elutes isocratically immediately following the flow through peak. Fractions containing GSΔTβR-I (16) were confirmed by SDS-PAGE and pooled for Ni^{2+} affinity purification. NaCl and imidazole were added to the protein at this stage to final concentrations of 500 mM and 15 mM, respectively. The sample was then applied to a 15 mL Ni^{2+}-NTA column (Qiagen) equilibrated with a buffer containing 20 mM Tris pH 8.5, 5 mM β-ME, 15 mM imidazole, 500 mM NaCl, and 8% glycerol (v/v). The column was washed and the salt concentration reduced to 50 mM by a linear gradient into a buffer containing 20 mM Tris pH 8.5, 5 mM β-ME, 15 mM imidazole, 50 mM NaCl, 8% glycerol (v/v). The protein was then eluted with a linear gradient into a buffer containing 20 mM Tris pH 8.5, 5 mM β-ME, 1 M imidazole, 50 mM NaCl, 8% glycerol (v/v). Fractions containing His-tagged GSΔTβR-I (16) were pooled and the approximate concentration of the sample assessed by an absorbance measurement at 280 nm. The His-tag was then removed by Factor Xa digestion in the presence of 2 mM dithiothreitol (DTT). Note, factor Xa cleaves GSΔTβR-I (16) nonspecifically in the absence of reducing agent. 15 units of factor Xa were added for every mg of GSΔTβR-I (16), and the digestion was allowed to proceed at room temperature for 7 hours at which point 1 mM PMSF was added to quench the reaction. The cleaved GSΔTβR-I (16) protein was subsequently purified from the protease using an ATP sepharose affinity column (prepared as described in *(Huse, 2000)*). The ATP column was equilibrated with a buffer containing 20 mM Tris pH 8.5, 10 mM $MgCl_2$, 2 mM DTT, 8% glycerol (v/v). Prior to loading, $MgCl_2$ was added to the sample to a final concentration of 10 mM. After loading, the column was washed with a buffer containing 20 mM Hepes pH 8.0, 2 mM DTT, 8% glycerol (v/v), and then eluted with a linear gradient into a buffer containing 20 mM Hepes pH 8.0, 1 M NaCl, 5 mM EDTA, 2 mM DTT, 8% glycerol (v/v). The purified protein was characterized as the desired product (16) by ESMS [observed mass =

35,263 ± 11.8 Da; expected mass (average isotopic mass) = 35259.6 Da]. Pure GSΔTβR-I (16) protein was concentrated using a 15 mL Biomax ultrafiltration device (Millipore) with a 30 kD molecular weight cutoff to a final concentration of 0.1 mM. mercaptoethanesulfonic acid (MESNA) and hepes pH 8.0 were added to final concentrations of 50 mM and 100 mM, respectively, and the protein solution was flash frozen in 100 μL aliquots. GSΔTβR-I (16) was stored at –78°C until further use.

Synthesis of GS-0BioTβR-I (17) and GS-4TβR-I (18) by EPL.

In general, ligation reactions were initiated by dissolving purified peptide (100 nmol) in a 100 mL aliquot of purified GSΔTβR-I (16) (10 nmol protein). For GS-4 ligations, the protein solution was brought to 1 M NaCl and 0.1 mM Na_3VO_4 prior to the addition of the peptide. Vanadate was added to inhibit any phosphatase activity that may have copurified with GSΔTβR-I (16), and the higher salt concentration reduced GSΔTβR-I (16) aggregation. Ligation reactions were performed at 4°C, and their progress monitored by SDS-PAGE. GS-0Bio reactions were generally stopped after 10-12 hours, GS-4 reactions after 24-48 hours. Reactions were stopped by gel filtration, which separated the protein components from the unreacted peptide. Gel filtration was performed using a Superdex 75 HR 10/30 column equilibrated in 20 mM hepes pH 8.0, 500 mM NaCl, 2 mM DTT, 8% glycerol (v/v). There are three major peaks observed by HPLC analysis of in TβR-I ligation reactions, a peak at the void volume consisting of aggregated protein, a second peak corresponding to monomeric protein, and a third peak of unreacted peptide. The second peak is enriched in ligation product, and was, in both cases, the sample used for subsequent purification.

GS-0BioTβR-I (17) Purification.

GS-0BioTβR-I (17) was purified from unreacted GSΔTβR-I (16) using monomeric avidin agarose affinity beads (Pierce). Avidin agarose (1 mL) was placed in a gravity flow column and equilibrated in the following equilibration buffer: 20 mM Tris pH 7.5, 150 mM NaCl, 4 mM DTT. After sample loading, the column was washed with equilibration buffer and eluted with a buffer containing 20 mM Tris pH 7.5, 150 mM NaCl, 2 mM biotin, 4 mM DTT. The elution fractions containing GS-0BioTβR-I (17)

31

were pooled and then concentrated by ultrafiltration using a centricon 30 (Millipore). The protein was concentrated to ~1 mg/mL. GS-0BioTbR-I (17) may be stored indefinitely at −78°C. The identity of GS-0BioTbR-I (17) was confirmed by ESMS [observed mass = 37698.0 ± 7.2 Da; expected mass (average isotopic mass) = 37695.2 Da].

GS-4TβR-I (18) Purification.

GS-4TβR-I (17) was purified from unreacted GSΔTβR-I (16) by anion exchange chromatography using a 1 mL Resource Q column (Pharmacia). The column was equilibrated in a buffer containing 20 mM Tris pH 8.5, 2 mM DTT, 8% glycerol (v/v) and eluted with a linear gradient into a buffer containing 20 mM Tris pH 8.5, 1 M NaCl, 2 mM DTT, 8% glycerol (v/v). The sample was diluted 5 fold in equilibration buffer prior to loading. Most of the unreacted GSΔTβR-I (16) flows through the column under these conditions. The eluted fractions containing GS-4TβR-I (18) were pooled and then concentrated by ultrafiltration using a centricon 30 (Millipore). The protein was concentrated to ~2 mg/ml and can also be stored indefinitely at −78°C. GS-4TβR-I (18) does not ionize under electrospray conditions (presumably a consequence of the 4 phosphates in the protein), but can be observed using MALDI-MS [observed mass = 37,731 ± 18 Da; expected mass (average isotopic mass) = 37,731.9 Da].

In Vitro Kinase Assays (performed by M. Huse).

The relative concentrations of the TβR-I fragments used in this assay were determined by HPLC peak integration and referenced to TβR-Icyt, whose concentration had been determined previously by absorbance at 280 nm. Kinase reactions were performed in a buffer containing 50 mM Tris pH 7.5, 60 mM NaCl, 5 mM $MgCl_2$, 5 mM $MnCl_2$, 2 mM DTT, 0.1 mM Na_3VO_4. Kinase activity is barely detectable in the absence of $MnCl_2$. While this requirement for Mn^{2+} is not fully understood, it is not unusual for kinases to require ions other than Mg^{2+} for full activity *(Amrein, 1995; Herrera, 1988)*. The substrate used in these assays is a complex between the C-terminal Smad2 MH2 domain and the 85 residue Smad binding domain of the Smad Anchor for Receptor Activation (SARA). SARA stabilizes Smad2, and renders it more soluble. The appropriate TβR-I fragment (10 pmole) and Smad2/SARA complex (80 pmole) were used in each reaction.

32

The total reaction volume was 15 mL. Reactions were initiated by the addition of ATP (200 mM final concentration) along with 7.5 mCi of $[\gamma\text{-}^{32}P]$ATP (Pharmacia), and were allowed to proceed at room temperature. They were quenched after 10 min by the addition of 5 mL SDS-PAGE sample buffer. Samples were subjected to SDS-PAGE, and the incorporation of radioactivity into each protein visualized and quantitated using a Storm 840 phosphorimager (Molecular Dynamics). All quantitated assays were performed in triplicate. TβR-Icyt and Smad2/SARA were expressed and purified as described previously (Huse, 1999; Wu, 2000).

2.4 Materials and Methods used in section 3.3

Synthesis of σ^{70} 4.2 benzophenone cross-linker peptides (19-25).

All derivatized peptides were synthesized on a 0.1 mmol scale using Boc SPPS. During the synthesis, a single Boc-Dapa(Fmoc)-OH amino acid was incorporated at positions corresponding to σ amino acids 571 (**20**) , 578 (**21**), 586 (**22**), 588 (**23**), 593 (**24**) or 597 (**25**). After the synthesis was complete, each Dapa(Fmoc)-containing peptide was treated with 2% DBU, 20% piperidine in DMF (3 x 5minute). The free β-amino group was then acylated (2 x 1hour) with p-benzoyl-benzoic acid (52.6mg, 0.23mmol) activated as its symmetrical anhydride using DIC (0.11 mmol, 15 μl) in DMF (3mL). After the second coupling the acylation reaction was > 99 % complete, as judged by ninhydrin. The peptide-resins were cleaved and deprotected by treatment with anhydrous HF, and the derivatized peptides were purified by HPLC using a linear gradient of 35-55% B. Peptide composition and purity were confirmed by analytical HPLC and ESMS (see Table 1 for masses).

Cloning and expression of AsiA protein.

The AsiA protein used in cross-linking experiments was obtained from Leonid Minakhin. The details of the cloning and expression are published *(Minakhin, 2001)*.

UV Photo Cross-Linking Experiments.

Each cross-linking reaction contained, in 15 μl of transcription buffer (50 mM Tris-HCl, pH 7.9, 10 mM $MgCl_2$, 40 mM KCI, 1 mM β-mercaptoethanol) 200 pmol of AsiA and 3 nmol of derivatized σ^{70} region 4.2 peptide. Reactions were incubated for 5 min at 37 °C and irradiated 20 times with 5-7 ns pulses of UV-laser (YAC Powerlite 7000, Continuum) at 355nm, 50 mJ per pulse. Reaction products were analyzed by SDS PAGE (Tris-glycine 10-20% gradient, Bio-Rad) and visualized by Coomassie blue staining.

Cloning and Expression of σ^{70} constructs: σ^{70}_{1-566}-VMA-intein-CBD (26) and $\sigma^{70}_{101-566}$-VMA-intein-CBD (27).

The constructs σ^{70}_{1-566}-VMA-intein-CBD (26) and $\sigma^{70}_{101-566}$-VMA-intein-CBD (27) were cloned and expressed in a manner similar to that described by Severinov et.al *(Severinov and Muir, 1998)*. The plasmid pCYB2-σ^{70}_{1-566}, which expresses regions 1 through region 4.1 of σ^{70} fused to VMA-intein-CBD from a trc promoter, was constructed by polymerase chain reaction amplification of the corresponding fragment of rpoD and recloning it in NdeI-SmaI-treated plasmid pCYB2 (New England Biolabs) as previously described *(Severinov and Muir, 1998)*. The DNA encoding $\sigma^{70}_{101-566}$ was isolated by PCR using the pTYB2 plasmid as a template. The 5' primer (5'-GCG TCT AGA AAT AAT TTT GTT TAA GAA GGA GAT ATA CAT ATG TAC ATG CGT GAA ATG GGC ACC G-3') and the 3' primer (5'- GCG CGC GGT ACC CTT GGC AAA GCA CAT ATC GAT ACC GAA ACG CAT ACG C -3') encoded XbaI and KpnI restriction sites, respectively. The PCR product was purified, digested simultaneously with XbaI and KpnI and then ligated into the XbaI-KpnI-treated plasmid pTYB2 (New England Biolabs). The resulting pCYB2-σ^{70}_{1-566} and pTYB2-$\sigma^{70}_{101-566}$ plasmids were shown to be free of mutations in the σ^{70} coding region by DNA sequencing. Six liters of *E. coli* BL21(DE3) cells containing the pCYB2-σ^{70}_{1-566} or pTYB2-$\sigma^{70}_{101-566}$ plasmid were grown to mid-log phase (OD_{590} = 0.6) in Luria-Bertani medium, induced with 1mM IPTG at 37°C for 6 hours and harvested by centrifugation. Cells were resuspended in 0.1 mM EDTA, 1 mM PMSF, 30 mM sodium phosphate, 250 mM NaCl buffer at pH 7.6 containing 5% glycerol and lysed using a French press. The lysate was centrifuged at

14,000 rpm for 30 minutes. The clarified supernatant was incubated for one hour with 5 mL chitin-beads (New England Biolabs), previously equilibrated in washing buffer (1 mM EDTA, 1 mM PMSF, 30 mM sodium phosphate, 250 mM NaCl buffer at pH 7.2). The beads were extensively washed with washing buffer (5 x 10 mL) and stored as a 50% suspension in wash buffer at 4 ^{0}C until further use. The loading of the column was determined by desorbing the fusion protein from a small aliquot of chitin beads with PBS containing 8 M GdmCl and quantifying the protein by UV absorbtion (σ^{70}_{1-566} ϵ_{280}= 38480 cm^{-1}mol^{-1}and $\sigma^{70}_{101-566}$ ϵ_{280}= 35920 cm^{-1}mol^{-1}). This analysis indicated a loading of ~ 10 mg/mL for both constructs.

Synthesis of fluorescently labeled σ^{70} region 4.2 peptides (region 4.2-fl (28)) and (region 4.2-dansyl (29)).

Synthesis of the fluorescently labeled σ^{70} region 4.2 peptides was conducted using 200 mg of stock σ^{70} region 4.2 peptide mutated at σ^{70} Lys^{578} that had been prepared for the cross-linking studies as described above. Each Dapa (Fmoc)-containing peptide was treated with 2% DBU, 20% piperidine in DMF (3 x 5minute). The free β-amino group was then acylated (1 x 2hour) with 100 mg (0.2 mmol) NHS-fluorescein or 120 mg (0.4 mmol) dansyl chloride dissolved in DMF (3 mL) and DIEA (100 mL). After coupling the acylation reaction was > 99 % complete, as judged by ninhydrin. The peptide-resins were cleaved and deprotected by treatment with anhydrous HF, and the fluorescently labeled peptides were purified by preparative HPLC using a linear gradient of 30-50% B (28) or 30-60% B (29) over 60 minutes. Peptide composition and purity were confirmed by analytical HPLC and ESMS [(28) observed mass = 4434.46 ± 1.0 Da; expected mass (average isotopic mass) = 4432 Da. (29) observed mass = 4307.52 ± 0.17 Da; expected mass (average isotopic mass) = 4306.0 Da].

Synthesis of peptide CGRAKG- (model-fl (30)) and (model-dansyl (31)).

CGRAKG was synthesized on 0.5 mmol scale using Boc SPPS. Orthogonal protection of the lysine ε-NH2 group with fluorenylmethoxycarbonyl allowed directed attachment of NHS-fluorescein or dansyl chloride in a manner similar to that described for σ^{70} region 4.2 peptides (28) and (29) described above. Following HF cleavage the model peptides

were purified by preparative HPLC using a linear gradient of 15-40% B over 60 minutes. Peptide composition and purity were confirmed by analytical HPLC and ESMS [(5) observed mass = 948.9 ± 3.46 Da; expected mass (average isotopic mass) = 950.0 Da. (4) observed mass = 823.9 ± 0.70 Da; expected mass (average isotopic mass) = 824.0 Da].

Synthesis of σ^{70} constructs: σ^{70}_{1-566}-4.2-fl (32), σ^{70}_{1-566}-4.2-fl (33), $\sigma^{70}_{101-566}$-4.2-dansyl (34), σ^{70}_{1-566}-model Fl (35), $\sigma^{70}_{101-566}$-model Fl (36), $\sigma^{70}_{101-566}$-model dansyl (37) using EPL.

In general purified peptide (28), (29), (30) or (31) (0.2 – 7.0 mg) was dissolved in buffer A (50μL – 1 mL). The solubilized peptide was then flash diluted into EPL buffer (0.1M sodium phosphate, 0.1M NaCl pH = 7.2) (200 μL - 4 mL). The dissolved peptide was added to a 50% slurry of σ^{70} VMA-intein fusion protein (26) or (27) (200 μL - 4 mL) in the presence of 3% ETSH and 100 mM mercaptoethanesulfonic acid (MESNA). The ligation reaction proceeded over night at room temperature with gentle mixing, at which point the ligation mixture was separated by centrifugation. The ligation supernatant and chitin beads were analyzed by SDS-PAGE electrophoresis and visualized by fluorescent imaging and coomassie staining. Characterization of the ligation supernatant by analytical HPLC and MALDI-TOF MS confirmed the desired σ^{70} ligation product (32), (33), (35) and (36) (See Table 3 for masses) [(32) observed mass = 69,112 ± 86.6 Da; expected mass (average isotopic mass) = 69,178.1 Da. (33) observed mass = 58,258.2 ± 1.8 Da; expected mass (average isotopic mass) = 58,261.2 Da. (35) observed mass = 65,730 ± 42 Da; expected mass (average isotopic mass) = 65,696.1 Da. (36) observed mass = 54,886.2 ± 97.4 Da; expected mass (average isotopic mass) = 54,799.2 Da]. Note, (34) and (37) did not ionize under MALDI-TOF MS conditions, but were confirmed by RP-HPLC and SDS-PAGE analysis.

Purification of σ^{70} ligation products.

Ligation reaction supernatants were purified by anion exchange using a 1 mL Q sepharose HiTrap HP column (Amerhsham Pharmacia). The column was equilibrated in a buffer containing 0.1M NaCL, 30 mM sodium phosphate pH= 7.2 and eluted with a

linear gradient (flow rate 0.5 mL/min) into a buffer containing 1M NaCl, 30 mM sodium phosphate pH 7.2 gradient. Most of the unreacted flourescently labeled region 4.2 (**28** or **29**) flows through the column under these conditions. The eluted fractions containing the pure σ^{70} constructs were then buffer exchanged into fluorescence buffer (0.1M sodium phosphate, 50 mM NaCl, pH = 7.2) using gel filtration with a Zorbax GF-250 column, running buffer 50mM sodium phosphate, 0.1M NaCl pH = 7.2 over 17 minutes (flow rate 1 mL/min). The fractions corresponding to the pure σ^{70} constructs were analyzed for purity using MALDI- TOF MS and SDS-PAGE analysis.

Fluorescence spectroscopy.

Spectroscopic analysis of the σ^{70} constructs was conducted at 25^0C in a stirred 0.5 cm-path length cuvette using a SPEX FL3-11C fluorimeter. Excitation was at 490 nm with a 1.5 nm slit width and the fluorescence emission was monitored at 520 nm through a 5 nm slit. In all cases, the protein concentration was kept at 5 nM in fluorescence buffer.

37

Chapter 3: Results and Discussions

"Luck is a large factor in the selection of a problem and the possibility that something new or unexpected will be found... In saying this, I realize how very lucky I have been-chance was on my side."

--Bruce Merrifield in a discussion about the work that led to his Nobel Prize *(Merrifield, 1993)*.

3.1 Site-Specific Incorporation of Fluorescent Probes into Abl-SH3 –3BP2 Receptor-Ligand Complex

Part of this work conducted in collaboration with Graham Cotton.

Introduction to Abl-SH3-3BP2 receptor-ligand complex – significance to EPL development

The Src homology 3 (SH3) domains were initially identified as conserved sequences in the N-terminal non-catalytic regions of cytoplasmic tyrosine kinases, phospholipase C, and vCRK *(Meyer, 1988)*. SH3 domains are found in a wide variety of proteins that are involved in several cellular functions including signal transduction (for review *(Mayer, 2001; Pawson, 2002)*). The Abl-SH3 domain was chosen as a model protein for the development of EPL for several reasons: it is well characterized both structurally and biochemically *(Gosser et al., 1995; Musacchio et al., 1992; Musacchio et al., 1994; Pisabarro and Serrano, 1996)*, it is well behaved under normal physiological conditions and it can be reversibly unfolded *(Xu et al., 1999)*. Also, the relatively small size of the Abl-SH3 domain, 57 residues, makes it well suited to analytical analysis methods like reverse-phase high performance liquid chromatography (RP-HPLC) and electrospray mass spectrometry (ESMS). Here the Abl-SH3 domain with its proline rich ligand 3BP2, was used as a model system to develop the tools that would allow for the site-specific incorporation of fluorophores into proteins using expressed protein ligation. In particular, an attempt was made to design a so-called contiguous receptor ligand biosensor for the Abl-SH3-3BP2 complex. Even though the efforts to manufacture a biosensor for the Abl-SH3-3BP2 complex were unsuccessful, the semi-synthetic chemical synthesis of the complex illustrated the power of the EPL technique for site-specifically incorporating fluorophores into proteins. Furthermore, model EPL studies with the Abl-SH3 domain indicated that ligations can be preformed in the presence of chaotropic and organic solvents.

Design of Abl-SH3-3BP2 contiguous- receptor ligand FRET biosensor

Biosensors can be the windows to the world of cell signaling *(D' Auria, 2001; Tsien, 1998)*. They can provide the ability to study *in vivo* the molecular events that regulate signal transduction. A biosensor detects physical changes in the cell and, for example, combines these changes with a specific bimolecular ligand binding event. It is this unique action and reaction relationship that makes biosensors effective signaling tools. Fluorescent proteins that involve simply tagging a protein of interest with the Green Fluorescent Proteins (GFP) provide useful information on cellular localization but not necessarily functional information. A solution to this problem has been to engineer biosensors that integrate signal transduction functions, thereby tailoring the sensor to a particular mechanism. The problem with this strategy is that a different biosensor might be needed for every distinct signal transduction mechanism. Thus, an important goal is the design of a general protein ligand biosensor that is applicable to the majority of signal transduction mechanisms.

In the quest for a general protein ligand biosensor design, the trend has been to exploit the ease of uv-visible spectroscopy to engineer biosensors with fluorescent probes *(Czarnik, 1995; Imperali, 1998; Still, 1998; Taylor, 1995)*. Biosensing is achieved when ligand binding corresponds to a change in the fluorescence intensity by the fluorescent probe. In such a design it is important to distinguish between specific ligand binding and nonspecific interactions. As a potential solution to this problem the concept of a contiguous receptor ligand biosensor was introduced. A contiguous receptor ligand biosensor has a low affinity ligand attached to the receptor via an amino acid tether, such that it can assume an intramolecular association. As illustrated in figure 3.1-1, a contiguous receptor ligand biosensor provides a fluorescent signal only as a consequence of a specific biological interaction; nonspecific interactions do not elicit a signal. In designing the Abl-SH3-3BP2 biosensor, it was decided that the use of fluorescence resonance energy transfer (FRET) would allow spatial, as well as, functional information to be obtained from the biosensor.
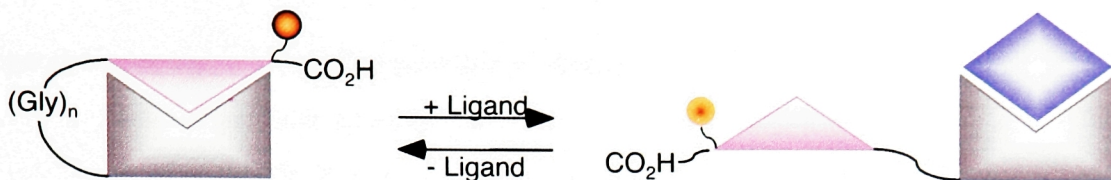
Figure 3.1-1. Concept of a contiguous receptor ligand biosensor. A change in ligand binding results in a change in fluorescence.

FRET, initially described by Forester in 1940, is a process where nonradiative electronic energy is transferred from an excited fluorophore donor (D) to a fluorophore acceptor (A) *(Forster, 1965; Forster, 1948)* (for review *(Dos Remedios, 1995))*. The excitation reaction involved in the reaction is as follows: D* + A => D + A*. Where D* and A* indicate the excited states of the donor and acceptor, respectively. As the reaction proceeds the donor molecule becomes quenched and the acceptor molecule begins to fluoresce. Thus, FRET has an inherent mode of detection, which makes it ideal for application to biosensors and signal transduction. Several biosensors have been designed that utilize FRET *(Cotton, 2000; Hofmann, 2001; Kurokawa, 2001; Miyawaki, 1997)*.

The intensity of FRET fluorescence is directly related to the spacing between donor and acceptor fluorophores. This is described by the following relationship:

$$R = R_0 \, (1/E-1)^{1/6} \quad \{equation \ 1\}$$

Where, R is the interchromophore distance between donor and acceptor, R_0, the Forster or critical transfer distance, is the distance between donor and acceptor when the efficiency of energy transfer is 50%, and E is the efficiency of energy transfer. Haugland and Stryer were the first to use FRET to calculate the distance between a fluorophore pair *(Stryer, 1967)*. Using a napthyl group and a dansyl group as a donor and acceptor pair, they calculated the distance from N to C terminus of a 12 amino acid poly-proline peptide, demonstrating that FRET can be used as a "spectroscopic ruler," where changes in fluorophore distance results in a change in fluorescence. Consequently, if FRET pairs are placed at either termini of the 3BP2 linker in the Abl-SH3-3BP2 contiguous receptor

ligand biosensor, it should be possible to distinguish a fluorescence change between the intra and inter-molecular associations of Abl-SH3 and its ligand (Figure 3.1-2). As the association of Abl-SH3 with its ligand changes from intra to inter, the distance between the FRET pair will change, thus changing the efficiency of FRET exchange, and hence creating effective biosensor. This type of chemical labeling is analogous to the incorporation of different green fluorescent protein (GFP) derivatives at the termini of recombinant proteins, by standard DNA cloning methodologies *(Tsien, 1998)*.
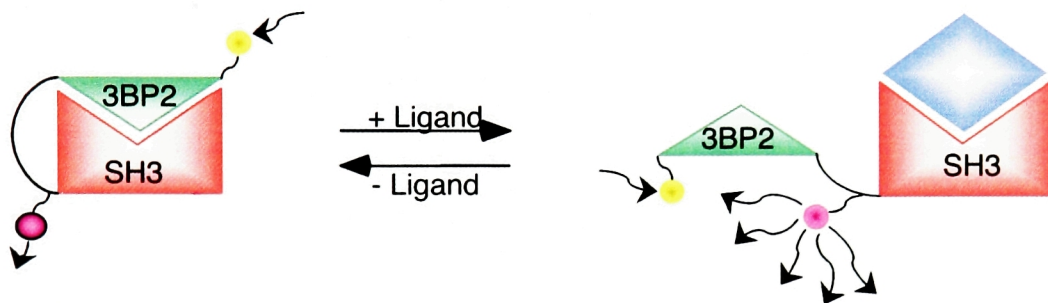
Figure 3.1-2. Design of Abl-SH3-3BP2 FRET biosensor. A change in ligand binding results in a change in the efficiency of FRET transfer.

Synthesis of the Abl-SH3-3BP2 contiguous FRET biosensor

There are four possible ways to ligate the Abl-SH3 domain to the 3BP2 sensor peptide: first, N-terminus to N-terminus; second, C-terminus to C-terminus; third, C-terminus Abl-SH3 to N-terminus 3BP2 sensor peptide; and fourth, N-terminus Abl-SH3 to C-terminus 3BP2 sensor peptide. Careful analysis of the crystal structure of the Abl-SH3 domain and its 3BP2 ligand (Figure 3.1-3) suggested that the optimal topology for a contiguous receptor ligand construct would be to link the two constructs via a poly-glycine linker using the first and third of the strategies listed above.

To determine the optimal length for the linker a region, a series of constructs were designed where the number of glycines, which constitute the linker, were either 2, 10, 12, or 14. The construct where n=2 acted as the control, since the linker length is too short to permit an intramolecular association. All the constructs had the potential to interact in an intra- or inter-molecular fashion, the former leading to self association. Size exclusion chromatography was used to establish the oligomerization states of the constructs at two different concentrations, 1 μM and 0.25 μM, and the data indicated that a linker region

containing 12 or 14 glycines is optimal for the desired intra-association between 3BP2 and Abl-SH3 (Cotton, unpublished). Thus, the synthesis strategy for the 3BP2 sensor peptide was as follows: H$_2$N-Lys[Rh]-(Gly)$_{11}$-(3BP2)-Ahx-Arg-Ala-Lys[Fl]-CONH$_2$ (1), where [Rh] and [Fl] are the fluorophores tetramethylrhodamine and fluorescein, FRET acceptor and donor pair, (Gly)$_{11}$ is the flexible linker, Ahx is aminohexanoic acid, used as a spacer, and (3BP2) corresponds to the amino acid sequence PPAYPPPPVP.
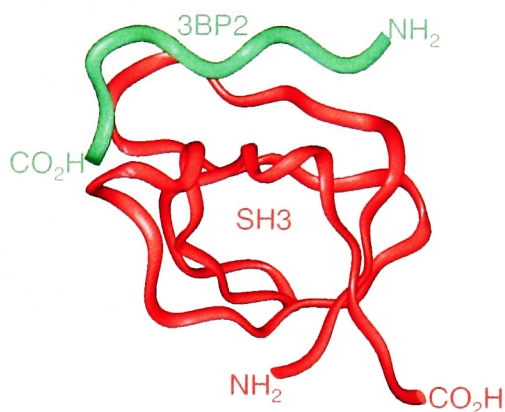
Figure 3.1-3. Crystal structure of Abl-SH3 complexed with 3BP2 ligand.

As was pointed out above, there are two efficient ways of ligating Abl-SH3 to the 3BP2 sensor peptide, N-terminus Abl-SH3 to N-terminus 3BP2 sensor peptide and C-terminus Abl-SH3 to N-terminus 3BP2 sensor peptide. The first product was generated via a purely synthetic thioether ligation strategy, and the second ligation product was prepared using the protein semi-synthesis strategy EPL.

Synthesis of Abl-SH3-3BP2 FRET Biosensor Using a Thioether Strategy

The $^\alpha$amino group of a peptide or protein has a significantly lower pKa (6.8-7.9) than side-chain amine groups, thus it can be chemoselectively alkylated and acylated. Pioneering work done by Wetzel and co-workers showed that at pH 6.5 peptides can be iodoacetylated specifically at the N-terminus *(Wetzel, 1990)*. The mechanism of a thioether ligation requires the presence of a nucleophilic thiol, which reacts with an electrophilic acetylated peptide (Figure 1-4). As illustrated in Figure 3.1-4A, the synthetic strategy called for the chemoselective thioether ligation between the N-iodoacteylated 3BP2 sensor peptide of sequence IAc-(Gly)11-3BP2 and a synthetic Abl-

SH3 mutant, containing a unique cysteine residue at the N-terminus (Cys-Abl-SH3 (**2**)). Both the 3BP2 sensor peptide (**1**) and synthetic Cys Abl-SH3 (**2**) constructs were synthesized using the *in situ* neutralization/HBTU activation protocol for Boc SPPS *(Schnolzer, 1992)*. The 3BP2 sensor peptide (**1**) was then idodactelyated in solution at the N-terminus using iodoacetic acid in tetra-hydrofuran (THF) at pH 6.3. The N-iodoacetylated 3BP2 peptide (IAc-3BP2) (**3**) and the synthetic Cys-Abl-SH3 (**2**) products were purified and characterized before ligation.
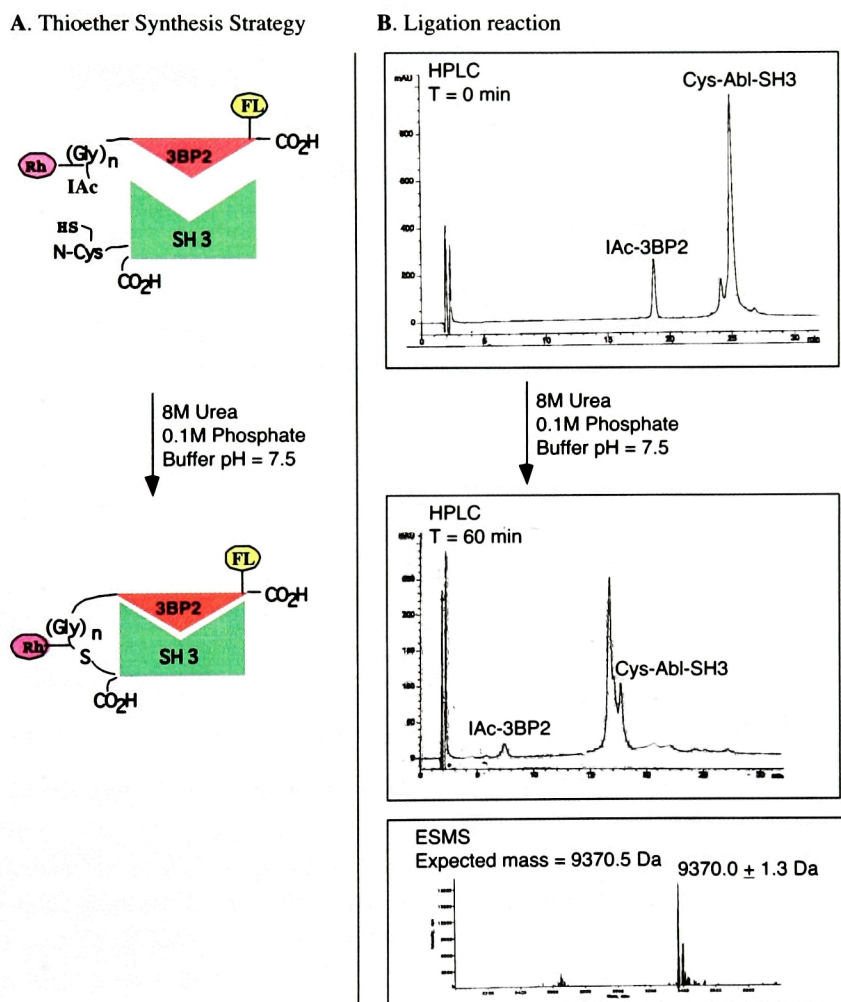
Figure 3.1-4. Synthesis and characterization of Abl-SH3-3BP2 FRET biosensor prepared by a thioether ligation approach (4). (A) Thioether synthesis strategy. **(B)** RP- HPLC and ESMS characterization of thioether Abl-SH3-3BP2 FRET biosensor. In the HPLC traces of the crude ligation reaction (**4**) indicates the ligation product. The HPLC gradient at T= 0 min is 15-55%B, while that at T = 60minis 30-55%B, both over 30 min. ESMS characterization of the ligation product (**4**) by ESMS gave the correct mass [observed mass = 9370 ± 0.8; expected average mass = 9368].

High concentrations of Cys-Abl-SH3 (**2**) and IAc-3BP2 (**3**) sensor peptide were ligated in 8M urea, 0.1M phosphate buffer, pH 7.5 for 1 hour. The thioether ligation product was then purified by preparative HPLC and characterized by electrospray mass spectrometry (ESMS). Figure 3.1-4B depicts the HPLC analysis of the ligation from time t = 0 minute to completion at t = 60 minutes, and ESMS characterization of the thioether ligation product.
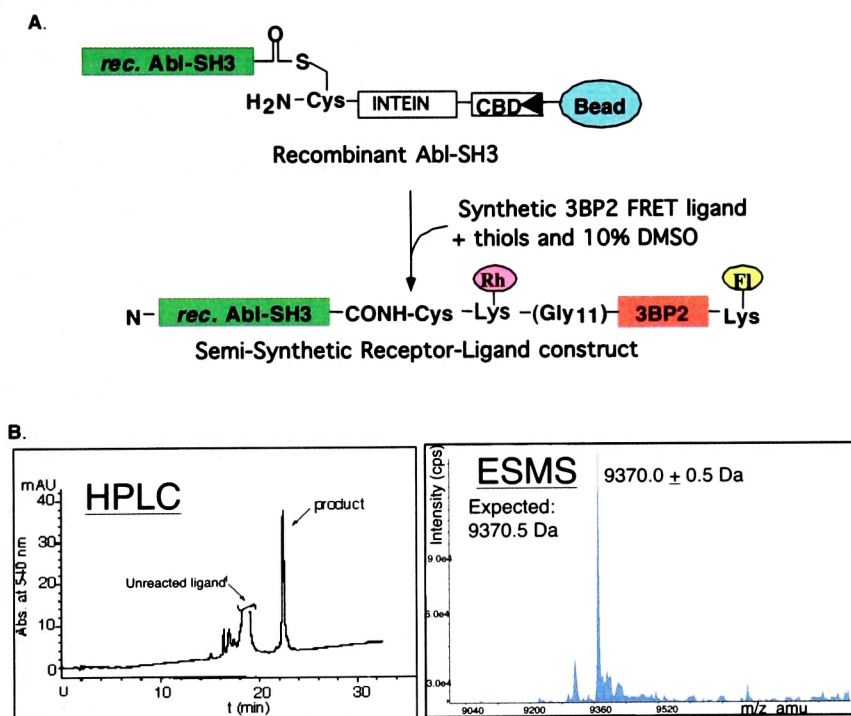
Figure 3.1-5. Synthesis and characterization of Abl-SH3-3BP2 FRET biosensor prepared by EPL. (A) EPL synthesis strategy. **(B)** HPLC and ESMS characterization of EPL Abl-SH3-3BP2 FRET biosensor (**7**). Addition of 10% DMSO to EPL ligation reaction resulted in formation of the semi-synthetic receptor-ligand construct, whose mass was confirmed by ESMS. A linear gradient of 0-73% over 30 min was employed in the HPLC of the crude ligation reaction.

Synthesis of Abl-SH3-3BP2 FRET biosensor using EPL

The protein semi-synthesis strategy Expressed Protein Ligation was used to ligate the C-terminus of Abl-SH3 to the N-terminus of the 3BP2 sensor peptide. As depicted in Figure 3.1-5A, Abl-SH3 is expressed recombinantly as a *Sce* VMA-intein-CBD fusion protein (**5**). The fusion protein (**5**) is then reacted with an N-terminal cysteine version of

the 3BP2 sensor peptide, Cys-3BP2 (**6**), in the presence of thiophenol, to generate the semi-synthetic Abl-SH3-3BP2 FRET biosensor (**7**). This procedure is depicted by the mechanism of EPL shown in Figure 1-5C and was discussed above.

Analysis of the two Abl-SH3-3BP2 biosensor constructs

Having successfully synthesized both Abl-SH3-3BP2 biosensor constructs, the next step was analyzing the structural and functional properties of the purified molecules. A series of circular diochroism (CD) spectroscopy experiments were conducted to determine if the ligation products were folded. The thioether ligation product had a CD characteristic of an Abl-SH3-3BP2 receptor ligand complex (**4**) (Figure 3.1-6) *(Viguera et al., 1994)*. In contrast, the CD of the EPL product could not be determined, due to the insoluble nature of the construct. While it is not clear why the EPL construct behaved differently from the thioether construct, one possibility is that the manner in which the ligation was performed, C-terminus of Abl-SH3 to N-terminus of 3BP2 ligand, affects the folding of the protein.
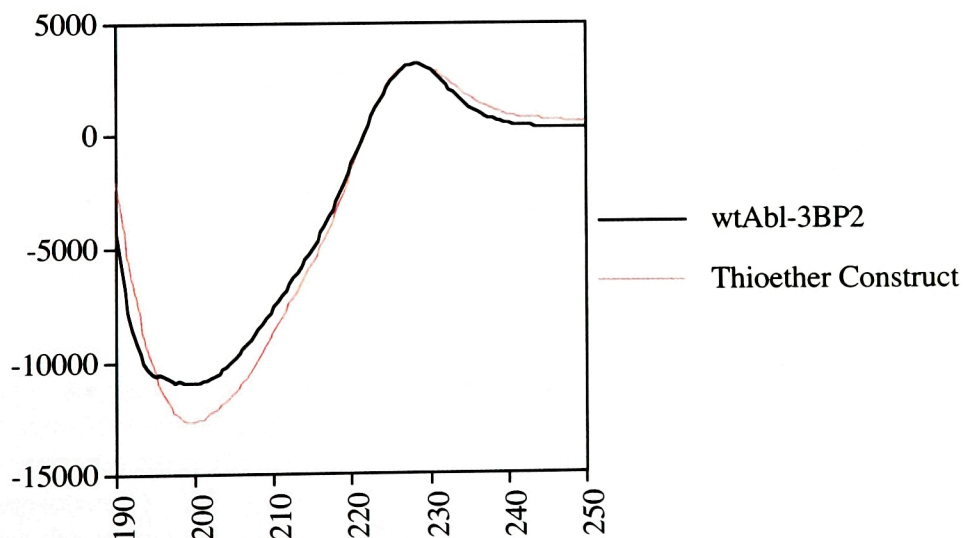
Figure 3.1-6. Far uv Circular Dichroism (CD) analysis of Abl-SH3-3BP2 thioether construct. The spectra of Abl-SH3-3BP2 thioether construct (**4**) as compared to wtAbl-SH3-3BP2: the fold of the thioether construct is similar to that of wildtype.

Encouraged by the CD analysis of the thioether ligation product, a decision was made to proceed with fluorescence spectroscopy analysis. The goal was to monitor the change in FRET between Fl and Rh as the intramolecular association between 3BP2 and Abl-SH3 is displaced by an exogenous competing ligand. To monitor FRET, the thioether construct was excited at 490nm (Fl absorption), and fluorescence emission was monitored at 520nm (Fl emission) and 580nm (Rh emission). In the absence of the competing ligand, FRET was observed between Fl and Rh. This is evidenced by the intense emission at 580nm compared to that at 520nm in Figure 3.1-7A.
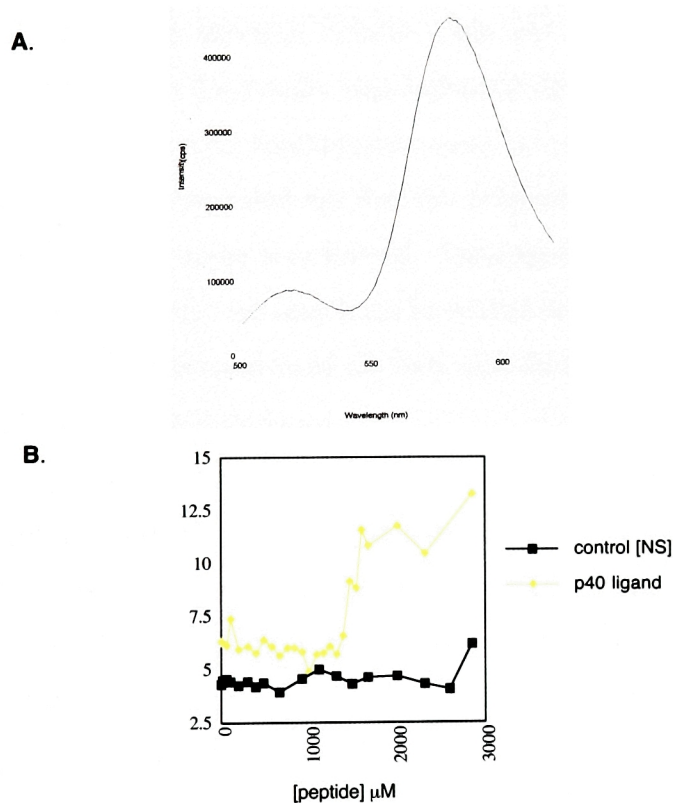
Figure 3.1-7. Analysis of Abl-SH3-3BP2 thioether construct. (A) Fluorescence spectrum of Abl-SH3-3BP2 thioether construct (**4**). Sample was excited at 490 nm (Fl absorption) and fluorescence emission is monitored from 500-600 nm (Fl and Rh emission, respectively). FRET exchange occurs between the FL and Rh fluorophores as evidenced by the intense emission at 580nm compared to that at 520nm. **(B)** FRET analysis of thioether ligation product. Plot of the concentration of competing ligands, P40 (**12**) and NS (**13**), versus 580:520 nm fluorescence intensity. There is an increase in FRET when P40 (**12**) ligand is used, but no change is observed when the non-specific ligand (NS (**13**)) is used.

47

The next step was to titrate a competing ligand into the thioether Abl-SH3-3BP2 biosensor construct (**4**). The titration study was conducted on a small scale, using a non-specific ligand (NS (**13**)) (amino acid sequence, YSPWTNF) that has no affinity for Abl-SH3, and a ligand P40 (**12**) (amino acid sequence, APTYSPPPPP) that is known to bind Abl-SH3 with a K_D of 7.0μM. Initial experiments were encouraging in that a dose dependent increase in the 580:520nm FRET ratio was observed using P40 (**12**) as the competing ligand, but no significant change was observed with NS (**13**) (Figure 3.1-7B). This result is consistent with the original biosensor design. Unfortunately, attempts to reproduce these data were largely unsuccessful, in most cases both the NS (**13**) and P40 (**12**) ligands yielded the same titration results (data not shown). It is certainly conceivable that the variation in the results was linked to differences in the Abl-SH3-3BP2 biosensor preparations (several batches were used in the course of these studies). However, at this stage it cannot be ruled out that the original design of the contiguous receptor-ligand construct was in some way flawed. Most notably, by covalently linking the 3BP2 ligand to its SH3 receptor we may have increased the affinity of the ligand for the receptor, making a tighter association of the two, and thus harder to illicit a change when the competing P40 ligand is introduced.

Although a biosensor for Abl-SH3 was never realized, Abl-SH3 did prove significant to the development of expressed protein ligation. As reported in a paper published in 1999, Abl-SH3 was used extensively to investigate the scope and limitations of EPL *(Ayers et al., 1999)*. Specifically, Abl-SH3 was used to investigate the effects of chaotropic agents on EPL.

Effect of chaotropic agents on EPL

An important feature of all chemical ligation approaches is that they can be performed in the presence of organic solvents or chaotropic agents such as guanidine hydrochloride (GdmCl) or urea *(Muir et al., 1994) (Muir et al., 1997)*. This frequently turns out to be of great practical value since many polypeptide reactants are insoluble in

aqueous buffers at neutral pH, particularly at the millimolar concentrations required for efficient ligation. Since the same solubility problems may be encountered in certain EPL-based syntheses (as was the case with the EPL ligation of Abl-SH3 fusion protein (5) and Cys-3BP2 (6) sensor peptide), it was decided to explore how tolerant the approach was to increasing amounts of GdmCl in the reaction buffer.

The model ligation reaction between Abl-SH3 fusion protein (5) and the synthetic peptide NH2-Cys-Gly-Arg-Gly-Arg-Gly-Arg-Lys[Fl]-CONH2 (8) was carried out in the presence of 0, 2, 3, 4, and 5.5M GdmCl at pH 7.0. The sequence of peptide (8) was designed to be highly soluble at neutral pH, and the fluorescent probe was incorporated to provide a convenient way of monitoring the ligation reactions by RP-HPLC using 490nm detection.

As summarized in Figure 3.1-8A, essentially no difference was observed in the yield of the ligation product of reactants (5) and (7), namely, Abl-SH3-Cys-Gly-Arg-Gly-Arg-Gly-Arg-Lys[Fl]-CONH$_2$ (9), on going from 0 to 2M GdmCl. However, a dramatic reduction in yield was observed on going from 2 to 3M denaturant. Since native chemical ligation reactions are unaffected by the presence of 6M GdmCl *(Muir et al., 1997)*, this result implies that the initial "thioester derivative of Abl-SH3 is not formed at GdmCl concentrations above 2M. This presumably translates to structural perturbation of the *Sce*-VMA intein at higher GdmCl concentrations. Although extremely high concentrations of chaotrope are not tolerated, the observation that EPL can be performed in 2M GdmCl is likely to be of considerable utility, and complements the use of 10% DMSO as a solubilizing agent for EPL.

The utility of 2M GdmCl was illustrated by the ligation of a synthetic 34-mer peptide derived from the C-terminal region of the *Esherichia Coli* σ^{70} specificity factor, NH2-CNTDYLEEVGKQFDVTRERIRQIEAKALRKLRH-CONH2 (10), to Abl-SH3 fusion protein (5) using EPL. Peptide (10) has poor solubility (low µM) in aqueous buffers at neutral pH, but can be solubilized at millimolar concentrations in the presence of 2M GdmCl. Attempts to ligate (10) to (5) in the absence of GdmCl led to no detectable product formation (data not shown), whereas (5) was clearly converted to the desired semi-synthetic protein Abl-SH3-CNTDYLEEVGKQFDVTRERIRQIEAKALRK LRH-CONH$_2$ (11) when 2M GdmCl was included in the ligation buffer (Figure 3.1-8B).

This result reflects the known relationship between high reactant concentrations and chemical ligation efficiency *(Muir et al., 1997)*.

A.

[Gdm.HCl]

B.
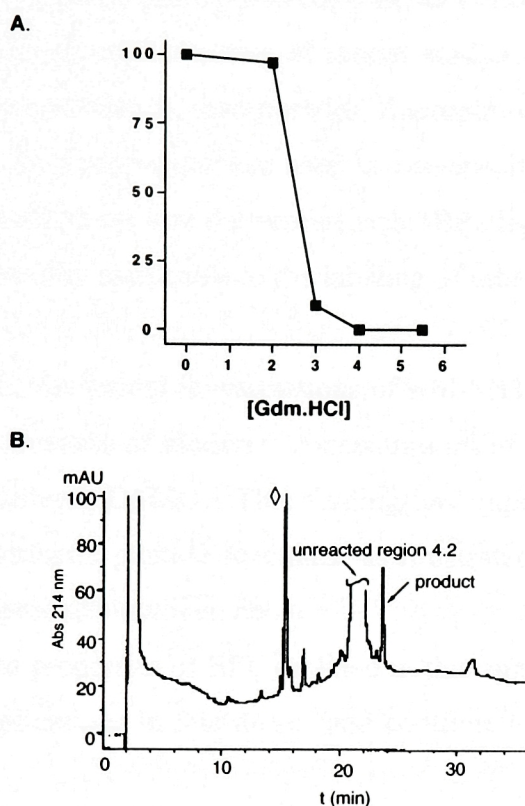
Figure 3.1-8. Effect of guanidine hydrochloride on EPL. (A) Ligation of Abl-SH3-intein fusion protein (**5**) to synthetic peptide (**8**) in the presence of increasing concentrations of GdmCl. The yield of each reaction is represented as a percentage of the total product formed in 0M GdmCl, and was determined by integration of the product peak in the corresponding HPLC chromatogram. **(B)** HPLC analysis of crude ligation mixture obtained from reaction with fusion protein (**5**) and σ^{70} region 4.2 (**10**) in the presence of 2M GdmCl. A linear gradient of 0-73% buffer B over 30min was employed. The ligation product elutes as a single peak at 23.60 minutes. The diamond indicates an unidentified nonpeptidic impurity.

Concluding remarks on the Abl-SH3-3BP2 biosensor - significance to EPL development

In the words of Bruce Merrifield when describing his attempt to synthesize the composite hormones angiotensinylbradykinin and bradykininylbradykinin, "the syntheses

50

were successful and chemically interesting, but the biological results were not as hoped for" *(Merrifield, 1993)*. The same can be said for the synthesis of the thioether and EPL based versions of an Abl-SH3-3BP2 biosensor. Although a contiguous receptor ligand biosensor for the Abl-SH3-3BP2 construct was not realized, the scope and limitations of EPL were explored through a series of model studies. Specifically, a paradigm was developed for site-specifically incorporating fluorophores into proteins using expressed protein ligation. This procedure was used to site-specifically incorporate a fluorescein and rhodamine fluorophore into the proline rich 3BP2 ligand of Abl-SH3. The utility of this procedure is readily applicable to the labeling of other protein constructs (see Section 3.3.).

In addition, the model investigations of Abl-SH3 have revealed that EPL can be performed in the presence of moderate concentrations of the chemical denaturant GdmCl and the organic solvent DMSO. This finding has important implications when using poorly soluble synthetic peptide reactants, as illustrated by the synthesis of the semi-synthetic protein products outlined above.

Finally, the properties of EPL outlined in this application will be used repeatedly in subsequent applications in this thesis, and continue to be used routinely in the Muir lab.

3.2 Site -Specific Incorporation of Post-Translational Modifications into Type I Transforming Growth Factor β Receptor

Part of this work conducted in collaboration with Morgan Huse

Introduction to TGFβ Type I receptor – Significance to EPL development

The synthesis of many proteins is not complete when the protein is released from the ribosome. A large number of proteins are modified after translation and these modifications play an important role in the function of the protein. Post-translational modifications are central to the molecular events controlling cellular signaling networks *(Lee, 2001; Massague, 2000; Nalivaeva, 2001)* and are, in effect, nature's method for conferring functional diversity onto the same translated sequence. It has been difficult, using standard recombinant DNA technology, to produce homogeneous samples of proteins that have defined and specific patterns of post-translational modifications. This inability has hindered biochemical and structural investigations into the role these modifications play in regulating structure and function. Protein chemical ligation could have a major impact in this area. An example of this is the regulation of the transforming growth factor β (TGFβ) receptor family through serine/threonine phosphorylation near the amino terminus of the cytoplasmic kinase domain *(Massague, 1998)*.

The TGFβ signaling pathway plays a prominent role in a variety of cellular functions including modulation of cell growth, extra-cellular matrix production, and cell death via apoptosis *(Massague, 1998)*. The protein growth factor, TGFβ transduces signals via a receptor complex composed of two transmembrane serine-threonine kinases, the TGFβ receptors I and II (TβR-I and TβR-II). Oligomerization of the two receptors induced by TGFβ binding causes TβR-II to hyperphosphorylate TβR-I within a highly conserved N-terminal regulatory region called the GS region, named for the highly conserved [185]TTSGSGSG[192] sequence contained within it. Activated TβR-I then phosphorylates members of the Smad family of transcription factors, which carry the signal to the nucleus (Figure 3.2-1).
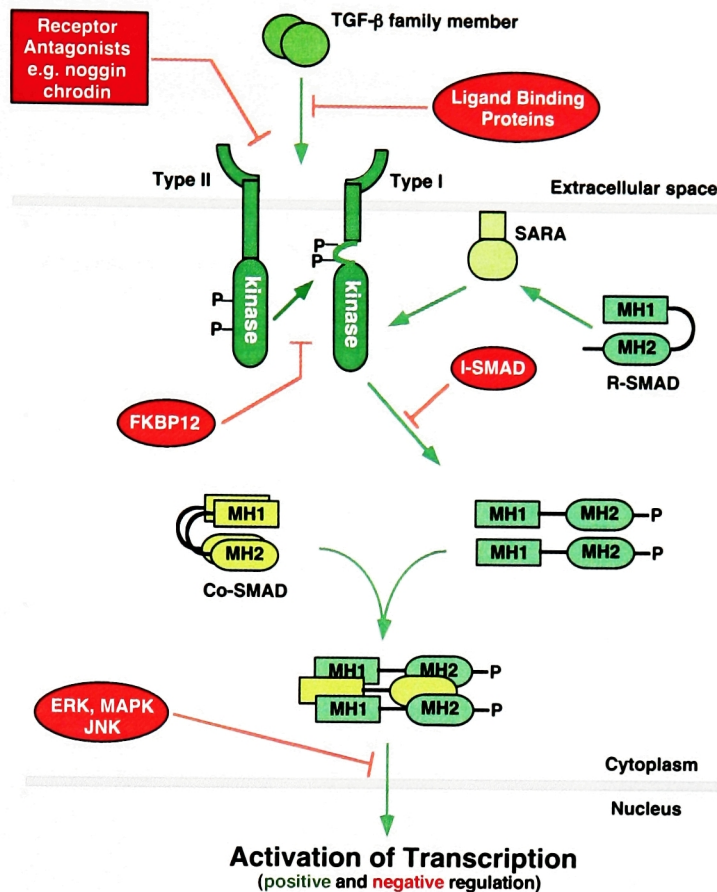
Figure 3.2-1. TGFβ signaling pathway.

Phosphorylation of four to five of the serines and threonines within the [185]TTSGSGSG[192] sequence is thought to be required for full activation of TβR-I signaling *(Wieser, 1995)*. However, the level to which hyperphosphorylation activates the kinase towards its natural substrate remains unknown, and the molecular mechanism of how this post-translational event leads to kinase activation has yet to be resolved. The crystal structure of the cytoplasmic domain of TβR-I (Figure 3.2-2A) *(Huse, 1999)* showed that, in its unphosphorylated state, the GS region adopts an inhibitory conformation that maintains the protein kinase domain in a catalytically inactive conformation. Thus, phosphorylation within the GS region would presumably activate the kinase by disrupting inhibitory interactions (a de-repression model for kinase activation) (Figure 3.2-2B). Mutagenesis data, however, suggest that the situation may be more complex *(Wieser, 1995; Willis, 1996)*. The phosphorylated GS region may also activate the receptor in a
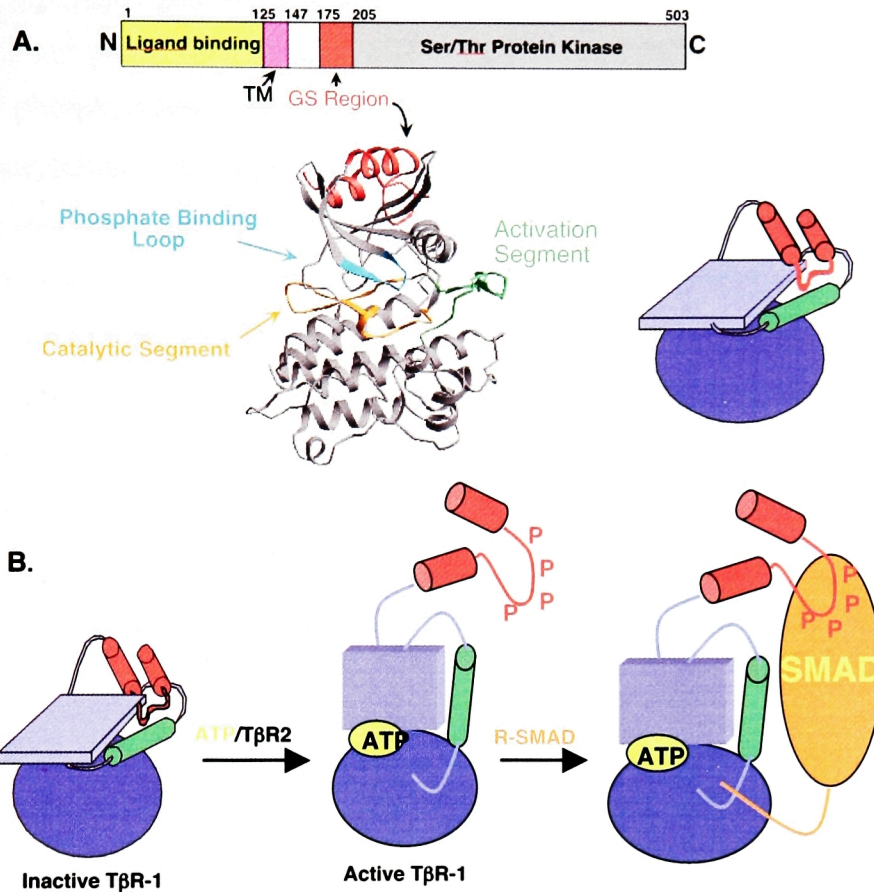
Figure 3.2-2. TβR-1 domain organization and activation model. (A) Overall domain architecture for TβR-1. Red denotes the GS region of TβR-1 and gray the Ser/Thr kinase domain. The crystal structure of the unphosphorylated cytoplasmic Ser/Thr kinase domain is shown adjacent to a schematic representation of it. In both the crystal structure and the schematic representation the GS region (in red) and activation segments (in green) are identified. (B) Model for TβR-1 activation. In this model hyperphoshorylation derepresses the kinase and creates a docking site for the R-SMAD substrate, which gets phosphorylated by TβR-I.

positive way, perhaps by adopting a new structure that complements kinase activity, or by recruiting substrate.

To fully understand the molecular mechanisms underlying signal transduction processes such as TβR-I activation, it is necessary to have access to chemically defined phospho-proteins for biochemical and biophysical analysis. This requirement is usually

unattainable using simple genetics due to the inability to encode for a phospho-amino acid using standard protein expression technologies. Several attempts to produce homogeneously phosphorylated samples of TβR-I by either coexpression with TβR-II, or *in vitro* phosphorylation were unsuccessful (M. Huse, personal communication). The field of protein semi-synthesis offers a solution to this problem in that it allows for the incorporation of a suitably modified synthetic cassette (which would presumably be homogeneous) into the protein. It was decided to exploit the protein semi-synthesis strategy EPL, to join a synthetic phospho-peptide segment to a recombinant TβR-I target protein, ΔTβR-I, producing a full-length homogeneously phosphorylated TβR-I receptor, residues 175-503 (Figure 3.2-3).
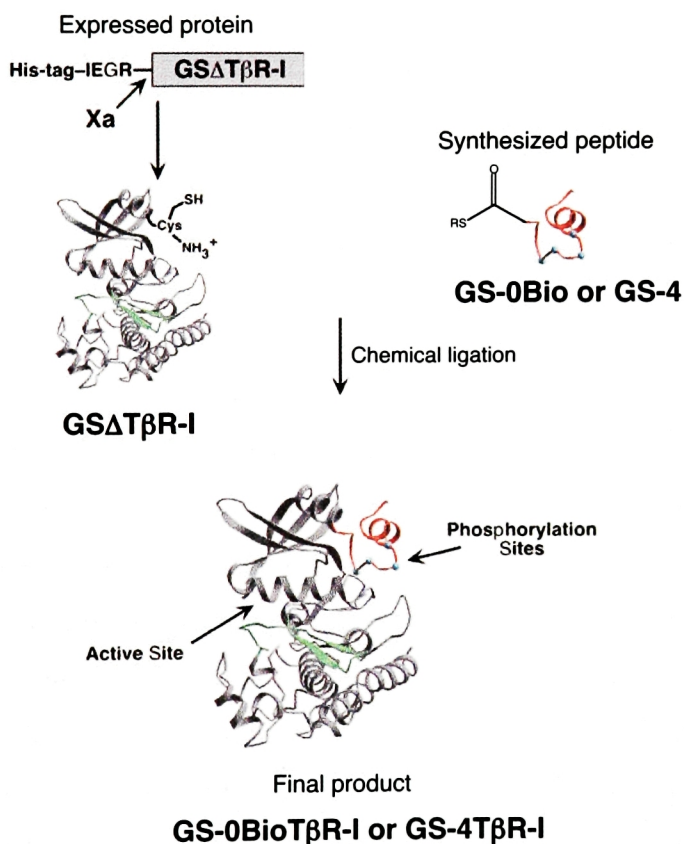
Figure 3.2-3. Protein semi-synthesis of TβR-I. The TβR-I cytoplasmic domain (GSΔTβR-I (**16**)), shown in gray ribbon, was expressed in SF9 cells downstream of a polyhistidine tag, which was later removed by proteolytic cleavage with Factor Xa. GSΔTβR-I (**16**) is then ligated to the synthetic GS peptide (GS-0Bio (**14**) or GS-4 (**15**)). The synthetic peptide is shown in red with the residues phosphorylated in this study depicted as blue spheres in the model. Following ligation the final product (GS-0BioTβR-I (**17**) or GS-4TβR-I (**18**)) is obtained.

55

In section 3.1, expressed protein ligation was defined as a protein semi-synthesis method for ligating synthetic fragments to the C-terminus of recombinant proteins. In the case of TβR-I, the GS region is located near the N-terminus of the kinase domain. Thus, in this case EPL must be used to ligate a synthetic peptide to the N-terminus of the recombinant fragment. Accordingly, this synthetic peptide must contain multi-phosphorylated amino acids with a C-terminal thioester. This work required the development of new chemistry, and signified a broadening of the scope of EPL to include the ligation of synthetic fragments to the N-terminus, as well as the C-terminus, of recombinant proteins *(Huse, 2000)*.

Design and Synthesis of hyperphosphorylated TβR-I

Currently there are two strategies for the synthesis of phosphopeptides; the SPPS approach using protected phosphoamino acids *(Perich, 1991; Perich, 1988)*, and the global phosphorylation approach, which involves post synthetic phosphorylation of unprotected hydroxyl groups following the chain-assembly *(Larsson, 1994; Perich, 1988)*. Due to the instability of phospho-amino acids during the final HF treatment in tert-butoxycarbonyl (Boc) SPPS, a 9-fluorenylmethoxycarbonyl (Fmoc) SPPS approach must be used for synthesis of these peptides. Unfortunately, the Fmoc approach has been incompatible with the generation of ᵅthioesters. This is because treatment of the resin bound peptide with piperidine during Fmoc deprotection at each cycle of peptide synthesis results in aminolysis of the thioester during the chain assembly. However, several advances have been made recently that to varying degrees address this problem. These include the use of modified Fmoc deprotection conditions *(Li, 1998)*, the use of Lewis-acid mediated thiolysis conditions *(Sewing, 2001; Swinnen, 2000)* and the use of specialized resin-linkers such as the backbone-amide linker *(Alsina, 1999)* and the sulfonamide safety-catch linker *(Ingenito. R., 1999; Quaderer, 2001; Shin, 1999)*. The last of these is of particular relevance to the generation of phosphopeptide ᵅthioesters.

The sulfonamide safety–catch linker system was first described in the early 1970's by Kenner and co-workers *(Kenner, 1971)* and was recently refined by Ellman

and colleagues *(Backes, 1996; Backes, 1999)*. In the safety-catch approach, the peptide is synthesized on the sulfonamide-resin using standard Fmoc SPPS protocols. Then, prior to cleavage of the peptide from the resin support, the sulfonamide linker is activated with a haloacetonitrile to form an extremely reactive N-cyanomethylacylsulfonamide, which when further reacted with a nucleophilic species such as a thiol, releases the peptide as a α‑thioester. The aim of this thesis section was to explore the safety-catch linker as a viable method for the generation of phospho-thioester peptides, and to determine if its use in conjunction with EPL could allow for the site-specific phosphorylation of proteins like TβR-I.

Two semi-synthetic TβR-I targets were selected: first, a TβR-I with an unphosphorylated GS region, to be used as control, and second, a TβR-I species in which Thr[185], Ser[187], Ser[189], and Ser[191] of the GS region were phosphorylated. For both TβR-I targets, the GS region, residues 175-195, were synthesized as a peptide α‑thioester (Figure 3.2-4), which was then joined to the remainder of the recombinantly expressed TβR-I cytoplasmic domain (GSΔTβR-I (**16**)) by EPL under physiological conditions.
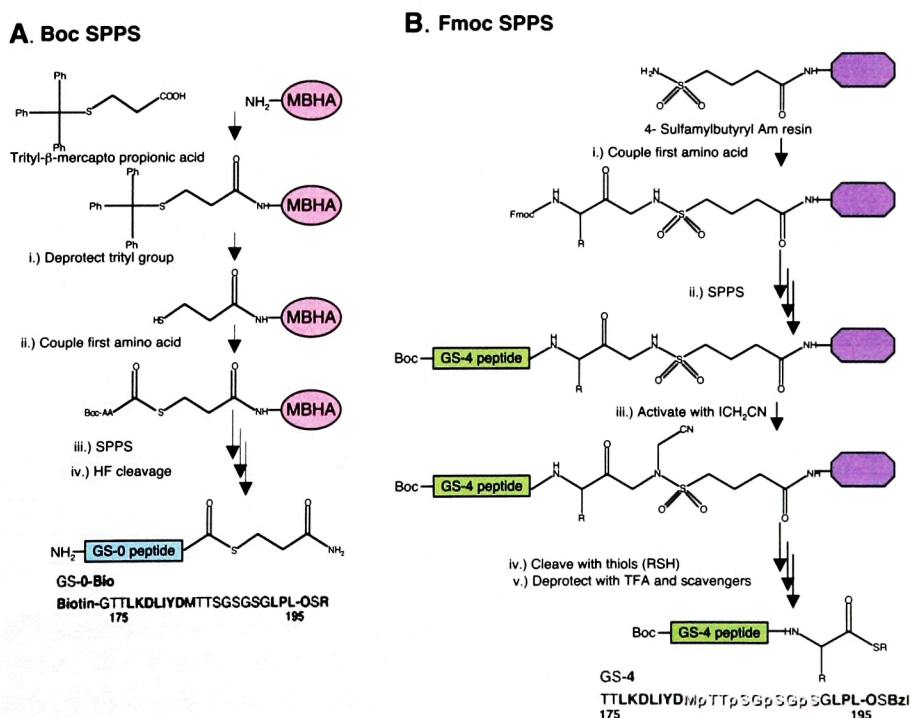
Figure 3.2-4. GS thioester peptide synthesis strategies. **(A)** Boc SPPS of non-phosphorylated GS thioester peptide (GS-0Bio (**14**)). **(B)** Fmoc SPPS of tetra-phosphorylated GS thioester peptide (GS-4 (**15**)).

Synthesis of the unphosphorylated GS thioester peptide

The unphosphorylated GS thioester peptide (GS-0Bio (**14**)) was prepared by Boc SPPS. The α-thioester moiety was introduced using trityl β-mercaptopropionic acid as a handle on an MHBA resin *(Camarero et al., 1998)* (Figure 3.2-4A). A biotin molecule and glycine spacer were included at the N-terminus of GS-0Bio (**14**) to facilitate purification of the ligation product from unreacted GSΔTβR-I (**16**). The peptide was purified by semi-preparative reverse phase HPLC and characterized as GS-0Bio by analytical HPLC and ESMS (Figure 3.2-5A).
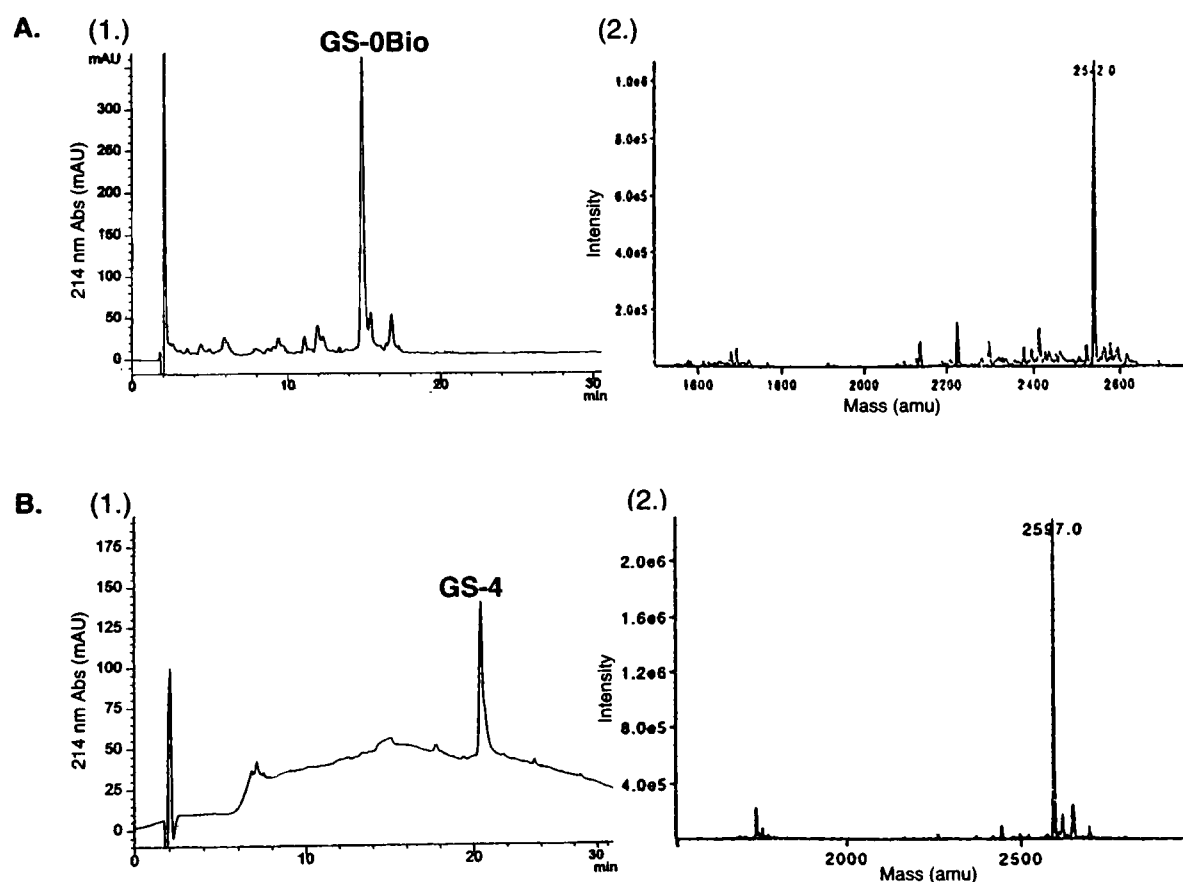
Figure 3.2-5. Characterization of GS peptides. (**A**) Analytical RP-HPLC (1) and ESMS (2) of crude GS-0Bio(**14**) peptide. RP-HPLC gradient is 30-50% B over 30 min. [Observed mass= 2542.0 ± 0.1Da; expected (average isotopic mass) = 2540.7 Da.]. (**B**) Analytical RP-HPLC (1) and ESMS (2) of purified GS-4 (**15**) peptide. RP-HPLC gradient is 30-50% B over 30 min. [Observed mass = 2597.0 ± 0.5 Da; expected (average isotopic mass) = 2596.5 Da].

58

Synthesis of the tetra-phosphorylated GS thioester peptide

The tetra-phosphorylated GS thioester peptide (GS-4 (15)) was made via Fmoc SPPS on the sulfonamide based safety-catch resin, 4-Sulfamylbutyryl Am resin (Figure 3.2-5B). The close apposition of protected phosphate groups during the GS-4 (15) synthesis could lead to slow acylation kinetics during this part of the chain assembly. Accordingly, residues from Met[184] through pSer[191] were double coupled with extended reaction times, and benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate (PyBOP) was used as the activation agent in all phospho-amino acid couplings. Elongation of the peptide continued until Met[184], at which point the synthesis was suspended in order to analyze the quality of the peptide after the addition of the four phosphorylated amino acids.

Following activation with iodoacetonitrile (ICH$_2$CN), a series of trial cleavage reactions were conducted in order to find the optimal conditions for cleaving the peptide from the sulfonamide resin. Specifically, the use of propylamine, ethanethiol and benzylmercaptan, in combination with various bases and organic solvents was investigated. Treatment with benzylmercaptan yielded the best results and thus was used to cleave a portion of the protected peptide from the sulfonamide resin. The cleaved peptide was then deprotected, under optimized conditions, using 95%TFA/p-cresol/H$_2$0 for 3 hours. Analysis of the crude product indicated the expected mass of the tetra-phosphorylated peptide (Figure 3.2-5B). It is worth noting that obtaining a mass of the peptide was not an easy task. Due to the overall negative charge of the peptide, it did not "fly" well (i.e., it did not ionize well) in the standard ESMS. After considerable effort, the mass of the peptide was determined using a Triple Stage Quadrupole (TSQ) mass spectrometer, courtesy of Julio Caesar Padavon and Brian Chait.

Confident about the correct synthesis of the first half of the peptide, the synthesis was completed using standard Fmoc SPPS procedures. The N-terminal Thr[175] was incorporated as an N$^\alpha$ Boc derivative to prevent its amino functionality from attacking the sulfonamide linker during cleavage. (Note, like GS-0Bio (14), GS-4 (15) was initially biotinylated at the N-terminus. Cleavage revealed a contaminating degradation product

59

144 Da less than the expected mass. Peptide mapping studies localized the modification to the N-terminus of the peptide. The contaminant did not bind avidin beads, and it was concluded that the biotin had fragmented at some point during the activation/cleavage process.) Full length GS-4 (15) peptide, as illustrated in figure 3.2-4B, was cleaved from the sulfonamide resin in a manner similar to that described for the truncated Met[184] peptide above, and analyzed by reverse phase HPLC and TSQ-MS.

Similar to the truncated Met[184] peptide, mass spectrometric analysis of the GS-4 (15) was not easy, again presumably because of the four phosphate groups. As a result, GS-4 (15) was ligated to a carrier protein in order to aid the mass analysis. The synthetic Cys-Abl-SH3 (2) construct made in Section 3.1 was used for this purpose. Ligation was preformed using the standard EPL buffer (0.2M sodium phosphate, 0.2M NaCl, pH 7.2) and allowed to proceed overnight. Analysis by ESMS yielded the expected mass of the ligated product (data not shown), which indicated the successful synthesized the GS region of TβR-I, with four site-specific phosphorylation site.

Ligation of GS-0Bio and GS-4 to GSΔTβR-I

TβR-I minus the GS region (GSΔTβR-I (16)), residues 196-503, with Leu[196] mutated to Cys to enable ligation, was expressed in SF9 cells downstream of a polyhistidine tag (Figure 3.2-3). The ligation site was chosen on the basis of analysis of the crystal structure of unphosphorylation TβR-I as well as inspection of the sequence conservation in that portion of the GS region. Following Ni^{2+} affinity purification, the affinity tag was cleaved with the protease Factor Xa to expose Cys[196], and further purified using γ-phosphate conjugated ATP sepharose. Ligation reactions were performed in the presence of 50mM mercaptoethanesulfonic acid (MESNA) as the thiol cofactor. GSΔTβR-I (16) displayed a tendency to aggregate, so ligations were carried out at low protein concentration (0.05-0.1mM) and at low temperature (4^0 C) in order to slow down this aggregation. Do note that denaturants could not used in this system, as their addition led to an irreversible loss in TβR-I kinase activity. Peptide concentrations were kept high, between 0.5 and 1mM, to generate pseudo-first-order reaction kinetics. Unreacted peptide was recovered by gel filtration and could be recycled for multiple ligations. The GS-0Bio (14) ligation reaction was efficient, proceeding to ~70%

completion after 10 hours, as judged by denaturing polyacrylamide gel electrophoresis. The GS-4 ligation was significantly slower, only 60% complete after 24 hours.

After gel filtration, the unphosphorylated TβR-I ligation product (GS-0BioTβR-I (17)) was separated from unreacted GSΔTβR-I (16) using monomeric avidin agarose. The purified GS-0BioTβR-I (17) product was analyzed by analytical HPLC and ESMS (Figure 3.2-6A). The tetra-phosphorylated TβR-I ligation product (GS-4TβR-I (18)) was purified from GSΔTβR-I (16) by anion exchange chromatography. Purified GS-4TβR-I (18) product was analyzed by analytical HPLC and MALDI MS (Figure 3.2-6B). Using these ligation and purification conditions, GS-0BioTβR-I (17) and GS-4TβR-I (18) could be routinely isolated in 10-20% overall yield. These amounts are sufficient for detailed biochemical and biophysical studies.
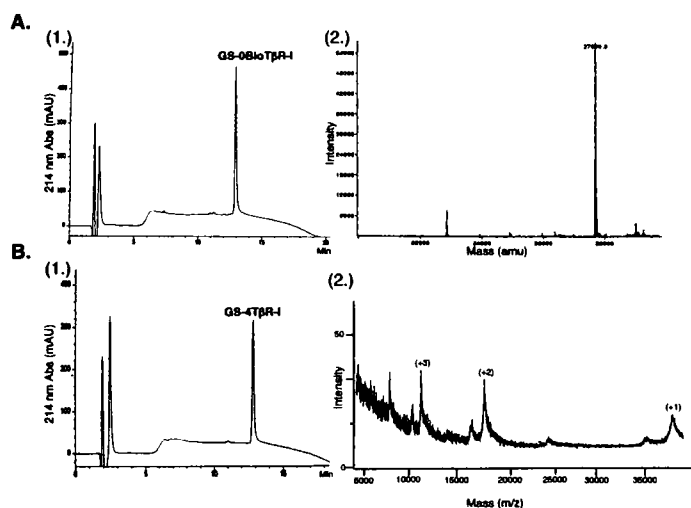
Figure 3.2-6. Characterization of TβR-I constructs. (A) Analytical RP-HPLC (1) and ESMS (2) of pure GS-0BioTbR-I (17). RP-HPLC gradient is 0-100% B over 30min. [Observed mass = 37,698 ± 7.2 Da; expected mass (average isotopic mass) = 37,695.2 Da]. (B) Analytical RP-HPLC (1) and MALDI-TOF MS (2) of pure GS-4TbR-I (18). RP-HPLC gradient is 0-100% B over 30min. [Observed mass = 37731 ± 18 Da; expected mass (average isotopic mass) = 37731.9 Da].

Biochemical analysis of the unphosphorylated and tetra-phosphorylated TβR-I constructs

To characterize the extent of TβR-I activation induced by hyperposhporylation, an *in vitro* kinase assay was performed, comparing GS-4TβR-I (18) and GS-0BioTβR-I (17)

with each other and against GSΔTβR-I (**16**). A slightly larger recombinant fragment of TβR-I, encompassing both the GS region and the kinase domain (termed TβR-Icyt and used in previous crystallographic analysis *(Huse, 1999)*), was also included for comparison. A C-terminal fragment of Smad2 was used as a substrate in this assay. This region, termed the MH2 domain, comprises residues 241-467 of Smad2 and contains both known phosphorylation sites, Ser^{465} and Ser^{467}, in the protein. Thus, this domain can be considered a physiological protein substrate for TβR-I. In a typical kinase reaction, 10pmol of TβR-I were mixed with 80pmol of Smad substrate in 15mL of kinase buffer containing trace amounts of [γ-^{32}P]ATP. Samples were subjected to gel electrophoresis and visualized by autoradiography.
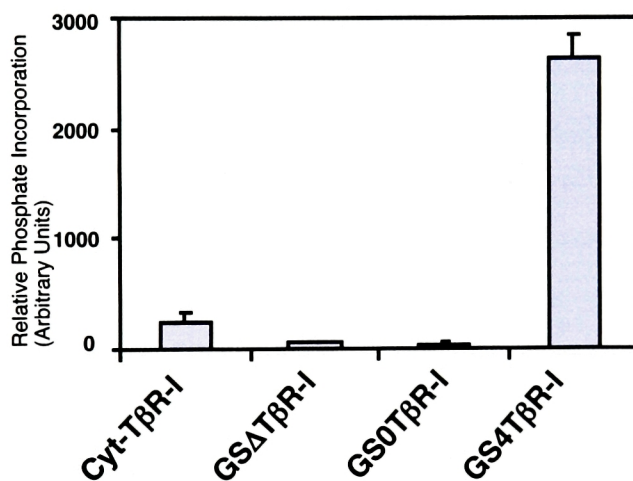
Figure 3.2-7. Phosphorylation of the Smad2 MH2 domain by TβR-I constructs. Smad2 phosphorylation in three independent kinase assays have been combined and quantified. GS4TβR-I (**18**) displays more than a 40-fold increase in kinase activity relative to both GS-0BioTβR-I (**17**) and GSΔTβR-I (**16**).

Under the conditions of this experiment, GS-4TβR-I (**18**) displayed more than a 40-fold increase in kinase activity relative to both GS-0BioTβR-I (**17**) and GSΔTβR-I (**16**) (Figure 3.2-7). GS-0BioTβR-I (**17**) did not demonstrate comparable activation relative to TβR-Icyt and GSΔTβR-I (**16**), indicating that the increase observed for GS-4TβR-I (**18**) is dependent upon phosphorylation, and not a result of the ligation chemistry or the Leu to Cys mutation. Furthermore, the observed increase in GS-4TβR-I (**18**)

kinase activity relative to GSΔTβR-I (16) indicated that the phosphorylated GS region does indeed play an active role in the upregulated state of the kinase, and argues against a simple de-repression model. In this assay, a 3-fold difference in activity is observed between TβR-Icyt and both GSΔTβR-I (16) and GS-0BioTβR-I (17). This difference could not be explained during this study, but we do note that it does not affect our conclusions with regard to GS-4TβR-I (18), which is still activated by a factor of 10 relative to TβR-Icyt.

Concluding remarks on the semi-synthesis of TβR-I – significance to EPL development

EPL has been effectively employed to produce homogeneous chemically defined samples of hyperphosphorylated TβR-I. Access to these molecules allowed the level to which hyperphosphorylation activates the kinase to be quantified for the first time and, in addition, has provided evidence that the phosphorylated GS region may also activate the receptor in a positive way. This work also represents the first example in which a peptide has been ligated to a protein expressed in eukaryotic cells. In subsequent studies performed by Morgan Huse, access to a homogeneous preparation of hyperphosphorylated TβR-I allowed the molecular mechanism of receptor activation to be studied for the first time. Accordingly, phosphorylation was shown to increase the binding affinity of TβR-I for its natural substrate, the Smad2 transcription factor, and decrease the affinity for the inhibitory protein FKBP12. These observations led to a new model of receptor activation in which phosphorylation of the receptor switches the protein from an inhibited state into an activated form capable of binding substrate.

An Achilles' heel of native chemical ligation and thus EPL, that has arguably restricted their application, is the technical difficulty associated with the preparation of synthetic peptide ᵅthioesters. Here it was demonstrated, with the successful synthesis of the tetraphosphorylated GS-4 (15) peptide on the alkanesulfonamide "safety catch" resin, an effective method to combat this problem. With the synthesis of the tetra-phosphorylated TβR-I receptor, it was also demonstrated that the alkanesulfonamide

resin, when used in conjunction with EPL, is a practical route to site-specifically incorporate multiple post-translational modifications into proteins. Previously, EPL had been used to attach a phospho-peptide to the C-terminus of a protein *(Muir et al., 1998)*. Here it is demonstrated that phospho-peptides can be linked to the N-terminus and, by extension, incorporated into the middle of a protein via a sequential ligation strategy *(Cotton et al., 1999)*. This EPL strategy may be readily adapted to study post-translational modifications in other systems.

3.3 Site-Specific Incorporation of UV-Photo Cross-Linkers and Fluorescent Probes into E. Coli Sigma Factor σ^{70}

Part of this work conducted in collaboration with Leonid Minakhin and Julio Camarero.

Introduction to prokaryotic transcription–significance to EPL development

In all organisms the enzyme RNA polymerase (RNAP) catalyzes the transcription of DNA into RNA. In prokaryotes, RNAP is made up of five subunits (α_2, β, β', ω) that together define the core RNAP. On its own, core RNAP cannot initiate transcription, as it cannot specifically bind promoter DNA to begin the synthesis of RNA. For promoter binding, core RNAP requires another subunit termed a sigma specificity factor (σ). When the sigma factor binds to core RNAP it converts the core complex into what is known as the holoenzyme (α_2, β, β', ω and σ), and it is this assembly that is able to initiate transcription *(Burgess, 1969; Burgess, 1970)*. In *Escherichia coli* the primary sigma factor, sigma 70 (σ^{70}), directs transcription of housekeeping genes. It is 70 kDa in mass and is comprised of 613 amino acids. σ^{70} has four conserved regions, each of which maps to a particular function in the transcription initiation process *(Lonetto, 1992; Severinova, 1996)* (Figure 3.3-1). In principle each stage of the transcription process can be regulated, but most regulation occurs at transcription initiation. In the case of sigma factors, regulation can occur either intra-molecularly, through auto-inhibition with σ^{70} region 1.1, or inter-molecularly, through interactions of σ^{70} with regulatory proteins like bacteriophage AsiA.

The bacteriophage T4 antisigma AsiA protein has been shown to inhibit *E. coli* σ^{70} dependent transcription by binding to region 4.2 of σ^{70} *(Severinov and Muir, 1998)* and interfering with its recognition of the –35 promoter element *(Colland, 1998; Severinova, 1998)*. AsiA is a 10 kDa protein expressed in the early stage of bacteriophage infection of *E. coli (Orsini, 1993)*. It binds to σ^{70} and efficiently inhibits
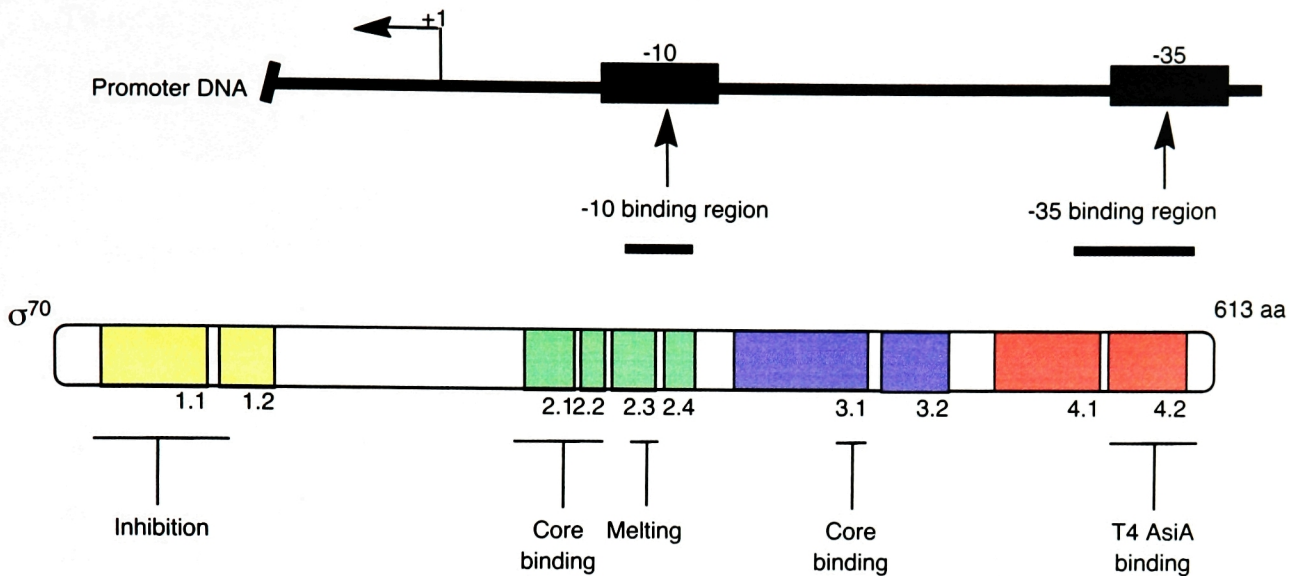
Figure 3.3-1. Domain organization of σ^{70}. The four conserved regions of σ^{70} (regions 1, 2, 3, 4) and their biochemical functions.

transcription from early phage and most *E. coli* σ^{70}-dependent promoters (Figure 3.3-2). At the same time AsiA collaborates with MotA, the product of another early gene, to activate *E. coli* σ^{70}-dependent transcription of viral middle genes *(Ouhammouch, 1994)*. At later stages of infection, the viral σ-like protein gp55, a product of a middle gene, replaces the host σ^{70}-AsiA complex and directs RNAP core to late promoters of the T4 phage *(Kolesky, 1999)*. Thus, AsiA abolishes transcription from most host and early T4 promoters and coactivates transcription of middle viral promoters *(Colland, 1998; Severinova, 1998)*.

Available data do not address whether the inhibitory effect of AsiA is direct (i.e. AsiA interacts directly with amino acids in σ^{70} region 4.2 that would normally interact with bases of the −35 element), or indirect (i.e. AsiA interacts outside the DNA-binding surface of σ region 4.2 and through a conformational change prevents *E. coli* σ^{70} binding to −10/-35 promoters). In the absence of core RNAP, σ^{70} region 1.1 is believed to bind directly to σ^{70} region 4.2, preventing region 4.2 from binding to the −35 promoter element, and thus inhibiting σ initiation of transcription *(Dombroski, 1992) (Dombroski, 1993)*. Recent work in the Muir lab used a combination of heteronuclear NMR spectro-
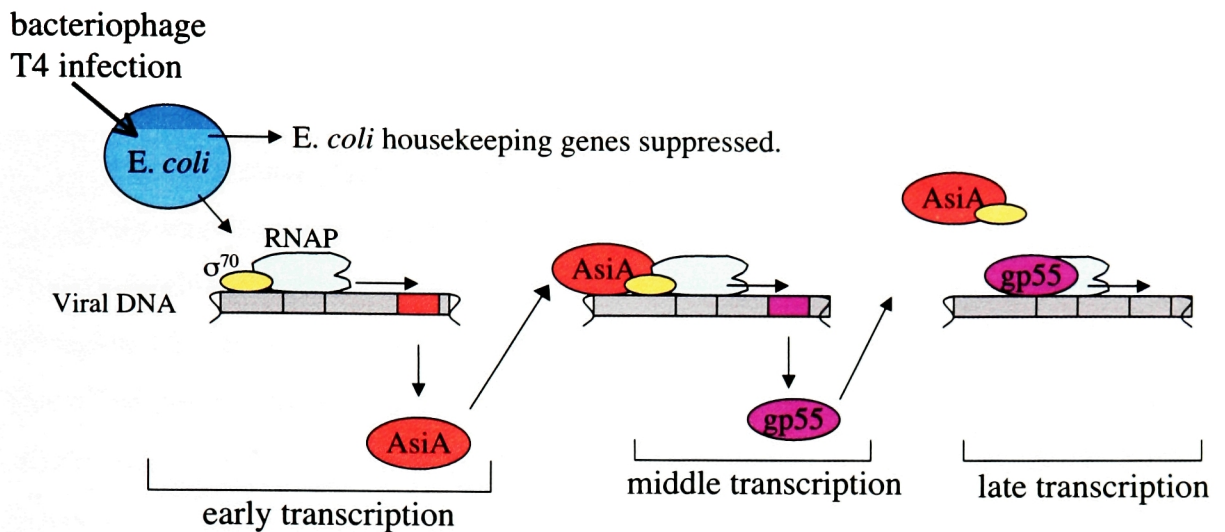
Figure 3.3-2. Bacteriophage T4 antisigma (AsiA) inhibition of σ^{70} dependent transcription. AsiA, which is expressed during early stage of bacteriophage infection of E. coli, binds to σ^{70} and inhibits σ^{70} dependent transcription. While inhibiting σ^{70} transcription, AsiA (with the aid of co-activators) also facilitates the transcription of alternative sigmas in middle transcription.

scopy and segmental labeling techniques to study the putative interaction between regions 1.1 and 4.2 of a σ^{70}-like subunit from *T. maritima* σ^{A} *(Camarero, 2002)*. Initial results of this study do not support the idea of a direct interaction of regions 1.1 and 4.2, but do suggest that region 1.1 can block the binding of region 4.2 to specific targets like the -35 promoter DNA sequence or bacteriophage AsiA. Specifically, addition of AsiA to truncated σ^{A} lacking region 1.1 resulted in dramatic chemical shift perturbations within labeled region 4.2. No such changes were observed when the same amount of AsiA was added to full-length σ^{A}. We were interested in exploring whether the inhibitory effect of region 1.1 also occurs in σ^{70}, the natural binding partner of AsiA. Also, we wanted to quantify the putative effect of region 1.1 on the σ^{70}-AsiA interaction.

Characterization of σ^{70} inhibition at a molecular level is a prerequisite for fully understanding the mechanism and regulation of *E.coli* transcription initiation. The site-specific incorporation of biophysical and biochemical probes into σ^{70} would aid the study of these complicated interactions. Using chemistry-driven approaches, such as expressed protein ligation, to site-specifically incorporate biophysical and biochemical probes into σ^{70}, it has been possibly to map the σ^{70}-AsiA interaction, determine that AsiA inhibition of σ^{70} is indirect, and provide the constructs for investigating whether region 1.1 influences the dissociation constant (K_d) of the σ^{70}-AsiA interaction.

AsiA inhibition of σ^{70} dependent transcription was addressed using a combination of genetic, biochemical, synthetic and combinatorial approaches. This has allowed us to extensively map the molecular interface between σ^{70} and AsiA. The results indicate that T4 AsiA inhibits *E. coli* σ^{70} transcription in an indirect manner, binding to amino acids in σ^{70} region 4.2 that are not involved in recognition of the -35 promoter *(Minakhin, 2001)*. The σ^{70} region 4.2-AsiA interface is extensive, with multiple amino acids of region 4.2 contributing to the interaction. In AsiA, however, the interaction site appears to be limited to the first 20 amino acids.

To examine if region 1.1 influences the K_d of the σ^{70}-AsiA interaction, EPL was used to site-specifically label several versions of σ^{70} with fluorophores for use in a fluorescence binding assay. The K_d of the σ^{70}-AsiA interaction cannot be determined by NMR experiments due to the sensitivity limits of the technique. For these reasons a fluorescence-based approach was sought. The constructs for the assay have been successfully synthesized and fully characterized, but, given the unanticipated difficulties in their synthesis (outlined below), they have not yet been applied in a fluorescence assay. However, the application of EPL to the semi-synthesis of σ^{70} demonstrated, once again, the way in which probes can be site-specifically incorporated into proteins using this method. Also, due to the insoluble nature of region 4.2, several chaotropic and organic solvents, other than and including GdmCl and DMSO, were used during EPL ligation reactions, serving as a model system to further outline the scope of this technique.

Mapping the molecular interface between σ^{70} and AsiA

As mentioned above, a tour de force of scientific techniques ranging from purely synthetic to purely biological was used to extensively map the interaction between σ^{70} and AsiA. Specifically, three different approaches were applied: 1. Photo cross-linking of region 4.2 to AsiA, where six versions of region 4.2 peptide were synthesized, each containing a site-specific photo-reactive benzophenone cross-linker. 2. Combinatorial

chemistry, where the novel technique, Encoded Amino Acid Scanning *(Camarero, 1998)* was employed to generate an alanine scanning library of region 4.2. 3. Mutagenesis of the AsiA protein, to confirm the results from the two synthetic methods. What follows below is a detailed description of the design, synthesis and analysis of σ^{70} region 4.2 benzophenone cross-linker peptides (my contributions to this work), a summary of the other two approaches, and a discussion of the results that were published *(Minakhin, 2001)*.

Design and Synthesis of σ^{70} region 4.2 benzophenone cross-linker peptides

Region 4.2 of *E. coli* σ^{70} is both necessary and sufficient to bind T4 AsiA. This was previously determined using EPL *(Severinov and Muir, 1998)*. Specifically, a semisynthetic version of σ^{70} was found to bind AsiA in a Ni^{2+}-NTA-agarose co-immobilization assay, whereas a fragment of the protein lacking region 4.2 did not. Moreover, a synthetic histidine-tagged σ^{70} fragment corresponding to region 4.2 only, was also able to bind wild type AsiA, demonstrating the interaction of these proteins. Therefore, to characterize the molecular contacts between σ^{70} and AsiA, several region 4.2 polypeptides were designed and synthesized containing the photoactivatable cross-linker, benzophenone. Benzophenone was chosen for several reasons: It is activated at a wavelength that will avoid damaging proteins (350-360 nm); it does not decompose into an inactive compound if cross-linking is not achieved *(Hermanson, 1996)*; finally, it is relatively small (277amu) and can be attached to the peptide backbone via a short spacer, thus limiting the distance between the cross-linker and the protein of interest. When benzophenone is activated with uv light it becomes a very reactive ketone radical intermediate that can react with C-H or N-H bonds to form stable covalent cross-links. If no covalent links are formed, the excited benzophenone species relaxes to the ground state, maintaining its binding and photoactivatable properties, and can then be recycled until a covalent modification is achieved. Thus, a benzophenone molecule has more than one chance to form a stable cross-link with the target protein.

At the time this study was conducted, the structure of region 4.2 of *E. coli* σ^{70} was unknown, however it was speculated to be similar to that of many other DNA binding

proteins, namely a helix-turn-helix (HTH) motif *(Helmann, 1988; Kim, 1995)*. Thus, in deciding where to site-specifically incorporate the benzophenone cross-linker, a secondary structure prediction analysis was performed using the structure prediction program PHDsec. Also the sequence of σ^{70} region 4.2 was traced through homologous HTH proteins using the HOMOLOGY program in the InsightII package. In performing these structure prediction analyses, the goal was to predict the residues in region 4.2 that could potentially participate in informative cross-links; namely, those residues that are oriented toward the solvent phase and away from the hydrophobic core of the protein. Using this approach, six residues were identified that spanned region 4.2 and were predicted to be solvent exposed on the surface of the HTH motif. These residues are not conserved in other σ factors, and are thus less likely to be required for the proper function of σ^{70}. The six residues are: Tyr^{571}(**20**), Lys^{578}(**21**), Arg^{586}(**22**), Arg^{588}(**23**), Lys^{593}(**24**), and Lys^{597}(**25**). These correspond to a benzophenone cross-linker at every 5.5 residues in region 4.2 (Figure 3.3-3).
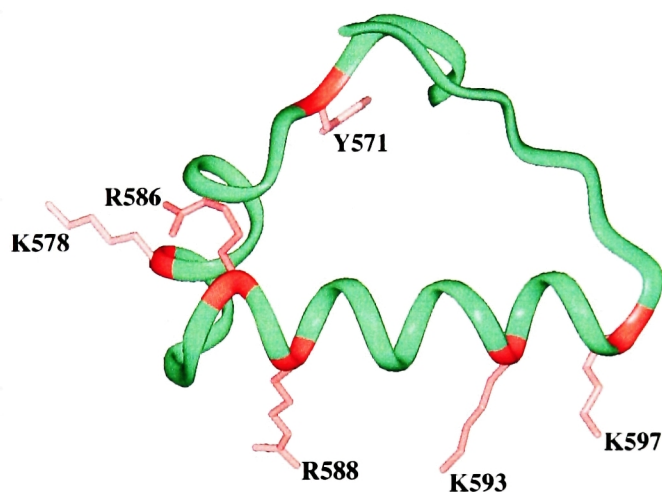
Figure 3.3-3. Homology model of Region 4.2. The six residues in region 4.2 that will be replaced are indicated.

Using optimized Boc-SPPS *(Schnolzer, 1992)*, a wild-type version of σ^{70} region 4.2 (**19**) and six synthetic region 4.2 analogues (**20-25**), each containing a benzophenone cross-linker in the six positions identified above were synthesized. The six synthetic peptide analogues each contained a single amino acid substitution where the wild-type residue was replaced by L-diaminopropionic acid (DAPA). The benzophenone cross-

N-[568]CNTDYTLEEVGKQFDVTRERIRQIEAKALRKLRH[600]-CONH$_2$

$\downarrow$ SPPS

NH$_2$～～～(MBHA)

$\downarrow$ Activated Boc-Dapa(Fmoc)-OH

NH-Fmoc
|
Boc-NH-Dapa～～～(MBHA)

$\downarrow\downarrow$ SPPS

MEB   NH-Fmoc
| |
Boc-HN-Cys～～～CONH-Dapa～～～(MBHA)

i) Deprotect Fmoc
ii) Add benzoylbenzoic acid
In form of anhydride

MEB
|
Boc-HN-Cys～～～CONH-Dapa～～～(MBHA)
                    NH

$\downarrow$ HF cleavage

SH
|
H$_2$N-Cys～～～CONH-Dapa～～～CONH$_2$
                    NH

Figure 3.3- 4. Introduction of benzophenone into σ[70] region 4.2 peptide by SPPS.

linker was then introduced via an orthogonal protection strategy, which allowed the β-amino group of the DAPA residue to be acylated with benzoylbenzoic acid. Figure 3.3-4 depicts the SPPS procedure. ESMS and analytical HPLC confirmed the composition and purity of the peptides (Table 1 for ESMS analysis).

71

Table 1. The wildtype and modified σ^{70} region 4.2 peptides synthesized are indicated along with their respective mass.

Region 4.2 Peptide	Expected Mass (Da)	Observed Mass (Da)
wt4.2 (**19**)	4116.2	4115.2 $\pm$ 0.5
Tyr^{571} (**20**)	4248.0	4248.2 $\pm$ 1.4
Lys^{578} (**21**)	4284.0	4284.8 $\pm$ 1.3
Arg^{586} (**22**)	4255.0	4255.9 $\pm$ 0.6
Arg^{588} (**23**)	4255.0	4254.7 $\pm$ 0.7
Lys^{593} (**24**)	4284.0	4283.8 $\pm$ 0.6
Lys^{597} (**25**)	4284.0	4284.8 $\pm$ 1.3

Cross-linking analysis of the region 4.2 benzophenone constructs

The purified σ^{70} region 4.2 peptide analogues were combined with recombinant AsiA and irradiated with high intensity 355 nm laser light. After irradiation, the samples were analyzed by SDS-PAGE (Figure 3.3-5). A new band with increased electrophoretic mobility corresponding to a molecular weight of ~14 kDa appeared in cross-linking reactions involving all of the σ^{70} region 4.2 peptides. For example, a cross-linked band was observed for the reaction involving AsiA and σ^{70} region 4.2 derivatized at the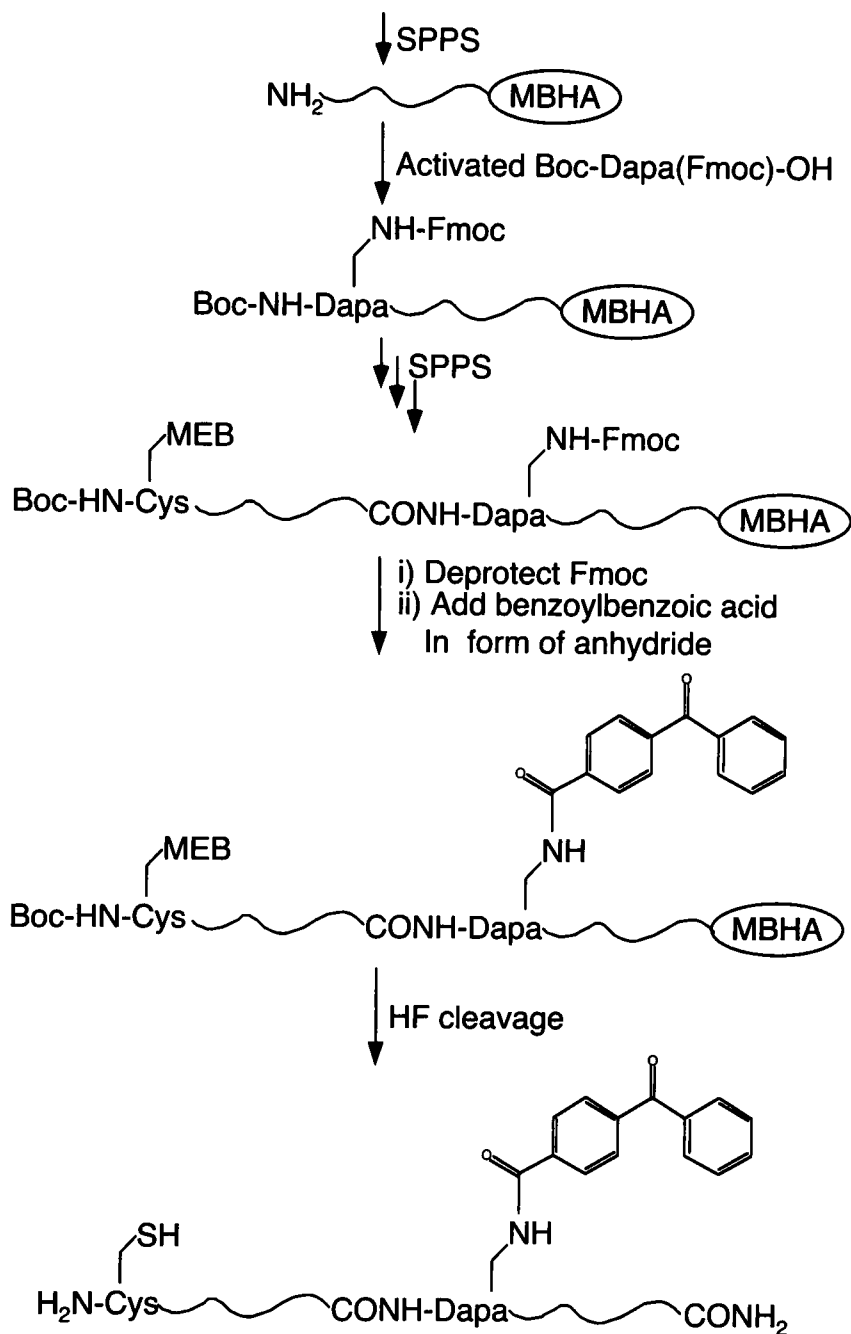 position corresponding to Lys^{593} (Figure 3.3-5A, lane 2). This band was absent from control reactions, that either contained AsiA or σ^{70} region 4.2 peptide alone, or contained both interaction partners but were not irradiated with laser light (lanes 1, 3, and 4, correspondingly). It was concluded that the low mobility band is the result of a protein-protein cross-link between the derivatized region 4.2 peptide and AsiA. This experiment was repeated for all six benzophenone region 4.2 peptide analogues and in all cases the results were the same. Representative data from these experiments where σ^{70} region 4.2 peptides derivatized at the positions corresponding to Lys^{578}, Arg^{588}, and Lys^{597} are shown (Figure 3.3-5B, lanes 10, 12, and 14).

The protein-protein cross-link between the derivatized σ^{70} region 4.2 peptides and AsiA was specific, since no cross-link was observed when reaction mixtures contained

myoglobin instead of AsiA (Figure 3.3-5A, lane 7); no cross-link was observed when underivatized σ^{70} peptide (**20**) was used with AsiA (Figure 3.3-5A, lane 5); and addition of excess of wild-type σ^{70} region 4.2 peptide (**20**) decreased the cross-link yield (Figure 3.3-5A, lane 6). Thus, it was concluded that the interface of the σ^{70}-AsiA complex is extensive, with spatially separated derivatized region 4.2 amino acids either in direct contact, or close to AsiA in the complex.
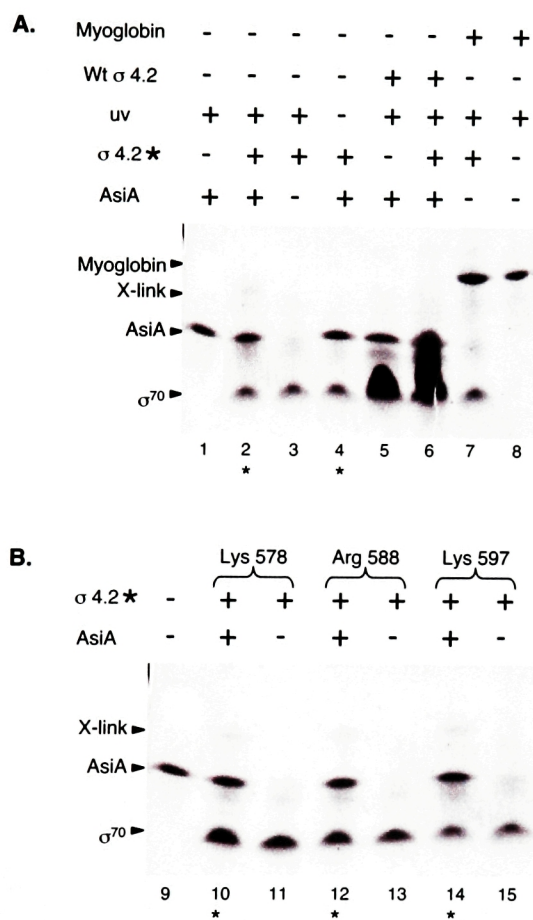
Figure3.3-5. **Protein-protein cross-linking between derivatized σ^{70} region 4.2 peptides and AsiA.** (A) σ^{70} region 4.2 peptide containing benzophenone probe incorporated at position Lys[593] was mixed with the indicated proteins in 20 μl of transcription buffer. Reactions were irradiated with 355 nm light from a laser source, and reaction products were resolved by SDS-PAGE and visualized by Coomassie staining. Lane 2 shows a high running band that corresponds to cross-linked product. This product is not present when derivatized region 4.2 peptide and AsiA are combined in the absence of uv light, see lane 4. Thus the high running band in lane 2 must be due to cross-linking of the two constructs. (B) Repeat of experiment in (A), this time using region 4.2 peptides derivatized at Lys[578], Arg[588], Lys[597]. All derivatives peptides gave the same results as Lys[593], i.e. protein-protein crosslinking is observed only when the peptides are combined with AsiA in the presence of uv light, see lanes 10, 12, and 14.

Combinatorial library and Mutagenesis results

The result described above, namely, that the σ^{70} region 4.2-AsiA interface involves interactions of multiple residues rather than a single key residue, was confirmed using a synthetic alanine scanning library of σ^{70} region 4.2. Using the encoded amino acid scanning technique *(Camarero, 1998)*, every residue in region 4.2 (residues 569-599) was mutated to an alanine residue and the ability of the mutant to bind AsiA was tested using an AsiA affinity column. Alanine substitutions at 6 positions corresponding to σ^{70} residues Thr[569], Val[576], Ile[590], Lys[593], Leu[595] and Arg[596] gave rise to the largest reduction of AsiA binding affinity. With the exception of the Thr[569] and Val[576] substitutions, mutations that affected AsiA binding are clustered together in a short motif near the C-terminus of σ^{70} region 4.2 *(Minakhin, 2001)* (Figure 3.3-6). Residues Lys[593], Leu[595] and Arg[596] have been shown previously to not be involved in the interaction with DNA, but rather appear to participate in protein-protein interactions with several types of transcription activators *(Lonetto, 1998)*.
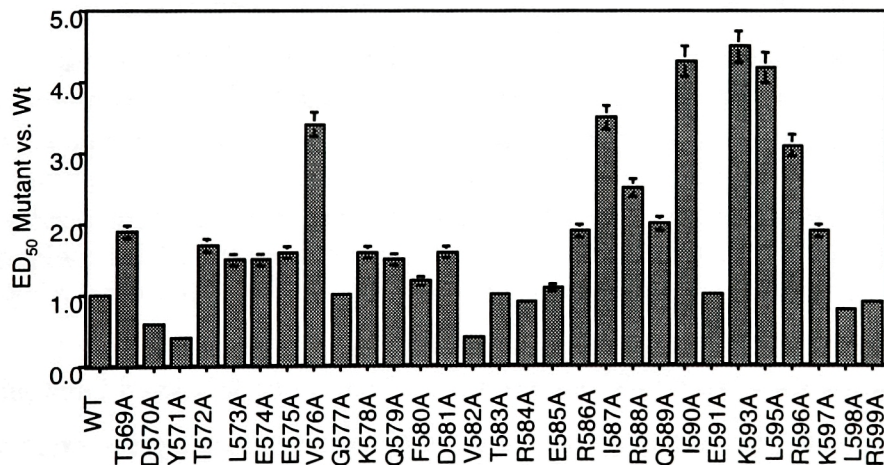
Figure 3.3-6. Relative affinities of the members of the σ^{70} region 4.2 alanine library for binding AsiA to wild type. The ED_{50} values were calculated by competitive elution from the AsiA column with untagged wild-type peptide from two independent experiments. With the exception of the Thr[569] and Val[576] substitutions, mutations that affected AsiA binding are clustered together in a short motif with the C-terminal of region 4.2.

In order to identify the AsiA amino acid residues that interact with σ^{70} region 4.2, five peptides, named Asi1 through Asi5, which correspond to AsiA amino acid residues 1-20, 21-40, 41-60, 61-80, and 71-90 were synthesized. Each of these peptides was tested for the ability to inhibit σ^{70}-dependent transcription (Figure 3.3-7).
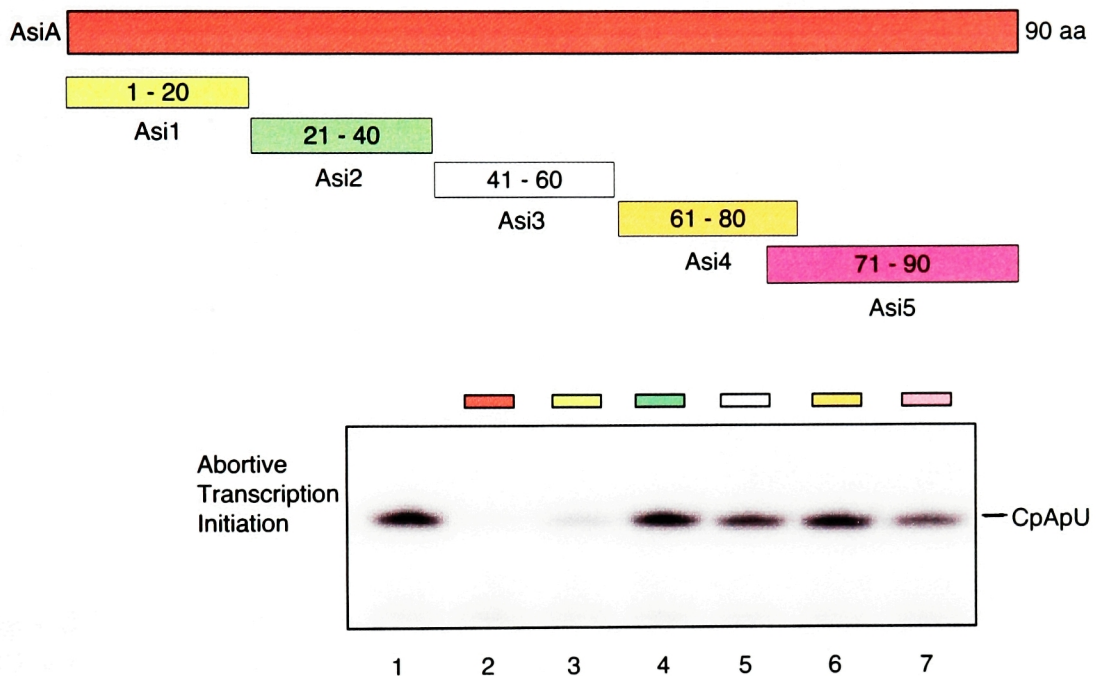
Figure 3.3-7. Inhibition of σ^{70} transcription by AsiA peptides. Abortive transcription initiation was performed in the absence of AsiA (lane 1), in the presence of 0.5 μM AsiA (lane 2), or in the presence of 25 μM of the indicated AsiA peptides using the T7 A1 promoter containing DNA fragments as a template, and CpA and [α-^{32}P]UTP as substrates. Reaction products were resolved on a 20% polyacrylamide gel and visualized by autoradiography. As the gel indicates, it appears as if the first twenty amino acids are necessary for AsiA inhibiton of σ^{70} dependent transcription.

Asi1 inhibited transcription on the T7 A1 promoter, while the other peptides had no effect. To further explore the notion that AsiA interacts with σ^{70} through its N-terminal 20 amino acid residues, three terminal recombinant deletions of cloned AsiA were constructed. The constructs, AsiA$^{\Delta N3}$, AsiA$^{\Delta N5}$, AsiA$^{\Delta N20}$, and AsiA$^{\Delta C20}$ (indicating that the first N-terminal three, five, and twenty residues were removed, and the last C-terminal twenty residues removed, respectively) were purified and their ability to inhibit σ^{70}-dependent transcription from the T7 A1 promoter was investigated. AsiA$^{\Delta N3}$ inhibited abortive transcription at levels that were similar to wild type (Figure 3.3-8,

lanes 1 and 2). AsiA$^{\Delta N5}$ and AsiA$^{\Delta C20}$ resulted in partial inhibition of transcription (Figure 3.3-8, lanes 3 and 5), while AsiA$^{\Delta N20}$ had little effect on transcription (Figure 3.3-8, lane 4). Thus, it was concluded that the first 20 amino acids of AsiA are necessary and sufficient for binding and inhibiting transcription of σ^{70} *(Minakhin, 2001)*.
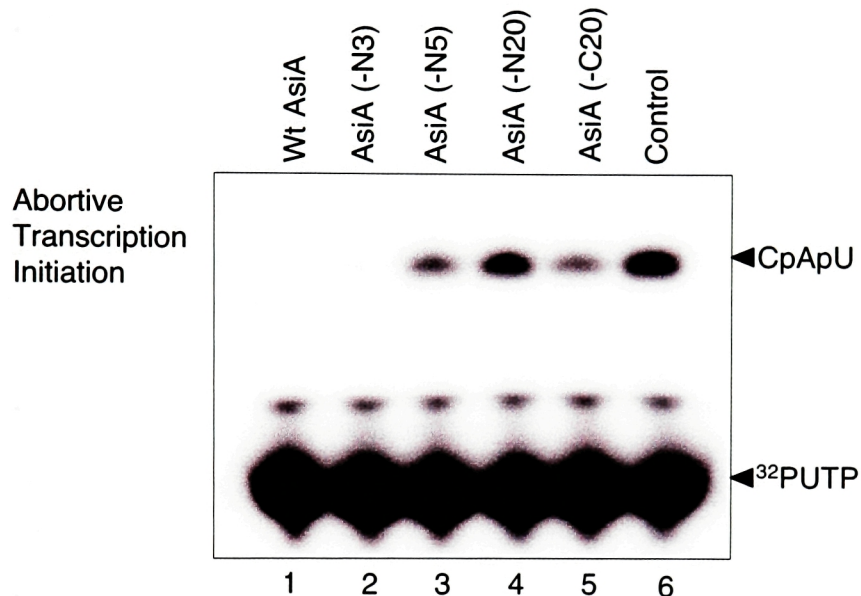
Figure3.3-8. Analysis of recombinant AsiA deletion mutants. Abortive transcription initiation was performed using σ^{70} dependent transcription in the absence of AsiA (lane 6), or in the presence of 5-fold excess of the wild type (lane 1), or indicated mutant derivatives of AsiA (lane2-5). The T7 A1 promoter containing DNA fragment was used as a template, and CpA and [α-^{32}P] UTP as substrates. Reaction products were resolved on a 20 % polyacrylamidegel and visualized by auto radiography. As shown in the gel, the AsiA (−N20) mutant has no affect on transcription initiation (lane 4).

Discussion of experimental results

Combining the results obtained from the UV photo cross-linking experiments, alanine library, and AsiA deletion mutants described above, a model was proposed in which AsiA inhibits σ^{70}-dependent transcription in an indirect manner with several region 4.2 residues contacting AsiA. Recent NMR structure studies of the bacteriophage T4 AsiA protein confirm the finding that C-terminal residues of region 4.2 identified by alanine scanning are involved in binding to AsiA *(Lambert, 2001; Urbauer, 2002)*. However, they suggest that more than the first 20 amino acids of AsiA are involved in binding σ^{70} *(Urbauer, 2001)*. One reason for the discrepancy between these and our results could be that the assays are different. Urbauer and co-workers used an NMR

minimal chemical shift perturbation assay to probe the binding of AsiA to the C-terminus of σ^{70}. In our work, an abortive initiation assay was used to test the function of the AsiA-σ^{70} interaction. An abortive initiation assay measures more than the interaction of AsiA and σ^{70}, it also gauges the ability of this interaction to inhibit transcription. Perhaps, in the context of transcription initiation the σ^{70} interactions with AsiA are shifted by the presence of core RNAP which σ^{70} also binds. Thus it is possible that more than the first 20 residues of AsiA are necessary for binding σ^{70} when σ^{70} is not in the context of core RNAP.

Towards determining the influence of region 1.1 on the σ^{70}–AsiA interaction

Design of the fluorescently labeled σ^{70} constructs

To examine the influence of region 1.1 on the σ^{70}–AsiA interaction, several versions of σ^{70} were synthesized. Using EPL, a strategy was devised to site-specifically label region 4.2 of σ^{70} with a fluorescent probe, then using fluorescence spectroscopy, monitor the change in fluorescence of the probe as increasing concentrations of AsiA protein are titrated. Subsequent analysis of the fluorescence data would yield the K_d of the σ^{70} AsiA interaction. If the inhibitory effect of region 1.1 observed in σ^A by NMR also occurs in σ^{70}, then full length and a truncated version of the protein, missing region 1.1, should yield slightly different K_d values for the σ^{70} AsiA interaction.

To this end, σ^{70} region 4.2 peptides were synthesized with fluorophores site-specifically incorporated at Lys[578]. This residue was selected based on results from the alanine scanning library experiment described above, which indicated that Lys[578] is located outside the binding region of AsiA to σ^{70} (Figure 3.3-6). Two different fluorophores were incorporated at Lys[578]: fluorescein or dansyl. These were chosen because they absorb at wavelengths that will not damage the protein (dansyl = 380 nm, fluorescein = 490 nm), and they have unique features that could be exploited. For example, dansyl intrinsic fluorescence is sensitive to environmental changes, and fluorescein fluorescence has a high quantum yield and the probe can be used in anisotropy measurements.

Two versions of the σ^{70}–VMA–intein-CBD fusion protein were constructed, a recombinant σ^{70} containing the first 566 amino acids (σ^{70}_{1-566}-VMA-intein-CBD (**26**)), and a truncated version of σ^{70} missing region 1.1 ($\sigma^{70}_{101-566}$-VMA-intein-CBD (**27**)). These constructs were then ligated to either the fluorescently labeled σ^{70} region 4.2 peptides or a small model peptide, CGRAKG, labeled with a fluorescent probe. The model peptide was constructed to examine the recent finding that region 4.1 is involved in AsiA binding. Previously, only region 4.2 was believed to be responsible for AsiA binding *(Severinov and Muir, 1998)*, however, it has recently been reported that region 4.1 may also be involved *(Urbauer, 2001)*. The two constructs, σ^{70}_{1-566}-VMA-intein-CBD (**26**) and $\sigma^{70}_{101-566}$-VMA-intein-CBD (**27**), both terminate at region 4.1. In a display of the robust and versatile nature of EPL, the thioester formed during thiol cleavage of the intein fusion proteins was exploited to introduce a small model peptide, CGRAKG, labeled with a fluorescent probe, thus yielding σ^{70}_{1-566} and $\sigma^{70}_{101-566}$ labeled at region 4.1. Applying these 4.1 labeled constructs to the fluorescence assay would enable us to investigate if AsiA does indeed bind to region 4.1 in σ^{70}, and if this binding is affected by region 1.1. Figure 3.3-9 depicts the EPL synthesis strategy for the fluorescently labeled σ^{70} constructs.
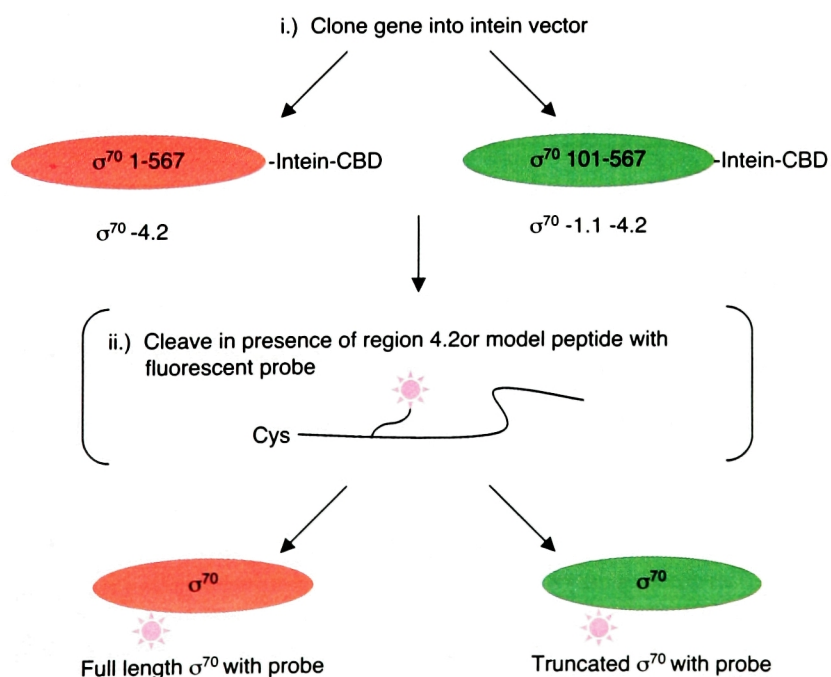

Figure 3.3-9. Synthesis strategy of fluorescently labeled σ^{70} constructs by EPL.

Synthesis of the fluorescently labeled σ^{70} constructs

The σ^{70} region 4.2 peptides labeled site-specifically at Lys[578] with a fluorescein (region 4.2-fl (**28**)) or dansyl (region 4.2-dansyl (**29**)) fluorophore were synthesized in a manner similar to that of the σ^{70} region 4.2 benzophenone peptides described above. In this instance, to attach the fluorescent probe, the DAPA mutant at Lys[578] was acylated with either a fluorescein succinimidyl ester or a dansyl chloride. The small model peptide CGRAKG was synthesized by standard Boc SPPS and acylated at the Lys residue with either a fluorescein (model-fl (**30**)) or dansyl (model-dansyl (**31**)) probe. The composition and purity of all peptides were confirmed by analytical HPLC and ESMS.

Table 2. List of chaotropic and organic solvents used in σ^{70} EPL trial ligation buffers.

Chaotropic and organic solvents used in σ^{70} EPL trial ligation reactions
Guanidine Hydrochloride
Urea
Dimethyl sulfoxide
Acteonitrile/ Water
Triton X detergent

Table 3: MALDI-TOF MS characterization of σ^{70} ligation products

Fluorescently labeled σ^{70} ligation constructs	Expected Mass (Da)	Observed Mass (Da)
σ^{70}_{1-566} -4.2-Fl (**32**)	69178.1	69112 ± 86.6
$\sigma^{70}_{101-566}$ -4.2-Fl (**33**)	58 261.2	58258.2 ± 1.8
σ^{70}_{1-566} -model-Fl (**35**)	65696.1	65730 ± 42.0
$\sigma^{70}_{101-566}$ -model-Fl (**36**)	54799.2	54886.2 ± 97.4

Note: σ^{70} constructs (**34**) and (**37**) do not ionize under MALDI-TOF MS.

In a manner similar to that reported by Severinov and Muir *(Severinov and Muir, 1998)*, standard molecular biology procedures were used to clone the σ^{70}_{1-566} and $\sigma^{70}_{101-566}$ fragments into the commercially available VMA-intein plasmid. Expression of the plasmids in *E.coli* BL21 cells gave the desired fusion proteins, σ^{70}_{1-566}-VMA-intein-CBD

(26) and $\sigma^{70}_{101-566}$-VMA-intein-CBD (27). Following cell lysis, the fusion proteins were purified from the soluble fraction by affinity chromatography on chitin beads.

The four fluorescent peptides, region 4.2-fl (28), region 4.2-dansyl (29), model-fl (30) and model-dansyl (31), were then linked to the recombinant constructs σ^{70}_{1-566}-intein-CBD (26) and $\sigma^{70}_{101-566}$-intein-CBD (27) using EPL. The ligation reactions involving peptides (28) and (29) were complicated by insolubility of these peptides in standard EPL buffer (0.2 M sodium phosphate, 0.2 M NaCl, pH 7.2). Due to this solubility issue it was explored whether EPL could be performed in the presence of different chaotropes and organic solvents (see Table 2 for list of conditions explored). Satisfactory results were eventually obtained by first dissolving peptides (28) or (29) in H_2O/0.1% TFA, then flash diluting the peptide into a slurry of immobilized fusion protein (26) or (27) in EPL buffer containing 3% ethanethiol and 100mM mercaptoethanesulfonic acid. Figure 3.3-10 shows a representative SDS-PAGE analysis of EPL ligation reactions between fusion proteins (26) and (27) and fluorescent peptides (28) and (29), indicating the appearance of bands corresponding to the desired products σ^{70}_{1-566}-4.2-fl (32), $\sigma^{70}_{101-566}$-4.2-fl (33) and $\sigma^{70}_{101-566}$-4.2-dansyl (34).
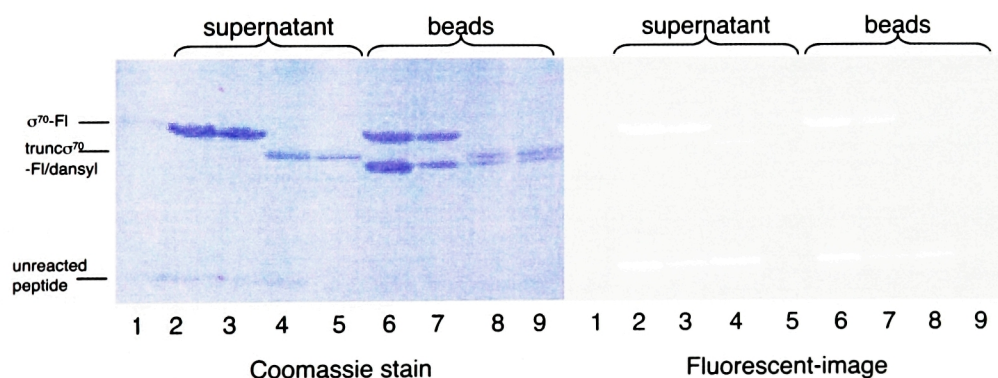
Figure 3.3-10. SDS-PAGE analysis of σ^{70} ligation reactions. 20ml of supernatant and chitin beads from crude EPL ligation reactions of (32), (33) or (34) with (28) or (29) were loaded on to 4-20% polyacrylamide gel and visualized by coomassie stain and fluorescent imaging. The lanes of the gel are as indicated:

Reaction supernatant
1. σ70 wild type.
2. Full length σ70 with a fluorescein probe.
3. Full length σ70 with a fluorecsein probe.
4. σ70 Δ region1.1 with a fluorescein probe.
5. σ70 Δ region 1.1 with a dansyl probe.

Reaction beads
6. Chitin beads of reaction in lane 2.
7. Chitin beads of reaction in lane 3.
8. Chitin beadsof reaction in lane 4.
9. Chitin beads of reaction in lane 5

The gel was visualized by coomassie staining and fluorescent imaging (Figure 3.3-10). Lanes 2-5 represent the ligation supernatant, and lanes 6-9 the chitin beads of the fusion proteins (26) or (27). As shown in lanes 6-9, much of the ligation product remained bound to the chitin beads, however, this product could be removed by washing the beads with standard EPL buffer. HPLC and MALDI-TOF mass spectrometry analysis confirmed the formation of the ligation products (Figure 3.3-11). Similar ligations were conducted with model peptides (30) and (31) to give the desired ligation products, see Table 3 for MALDI-TOF MS characterization of the σ^{70} ligation products. Since comparison of several fluorescein and dansyl constructs indicated that fluorescein labeled constructs allowed for better characterization of the protein, the σ^{70}_{1-566}-4.2-dansyl was not generated. MALDI-TOF analysis revealed that in each of the σ^{70} ligation
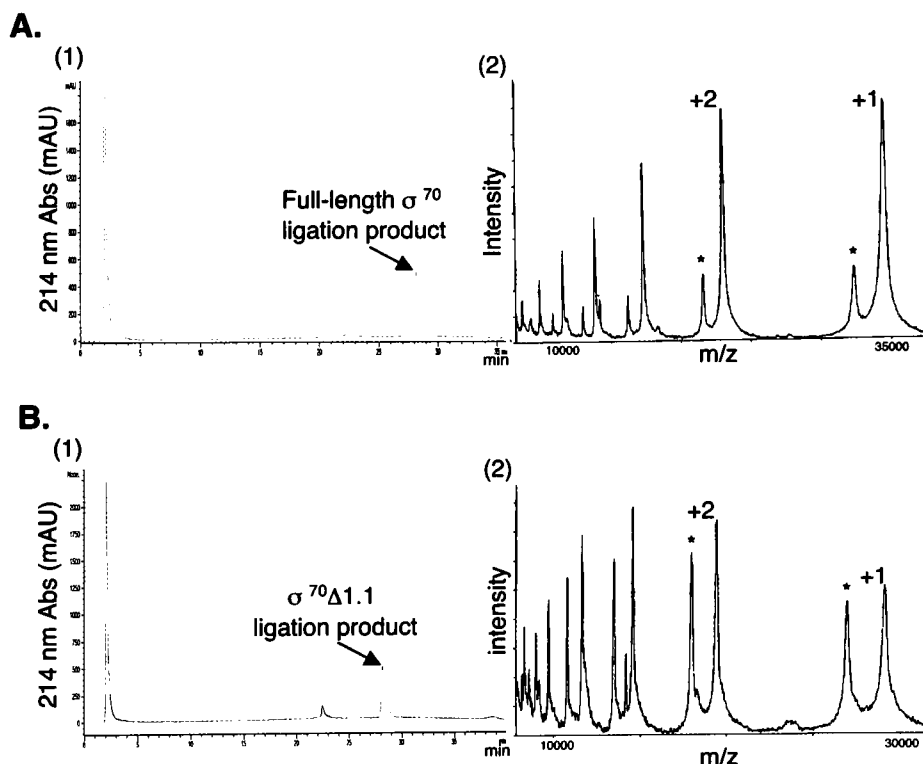
Figure 3.3-11. RP-HPLC and MALDI-TOF MS analysis of fluorescent σ^{70} constructs. (A.) RP-HPLC (1) and MALDI-TOF MS (2) of crude EPL reaction: σ^{70}_{1-566} + 4.2 fluorescein peptide. RP- HPLC gradient is 0-73% B. over 30 min. [Observed mass =69,112 ± 86.6 Da; expected mass (average isotopic mass) = 69,178.1 Da]. **(B.)** RP-HPLC (1) and MALDI-TOF MS (2) of crude EPL reaction: $\sigma^{70}_{101-566}$+ 4.2 fluorescein peptide. RP- HPLC gradient is 0-73% B. over 30 min. [Observed mass =58,258.2 ± 1.8 Da; expected mass (average isotopic mass) = 58,261.2 Da]. MALDI-TOF MS anaylsis indicates two species present in ligation product peak from RP-HPLC, the desired ligation product and a hydrolysis product (*).
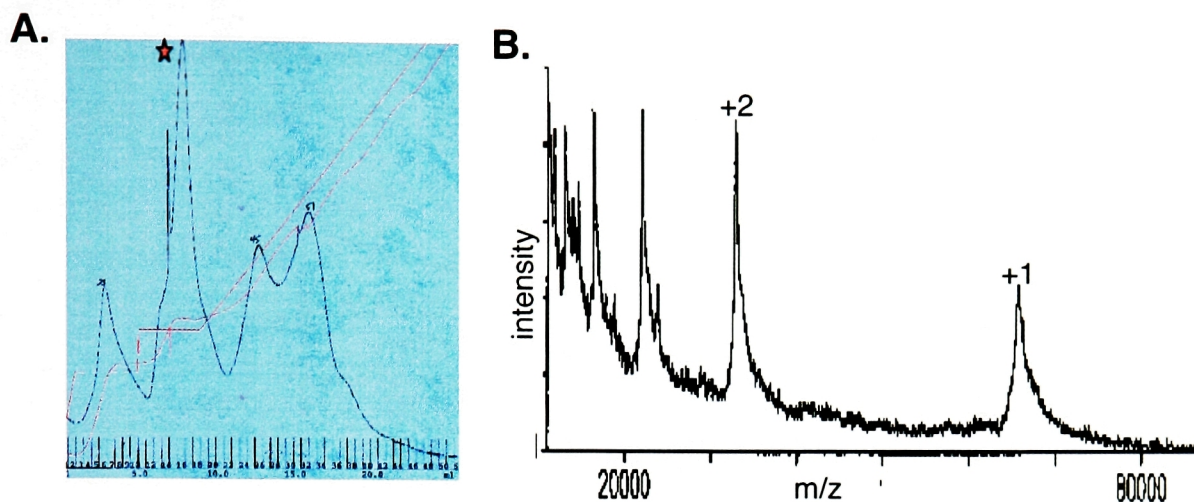
Figure 3.3-12. Anion exchange purification of fluorescently labeled σ⁷⁰ EPL construct (32). (A) FPLC trace of crude EPL ligation supernatant from reaction of σ⁷⁰-model-Fl construct (36). The ligation supernatant was purified by anion exchange (Q column) using 0.1-1M NaCl, 30mM sodium phosphate buffer, pH 7.2. The * indicates pure ligation product (36). (B) MALDI-TOF MS analysis of starred peak from FPLC trace. [Observed mass = 65,730 + 42 Da; expected mass (average isotopic mass) = 65,696.1 Da]. MALDI-TOF indicates a single species, the desired ligation product (32).

reactions both the desired ligation product and a hydrolysis product were present (Figure 3.3-11).

Purification by anion exchange using a sepharose stationary phase allowed these products to be resolved to a homogeneous ligation product (Figure 3.3-12). The purified protein was then exchanged to fluorescence buffer (0.1M sodium phosphate, 50 mM NaCl, pH 7.2) by gel filtration. This additional buffer exchange step significantly diminished the yield of the σ⁷⁰ constructs. Therefore, to increase yields the anion exchange step was later omitted from the purification process and only gel filtration was used. Gel filtration was effective in separating the σ⁷⁰ constructs from unreacted peptide, but it did not separate the hydrolysis σ⁷⁰product from the desired σ⁷⁰ligation product (Figure 3.3-13). While it is always desirable to use a homogeneous sample for biochemical/biophysical studies, here the contaminant protein is non-fluorescent and thus transparent to our binding assay.
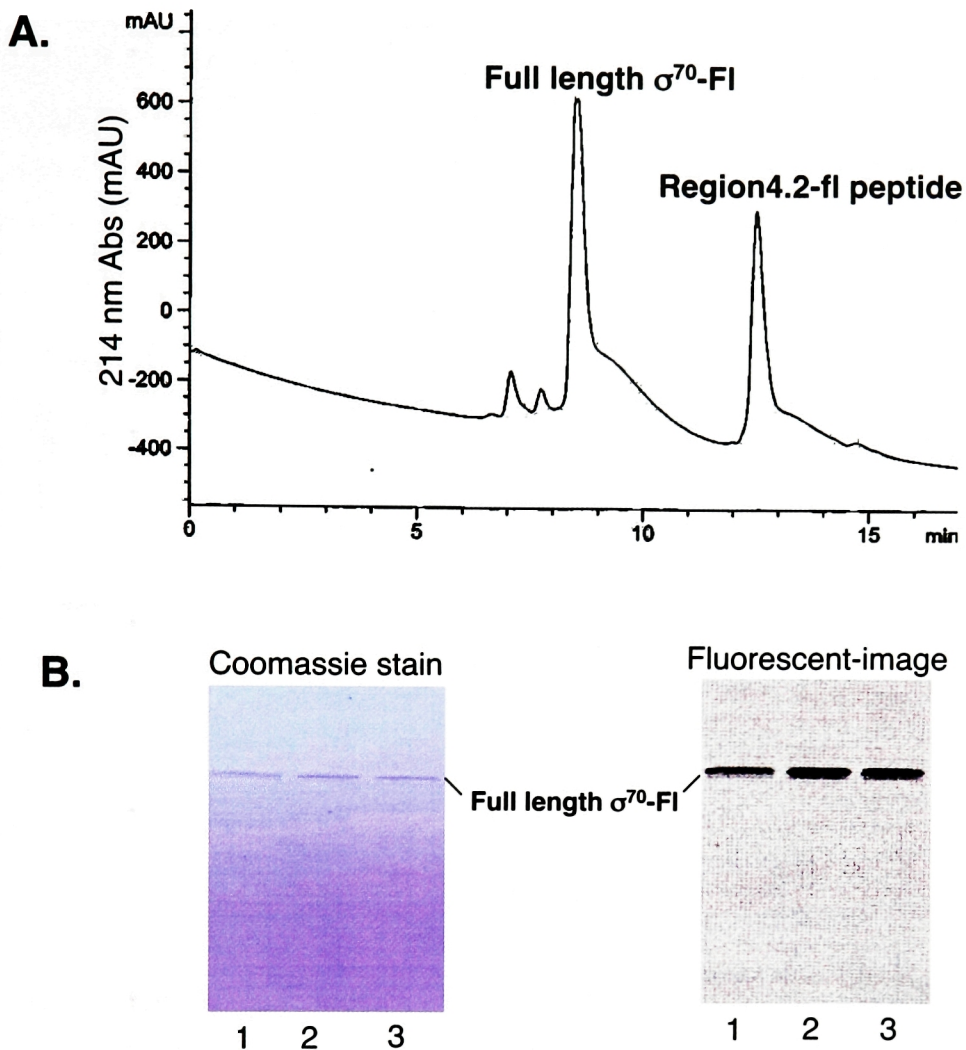
Figure 3.3-13. Gel filtration purification of σ[70] EPL construct (32). (A) Gel filtration HPLC trace of crude ligation reaction of σ[70]$_{1-566}$-VMA-intein-CBD (**26**) and region 4.2-FL (**28**) using Zorbax gel filtration GF-250 column with running buffer 50 mM sodium phosphate, 0.1M NaCl pH 7.2 over 30 min. (B) SDS-PAGE analysis of the full-length σ[70]$_{1-566}$–Fl product (**32**) collected from gel filtration purification. The gel was visualized by coomasie stain and fluorescent imaging. No unreacted 4.2 peptide (**28**) is present.

Spectroscopic analysis revealed that the labeled σ[70] constructs were fluorescent (Figure 3.3-14). Thus a semisynthetic route to fluorescent σ[70] analogues that may be useful for examining the influence of region 1.1 on the σ[70] AsiA interaction was successfully developed. As noted above, several technical problems had to be resolved during the course of this work. These have now been addressed (which allowed the scope of EPL to be extended), however, due to time constraints brought on by these problems, the fluorescence binding assays on AsiA were unable to be performed. However, a robust semisynthetic route to σ[70] analogues is now in place, and this impor-
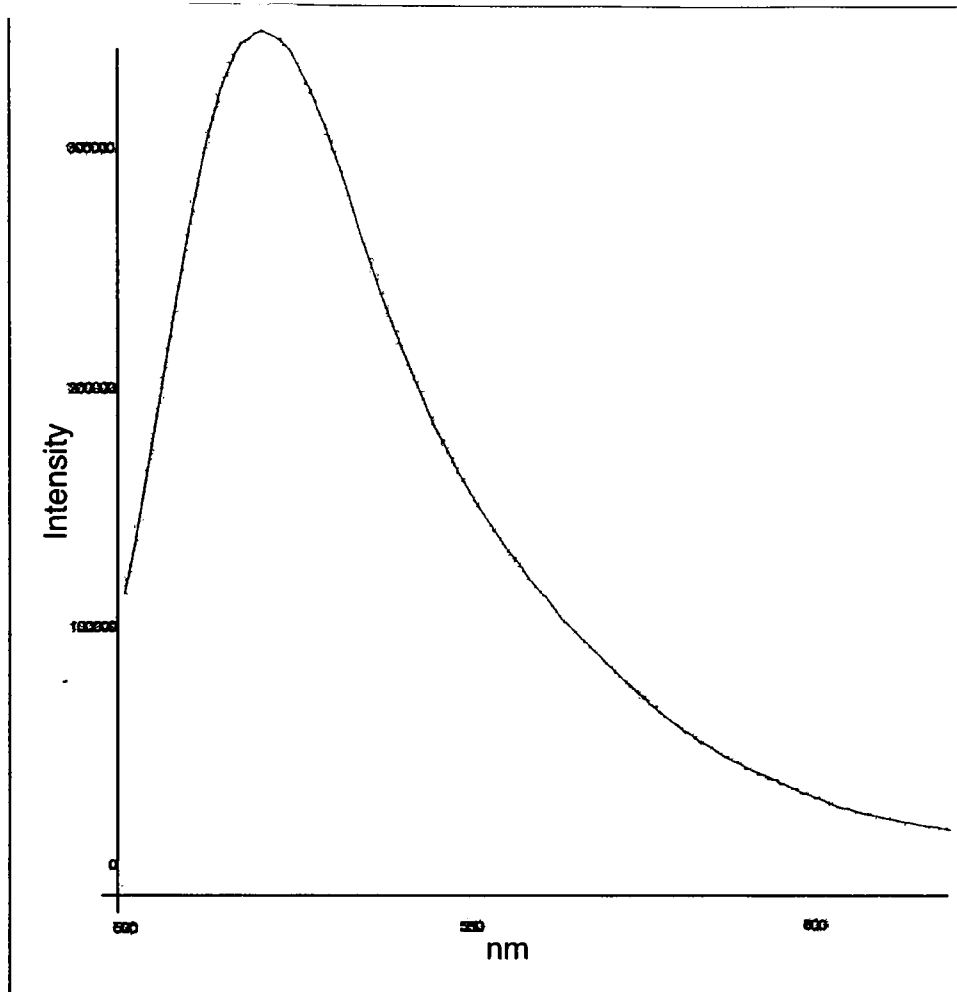
Figure 3.3-14. Fluorescence emission spectrum of σ⁷⁰ construct (32). The σ ^{70} construct **(32)** labeled with fluorescein was excited at 490nm (Fl absorption) and fluorescence emission was monitored from 500-600nm. **(32)** (5nM) was dissolved in 200 μL of 0.1M sodium phosphate, 50mM NaCl, pH 7.2. Emission maximum at 520nm is characteristic of a fluorescein containing compound.

tant σ^{70} function will be addressable in future studies within the Muir group.

Concluding remarks on the incorporation of biophysical and biochemical probes into σ^{70} - significance to EPL development

This work illustrates the enormous potential of expressed protein ligation for exploring the mechanism and regulation of complex biomolecular machines like *E. coli* σ^{70}. Specifically, the molecular nature of AsiA inhibition of *E. coli* σ^{70}-dependent transcription was investigated, and the constructs for determining the influence of region 1.1 the σ^{70}-AsiA interaction were synthesized.

Using a combination of protein-protein photo cross-linking, combinatorial peptide chemistry and mutational analysis to extensively map the σ^{70}-AsiA interactions, it was determined that AsiA inhibits *E.coli* σ^{70}-dependent transcription in an indirect manner. AsiA appears to cover most of the region 4.2 surface, including the putative DNA binding helix, but does not directly interact with the amino acids involved in binding the of −35 promoter DNA sequence. It was also determined that the first 20 amino acids of AsiA were sufficient for its binding to σ^{70}. While it is possible that other residues within AsiA are involved in the interaction *(Lambert, 2001; Urbauer, 2001; Urbauer, 2002)*, these interactions are clearly dispensable for the transcription inhibition activity of AsiA.

Several σ^{70} constructs site-specifically labeled with a fluorescein or dansyl fluorescent probe were synthesized using EPL. These constructs could be used in future studies to investigate the influence of region 1.1 on the σ^{70}-AsiA interactions. Due to the insoluble nature of region 4.2, several chaotropic and organic solvents were employed during the EPL ligation reactions, further outlining the range of this methodology. Furthermore, the successful synthesis of the 70 kDa σ^{70} subunit, all but eliminates the size barrier issue that affected other protein synthesis techniques, demonstrating that EPL can be used to generate large macromolecules for use in structure-function analysis experiments.

Chapter 4: Conclusion

"All technical evolution has a fundamental behavior pattern. First there is scientific discovery of a generalized principle... Next comes objective employment of that principle in a special case invention. Next the invention is reduced to practice... If successful as a tool of society, the invention is used in bigger, swifter, and everyday ways..."

--R. Buckminster Fuller*(Fuller, 2002)*

Summary of thesis work

The goal of this thesis was to illustrate the development and application of the protein semi-synthesis technique Expressed Protein Ligation (EPL) as a readily accessible and effective tool for the incorporation of unnatural amino acids and physical probes into proteins. The scope and versatility of EPL was demonstrated using three distinct applications.

In the first application, section 3.1, a successful strategy was developed for the site-specific incorporation of fluorophores into proteins. Using Abl-SH3 as a model protein, a method was devised for site-specifically incorporating fluorescein and rhodamine fluorophores into a contiguous receptor-ligand designed to act as a biosensor for macromolecular interactions. The biosensor was not realized, but what is striking about this example, is that it identified a method for incorporating, not only one, but two fluorophores into the protein at the specific locations of interest. In addition, our work with the Abl-SH3 domain demonstrated the robust nature of the EPL reaction to organic and chaotropic solvents. Specifically, EPL can be preformed in the presence of 10% DMSO and up to 2M concentrations of GdmCl *(Ayers et al., 1999)*. The tolerance of EPL to these agents has important implications when dealing with poorly soluble synthetic peptides in ligation reactions, and has been used routinely in the Muir lab to aid in the successful synthesis of a number of protein targets *(Blaschke, 2000; Camarero, 2002; Cotton, 2000; Valiyaveetil, 2002)*.

In the second application, section 3.2, a general strategy was developed to site-specifically introduce phospho-amino acids into proteins for analysis of signal transduction pathways. Using the transforming growth factor β receptor I (TβR-I) as a model, the effects of phosphorylation on the activity of the kinase domain was investigated. To this end, a tetra-phosphorylated peptide $^\alpha$thioester was synthesized by Fmoc SPPS. Significantly, this was the first example of the chemical synthesis of a phosphopeptide $^\alpha$thioester, and new synthetic methodology was implemented for this purpose. The peptide $^\alpha$thioester was then successfully ligated to the N-terminus of the kinase domain of TβR-I. Initially, EPL was described as a semi-synthesis strategy for ligating the C-terminus of recombinant proteins to the N-terminus of synthetic constructs

(Evans et al., 1998; Muir et al., 1998; Severinov and Muir, 1998), however, with this application, it has been demonstrated that EPL can also be used at ligate the N-terminus of recombinant proteins to the C-terminus of synthetic constructs. This achievement broadened the range of the EPL technique. Also, the ability to obtain a homogeneous preparation of tetra-phosphorylated TβR-I allowed the level to which hyperphosphorylation activates the kinase to be quantified for the first time. As a result, a model was suggested in which the phosphorylated GS region may activate the receptor in a positive manner *(Huse, 2000)*. These findings could not have been achieved with out EPL. Thus, it has been effectively demonstrated that EPL is a powerful tool for investigating the effects of post-translational modifications on cellular signaling proteins.

In the third and final application, section 3.3, the enormous potential of expressed protein ligation for exploring the mechanism and regulation of complex macromolecular machines was illustrated. Several versions of *E. coli* σ^{70} were designed and synthesized in order to map and quantify the inhibition of σ^{70} transcription initiation. Using a combination of genetic, biochemical, synthetic and combinatorial chemical approaches it was determined that the AsiA protein contacts several amino acids in region 4.2, and completely occludes the DNA-binding surface of region 4.2, thus acting in an indirect manner to inhibit σ^{70} dependent transcription *(Minakhin, 2001)*. In an effort to examine the influence of region 1.1 on the K_d of the σ^{70} AsiA interaction, several versions of σ^{70} (full length, truncated Δ region 1.1, and truncated Δ region 4.2) containing a fluorescein or dansyl fluorophore were synthesized. Due to the insoluble nature of region 4.2, several chaotropes and organic solvents were examined for their use in EPL reactions (see Table 3). These conditions further define the scope of the EPL method. The successful synthesis of the fluorescently labeled 70 kDa σ^{70} constructs illustrates the use of EPL as a tool for engineering large proteins.

In summary, expressed protein ligation was successfully developed and applied to the semi-synthesis of three very distinct macromolecules, ranging in size (from 9 kDa (Abl-SH3) to 70 kDa (σ^{70})) and function. These achievements effectively demonstrate that EPL is a powerful tool for chemically manipulating proteins by allowing the introduction of a host of alterations that range from biochemical and biophysical probes to post-translational modifications. In 1997, when EPL was initially conceived, the

88

impact of the technology could not be predicted. In 2002, just five years later, EPL has been established as a potent tool for engineering proteins. A review of the literature indicates that over forty different proteins have been studied by EPL to date (Table 4), with multiple semi-synthetic analogues of the same protein having been prepared in several instances. This analysis also reveals that the number of EPL applications is increasing rapidly each year, from half a dozen reports in the first year, to over 30 in the last 12 months. The work carried out this thesis has helped to bring about this transformation by providing the fundamental technical guidelines with which to apply EPL to diverse systems. This has allowed us to go from the proposed applications depicted in Figure 1.6 to the commanding list of successful applications shown in Table 4.

Concluding remarks about the future of EPL

Understanding protein function is at the heart of experimental biology. This requires a full description of a protein's post-translational modifications and how these affect intrinsic function and three-dimensional structure, as well as investigations of protein-protein interactions. To achieve this at the level of a proteome is a heart-stopping proposition, yet the assembly of this encyclopedia must be one of the long-term goals of biology. The magnitude of this task is fueling considerable interest in the development of new technologies that allow the acquisition of information regarding protein function to be acquired at an accelerated rate relative to traditional genetic or biochemical approaches. The emergence of dynamic new fields such as Proteomics, Structural Genomics, Bioinformatics and Chemical Biology illustrates how this problem is attracting the attention of researchers from a broad range of disciplines and expressed protein ligation is poised to make a huge impact in these emerging fields.

EPL is the chemical protein engineering equivalent to a jack-of-all-trades. Presented in Figure 4.1 are the different EPL strategies for synthesizing a target protein. The technology has advanced to the point where the original description of EPL, a semi-synthetic strategy for attaching synthetic constructs to the C-terminus of recombinantly

expressed proteins (Figure 4.1, ligation product (i)), is only one of three possible outcomes that can be achieved. Currently EPL can be used to combine two fragments of a recombinant protein (Figure 4.1, ligation product (ii)), or to attach a synthetic construct to the N-terminus of a recombinantly expressed protein (Figure 4.1, ligation product (iii)). Although, not depicted, it is also possible to cyclize a protein and insert synthetic fragments into a protein's core using EPL. This versatility makes EPL a potent tool for addressing proteome type questions. In fact, recent publications have examined the use of expressed protein ligation to the generation of proteins for use in protein microarrays (*Lesaichere, 2002; Tolbert, 2002*). The ease and applicability of the EPL technology to protein engineering suggests that it will be used increasingly in more areas of biology.
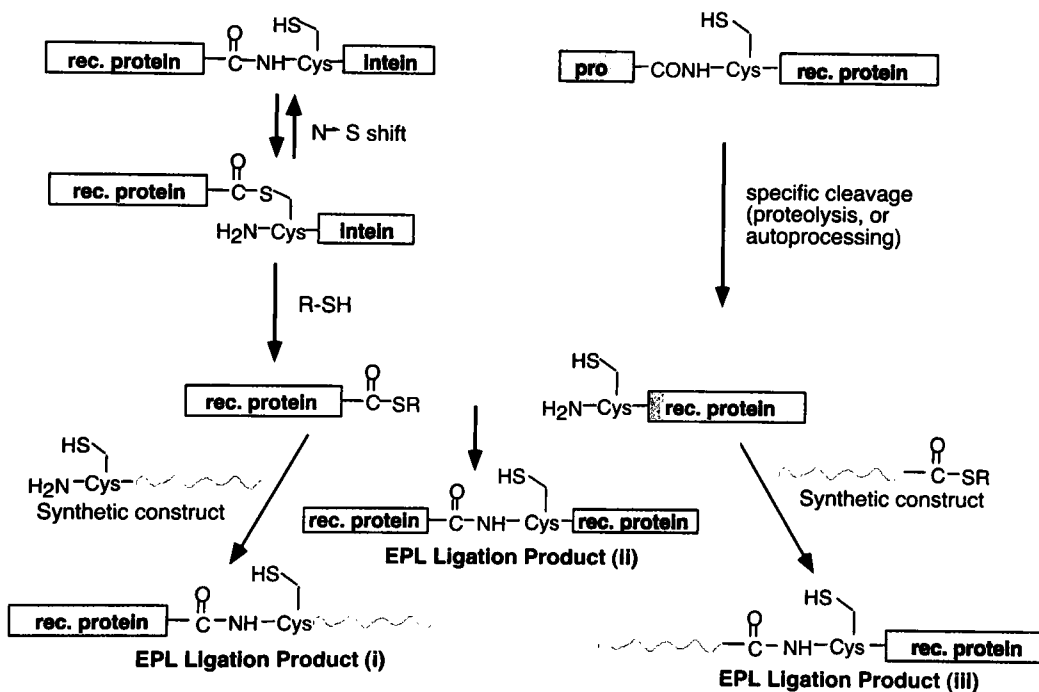
Figure 4.1. Current EPL ligation strategies. (i) Ligation of a synthetic construct to the C-terminus of a recombinant fragment. (ii) Ligation of two recombinant fragments. (iii) Ligation of a synthetic construct to the N-terminus of a recombinant protein.

Herein, I have described in vitro applications of EPL. This is scratching the surface of EPL's abilities, for as the technology improves for introducing semi-synthetic proteins into cells *(Morris, 2001; Schwarze, 2000)* one can only anticipate seeing an increase in the number of *in vivo* studies using proteins created by expressed protein

ligation. Such an application would truly signal the dawning of a new millennium in the field of chemical protein engineering.

Table 4. Proteins engineered using EPL.

Protein	Size	Manipulation	References
AP-1	70	Fe-EDTA probe.	*(Chytil. M, 1998; Erlandson et al., 1996)*
c-Abl	160	Segmental isotopic labeling, fluorescent probe.	*(Cotton et al., 1999; Xu et al., 1999)*
Calcitonin	32	C-terminal amidation.	*(Cottingham, 2001)*
c-Crk	304, 56	Fluorescent probes, backbone cyclization.	*(Camarero, 1999; Camarero, 2001; Camarero, 2001; Cotton, 2000; Hofmann, 2001)*
Csk	450	Fluorescent probes.	*(Muir et al., 1998)*
GlyCAM	77	Carbohydrate.	*(Macmillan, 2000)*
Hedgehog	~170	Lipidation.	*(Porter, 1996; Taylor, 2001)*
Hpal	254	Semi-synthesis.	*(Evans et al., 1998)*
KcsA	135	Semi-synthesis.	*F. Valiyaveetil, TWM Unpublished*
β-lactamase	278	Backbone cyclization.	*(Iwai, 1999)*
Leptin	146	Semi-synthesis.	*E. Asilmaz, J. Friedman, I. Giriat, T.W. Muir, Unpublished*
MBP	395	Various C-terminal modifications.	*(Chong et al., 1997; Tolbert, 2000)*
McbA	40	Unnatural amino acids.	*(Roy, 1999)*
P21Max	101	Phosphorylation.	*(Kawakami, 2001)*
Rab	203	Fluorescent probe.	*(Iakovenko, 2000)*
RNAP β'	1407	Fluorescent probe.	*(Mukhopadhyay, 2001)*
RNase A	124	Semi-synthesis and Selenocysteine ligation.	*(Evans et al., 1998; Hondal, 2001)*
SHP-2	568	Phosphonate.	*(Lu, 2001)*
σ^{70}	613	Semi-synthesis.	*(Severinov and Muir, 1998)*
SigmaA	400	Segmental labeling.	*(Camarero, 2002)*
Smad2	221, 462	Phosphorylatio.n	*(Huse, 2001)*
Smad3	125, 425	Phosphorylation.	*(Huse, 2001)*
Src	441	Tyrosine analogs.	*(Wang, 2001)*
STAT1	713	Fluorescent probe.	*Y. Lui, TWM, Unpublished*
TβR-I	303	Phosphorylation.	*(Flavell, 2002; Huse, 2000)*
T4 DNA ligase	487	Semi-synthesis.	*(Evans Jr. et al., 1999)*
T7 RNAP	887	C-terminal modifications.	*(Lykke-Andersen, 1998)*
Thioredoxin	135	Backbone cyclization, polymeric versions prepared.	*(Evans, 1999)*
ThiS	66	Semi-Synthesis.	*(Kinsland, 1998)*
VEGF	110	C-terminal modifications.	*R. Flavell, I. Giriat, T.W. Muir, Unpublished*

"It has all been very interesting and rewarding, and the opportunity of a lifetime. But now it is time to return to reality and resume my [life] because so many things have yet to be done."

--Bruce Merrifield on returning to normalcy after his Nobel Prize publicity tour *(Merrifield, 1993).*

Courses taken during PhD:

Chemical Biology.

Biophysical Chemistry.

Introduction to Protein Structure.

Advanced Organic Chemistry (Columbia University).

The Application of NMR Spectroscopy to Proteins (Columbia University).

Cell Biology.

Publications:

Minakhin, L., Camaero, J. A., **Holford, M.**, Parker, C., Muir, T. W., Severinov, K. (2001) *Mapping the Molecular Interface Between the Sigma-70 subunit of E. coli RNA Polyermase and T4 AsiA,* J.Mol. Biol., 306, 631-42.

Huse, M., **Holford, M. N.**, Kuriyan, J., Muir, T. W. (2000) *Semisynthesis of hyperphosphorylated type I TGFb receptor: Adressing the mechanism of kinase activation.* J. Am. Chem. Soc., **122**, 8337-8.

Ayers, B., Blaschke, U. K., Camarero, J. A., Cotton, G. J., **Holford, M.**, Muir, T. W. (1999) *Introduction of unnatural amino acids into proteins using expressed protein ligation.* Biopolymers, **51**, 343-54.

Holford, M. & Muir, T. W. (1998) *Adding Splice to Protein Engineering,* Structure, **6**, 951-4.

References

Akaji, K., Fujii, N., Yamjima, H., Hayashi, K., Mizuta, K., et.al (1985). Chem. Pharm. Bull. *33*, 184-201.

Alsina, J., Yokum, T.S., Albericio, F., Barany, G. (1999). J. Org. Chem. *64*, 8761-8769.

Amrein, K. E., Tackas, B., Stieger, M., Molonos, J., Flint, N. A., Burn, P. (1995). Proc. Natl. Acad. Sci. USA *92*, 1048-52.

Atherton, E., Fox, H., Harkiss, D., Logan, C.J., Sheppard, R.C., Williams, B. J. (1978). J. Chem. Soc. Chem. Commun, 537.

Ayers, B., Blaschke, U. K., Camarero, J. A., Cotton, G. J., Holford, M., and Muir, T. W. (1999). Biopolymers *51*, 343-54.

Baca, M., Muir, T. W., Schnolzer, M., and Kent, S. B. H. (1995). J. Am. Chem. Soc. *117*, 1881-1887.

Backes, B. J., Virgilio, A. A., Ellman, J. A. (1996). J. Am. Chem. Soc. *118*, 3055-56.

Backes, B. J. E., J. A. (1999). J. Org. Chem *64*, 2322.

Bergmann, M. Z., L (1932). Ber. Dtsch. Chem. Ges *65*, 1192.

Blake, J., and C.H., L. (1981). Proc. Natl. Acad. Sci. *78*, 4055-4058.

Blaschke, U. K., Cotton, G. J., Muir, T. W. (2000). Tetrahedron *56*, 9461-9470.

Bodanszky, M. (1955). Nature (London) *175*, 685.

Burgess, R. R., Travers, A. A., Dunn, J.J., Bautz, E. K. F. (1969). Nature *221*, 43.

Burgess, R. R. T., A. A. (1970). 29, 1164.

Camarero, J. A., Cotton, G. J., Adeva, A., Muir, T. (1998). J. Pep. Res *51*, 303-16.

Camarero, J. A., Shektman, A., Campbell, E., Darst, S. A., Cowburn, D., Muir, T. W. (2002). Proc. Natl. Acad. Sci. *99*, 8536-8541.

Camarero, J. A., Muir, T. W. (1999). J. Am. Chem. Soc. *121*, 5597-8.

Camarero, J. A., Fushman, D., Cowburn, D., Muir, T. W. (2001). Bioorg. Med. Chem. *9*, 2479-84.

95

Camarero, J. A., Fushman, D., Sato, S., Giriati, I., Cowburn, D., Raleigh, D. P., Muir, T. W. (2001). J. Mol. Biol. *308*, 1045-62.

Camarero, J. A., Ayers, B., Muir, T. W. (1998). Biochemistry *37*, 7487-95.

Camarero, J. A., Cotton, G. C., Adeva, A., and Muir, T. W. (1998). J. Peptide Res. *51*, 303-316.

Carpino, L. A. (1957). J. Am. Chem. Soc. *79*, 4427.

Chang, T. K., Jackson, D. Y., Burnier, J. P., and Wells, J. A. (1994). Proc. Natl. Acad. Sci. USA *91*, 12544-12548.

Chong, S., Mersha, F. B., Comb, D. G., Scott, M. E., Landry, D., Vence, L. M., Perler, F. B., Benner, J., Kucera, R. B., Hirvonen, C. A., Pelletier, J. J., Paulus, H., and Xu, M.-Q. (1997). Gene *192*, 271-281.

Chong, S., Shao, Y., Paulus, H., Benner, J., Perler, F. B., and Xu, M.-Q. (1996). J. Biol. Chem. *271*, 22159-22168.

Chytil. M, P., B. R., Erlanson, D. A., Verdine, G. L. (1998). Natl. Acad. Sci. USA *95*, 14076-81.

Colland, F., Orsini G., Brody, E. N., Buc, H., Kolb, A. (1998). Mol Microbiol. *27*, 819-29.

Cornish, V. W., Mendel, D., and Schultz, P. G. (1995). Angew. Chem. Int. Ed. Engl. *34*, 621-633.

Cottingham, I. R., Millar, A., Emslie, E., Colman, A., Schnieke, A. E., McKee, C. (2001). Nat. Biotechnol. *19*, 974-7.

Cotton, G. J., Ayers, B. A., Xu, R., and Muir, T. W. (1999). J. Am. Chem. Soc. *121*, 1100-1101.

Cotton, G. J. M., T. W. (2000). Chem. & Biol. *7*, 253-61.

Curtius, T. (1881). J. Pract. Chem. *24*, 239.

Czarnik, A. W. (1995). Chem. Biol. *2*, 428-32.

D' Auria, S. L., J. R. (2001). Curr. Opin. Biotechnol. *12*, 99-104.

Dawson, P. E., Kent, S. B. (2000). Annu. Rev. Biochem. *69*, 923-60.

Dawson, P. E., Churchill, M. J., Ghadiri, M. R., and Kent, S. B. H. (1997). J. Am. Chem. Soc. *119*, 4325-4329.

Dawson, P. E., Muir, T. W., Clark-Lewis, I., and Kent, S. B. H. (1994). Science *266*, 776-779.

Dombroski, A. J., Walter, W. A., Gross, C.A. (1993). Genes Dev. *7*, 2446-55.

Dombroski, A. J., Walter, W.A., Record, M. T., Jr., Siegele, D. A., Gross, C. A. (1992). Cell *70*, 501-12.

Dos Remedios, C. G. M., P. D. J. (1995). J. Struct. Biol. *115*, 175-85.

du Vigneaud, V., Ressler, C., Swan, J. M., Roberts, C. W., Katsoyannis, P.G., Gordon, S. (1953). J. Am. Chem. Soc. *75*.

Duan, X., Gimble, F. S., and Quiocho, F. A. (1997). Cell *89*, 555-564.

Erlandson, D. A., Chytil, M., and Verdine, G. L. (1996). Chem. Biol. *3*, 981-991.

Evans, J., T.C, Benner, J., and Xu, M.-Q. (1998). Protein Science *7*, 2256-64.

Evans Jr., T. C., Benner, J., and Xu, M.-Q. (1999). J. Biol. Chem. *274*, 3923-3936.

Evans, T. C., Benner, J., Xu, M. Q. (1999). J. Biol. Chem. *274*, 18359-81.

Fischer, E. F., E. (1901). Ber. Dtsch. Chem. Ges *34*, 2868.

Flavell, R. R., Huse, M., Goger, M., Trester-Zedlitz, M., Kuriyan, J., Muir, T .W. (2002). Org Lett *4*, 165-68.

Forster, T. (1965). Delocalized excitation and excitation transfer, O. Sinanoglu, ed. (New York: Academic Press).

Forster, T. (1948). Ann. Phys. *2*, 55-75.

Fruton, J. S. (1985). Proc. Amer. Phil. Soc. *129*, 313-370.

Fuller, R. B. (2002). Critical Path: St. Martin's Press).

Gaertner, H. F., Rose, K., Cotton, R., Timms, D., Camble, R., & Offord, R. E. (1992). Bioconjugate Chem. *3*, 262-68.

Gosser, Y. Q., Zheng, J., Overduin, M., Mayer, B. J., and Cowburn, D. (1995). Structure *3*, 1075-1086.

Hall, T. M., Porter, J.A., Young, K. E., Koonin, E. V., Beachy, P.A., Leahy, D. J. (1997). Cell *91*, 85-97.

Hashimoto, H., Takahashi, H., Nishioka, M., Fujiwara, S., Takagi, M., Imanaka, T., Inoue, T., & Kai, Y. (2000). Acta Crystallogr. D. Biol. Crystallogr. *56*, 1185-86.

Heath, W. F. M., R.B. (1988). Proc. Natl. Acad. Sci. USA *83*, 6367-71.

Helmann, J. D. C., M. J. (1988). J. Annu. Rev. Biochem. *57*, 839.

Hermanson, G. T. (1996). Bioconjugate Techniques: Academic Press).

Herrera, R., Lebwohl, D., Herreros, A. G. D., Kallen, R. G., Rosen, O. (1988). J. Biol. Chem. *263*, 5560-68.

Hirata, R., Ohsumi, Y., Nakano, A., Kawasaki, H., Suzuki, K., Anraku, Y. (1990). J. Biol. Chem. *265*, 6726-6733.

Hofmann, R. M., Cotton, G. F., Chang, E. J., Vidal, E., Veach, D., Bornmann, W., Muir, T. W. (2001). Bioorg. Med. Chem. Lett. *11*, 3091-94.

Hofmann, R. M. Muir, T. W. (2002). Curr. Opin. Biotech. *13*, 297.

Holford, M. Muir, T. W. (1998). Structure *6*, 951.

Homandberg, et. al. (1978). Biochem. *17*, 5220.

Hondal, R. J., Nilsson, B. L., Raines, R. T. (2001). J. Am. Chem. Soc. *123*, 5140-1.

Huse, M., Chen, Y.-G, Massague, J, Kuriyan, J. (1999). Cell *96*, 425-436.

Huse, M., Holford, M. N., Kuriyan, J., Muir, T. W. (2000). J. Am. Chem. Soc. *122*, 8337-8338.

Huse, M., Muir, T. W., Xu, L., Chen Y. G., Kuriyan, J., Massague, J. (2001). Mol. Cell *8*, 671-82.

Iakovenko, A., Rostkova, E., Merzlyak, E., Hillebrand, A. M., Thoma, N. H., Goody, R. S., Alexandrov, K. (2000). FEBS Lett. *468*, 155-8.

Ichiyanagi, K., Ishino, Y., Ariyoshi, M., Komori, K., & Morikawa, K. (2000). J. Mol. Biol. *300*, 889-901.

Imperali, B. (1998). J. Amer. Chem. Soc. *120*, 609-10.

Ingenito. R., B. E., Fattori, D. Pessi. A. (1999). J. Am. Chem. Soc *121*, 11369-74.

Iwai, H. P., A. (1999). FEBS Lett., 166-72.

Jackson, D. Y., Burnier, J., Quan, C., Stanley, M., Tom, J., and Wells, J. A. (1994). Science *266*, 243-247.

Jones, J. (1994). The chemical Synthesis of Peptides: Clarendon Press).

Kane, P. M., Yamashiro, C. T., Wolczyk, D. F., Neff, N., Goebl, M., and Stevens, T. H. (1990). Science *250*, 651-657.

Kawakami, T., Hasegawa, K., Teruya, K., Akaji, K., Horiuchi, M., Inagaki, F., Kurihara, Y., Uesugi, S., Aimoto, S. (2001). J. Pept. Sci *7*, 474-87.

Kemp, D. S., and Carey, R. I. (1993). J. Org. Chem. *58*, 2216-2222.

Kenner, G. W., McDermott, J. R., Sheppard, R. C. (1971). J. Chem. Soc. Chem. Commun., 636-37.

Kent, S. B. H. (1988). Ann. Rev. Biochem. *57*, 957-989.

Kim, S., Makino, K., Amemura, M., Nakata, A., Shinagawa, H. (1995). Mol. Gen. Genet. *248*, 1.

Kinsland, C., Taylor, S. V., Kelleher, N. L., McLafferty, F. W., Begeley, T. P. (1998). Protein Sci. *7*, 1839-42.

Kiso, Y., Kimura, T., Fujiwara, Y., Sakikawa, H., Akaji, K. (1990). Chem. Pharm. Bull. *38*, 270.

Klabunde, T., Sharma, S., Telenti, A., Jacobs Jr., W. R., and Sacchettini, J. C. (1998). Nature Struct. Biol. *5*, 31-36.

Kolesky, S., Ouhammouch, M., Brody, E. N., Geiduschek, E. P. (1999). J. Mol. Biol. *291*, 267-81.

Konig, W. a. G., R. (1970). Chem. Ber. *103*, 788.

Kurokawa, K., Mocchizuki, N., Ohba, Y., Mizuno, H., Miyawaki, A., Matsuda, M. (2001). J. Biol. Chem. *13*, 13.

Lambert, L. J., Schirf, V., Demeler, B., Cadene, M., Werner, M .H. (2001). Embo *20*, 7149-59.

Larsson, E. L., B. (1994). Terta. Lett. *35*, 2737.

Lee, M. G., S. (2001). Essays. Biochem. *37*, 71-85.

Lesaichere, M.-L., Lue, R. Y. P., Chen, G. Y. J., Zhu, Q., Yao, S. Q. (2002). J. Am. Chem. Soc. *124*, 8768-69.

Li, X., Kawakmi, T., Aimoto, S. (1998). Tet. Lett. *39*, 8669-72.

Liu, C.-F., and Tam, J. P. (1994). J. Am. Chem. Soc. *116*, 4149-4153.

Lloyd-Williams, P., Albericio, F., and Giralt, E. (1993). Tetrahedron *48*, 11065-11133.

Lonetto, M. A., Gribskov, M.m Gross, C. A. (1992). J. Bacteriol. *174*, 3843.

Lonetto, M. A., Rhodius, V., Lamberg, K., Kiley, P., Busby, S., Gross, C. (1998). J. Mol. Biol. *284*, 1353-1365.

Lu, W., Gong, D., Bar-Sagi, D., Cole, PA (2001). Mol Cell *8*, 759-69.

Lykke-Andersen, J. C., J. (1998). Nucleic Acids Res *26*, 5631-35.

Macmillan, D. B., C. R. (2000). Tetrahedron *56*, 9515-25.

Massague, J. (1998). J. Ann. Rev. Biochem. *67*, 753-91.

Massague, J. (2000). Nat. Rev. Mol. Cell. Biol. *1*, 169-78.

Mayer, B. J. (2001). J. Cell. Sci. *114*, 1253-63.

Merrifield, B. (1993). Life During a Golden Age of Peptide Chemistry, J. I. Seeman, ed. (wahington, DC: American Chemical Society).

Merrifield, R. B. (1963). J. Am. Chem. Soc. *85*, 2149-2154.

Meyer, B. J., Hamaguchi, M., & Hanafusa, H. (1988). Nature *332*, 272-275.

Miller, M., Schneider, J., Sathyanarayana, B. K., Toth, M. V., Marshall, G. R., Clawson, L., Selk, L., Kent, S. B. H., and Wlodawer, A. (1989). Science *246*, 1149-1152.

Minakhin, L., Camarero, J. A., Holford, M., Parker, C., Muir, T. W., Severinov, K. (2001). J. Mol. Biol. *306*, 631-642.

Miyawaki, A., et. al., & Tsien, R. Y. (1997). Nature *388*, 882-886.

Morris, M. C., Depollier, J., Mery, J., Heitz, F., Divita, G. (2001). Nat. Bitotechnol. *19*, 1173-6.

Muir, T. W. (1995). Structure *3*, 649-652.

Muir, T. W., Williams, M. J., Ginsberg, M. H., & Kent, S. B. H. (1994). Biochem. *33*, 7701-08.

Muir, T. W. (2001). Synlett *6*, 733-40.

Muir, T. W., Dawson, P. E., and Kent, S. B. H. (1997). Methods Enzymol. *289*, 266-298.

Muir, T. W., Sondhi, D., and Cole, P. A. (1998). Proc. Natl. Acad. Sci. USA *95*, 6705-6710.

Muir, T. W., Williams, M. J., Ginsberg, M. H., and Kent, S. B. H. (1994). Biochemistry *33*, 7701-7708.

Mukhopadhyay, J., Kapanidis, A. N., Mekler, V., Kortkhonjia, E., Ebright, Y. W., Ebright, R. H. (2001). Cell *106*, 453-63.

Musacchio, A., Noble, M., Paupit, R., Wierenga, R., and Saraste, M. (1992). Nature *359*, 851-855.

Musacchio, A., Saraste, M., and Wilmanns, M. (1994). Nature Struct. Biol. *1*, 546-551.

Nakatsuka, T., Sasaki, T., and Kaiser, E. T. (1987). J. Am. Chem. Soc. *109*, 3808-3810.

Nalivaeva, N. N., Turner, A. J. (2001). Proteomics *1*, 735-47.

Nefzi, A., Sun, X., and Mutter, M. (1995). Tetrahedron Letters *36*, 229-230.

Offer, J. D., P. E. (2000). Organic Letters *2*, 23.

Orsini, G., Ouhammouch, M., Le Caer, J. P., Brody, E. N. (1993). J. Bact. *175*, 85-93.

Ouhammouch, M., Orisini, G., Brody, E. N. (1994). J. Bacteriol. *176*, 3956-3965.

Pawson, T. R., M., Nash, P. (2002). FEBS Lett. *513*, 2-10.

Perich, J. W., & Reynolds, E. C. (1991). Tetra. Lett. *37*, 572.

Perich, J. W. J., R. B. (1988). J. Org. Chem. *53*, 4103.

Perich, J. W. J., R. B. (1988). Tetra. Lett. *29*, 2369.

Perler, F. B. (2000). Nucl. Acids Res. *28*, 344-345.

Perler, F. B. (1999). Nucl. Acids Res. *27*, 346-347.

Perler, F. B., Davis, E. O., Dean, G. E., Gimble, F. S., Jack, W. E., Neff, N., Noren, C. J., Thorner, J., and Belford, M. (1994). Nucleic Acids Res. *22*, 1125-1127.

Perler, F. B., Xu, M.-Q., and Paulus, H. (1997). Curr. Opin. Chem. Biol. *1*, 292-299.

Pisabarro, M. T., and Serrano, L. (1996). Biochemistry *35*, 10634-10640.

Poland, B. W., Xu, M.Q., Quiocho, F. A. (2000). J. Biol. Chem *275*, 16408-16413.

Porter, J. A., Ekker, S.C., Park, W-J, von Kessler, D. P., Young, K. E., Chen, C. H., Ma, Y., Woods, A. S., Cotter, R. J., Koonin, E. V., Beachy, P. A. (1996). Cell *86*, 21-34.

Quaderer, R., Hilvert, D. (2001). Organic Lett. *3*, 3181-84.

Rose, K. (1994). J. Am. Chem. Soc. *116*, 30-33.

Roy, R. S., Allen, O., Walsh, C. T. (1999). Chem. Biol. *6*, 789-99.

Schneider, J., and Kent, S. B. H. (1988). Cell *54*, 363-368.

Schnolzer, M., Alewood, P., Jones, A., Alewood, D., Kent, S.B.H. (1992). Int. J. Pept. Protein Res. *40*, 180.

Schnolzer, M. K., S. B. H. (1992). Science *256*, 221-25.

Schwarze, S. R., Hruska, K. A., Dowdy, S. F. (2000). Trends Cell Biol. *10*, 290-5.

Severinov, K., and Muir, T. W. (1998). J. Biol. Chem. *273*, 16205-16209.

Severinova, E., Severinov, K. Fenyo, D., Marr, M., Brody, E. N., Roberts, J. W., Chait, B. T., Darst, S. (1996). J. Mol. Biol *263*, 637-647.

Severinova, E., Severinov, K., Darst, S. A. (1998). J. Mol. Biol. *279*, 9-18.

Sewing, A. H. D. (2001). Angew. Chem. Int. Ed. *40*, 3395-6.

Shao, Y., Xu, M.-Q., and Paulus, H. (1996). Biochemistry *35*, 3810-3815.

Sheehan, J. C. a. H., G. P. (1955). J. Am. Chem. Soc. *77*, 1067.

Shin, Y., Winans, K.A., Backes, B.J., Kent, S.B.H., Ellman, J.A., Bertozzi, C.R. (1999). J. Am. Chem. Soc. *121*, 11684-11689.

Sieber, P., Eisler, K., Kamber, B., Riniker, B., Rittel, W., et. al. (1978). Hoppe Seyler's Z. Physiol. Chem. *359*, 113-23.

Soderbaum, H. G. (1916). Jac. Berzelius Lettres 5.

Southworth, M. W., Benner, J., Perler, F. B. (2000). EMBO J *19*, 5019-5026.

Still, W. (1998). Science *279*, 851-853.

Stryer, L., & Haughland, R.P. (1967). Proc. Natl. Acad. Sci. USA *58*, 716-26.

Swinnen, D. H., D. (2000). Org. Lett. *2*, 2439-42.

Tam, J. P., Heath, W.F., Merrifield, R. B. (1983). J. Am. Chem. Soc. *105*, 6442-55.

Tam, P., Lu, Y. A., Liu, C. F., Shao, J. (1995). Proc. Natl. Acad. Sci. USA *92*, 12485-89.

Tam, P. J., Sheikh, M.A., Solomon, D.S., Ossowski, L., (1986). Proc. Natl. Acad. Sci. USA *83*, 8082-86.

Tanford, C., & Reynolds, J. (2001). Nature's robots (New York: Oxford University Press).

Taylor, D. (1995). Annu. Rev. Biophys. Biomol. Struct. *24*, 405-34.

Taylor, F. R., Wen, D., Garber, E. A., Carmillo, A. N., Baker, D. P., Ardunin, R. M., Williams, K. P., Weinreb, P. H., Rayhorn, P., Hronowski, X. et. al. (2001). Biochemistry *40*, 4359-71.

Tolbert, T. J. W., C.-H. (2002). Angew. Chem. Int. Ed. *41*, 2171-74.

Tolbert, T. W., C-H (2000). J. Am. Chem. Soc. *122*, 5421-28.

Tsien, R. Y. (1998). Annu. Rev. Biochem. *67*, 509-44.

Tsien, R. Y. M., A. (1998). Science *280*, 1954-5.

Urbauer, J. L., Adelman, K., Urbauer, R. J. B., Simeonov, M. F., Gilmore, J. M., Zolkiewski, M., Brody, E. N. (2001). J. Biol. Chem. *276*, 41128-32.

Urbauer, J. L., Simeonov, M. F., Urbauer, R. J. B., Adelman, K., Gilmore, J. M., Brody, E. N. (2002). PNAS *99*, 1831-35.

Valiyaveetil, F. I., MacKinnon, R., Muir, T. W. (2002). J. Am. Chem. Soc. *24*, 9113-9120.

Viguera, A. R., Arrondo, J. L. R., Musacchio, A., Saraste, M., and Serrano, L. (1994). Biochemistry *33*, 10925-10933.

Wallace, C. J. A. (1995). Curr. Opin. Biotech. *6*, 403-410.

Wallace, C. J. A., and Clark-Lewis, I. (1992). J. Biol. Chem. *267*, 3852-3861.

Wang, D., Cole, P. A. (2001). J. Am. Chem. Soc. *123*, 8883-6.

Wetzel, R., Halualani, R., Stults, T., Quan, C. (1990). Bioconjugate Chem *1*, 114-22.

Wieland, T., Schafer, W., Bokelmann, E. (1951). Liebigs Ann. Chem *573*, 99.

Wieland, T. B., H. (1951). Liebigs Ann. Chem *572*, 190.

Wieser, R., Wrana, J. L., Massague, J. (1995). EMBO J. *14*, 2199-2208.

Williams, M. J., Muir, T. W., Ginsberg, M. H., & Kent, S. B. H. (1994). J. Am. Chem. Soc. *116*, 10797-98.

Willis, S. A., Zimmerman, C. M., Li, L., Matthews, L. S. (1996). Mol Endocrinol. *10*, 367-379.

Woods, A. C., Guillemette, J. G., Pasrrish, J. C., Smith, M., and Wallace, C. J. A. (1996). J. Biol. Chem. *271*, 32008-32015.

Wu, G., Chen, Y.-G., Ozdamar, B, Gyuricza, C. A., Chong, P. A., Wrana, J. L., Massague, J., Shi, Y. (2000). Science *287*, 92-97.

Xu, M.-Q., and Perler, F. B. (1996). EMBO J. *15*, 5146-5153.

Xu, R., Ayers, B., Cowburn, D., and Muir, T. W. (1999). Proc. Natl. Acad. Sci. USA. *96*, 388-393.

Yamashiro, D., and Li, C. H. (1988). Int. J. Pept. Protein. Res. *31*, 322-334.

Zawadzke, L. E., Berg, J. M., (1993). Proteins *16*, 301-5.