2003

Protein Semi-Synthesis in Vivo

Izabela Giriat Stankiewicz

Protein Semi-Synthesis

in vivo

by

Izabela Giriat Stankiewicz

Submitted in Partial Fulfillment

of the Requirement for the

Degree of Doctor of Philosophy

March, 2003

THE ROCKEFELLER UNIVERSITY
1230 YORK AVENUE • NEW YORK, NEW YORK 10021-6399

OFFICE OF GRADUATE STUDIES
Tel : 212-327-8086
Fax: 212 327-8505
E-mail: phd@rockefeller.edu

Dr. Sidney Strickland
Dean and Vice President for Academic Affairs
The Rockefeller University
1230 York Avenue, Box 177
New York, NY 10021

Dear Dr. Strickland:

As members of the committee appointed to evaluate the thesis of Izabela Giriat Stankiewicz, we have read her thesis and have found it to be acceptable with revision in partial fulfillment of the University's requirements for the granting of the doctoral degree. Therefore, we, the undersigned, are satisfied that Ms. Izabela Giriat Stankiewicz has met the requirements for the Ph.D. degree with respect to her written thesis, thesis presentation and examination by this committee.

Chairman

Faculty Advisor

Faculty Member

Name & Affiliation of
External Member

Faculty Member

Dedicated to my parents,

Witold Giriat and Jolanta Stankiewicz

Acknowledgements

I am indebted to all those that one way or another taught me, helped me, encouraged me and made me grow as a person in the years I did my doctoral work. To all, my most sincere thanks.

Specifically, I would like to thank Dr. Tom Muir for having me as a graduate student. Many of the members in the Muir laboratory had a great influence on me, and I would like to name in particular Mande Holford (who recommended the lab and became a close friend), Graham Cotton, Julio Camarero and Brenda Ayers (for helping me adjust to the new lab), Gholson Lyon (for making the lab interesting), Rob Flavell (for allowing me to train him and for working with me on my secondary project) and Vasant Muralidharan and Mike Hahn (for giving me a helping hand with this document).

I would also like to thank my thesis committee: Dr. Brian Chait, Dr. Sandy Simon and Dr. Jeff Friedman. They were encouraging through out, and gave me good advice at times. Dr. Phil Cole I would like to thank for taking the time to be my external reviewer.

Finally I would like to thank my immediate family: my mother (Jolanta Stankiewicz), my father (Witold Giriat), my brother (Constantino Giriat) and my life partner (Ali K. Arsan).

Table of Content

List of Figures

List of Schemes

List of Tables

Abbreviations

ANTP	Homeotic transcription factor of *Drosophila*.
Boc	tert-butyloxycarbonyl.
CBD	Chitin Binding Domain.
CFP	Cyan Fluorescent Protein.
CMV	Human cytomegalovirus
CPM	Counts per Minute.
DM	Docecylmaltoside.
DOD	LAGLIDADG homing endonuclease.
DsRed	Red fluorescent protein.
E. coli	*Escherichia coli*.
EDT	1,2-ethanedithiol.
EM	Expected Mass.
EPL	Expressed Protein Ligation.
ER	Endoplasmic Reticulum.
ES-MS	Electro-spray Mass Spectrometry.
FlAsH	4', 5'-bis(1,3,2-dithioarsolan-2-yl) fluorescein.
Fm	9-fluoronylmethyl.
Fmoc	9-fluoremethyloxycarbonyl.
FRAP	Fluorescent Recovery After Photo-bleaching.

FRET	Fluorescence Resonance Energy Transfer
GFP	Green Fluorescent Protein.
GI	Gastric Intestine.
GPCR	G-Protein-Coupled Receptor.
GST	Glutathione S-transferase.
GyrA	*Mycobacterium xenopi* DNA Gyrase A.
hAGT	O^6-alkylguanine-DNA alkyltransferase.
HBSS	Hank's Balanced Salt Solution.
HS	Heparin Sulfate.
HSV-1	Herpes Simplex Virus type 1.
HUVEC	Human Umbilical Vein Endothelial Cells.
I_C	Second half of the naturally occurring split intein DnaE.
I_N	First half of the naturally occurring split intein DnaE.
IP	Immuno-Precipitation.
IPTG	Isopropyl-β-D-thiogalactopyranoside
kb	Kilo bases.
K_D	Dissociation constant.
MAI	Minutes After Injection.
MESNA	2-Mercaptoethane sulfonic acid.
mRNA	messenger RNA.
Msc	2-(methylsulfonyl) ethyloxycarbonyl.

NCL Native Chemical Ligation.

NIR Near-Infra Red.

NLS Nuclear Localization Sequence.

NPyS 3-nitro-2-pyridinesulfenyl.

OM Observed Mass.

PBS Phosphate Buffer Saline.

PTD Protein Transduction Domain.

Rh Tetramethylrhodamine.

RP-HPLC Reverse Phase High Pressure Liquid Chromatography.

Sce. VMA *Saccharomyces cerevisiae* vacuolar ATPase.

SPPS Solid Phase Peptide Synthesis.

Ssp. *Synechocystis* sp., strain PCC6803.

TnfR Transferrin Receptor.

tRNA Transfer RNA.

VEGF Vascular Endothelial Growth Factor.

VEGFR VEGF Receptor.

VP22 Herpes simplex virus type 1 (HSV-1) VP22 transcription factor.

Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

Abstract.

Incorporation of chemical probes into proteins is a powerful way to elucidate biological processes and to engineer novel function. This thesis describes the ligation of synthetic molecules to target proteins in an intracellular environment, and the use of semi-synthetic proteins for whole animal studies.

In the first approach a cellular protein is genetically tagged with one half of a split-intein. The complementary half is linked *in vitro* to the synthetic probe and this fusion is delivered into cells using a transduction peptide. Association of the intein halves in the cytosol triggers protein trans-splicing, resulting in the ligation of the probe to the target protein through a peptide bond. This process is specific and applicable to cytosolic and integral membrane proteins. The technology should allow cellular proteins to be elaborated with a variety of abiotic probes.

In the second approach a sequential expressed protein ligation was used to link three fragments: a targeting protein, an affinity tag, and an imaging moiety. Specifically, vascular endothelial growth factor (VEGF) was ligated with a 6xHis tag and a synthetic high affinity ^{99m}Tc chelator. Following protein semi-synthesis, the ^{99m}Tc was incorporated into the chelator and the ^{99m}Tc-labeled protein was injected into mice bearing a tumor. The semi-synthetic VEGF localized onto the vascularized tumor and allowed it to be imaged with a γ camera. The strategy developed is modular, permitting the use of multiple chemical moieties, purification methods, and targeting proteins. This methodology could be used to provide insights into biological processes in normal development and homeostasis, as well as during pathogenic events such as cancers.

Chapter 1: Introduction.

Proteins account for more then 50% of the dry weight of most cells and they are viewed as the basic machinery inside the cell. They are involved in structural support, defense against foreign agents, acceleration of chemical reactions (enzymatic activity), signaling, movement and the storage and transport of substances within the cell.[85] Historically, proteins have been studied using biochemistry, a discipline that distinguishes itself by the first act of its practitioners: converting cells into a soup. Studies of the molecules purified from this soup serve to define the spectrum of possible functions that proteins could carry out within the living cell. Biochemistry remains a vibrant field but the reductionism at its heart creates limitations. Proteins are in an unnatural environment and under such artificial conditions, biochemists have to continually question the relevance of their *in vitro* systems to the cellular environment.

> *"A clever engineer can make a vacuum cleaner from the wreck of an automobile, but this does not show that cars contain vacuum cleaners."* Green.[45]

Geography, movement and chemistry are what provide proteins with their extraordinary capability to regulate processes in living cells, such that only *in vivo* experiments can determine where, within the spectrum of possible functions, the actual biological functions of proteins reside. Thus, manipulating proteins *in vivo*, in the context of relevant physiochemical conditions, is a necessary approach to complement biochemistry.

Classical genetics with techniques such as mutagenesis, over-expression, knock-out, conditional alleles and others,[66] has proven to be a powerful tool for the study of proteins inside cells. Recently, these approaches have been complemented by the discovery of fluorescent proteins.

Green Fluorescent Proteins.

The green fluorescent protein, isolated from *Aequorea victoria*, is a β-barrel-shaped protein that contains an amino acid triplet (Ser-Tyr-Gly) which undergoes a chemical rearrangement to form a fluorophore.[69] GFP is most celebrated for its ability to act as a genetic fluorescent marker, being directly visible in the living cell and requiring no fixation, substrates or co-enzymes.[137] This characteristic has allowed for visualizing the location and trafficking of proteins *in vivo*,[137] labeling of many sub-cellular structures[120] and discovery of genes whose products have a specific intracellular location.[63] Because of this success, both random and rational mutagenesis have been

used to create GFP variants with new colors, improved folding, varied extinction coefficients and quantum yields[137] which, together with orthologues fluorescent proteins such as the sea coral protein DsRed,[11 98] have increased the breadth of genetic fluorescent markers.

New GFPs can now report on the chemical environment of a protein thanks to their engineered sensitivity to acidity,[82 89] or concentration of ions such as Cl^-,[73] Ca^{2+}[101] and Zn^{2+}.[10] Importantly, the availability of fluorescent proteins with diverse colors permits the study of protein-protein interactions through the use of fluorescence resonance energy transfer (FRET).[94 101]

One of the most recent development is a fluorescent protein whose spectral properties change with time. The "fluorescent timer" protein was generated by random mutagenesis of the red fluorescent protein DsRed.[130] This protein initially produces a green-emitting fluorophore (similar to that of GFP), which over time undergoes an oxygen-dependent autocatalytic reaction to create a red-emitting fluorophore. The rate of color conversion is independent of protein concentration and can be used to trace time-dependent activation and down-regulation of the promoters of target proteins.[130]

Although powerful, there are limitations to the use of GFP-fusion reporters. These include the potential perturbation to the protein of interest (due to GFPs' size of 238 amino acids), significant delay between protein

synthesis and the development of fluorescence (~1 hour), limited points of attachment, a complex photoisomerization and a small spectral range.[137] The introduction of unnatural moieties into proteins *in vivo* could overcome some of these limitations and open the possibility of exploiting the orthogonality and selectivity that chemistry has to offer. A unique probe on a protein of interest allows it to be easily identified from a multitude of similar species, and permits specific reactions between proteins to take place.

Chemical modification of proteins in vivo.

Chemical modification of proteins in living cells originally involved the non-specific modification of membrane proteins with fluorescent molecules that were covalently attached to these proteins through their side chain amino or sulfhydryl groups. Initially such experiments were performed to determine what parts of the membrane proteins were extra-cellular,[40] but later they permitted the study of protein diffusion constants in the plasma membrane using fluorescent recovery after photo-bleaching (FRAP).[6]

Selective Pressure Incorporation.

More recently, an *in vivo* method termed selective pressure incorporation[18] was developed to exploit the promiscuity of wild-type bacterial synthetases. Inducing expression of a recombinant protein in an auxotrophic

strain grown in minimal media with an unnatural amino acid analog, results in the production of the protein containing this moiety. In general this method is simple to carry out, and can provide large quantities of engineered proteins. However, major limitations include the replacement of all sites corresponding to a particular natural amino acid throughout the protein, the variation in the extent of incorporation of the natural and unnatural amino acid, the toxicity of some analogues and the use of only close structural analogs of the common amino acids. Moreover, to date it has been applicable only to bacteria.

Unnatural Amino Acid Incorporation Through Nonsense Suppression.

Many of the limitations of selective pressure incorporation were by-passed with the creation of unnatural amino acid incorporation through nonsense suppression.[106] This technique uses a suppressor transfer RNA, chemically acylated with an unnatural amino acid, that is inserted into proteins in response to an amber stop codon. This new technology was introduced into an *in vivo* setting using *Xenopus* oocytes for protein synthesis[107] (Figure 1.1).

Nonsense suppression methodology has permitted *in vitro* and in some cases, *in vivo* analysis of structure-function relationships,[33 41] incorporation of biophysical probes,[32 78] incorporation of photo-reactive side chains and

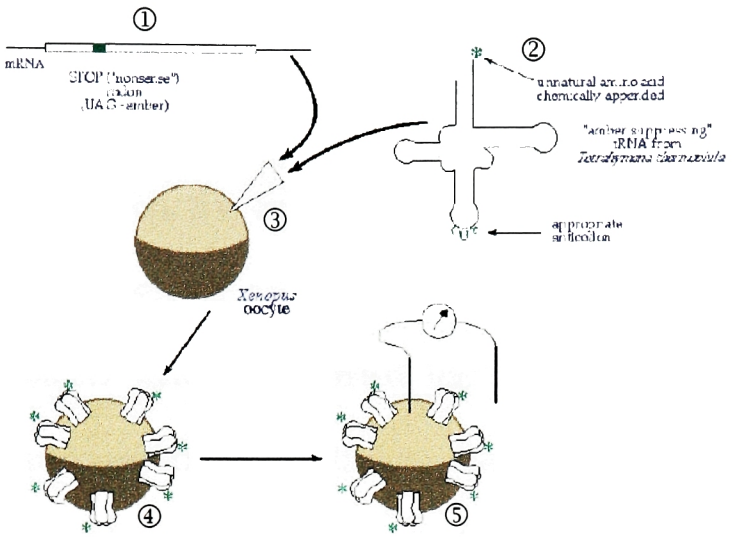


Figure 1.1. General outline for the incorporation of unnatural amino acids through nonsense suppression into proteins in vivo. ① The gene of the protein is cloned with a "stop" codon at the site of interest. ② A tRNA with the appropriate anti-codon and unnatural amino acid (*) is synthesized. ③ Both are co-injected into a cell, most commonly an oocyte. ④ The cell produces the protein with the unnatural amino acid incorporated, and in the case of ion channels, the protein is transported to the cell surface ⑤ where it can be studied by electrophysiology.

backbone mutations.[34 83] Almost 100 unnatural amino acids have been integrated into proteins, but the technique remains challenging. Non-sense suppression requires stoichiometric amounts of the amino acylated tRNA (which is difficult to synthesize), the suppression efficiency is imperfect, the yields are poor and the *in vivo* system used has to be amicable to the methodology (which requires injection of both the tRNA and mRNA).[42] The latter draw-back has impeded the study of a protein within its natural

environment (*e.g.* a mammalian protein in mammalian cells). However, in a monumental achievement, the Schultz group has made progress towards overcoming this limitation, by expanding the genetic code of *Escherichia coli* (*E. coli*).[145] Three additional and orthogonal components for the protein biosynthetic machinery were incorporated into the bacteria: a novel tRNA-codon pair, an aminoacyl-tRNA synthetase, and an amino acid. This new set of components did not cross talk with the natural machinery of *E. coli*, and it had a fidelity greater then 99% when incorporating the synthetic amino acid O-methyl-L-tyrosine into proteins in response to an amber nonsense codon.[145] Since this first report, a series of unnatural amino acids have been incorporated into the genetic code of *E. coli*: O-allyl-L-tyrosine,[159] p-azido-L-phenylalanine,[25] L-3-(2-naphthyl)alanine[146] and p-benzoyl-L-phenylalanine.[24] The latter was incorporated into the dimeric protein glutathione S-transferase (GST) and efficient cross-linking (>50%) of the protein subunits resulted upon irradiation of the bacteria. Such results start to shed light on the possibilities that could emerge when increasing the repertoire of functional groups that can be introduced *in vivo*. Although, nonsense suppression has made great advances, it is still unavailable in mammalian cells, and is restricted to the use of only two redundant stop codons.

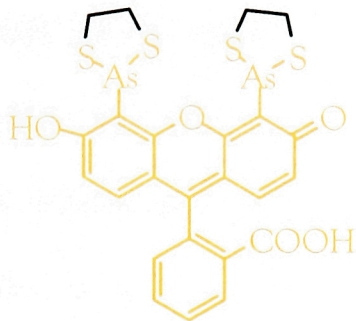


Figure 1.2. Structure of FlAsH-EDT$_2$. The structure highlighted in orange attaches to proteins tagged with the tetra-cysteine motif.

FlAsH

An alternative way to attach unnatural chemical moieties onto proteins was recently developed.[65] A small membrane permeable ligand FlAsH (4', 5'-bis(1,3,2-dithioarsolan-2-yl) fluorescein) (Figure 1.2) that binds the four cysteines in the sequence CysCysXxxXxxCysCys (where Xxx is any amino acids except cysteine) was developed.[65] The small polypeptide to which FlAsH binds can be incorporated into proteins.

FlAsH-EDT$_2$ (1,2-ethanedithiol) is practically non-fluorescent becoming over 50,000 times more fluorescent upon exchanging the EDT moieties for the tetra-cysteine sequence. This polypeptide is believed to be in a β-hairpin conformation[1] and the rigid spacing of the two arsenates in FlAsH enables it to bind with high affinity and specificity to the motif introduced in target proteins. Addition of micro-molar levels of 1,2-dithiols, such as EDT, can minimize FlAsH toxicity and nonspecific labeling, while milli-molar

concentrations can strip FlAsH off the target proteins. The small size of the genetic tag (6-10 amino acids) and relative ease of the synthesis of FlAsH has encouraged the synthesis of new derivatives with different properties. To date FlAsH has been modified to have red and blue emissions, photosensitize the polymerization of diaminobenzidine (to produce a localize precipitate useful for electron microscopy) and be membrane impermeable.[1 105] FlAsH can also survive analysis of protein by SDS-PAGE, can serve as a protein purification tool analogous to a His-tag[1 132] and the conformational rigidity the FlAsH-peptide adducts should make them ideal for anisotropy studies.[1] Recently, another methodology to chemically modify proteins *in vivo* was reported. In this strategy, covalent labeling with a small molecule is achieved through the use of an engineered human DNA repair protein O^6-alkylguanine-DNA alkyltransferase (hAGT), which irreversibly transfers the alkyl group from its substrate (O^6-alkylguanine-DNA) to one of its cysteines. The substrate specificity of hAGT was reduced, allowing the ligation of unnatural alkyl groups onto hAGT fusion proteins.[79]

The appeal of all these methodologies is the ability to incorporate non-genetic coded molecules into proteins *in vivo*, maximizing chemically favorable properties and minimizing possible disruption in normal protein behavior. A novel approach in which a synthetic molecule is ligated to a protein (i.e. semi-synthesis) inside living cells could increase the diversity of

chemical modifications for probing cellular proteins and should complement methods already in existence. The bases for such an approach are protein splicing and protein transduction domain (PTD) peptides.

Protein Splicing.

Protein splicing is an autocatalytic, post-translational process in which an intervening sequence, the intein, is removed from a host protein, the extein.[115] The first intein was discovered in 1990 as a "spacer" within the *Saccharomyces cerevisiae* vacuolar ATPase (*Sce* VMA) gene that was absent in other VMA homologs.[70 76 127] Since then, over 100 putative inteins have been identified and cataloged in InBase, the on-line intein registry and database (http://www.neb.com/neb/inteins.html).[114] Formation of a native peptide bond between the ligated exteins is a key distinguishing feature of intein-mediated protein splicing. To date, only a small number of inteins

Figure 1.3. General layout of an intein. Mini-inteins have a small linker region instead of the endonuclease. Conserved residues are shown below in bold. The intein splicing regions (blue boxes) form a single domain that mediates protein splicing.

have been experimentally shown to splice, while the majority have been found by homology searches of whole genome sequences (Figure 1.3).[114 116] The majority of inteins are bi-functional proteins that also encode homing endonuclease signature motifs. Homing endonucleases were originally shown to initiate mobility (or homing) of intron genes into sites in homologous genes by creating double-stranded breaks at the homing site.[43 60] Several inteins have demonstrated endonuclease activity and the *Sce* VMA intein gene has been shown to be mobile in yeast.[61] Biochemical analysis indicates that splicing and endonuclease active sites reside in different regions of the intein.[71] This has been further established with the crystal structure of the *Sce*. VMA intein, which shows separately folded sub domains (Figure 1.4).[117]

Inteins are present in unicellular organisms from all 3 phylogenetic domains: Archaea, Eukarya and Eubacteria. The absence of inteins in multicellular organisms may be limited by the inability of mobile intein genes to invade germ line cells. Although the role of inteins in early evolution is unknown, several groups have suggested that they are selfish DNA elements, or could be involved in enzyme evolution by peptide or domain ligation and shuffling.[113] It has also been proposed that inteins could control maturation of the protein wherein they reside,[118] however this is unlikely, as splicing is rapid under physiological conditions, and precursors cannot be isolated in organisms that have inteins.

Figure 1.4. Crystal structure of Sce. VMA. The endonuclease is represented in red, and the intein splicing regions in blue (see also Figure 1.3 - note, in Figure 1.3 the endonuclease is presented in black).

The standard protein splicing mechanism was first defined using the Psp-GBD Pol intein[71 154] and was later confirmed with the *Sce* VMA intein.[28 29] Protein splicing requires four nucleophilic displacements mediated by three splice junction residues (Figure 1.5). The standard protein splicing reaction begins when the Ser or Cys at the intein amino-terminus undergoes an acyl rearrangement, replacing the amino-terminal splice junction peptide bond with an α-(thio)-ester bond (Figure 1.5 ①). In the second step, 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

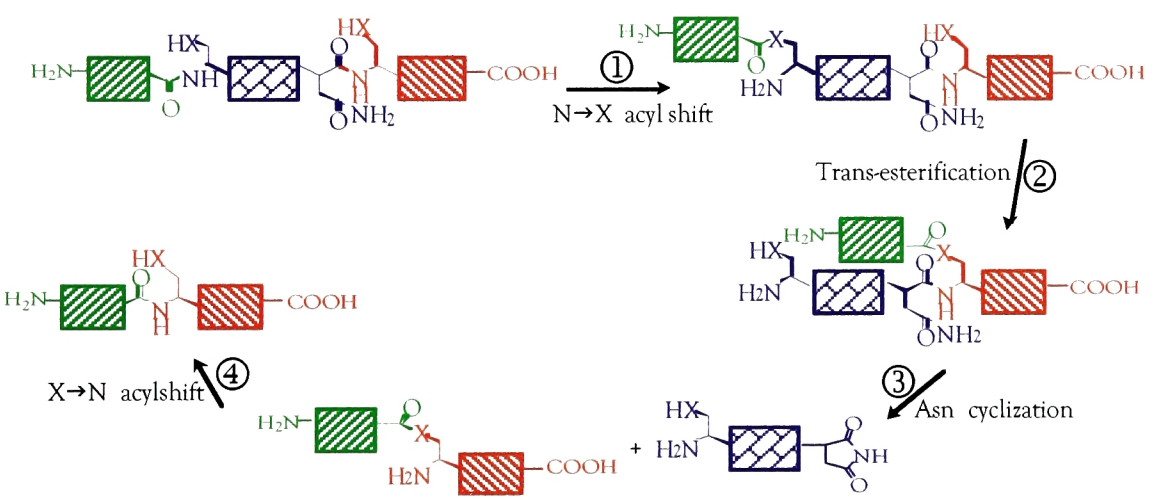


Figure 1.5. Standard protein splicing mechanism. "X" represents the oxygen or sulfur atom in Ser, Thr or Cys and the steps are explained in the text. The succinimide ring may be hydrolyzed to reform Asn or isoasparagine. The N-terminal extein, C-terminal extein and the intein are shown in green, red and blue respectively.

the exteins and formation of a branched protein intermediate (Figure 1.5 ②). The intein is cleaved away from this intermediate when cyclization of the intein carboxy-terminal Asn breaks the adjacent peptide bond at the carboxy-terminal splice junction (Figure 1.5 ③). In the absence of the intein, a spontaneous rearrangement of the α-(thio)-ester linkage between the exteins occurs, resulting in the formation of the more stable peptide bond (Figure 1.5 ④). No cofactors or exogenous proteins are required for splicing activity. An alternative mechanism for protein splicing has also been found for three families of inteins that begin with Ala. In this non-canonical pathway the C-extein nucleophile attacks the peptide bond at the N-terminal splice junction directly.[117 153]

Inteins function like regular enzymes, requiring facilitating groups positioned in three dimensional space to activate nucleophiles and electrophilic centers. However, unlike normal enzymes they only act on a single substrate, the precursor in which they exist. This is further highlighted when the precursor is split and splicing occurs in *trans* (see below), unequal molar ratios of fragments result in the accumulation of the fragment in excess and not in turnover (the limiting fragment does not react with multiple partners).[49 87 99 128 152]

Several groups have independently demonstrated that protein splicing precursors can be cut into two pieces that individually have no activity, but

when combined they will non-covalently associate to give a functional intein (Figure 1.6).[87 99 128 152] However, *in vitro trans*-splicing often requires denaturation and renaturation of the isolated precursor fragments.[99 128] The denaturants may be required for multiple reasons: many intein fragments are insoluble when expressed separately,[99 128] possibly due to improper folding or exposure of hydrophobic residues in the absence of the complete intein, alternatively, aggregation or insolubility could be due to the extein portion of the precursor fragment.[128]

Protein *trans*-splicing is a bimolecular reaction. However, once the intein fragments reassemble, the reaction is, in effect, a uni-molecular reaction. Since non-allelic inteins have little sequence identity, different split inteins only re-associate with their homologous partners and not with

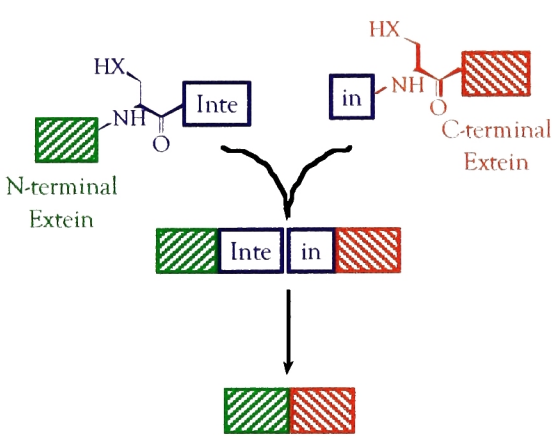


Figure 1.6. Schematic representation of protein trans-splicing. The N-terminal extein, C-terminal extein and the intein are shown in green, red and blue respectively. The intein is fragmented and inactive, but upon association activity is regained.

heterologous partners.[109] Most recently, the protein *trans*-splicing area has received a considerable boost through the discovery of the naturally occurring split *Synechocystis* sp. (*Ssp*) DnaE intein.[151] The *dnaE* gene encodes the catalytic subunit of the replicative DNA polymerase and is an essential gene. The two *dnaE* gene fragments are over 700 kilo bases (kb) apart on opposite strands of the *Ssp*. PCC6803 chromosome. Significantly, this split intein does not require denaturants for reconstitution to initiate splicing. This feature makes the *Ssp*. DnaE intein especially well suited for protein engineering applications.

Applications of Protein Splicing.

Understanding the protein splicing mechanism has led to the design of a number of mutant inteins which have significant utility in protein science and in particular, peptide ligation.[62] Proteins expressed in-frame and amino-terminal to a mutated intein (Ans→Ala) that can only participate in the first step of splicing (*i.e.* the N→S acyl shift, Figure 1.5 ①), can be cleaved by thiols via an intermolecular *trans*-thio-esterification reaction.[27] It is also possible to mutate an intein such that after cleavage at the carboxy-terminal splice junction one obtains the protein of interest with an amino-terminal Cys.[97] Importantly, the use of these engineered inteins permits the production of large recombinant building blocks for use with native chemical

ligation, and allows for mixing of recombinant and synthetic building blocks in chemical protein synthesis.

Native chemical ligation is a technique that allows the chemo-selective addition of two unprotected peptides at physiological pH, with the resulting polypeptide containing a native backbone at the point of ligation.[36] In this technique a peptide containing a C-terminal α-thioester group reacts with a second peptide containing an amino-terminal cysteine. Thiol exchange yields an intermediate linked through an α-thioester, which spontaneously rearranges to a peptide bond. Native chemical ligation can be performed in the presence of all the functionalities commonly found in proteins, including free cysteine sulfhydryls. Cysteine residues that are not N-terminal cannot spontaneously rearrange to give an amide bond, and due to the reversible nature of the initial *trans*-thio-esterification step, these side-products are simply converted back to starting materials in the presence of exogenous thiols. This advantage has made native chemical ligation the most successful and commonly used chemical ligation approach.[35] Native chemical ligation (Figure 1.7) is the last step of protein splicing (Figure 1.5 ④) and their combination allowed the development of the protein semi-synthesis technique expressed protein ligation (EPL).

EPL permits the ligation of two polypeptides, at least one of which is recombinant. This methodology is very flexible, allowing ligation of

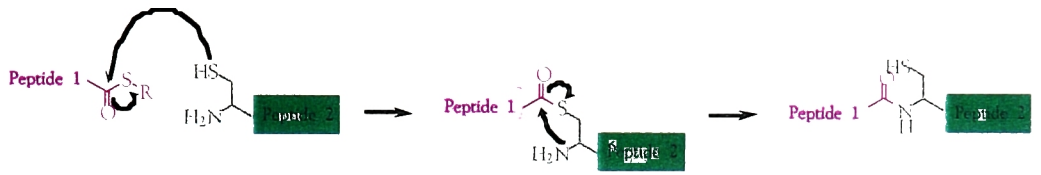


Figure 1.7. Mechanism of native chemical ligation. The mechanism is explained in the text.

polypeptides to take place under a variety of conditions, *e.g.* in the presence of denaturants, organic solvents, and detergents.[7] In addition, the amino-terminal α-thioester fragment can contain essentially any residue next to the α-thioester.[67] EPL has been used for the incorporation of probes into proteins *in vitro* (*e.g.* unnatural amino acids, fluorophores, post-translational modifications),[7] the isotopic segmental labeling of multi-domain proteins,[155] and the backbone cyclization of peptides and proteins.[20-22]

Protein *trans*-splicing has been used in a similar manner to EPL. The process has found a variety of applications *in vitro*, including protein semi-synthesis,[87] segmental isotopic labeling,[157] control of protein function[102] and cyclization or polymerization of proteins.[48] In addition, split-inteins have been used to cyclize proteins *in vivo*[48 126] and to study protein-protein interactions in prokaryotic and eukaryotic cells.[111 112] The latter approach was first described in 2000 and allows protein-protein interactions to be identified in a manner reminiscent of the yeast two-hybrid system. Two proteins known to interact (calmodulin and its target peptide M13) were

expressed with one half of the *Sce* VMA intein and one half of GFP. Interaction of calmodulin and M13 resulted in the generation of GFP via protein *trans*-splicing and hence the interaction could be detected by fluorescence techniques.[111] This technique has been applied to other systems[110] and adds to the growing number of reporter protein complementation approaches designed to rapidly identify novel protein-ligand interactions.[133] Furthermore, it is attractive because it should be amenable to any cell type.

Protein splicing has many features that make it amenable for the semi-synthesis of proteins inside cells, namely; inteins are promiscuous with respect to their flanking sequences,[72] and have demonstrated activity in cells.[111] In addition, EPL is a powerful technique for *in vitro* manipulation of proteins,[72] and one could expect an analogous technique *in vivo*, where a chemical probe is ligated to a protein inside cells, should also prove useful.

Protein Transduction Domains.

Passive diffusion of chemical probes into the cell is limited by the physical properties of the compound, such as charge and lipid solubility. There also exists a molecular weight cut-off that limits the size of the largest lipid-soluble substances able to cross the membrane to ~500 Da.[86] Though a variety of methods exist to overcome this limitation (*e.g.* electrophoresis,

microinjection, vesicle delivery), they have been difficult to work with and can damage cells. It was recently discovered that several small regions of proteins called protein transduction domains (PTD) possess the ability to transverse biological membranes efficiently.

Two groups independently reported protein transduction, demonstrating that the Tat protein from the HIV-1 virus was able to enter cells when added to the surrounding media.[56 64] Subsequently, several other proteins with transducing capabilities were identified, including the *Drosophila* homeotic transcription factor ANTP (encoded by the *antennapedia* gene)[75]and the herpes simplex virus type 1 (HSV-1) VP22 transcription factor.[44] These are currently are the most popular.

Table 1.1. Amino acid sequence of the first three characterized PTDs.

PTD	Amino acid sequence
HIV-1 Tat	^{47}YGRKKRRQRRR^{57}
ANTP	^{43}RQIKIWFQNRRMKWKK^{58}
HSV VP22	^{207}DAATATRGRSAASRPTERPRAPARSASRPRRPE^{300}

The domains responsible for transduction in Tat, VP22 and ANTP have been identified. The minimal Tat transduction domain consists of residues 47-57,[50 52 141] whereas in VP22 and ANTP, residues 267-300[44] and

residues 43-58,[39] respectively, are required for transduction (Table 1.1). Interestingly, all three of these proteins are involved in transcriptional regulation and it is the PTDs that make contact with nucleic acids. The commonality between each PTD is the presence of basic amino acids (arginine and lysine), which might be important for lipid interactions, penetration of the membrane, or both. Protein-structure-predicting algorithms strongly suggest that the Tat PTD can adopt an alpha helix conformation,[91] and the ANTP PTD is known to be helical when present in the core homeodomain. However, circular dichroism and NMR spectroscopy have indicated that the Tat basic domain is unstructured,[19 92] and the insertion of prolines into ANTP does not appear to inhibit transduction,[37] suggesting that the helical conformation is not required for the PTDs activity.

Although the mechanism of PTD-mediated transduction is unknown, it is clear that it does not occur through a classical receptor-, transporter- or endosome-mediated fashion.[95] Indeed, both full-length PTD fusion proteins and PTD peptides rapidly and efficiently transduce into cells at 4°C, suggesting that a form of absorptive endocytosis is not responsible.[37 141] Additionally, treatment of cells with drugs that impair cellular transport processes does not inhibit transduction.[44] The mechanism of protein transduction remains a difficult question to address experimentally, but progress has recently been made. In general, transduction efficiency appears

to correspond to the number and location of arginine residues in the PTD sequence. When a series of arginine-rich peptides (derived from 14 RNA- and DNA-binding proteins) were screened, they were all able to *trans*-locate through the cell membrane.[57] Substitution of basic residues in Tat (49–57) produces a significant (70–90%) decrease in cellular uptake.[148] Furthermore, the guanidine head-group of the arginine side-chain seems to be a critical structural feature for transduction. Homo-polymers of citrulline, an arginine isostere in which the nitrogen of the guanidine group is replaced with oxygen, show no transduction activity.[148] Moreover, increasing the distance between the guanine group and the peptide backbone had a positive effect on cellular uptake.[100 148] These observations suggest that the guanidine group is involved in the interaction of the PTD with the cell membrane. The guanidine group could be forming stable bi-dendate hydrogen bonds with anions such as phosphate or sulfate[100] or the delocalized positive charge could be making membrane interactions more amicable. Thus, the internalization of arginine-rich PTDs seems to occur through a specific interaction between the arginine head-group and charged members of the cell membrane. These experiments are in accordance with the recent finding that cell-surface heparin sulfate (HS) proteoglycans are required for protein translocation (cells genetically impaired in the biosynthesis of fully sulfonated HS proteoglycans were selectively impaired for transduction).[138] The

ubiquitous presence of HS could explain why Tat peptides and conjugated proteins are able to enter a wide variety of cells. However, it is still unknown how PTD-linked macromolecules are internalized following the interaction with the cell surface, and furthermore the HS PTD interaction fails to explain the ability of these peptides to shuttle out of cells (no HS are present in the cytosolic side of the membrane).

A significant barrier to the uptake of hydrophilic proteins is imposed by the hydrophobic interior of the lipid membrane, and the uptake mechanism of PTDs must overcome this obstacle. The formation of membrane channels is unlikely to account for the internalization of large PTD-conjugated macro-molecules. In addition, Tat-mediated transduction does

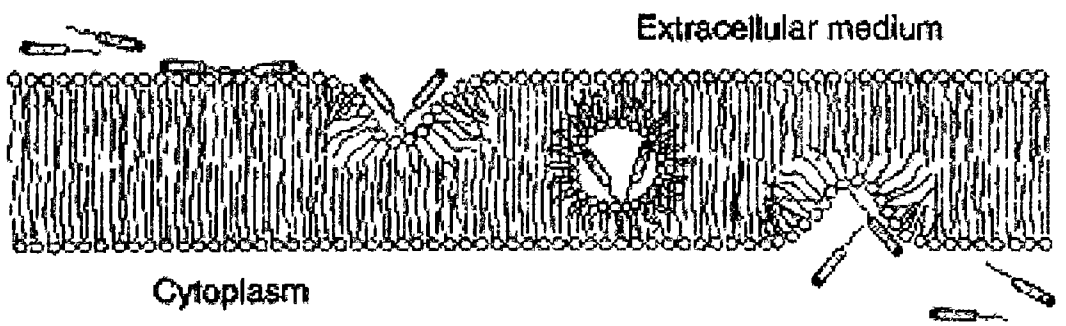


Figure 1.8. Inverted micelle formation model for the translocation of PTD across membranes. The peptides interact with the charged phospholipids on the outer side of the membrane. Destabilization of the bilayer results in the formation of an inverted micelle that travels across the membrane and eventually opens on its cytoplasmic side. In such a model the peptides never leave an aqueous environment.

not appear to involve any disruption of the plasma membrane, as uptake of unrelated non-conjugated peptides present in the incubation media could not be observed.[141] The latter observation demonstrates a key asset of PTD-mediated transduction: the ability to deliver a cargo, avoiding the movement of non-linked molecules into or out of the cell. NMR data has also demonstrated that the backbones of PTD peptides are flexible, presumably allowing them to adapt to the concave surfaces of micelles.[16]. Though a few models for translocation have been suggested, one of the most commonly cited is the inverted micelles translocation model[38] (Figure 1.8).

PTDs that are chemically cross-linked to heterologous proteins, such as antibodies and enzymes, have been shown to transduce into cells.[3 53] In addition, the molecules that PTDs can deliver are not limited to proteins[2] (Figure 1.9). In 1991 Frankel stated:

> *"Beyond the possible relevance for HIV infection, the efficiency of uptake suggests that Tat might prove a useful vehicle for delivering proteins or peptides into cells"* Frankel.[95]

In the PTD based delivery system, intracellular concentrations of the transduced protein are almost identical between cells and can be readily controlled simply by changing the amount of protein added to the culture media. Also, PTD fusion proteins are rapidly internalized allowing for

temporal control.[143] For example, the fusion protein can be added to cells in a manner that is synchronized with a particular phase of the cell cycle with activity occurring within minutes of protein addition. The same experiment performed with plasmid DNA transfections would be impossible because of the large delay between DNA uptake and protein expression. To date, a large number of proteins have been successfully transduced (Table 1.2).

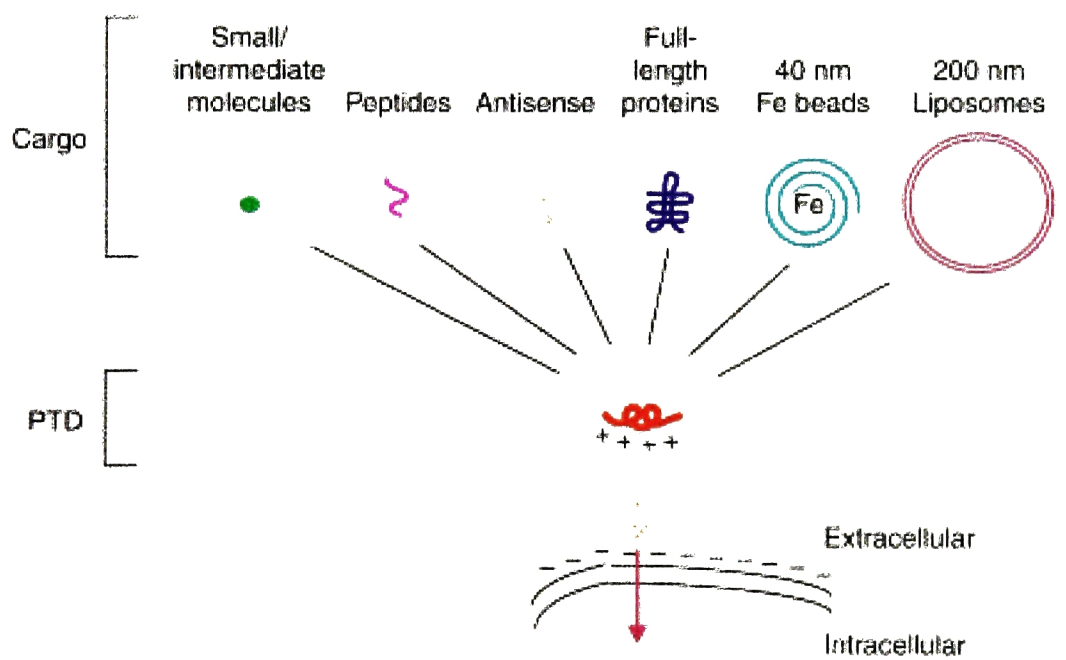


Figure 1.9. Utility of protein transduction technology. A wide variety of cargo has been covalently linked to arginine-dependent protein transduction domains (PTDs).

Table 1.2. Examples of proteins transduced with PTDs.

Protein	Reference
p16	52
p27	104
E2F	88
β-galactosidase	125
Cdk2	51
Cu/Zn superoxide dismutase	84
enhanced GFP	23
Cre recombinase	74
p53	150
caspase-3	142
PEA-15	46

Protein transduction has also been shown to be advantageous over viral transgene delivery *in vivo* (*e.g.* whole animals). Injection of a Tat-conjugated protein results in transduction of 100% of the cells in a concentration-dependent manner, whereas viral delivery achieves only 30-50% efficiency.[13 14]. Although PTDs have been used extensively, they are not silver bullets. For example, proteins transduced with TAT are denatured,[23] and the translocation activity of ANTP is greatly reduced with increased cargo size.[119] These latter characteristic may not be ideal for all applications, but PTDs are well suited to deliver small, un-structured synthetic molecules

into cells. In conjunction with protein splicing, PTDs presented a viable route to investigate semi-synthetic protein synthesis *in vivo*.

Thesis Objectives

The objective of this thesis was to explore the union of PTD and protein splicing in order to achieve protein semi-synthesis in living cells. First, PTD peptides were evaluated as a means to introduce *in vitro* manipulated peptides and proteins into cultured eukaryotic cells. Then, protein *trans*-splicing was explored to address the possibility of *in vivo* protein semi-synthesis in cells. Lastly, there was a slight divergence from the main focus of the thesis for the exploration of the utility of semi-synthetic proteins as a tool to answer biological questions in the context of whole animals.

Chapter 2: Results and Discussion.

Incorporation of chemically manipulated peptides and proteins into cultured mammalian cells using PTDs.

A key feature for the development of protein semi-synthesis inside cells is the need to shuttle the synthetic component into the cytosol. PTD peptides have the ability to translocate through cellular membranes and transport cargo molecules into the cytosol of cells (see chapter 1).[143] ANTP was selected for this thesis because its small size, 16 amino acids, makes it amenable for SPPS.

Three methods are typically used for coupling the cargo molecule to the PTD. The cargo may be attached via a disulfide bond[136], chemically synthesized in tandem[68] or a fusion protein comprising the PTD and cargo can be expressed in bacteria and purified[15] (Figure 2.1).

To explore the translocation activity of ANTP in preliminary experiments, the PTD was synthesized in tandem within a 70 amino acids model polypeptide, derived from DnaE, by Boc SPPS. The fluorophore tetramethylrhodamine (Rh) was attached to the ε-NH_2 group of Lys^{69} using an

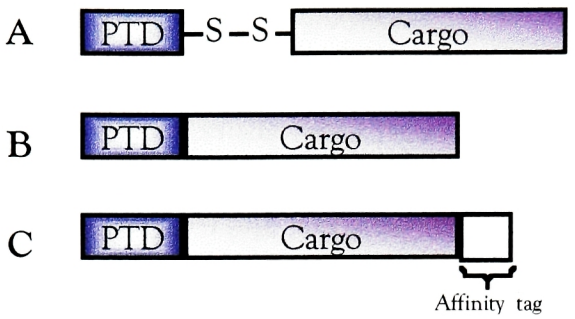


Figure 2.1. Methods for coupling PTD peptides to a cargo. (**A**) The cargo and PTD can be synthesized separately and attached through a disulfide bond. (**B**) The cargo and PTD can be chemically synthesized in tandem. (**C**) A fusion protein comprising the PTD and cargo can be expressed in bacteria and purified.

orthogonal (Fmoc/Boc) protection strategy. (Peptide 1, ANTP is highlighted in blue and Rh in red).

Peptide 1:

H-MVKVIGRRSLGVQRIFDIGLPQDHNFLLANGAIAANCFGCGRQIKI

WFQNRRMKWKKGGDYKDDDDKGK(Rh)G-NH_2

The purified peptide was added into the media of HeLa cells growing in a microscopy chamber to a final concentration of 1 nM. After 30 minutes incubation, the cells were washed with Hanks' balanced salt solution (HBSS) and imaged live with a fluorescence microscope. ANTP was able to translocate the 70mer (Peptide 1) into cells (Figure 2.2 **A**) and over a period of an hour it accumulated in the nucleus (Figure 2.2 **B**). This nuclear

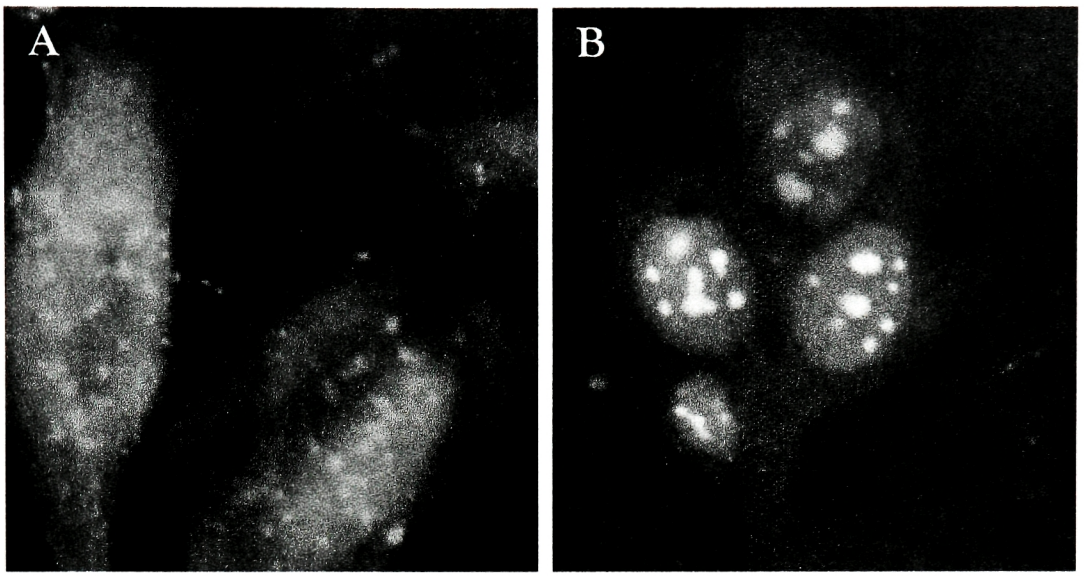


Figure 2.2. Localization of peptide 1. (**A**) Microscopy imaging of HeLa cells after 30 minutes incubation with 1 nM peptide 1 confirmed translocation. (**B**) An hour after removing peptide 1 from the media, the internalized peptide accumulates in the nucleus.

accumulation is well documented for PTDs, including ANTP,[119] most of which have nuclear localization sequences (NLS).

The nuclear localization of peptide 1, due to the tandem inclusion of ATNP, would ultimately lead to the separate compartmentalization of the components of semi-synthesis (nucleus *vs.* cytosol) and could prevent the process. Previous studies have shown that PTDs can internalize disulfide bound cargo and, once inside the reductive environment of the cytoplasm, the cargo is released from the PTD.[134 135] To verify this, a protocol was developed that allowed a cargo peptide, derived from DnaE (Peptide 2), to be coupled to ANTP through a disulfide bond (Figure 2.3 **A**).

Peptide 2 (later referred to as I_C-FLAG):

H-MVKVIGRRSLGVQRIFDIGLPQDHNFLLANGAIAANCFN

DYKDDDDKG-NH_2

ANTP and peptide 2 were separately synthesized by Boc SPPS and purified. As shown in figure 2.3 **A**, a protected cysteine residue was added at the N-

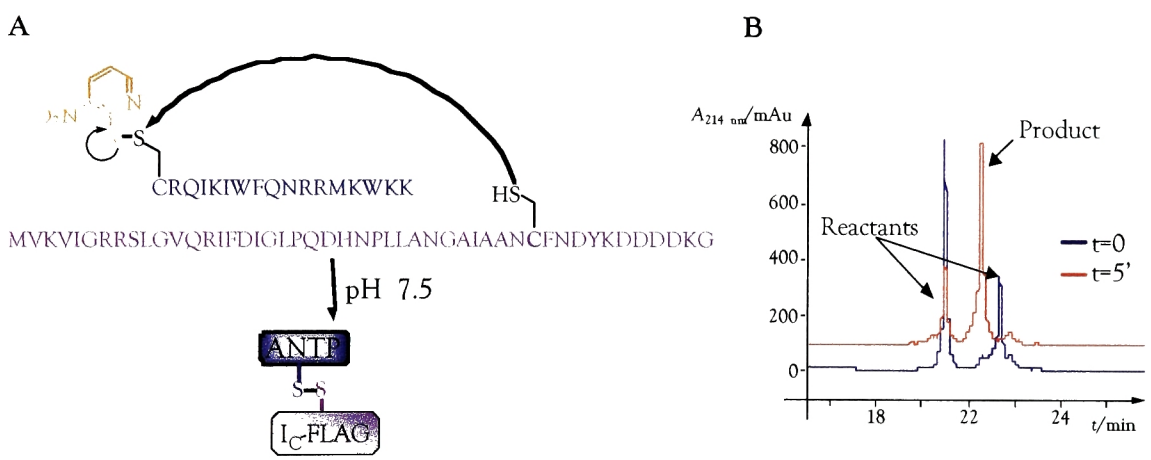


Figure 2.3. Coupling of peptide 2 to Cys(NPyS)ANTP. (**A**) General scheme for the synthesis of ANTP-S-S-Peptide 2 via a nucleophilic disulfide exchange. The amino acid sequence of peptide 2 and ANTP are depicted in purple and blue respectively. Highlighted in orange or in bold are he NPyS group on ANTP and the single Cys in peptide 2, respectively. (**B**) RP-HPLC analysis of the conjugation reaction between Cys(NPyS)ANTP and peptide 2. Linear gradient: 0-73% B (90% acetonitrile and 0.1% TFA in water) over 30 minutes.

terminus of ANTP. The sulfhydryl group was protected with a 3-nitro-2-pyridinesulfenyl group (NPyS, highlighted in orange in Figure 2.3 **A**), which acts as a leaving group upon nucleophilic disulfide exchange with the unique cysteine within peptide 2 (highlighted in bold). After purification, the peptides were dissolved together in degassed 50 mM Tris/HCl buffer pH 7.5 to a concentration of 1 mM in each. The nucleophilic disulfide exchange reaction was allowed to proceed for 5 minutes at room temperature (becoming visibly yellow as the NPyS group was being displaced), and then quenched with acid. This reaction allowed highly efficient (>95%) conjugation of the cargo to ANTP (Figure 2.3 **B**), and un-reacted starting material could be recovered for re-use.

Two disulfide conjugates were used to study if the disulfide conjugation system allowed delivery of the cargo into the cytoplasm; one bearing Rh tethered to the ANTP component (Rh-ANTP), and the other with Rh linked to peptide 2 (Peptide 2-Rh). These peptides were independently added onto HeLa cells (to a final concentration of 2 nM), which were then incubated for 30 minutes before being washed twice with PBS and fixed with 3.7% formaldehyde, followed by BSA blocking and nucleic acid staining with Hoechst. The fixed cells were then analyzed by fluorescent microscopy. As shown in Figure 2.4 **A**, the Rh-ANTP peptide was localized throughout the cells, including the nucleus, whereas peptide 2-Rh was predominantly localized in the cytoplasm (Figure 2.4 **C**). This result confirmed that the disulfide reduction was taking place inside the cells, dissociating

the cargo from the PTD and leaving it behind in the cytosol, where it can later react with other molecules.

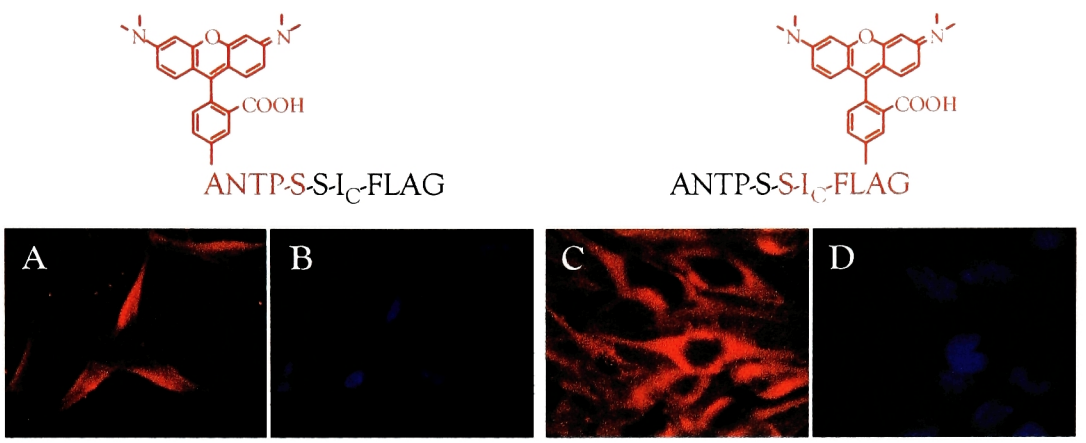


Figure 2.4. Cellular distribution of ANTP-S-S-Peptide 2 Rh-labeled peptides. HeLa cells were incubated with either Rh-ANTP-S-S-Peptide 2 (**A** & **B**) or ANTP-S-S-Peptide 2-Rh (**C** & **D**) at 2 nM for 30 minutes before being washed and fixed. Cells were then analyzed by fluorescence microscopy: Rh (**A** & **C**) and Hoechst (**B** & **D**) signals are shown.

Translocation and delivery of short synthetic peptides into cells was achieved using ANTP as the PTD. ANTP was able to transverse the membrane of mammalian cells imbedded in a 70mer. Furthermore, ANTP could translocate a disulfide linked 48mer, and dissociate from it inside the reductive environment of the cytosol. The cargo peptide was left in the cytoplasm of living cells to undergo further reaction (*e.g.* protein semi-synthesis).

Protein semi-synthesis in vivo using protein trans-splicing.

Protein *trans*-splicing has many of the attributes necessary for the semi-synthesis of proteins inside cells: inteins are promiscuous with respect to their flanking sequences,[72] split inteins are inactive when expressed individually,[87 99 128 152] they associate exclusively with their corresponding partner,[109] and, specifically for the *Ssp* DnaE system, the components have high affinity for each other.[96]

The methodology summarized in Figure 2.5 was envisioned to produce semi-synthetic proteins in living cells using protein *trans*-splicing.

The protein of interest is expressed in cultured cells with the first half of the naturally occurring *Ssp* DnaE split intein (I_N) fused to its C-terminus (Figure 2.5 ①). Then, a semi-synthetic polypeptide comprised of the second half of the intein (I_C) covalently linked to a synthetic probe and a PTD peptide, is added to the cell media (Figure 2.5 ②). The PTD peptide would deliver the I_C-probe into the cells, whereupon it can associate with its complementary half, I_N, to trigger protein splicing (Figure 2.5 ③). This would result in the removal of the intein and ligation of the probe to the selected protein through a normal peptide bond (Figure 2.5 ④). To explore this methodology, a model system was designed in which the protein of

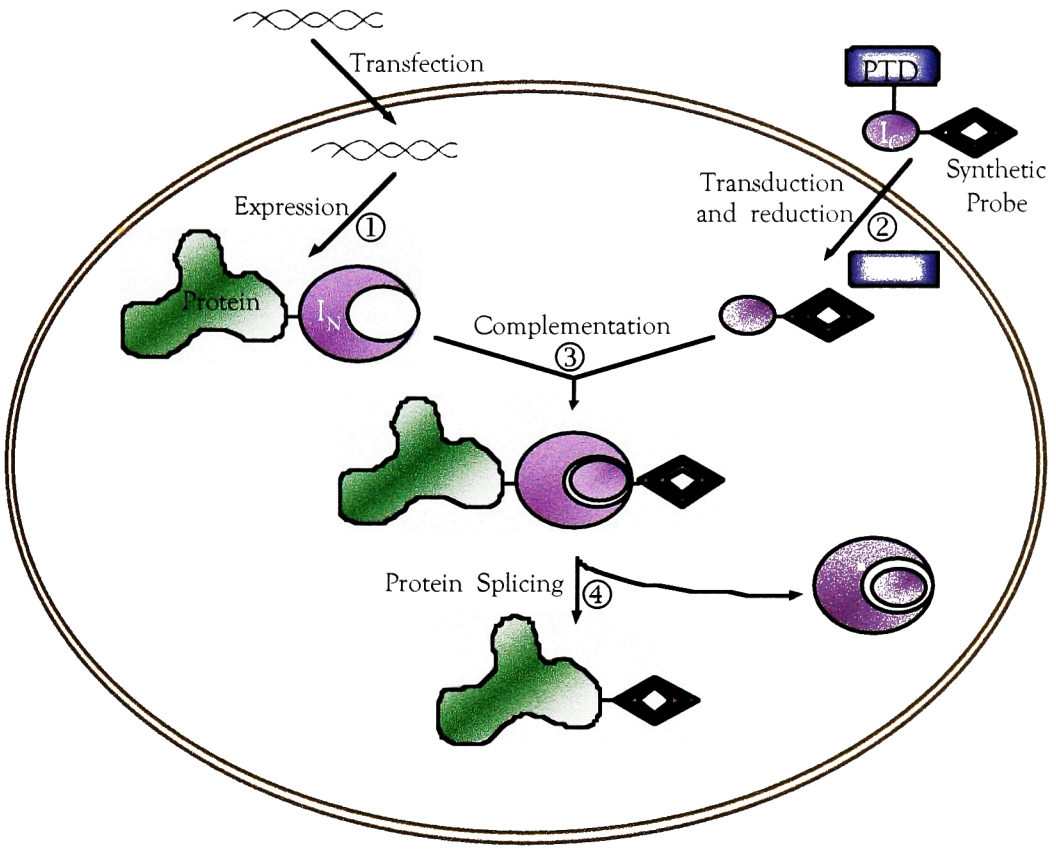


Figure 2.5. Principle of semi-synthetic protein trans-splicing in living cells. I_N is the first half of the naturally occurring *Ssp* DnaE intein, I_C is its second half and PTD is protein transduction domain.

interest was GFP and the probe was a short synthetic peptide, based on the FLAG epitope.

Initially, a bacterial expression plasmid was constructed encoding 6xHis-GFP-I_N, in order to produce large enough quantities to perform preliminary *in vitro trans*-splicing studies. A standard His-tag purification protocol was used to obtain the protein in relatively pure form (Figure 2.6).

To confirm that protein *trans*-splicing could take place between the components of the chosen model system, the bacterially expressed 6xHisGFP-I_N was reacted with I_C-FLAG at a final concentration of 100 μM. (I_C-FLAG = peptide 2. I_C and FLAG are highlighted in purple and red, respectively)

I_C-FLAG: (previously referred to as peptide 2)

H-MVKVIGRRSLGVQRIFDIGLPQDHNFLLANGAIAANCFN

DYKDDDDKG-NH_2

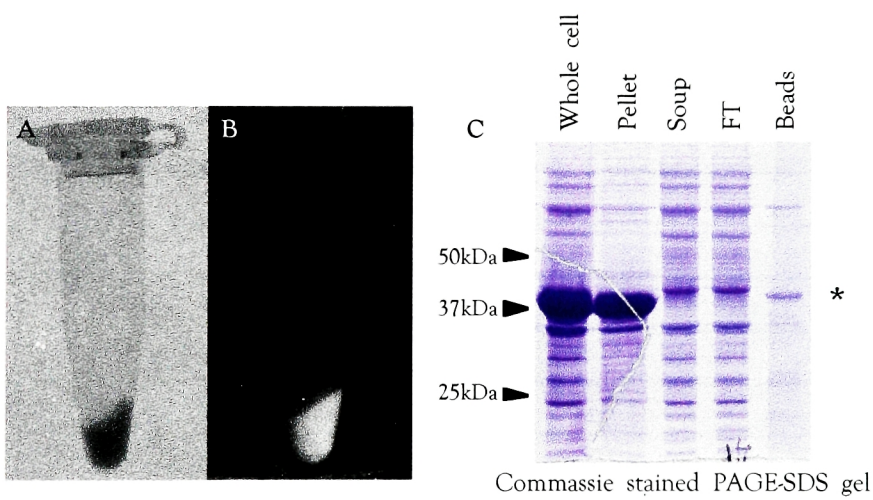


Figure 2.6. Expression and purification of 6xHis-GFP-I_N from E. coli. The level of expression induced in bacteria was very high. The bacterial pellet (**A**) was bright green when illuminated with UV light (**B**). (**C**) Analysis of the purification procedure by SDS-PAGE gel showed that most (>90%) of the protein was insoluble (6xHis-GFP-I_N is denoted by *), but enough was recovered to perform *in vitro* experiments.

The reaction mixture was incubated at room temperature for 5 hours and was then analyzed by SDS-PAGE (12%) followed by an anti-FLAG western blot (Figure 2.7). As shown in the figure, *trans*-splicing takes place between 6xHis-GFP-I_N and I_C-FLAG, yielding 6xHisGFP-FLAG as the only splice product (Figure 2.7 lane 2).

Having thus confirmed that semi-synthetic protein *trans*-splicing was taking place *in vitro*, and that I_C-FLAG could enter cells (see previous section), experiments to address semi-synthetic protein *trans*-splicing *in vivo* were undertaken. A mammalian expression plasmid was constructed encoding GFP-I_N driven by the human cytomegalovirus (CMV) constitutively active promoter (pGFP-I_N). To confirm expression in mammalian cells, CHO cells were transiently transfected with pGFP-I_N using Fugene. After 24 hours of

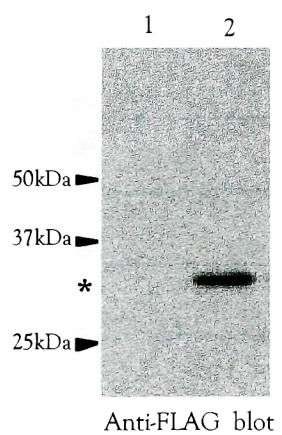


Figure 2.7. Anti-FLAG western of in vitro semi-synthetic protein trans-splicing. 6xHisGFP-I_N with no peptide added (lane 1) and 6xHisGFP-I_N with 100 μM of I_C-FLAG present (lane 2). The product, 6xHisGFP-FLAG, is denoted by *, and its expected mass is ~29 kDa.

expression the cells were washed twice with PBS and harvested. The cell pellet was resuspended in SDS sample buffer and loaded onto an SDS-PAGE gel (12%) for blotting against GFP (Figure 2.8), which verified the expression of GFP-I_N.

HeLa cells were also transiently transfected with pGFP-I_N to visualize the localization of the fusion protein. After 24 hours of expression the cells were washed twice with PBS and fixed with 3.7% formaldehyde, followed by BSA blocking and nucleic acid staining with Hoechst. As seen in figure 2.9 GFP-I_N was distributed throughout the cell.

With the individual pieces in place, semi-synthetic *trans*-splicing inside cells was investigated. 10 cm plates of CHO cells were transfected with pGFP-I_N. After

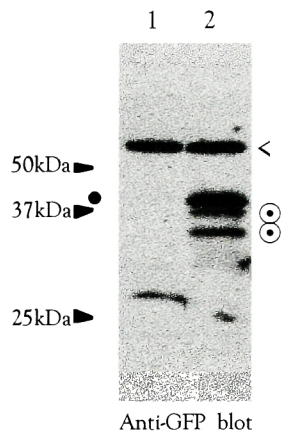


Figure 2.8. Mammalian expression of GFP-I_N. Anti-GFP western blot of un-transfected CHO cells (lane 1) and CHO cells transfected with pGFP-I_N (lane 2). GFP-I_N is denoted by •, and its expected mass is ~42 kDa. Background bands are denoted by < and proteolysis products by ⊙.

24 hours of transient expression, the cells were washed twice with 10 mL of PBS and fresh media (2.5 mL) containing 0.5% DMSO or 0.5% DMSO with the ANTP-S-S-I_C-FLAG peptide added to a final concentration of 100 μM. The cells were then incubated for three hours before being harvested. The cell pellet was resuspended in 1 mL of lysis buffer: PBS containing 1% (w/v) docecylmaltoside (DM), 10 mM iodoacetamide, 2 mM N-ethyl-maleamide and protease inhibitors. Lysis under alkylating conditions blocks all cysteine sulfhydryl groups, therefore preventing splicing *in vitro*. The soluble fraction was immuno-precipitated with anti-GFP or anti-FLAG beads and the captured proteins were loaded onto an SDS-PAGE gel and blotted against GFP or FLAG. This procedure allowed us to detect the presence of proteins containing either GFP or FLAG alone (i.e. the reactants or potential side-products) or both GFP and FLAG (i.e. the desired product). The results of this experiment are shown in Figure 2.10.

Figure 2.9. Cellular localization of GFP-I_N in HeLa cells. The image shows that GFP-I_N is distributed throughout the cell.

A band with the expected molecular weight for the semi-synthetic product (~29 kDa) appeared when transfected cells were treated with the ANTP-S-S-I_C-FLAG peptide (lanes 4, 8, 12 and 16). Importantly, this new band cross-reacted with both the anti-GFP and anti-FLAG antibodies confirming it was the desired product, GFP-FLAG. In contrast, this new band was not present in non-transfected and untreated cells (lanes 1, 5, 9 and 13), non-transfected cells treated with the ANTP-S-S-I_C-FLAG peptide

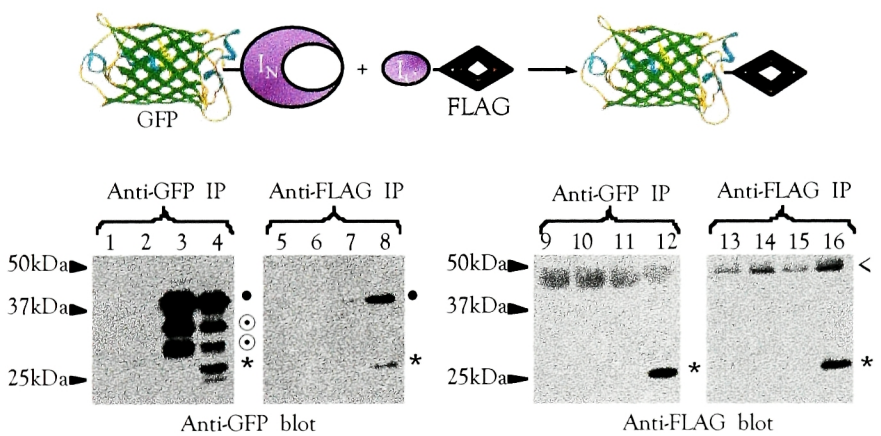


Figure 2.10. Semi-synthetic trans-splicing between GFP-I_N and I_C-FLAG in CHO cells [schematically shown in top panel]. Cells treated as indicated below were lysed and the soluble fraction immuno-precipitated and blotted as indicated in the panels. Non-transfected and untreated cells (lanes 1, 5, 9 and 13), non-transfected cells treated with 100 μM ANTP-S-S-I_C-FLAG (lanes 2, 6, 10 and 14), transfected cells expressing GFP-I_N with no peptide added (lanes 3, 7, 11 and 15) and transfected cells expressing GFP-I_N treated with 100 μM ANTP-S-S-I_C-FLAG (lanes 4, 8, 12 and 16). The ligation product GFP-FLAG is denoted by *, GFP-I_N is denoted by •, background bands are denoted by < and proteolysis products are denoted by ⊙.

(lanes 2, 6, 10 and 14) and transfected cells with no peptide added (lanes 3, 7, 11 and 15).

Interestingly, GFP-I_N co-immuno-precipitated with I_C-FLAG (lane 8) indicating a non-covalent complex between the two components. This suggests that protein splicing, and not split-intein association, is the rate-determining step in the process, which is in accordance with previous studies on *Ssp* DnaE-mediated protein trans-splicing.[96] Furthermore, co-immuno-precipitation of GFP-I_N but not of the proteolysis products, suggested that the

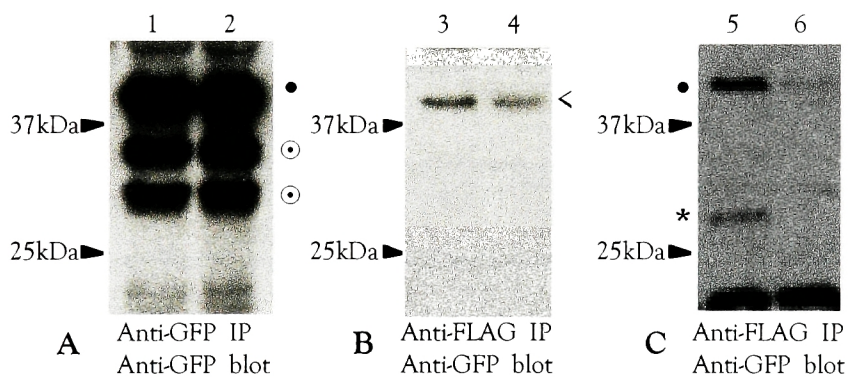


Figure 2.11. Semi-synthetic protein trans-splicing takes place in cells. **(A & B)** Transfected cells expressing GFP-I_N with no peptide added at the time of lysis (lane 1 & 3) and transfected cells expressing GFP-I_N treated with 50 μL of 100 μM I_C-FLAG at the time of lysis (lane 2 & 4). **(C)** Transfected cells expressing GFP-I_N treated with 100 μM ANTP-S-S-I_C-FLAG (lane 5) and transfected cells expressing GFP-I_N mixed at the time of lysis with non-transfected cells treated with 100 mM ANTP-S-S-I_C-FLAG (lane 6). The ligation product GFP-FLAG is denoted by *, GFP-I_N is denoted by •, proteolysis products are denoted by ⊙ and background bands are denoted by <.

degradation of the over-expressed protein was taking place C-terminally (at I_N) (lane 3 & 4, indicated by ⊙); proteolytic fragments were not associating or reacting with I_C-FLAG.

The anti-FLAG IP analysis also demonstrated that the *trans*-splicing reaction is specific as only one protein was tagged with FLAG (lane 16). The efficiency of the semi-synthetic protein *trans*-splicing reaction could not be easily determined, because GFP-I_N was being continually expressed (and so therefore not depleted) and because the ANTP-S-S-I_C-FLAG peptide was added in large excess. Finally, the generation of the semi-synthetic product provides

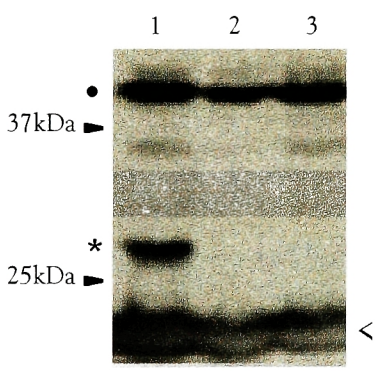


Figure 2.12. FLAG labeling of GFP is due to protein splicing. Transfected cells expressing GFP-I_N treated with 100 μM ANTP-S-S-I_C-FLAG (lane 1), transfected cells expressing GFP-I_N treated with 100 μM ANTP-S-S-I_C(S-R)-FLAG (lane 2) and transfected cells expressing GFP-I_N(Ala^{250}) treated with 100 μM ANTP-S-S-I_C-FLAG (lane 3). The ligation product GFP-FLAG is denoted by *, GFP-I_N is denoted by • and background bands are denoted by <.

unequivocal proof that I_C-FLAG was released from ANTP since this process (i.e. disulfide reduction) unmasks the reactive cysteine within I_C required for splicing.

Further control experiments were performed to completely rule out the possibility that splicing was taking place *in vitro* (*i.e.* after lysis). The issue was addressed in two ways. The first experiment consistent in adding 50 μL of 100 μM of I_C-FLAG onto the cells at the time of lysis (Figure 2.11 **A** & **B**) .The second experiment consistent in mixing transfected cells expressing GFP-I_N, but not treated with the peptide, with non-transfected cells treated with the peptide at the time of lysis (Figure 2.11 **C**). The two components of the *trans*-splicing system were present throughout the *in vitro* manipulations in both cases, such that, if *trans*-splicing was taking place after lysis, the activity should be detected. As shown in Figure 2.11, no splicing or split-intein association was detected (compare lanes 1 & 2, 3 & 4 and 5 & 6).

To confirm that product formation was due to protein splicing, an experiment was performed using inactive forms of I_N or I_C. The reactive cysteine in I_N was mutated to an alanine (GFP-I_N(Ala^{250}) and reactive cysteine in I_C was alkylated and an extra cysteine added onto the N-terminus of the peptide for disulfide linkage to ANTP (I_C(S-R)-FLAG). Although the inactive I_N and I_C analogs were able to associate with their complementary component, in neither case was *trans*-splicing observed (Figure 2.12, lanes 2

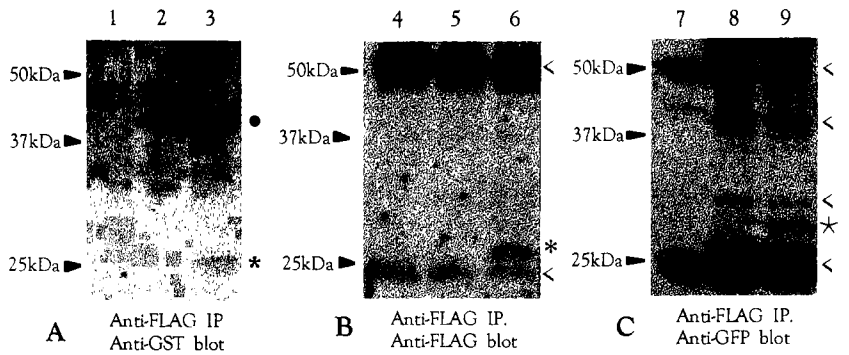


Figure 2.13. Generality of protein semi-synthesis in vivo. (A) Anti-GST western blot of 12% SDS-PAGE gel of un-transfected and untreated cells (lane 1), cells transfected with pGST-I_N left untreated (lane 2) and cells transfected with pGST-I_N treated with peptide (lane 3). The product GST-FLAG is denoted by *, and its expected mass is ~28 kDa. The precursor GST-I_N is denoted by •. (B) Anti-FLAG western blot of a 15% SDS-PAGE gel of un-transfected and untreated cells (lane 4), cells transfected with pDHFR-I_N left untreated (lane 5) and cells transfected with pDHFR-I_N treated with peptide (lane 6). The product DHFR-FLAG is denoted by *, and its expected mass is ~24.5 kDa. (C) Anti-GFP western blot of a 15% SDS-PAGE gel of un-transfected and untreated cells (lane 7), cells transfected with pYFP-I_N-CFP left untreated (lane 8) and cells transfected with pYFP-I_N-CFP treated with peptide (lane 9). The product YFP-FLAG is denoted by ✶ and its expected mass is ~29 kDa. Background bands are denoted by <.

& 3). Thus product formation occurs via protein *trans*-splicing in the living cells.

To explore the generality of protein semi-synthesis *in vivo*, I_C-FLAG was reacted with three other model cytosolic proteins: GST-I_N, DHFR-I_N and YFP-I_N-CFP and in each case the desired semi-synthetic product was obtained

(Figure 2.13). Interestingly, the later construct (YFP-I_N-CFP) showed improved cytosolic expression as compared to GFP-I_N (Figure 2.14), suggesting that inserting I_N between two domains may help soluble expression of the fusion protein. In addition, trans-splicing resulted in cleavage of YFP from I_N-CFP, indicating that the methodology may also serve as a tool to trigger the un-coupling of fusion proteins *in vivo*.

In principle, the protein semi-synthesis approach based on protein *trans*-splicing should also be applicable to integral membrane proteins. This possibility was investigated by preparing a semi-synthetic version of the G-

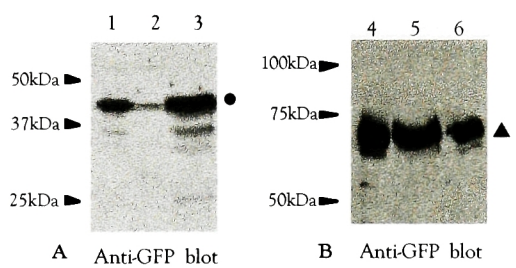


Figure 2.14. Comparison of whole cell vs. detergent soluble or insoluble expression of GFP-I_N and YFP-I_N-CFP. CHO cells were transiently transfected with either pGFP-I_N (**A**) or pYFP-I_N-CFP (**B**) and after 24 hours of expression the cells were lysed in PBS + 1% DM. Whole cell (lanes 1 & 4), soluble (lanes 2 & 5) or detergent insoluble fraction (lanes 3 & 6) samples were loaded for analysis onto a 12% PAGE-SDS gel and blotted against GFP. In the case of GFP-I_N >90% of the protein is insoluble (compare lanes 2 & 3) vs. ~50% in the case of YFP-I_N-CFP (compare lanes 4 & 5). GFP-I_N is denoted by •, and its expected mass is ~42 kDa. YFP-I_N-CFP is denoted by ▲, and its expected mass is ~74 kDa.

protein-coupled receptor (GPCR), Opsin. As shown in Figure 2.15, Opsin-GFP-FLAG was generated by treating transfected cells expressing Opsin-GFP-I_N with ANTP-S-S-I_C-FLAG (compare lane 3 with control lanes 1 & 2). Importantly, the smearing of the band corresponding to the ligation product, Opsin-GFP-FLAG, is consistent with glycosylation of this protein, known to

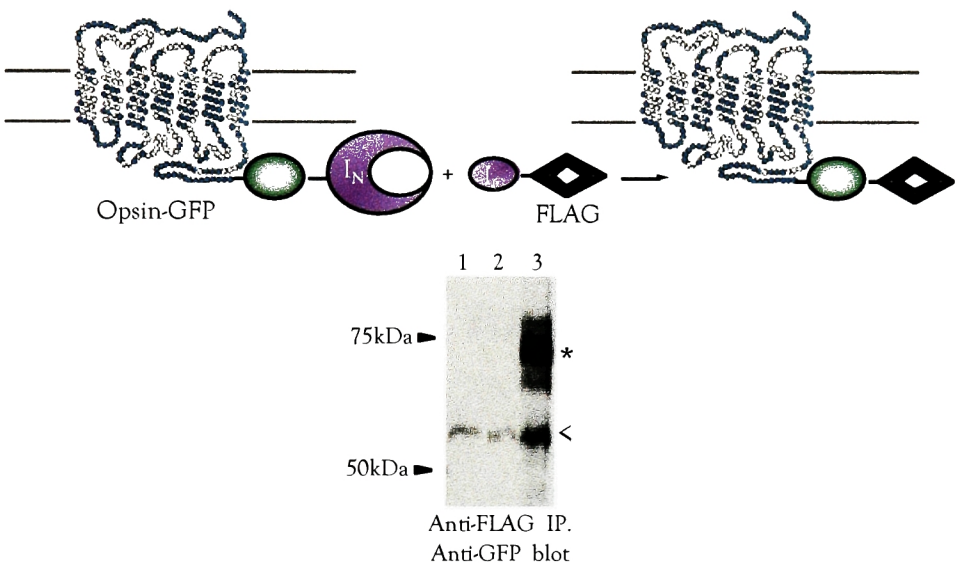


Figure 2.15. Semi-synthetic trans-splicing between Opsin-GFP-I_N and I_C-FLAG in CHO cells [schematically shown in top panel]. Cells treated as indicated below were lysed and the soluble fraction immuno-precipitated and blotted as indicated in the panel. Non-transfected and untreated cells (lane 1), transfected cells expressing Opsin-GFP-I_N with no peptide added (lane 2) and transfected cells expressing Opsin-GFP-I_N treated with 100 μM ANTP-S-S-I_C-FLAG (lane 3). The ligation product (Opsin-GFP-FLAG) is denoted by *, background bands are denoted by <, and the smearing in the band is due to the glycosylation of Opsin. The expected molecular weight of Opsin-GFP-FLAG is ~70 kDa.

occur during membrane integration.[77]

Preliminary experiments were performed to determine the minimum concentration of ANTP-S-S-I_C-FLAG needed to observe semi-synthetic *trans*-splicing within 3 hours of incubation (Figure 2.16). CHO cells expressing GFP-I_N for 24 hours were treated for 3 hours with varying concentrations of ANTP-S-S-I_C-FLAG. Following harvest and immuno-precipitation against FLAG, samples were analyzed by western blotting against GFP. As seen in figure 2.16, a minimum of 50 μM was required to observe activity or association of the fragments at this time point. This observation is rather

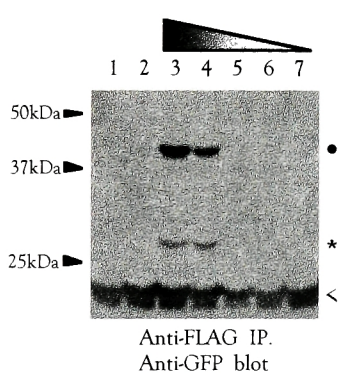


Figure 2.16. Minimum concentration of I_C-FLAG necessary to observe semi-synthetic protein trans-splicing in cells. Un-transfected and non-treated cells (lane 1), transfected cells expressing GFP-I_N with no peptide added (lane 2) and transfected cells expressing GFP-I_N treated with 100 μM (lane 3), 50 μM (lane 4), 10 μM (lane 5), 5 μM (lane 6) or 1 μM (lane 7). Product formation or intein association is only seen at 100 or 50 μM (lanes 3 & 4). The ligation product GFP-FLAG is denoted by *, GFP-I_N is denoted by • and background bands are denoted by <.

surprising, as literature reports have suggested that the dissociation constant (K_D) between I_N and I_C is in the nano-molar range.[96] The possibility that only a small percentage of the I_C-FLAG is being translocated is unlikely, because transduction of both ANTP and I_C-FLAG was observed at 2 nM (see previous sub-section). Future experiments will be required before the origins of this observation are known.

Although protein semi-synthesis based on *trans*-splicing allows ligation to both cytosolic and membrane bound proteins *in vivo*, the technique is not without drawbacks. Addition of I_N to the C-terminus of proteins affects their solubility, and in the case of membrane bound proteins, their localization. Solubility problems with split inteins have been previously reported and can possibly be due to improper folding or exposed hydrophobic regions.[99 128] As seen in figure 2.14 (compare lanes 2 & 3) most of the GFP-I_N is expressed in an insoluble form. To partially overcome this issue, I_N was inserted between two GFP proteins to stabilize it and increase its solubility (Figure 2.14, compare lanes 2 & 3 with 5 & 6). This protocol also increased expression of the membrane bound transferrin receptor (TnfR) tagged with GFP-I_N (Figure 2.17, compare **B** & **C**), but did not overcome the undesired (*e.g.* endoplasmic reticulum (ER)) localization observed with the tandem expression of TnfR and I_N. As seen in figure 2.17, TnfR-GFP localizes to the cell surface of mammalian cells as desired

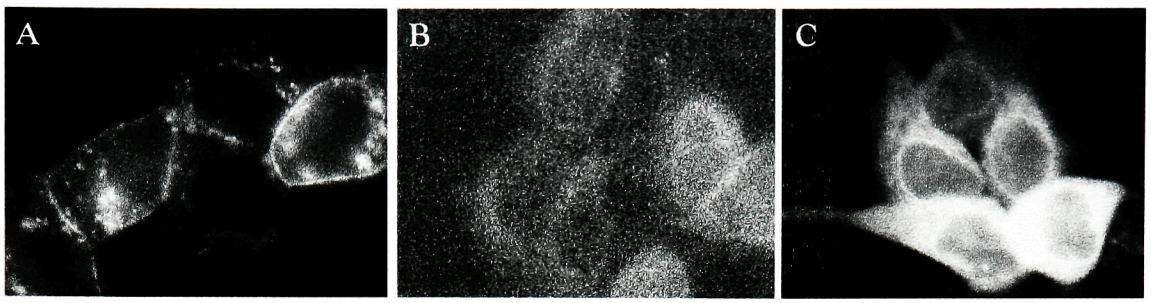


Figure 2.17. Localization of TnfR-GFP, TnfR-GFP-I_N and TnfR-YFP-I_N-CFP in MCF-7 cells. Cells were transiently transfected and allowed to express the proteins for 24 hours, before being fixed. **(A)** TnfR-GFP is localized to the plasma membrane. **(B)** TnfR-GFP-I_N does not localize to the plasma membrane, and is scarcely visible. **(C)** TnfR-YFP-I_N-CFP does not localize to the plasma membrane, but it expresses better then TnfR-GFP-I_N. GFP (**A** & **B**) or YFP (**C**) fluorescence signal is shown.

(Figure 2.17 **A**), contrary to both TnfR-GFP-I_N and TnfR-YFP-I_N-CFP, which appeared to be largely retained in the ER (Figure 2.17 **B** & **C**). This type of undesired localization was also observed with Opsin-GFP-I_N (Figure 2.18 compare **A** & **B**).

In a previous study, Ozawa *et. al.* attached I_N to membrane bound proteins, but did not visualize their localization.[110] Improper folding or exposure of hydrophobic residues of I_N may also be responsible for the inadequate delivery of membrane proteins. Additionally, the size of I_N (123 amino acids) could have an effect, analogous to that observed in some cases with GFP tagging. Further studies on I_N may reveal a way to make this intein fragment behave better. Mutagenesis, or minimization of I_N could improve its characteristics. Alternatively, inverting the system (*e.g.* I_C tagged

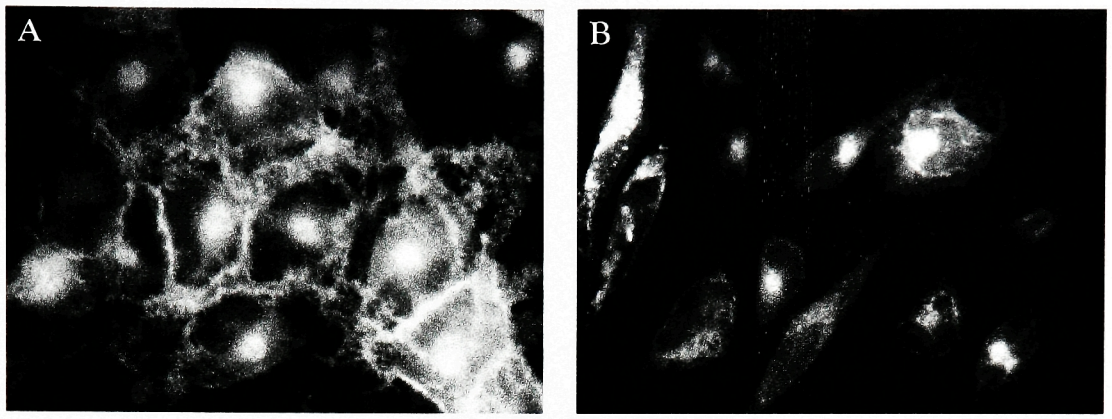


Figure 2.18. Localization of Opsin-GFP and Opsin-GFP-I_N in HeLa cells. Cells were transiently transfected and allowed to express the proteins for 24 hours, before being fixed. **(A)** Opsin-GFP is localized to the plasma membrane. **(B)** Opsin-GFP-I_N does not localize to the plasma membrane. GFP fluorescence signal is shown.

to the protein of interest and I_N used to deliver the probe) could all together by-pass these solubility and localization problems as I_C is small and seemingly well behaved (see below).

While I_C-FLAG was synthesized by SPPS for this study, a semi-synthetic protocol for the addition of probes onto I_C was also developed for ease and flexibility. The approach, based on EPL, is summarized in Figure 2.19.

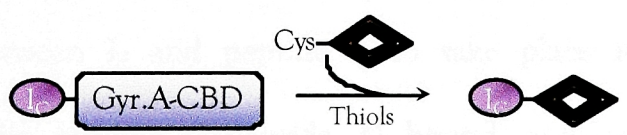


Figure 2.19. General scheme for the addition of a probe onto I_C. I_C-GyrA-CBD is expressed in bacteria and purified using the chitin binding domain (CBD) or other affinity tag. Once purified it can react with a N-terminal cysteine containing probe to give the desired product: I_C-probe.

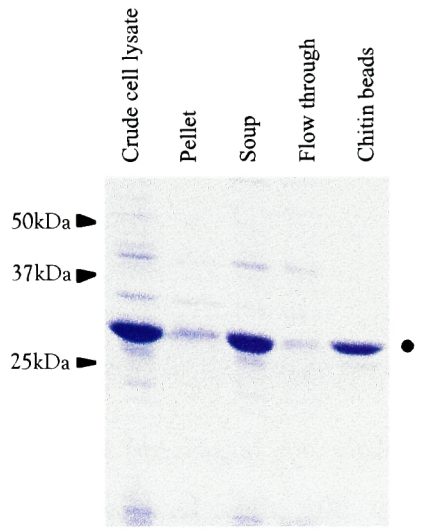


Figure 2.20. Bacterial expression and purification of Ic-GyrA-CBD. The fusion protein was soluble, and was induced to very high levels (~50% of the total bacterial proteins). Ic-GyrA-CBD is denoted by •, and its expected mass is 32 kDa.

A bacterial expression plasmid was constructed with Ic N-terminally fused to the mutated Gyrase A (GyrA) intein followed by a chitin binding domain (CBD) (pIc-GyrA-CBD). Expression was induced to high levels in *E. coli*, and the soluble fusion protein was purified over chitin beads (Figure 2.20).

The GyrA intein within the purified fusion protein was functional and permitted EPL between Ic and peptide 3 to take place in yields of >90% (Figure 2.21). The product (Ic-Peptide 3) bound non-specifically to chitin beads, but was effectively eluded with a solution containing 8 M urea.

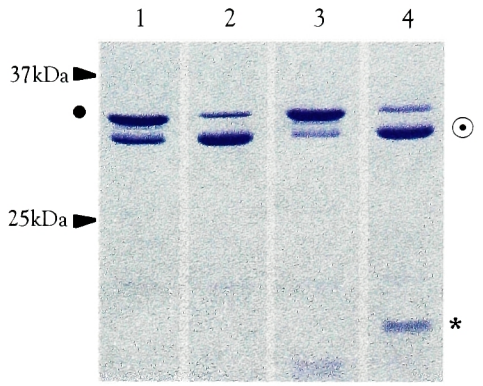


Figure 2.21. Ic-GyrA-CBD is functional and undergoes EPL. Untreated protein (lane 1), protein cleaved with 100 mM MESNA (lane 2), protein treated with 1 mM of peptide 3 is unmodified (lane 3) and protein treated with 100 mM MESNA and 1 mM of peptide 3 produces the expected product Ic-Peptide 3. Full length Ic-GyrA-CBD is denoted by •, cleave GyrA-CBD is denoted by ⊙, and the EPL product Ic-Peptide 3 is denoted by *.

Peptide 3:

H- CGRGK(Rh)G-NH_2

Protein semi-synthesis *in vivo* based on protein *trans*-splicing was successful. In principle, any synthetic molecule - a polypeptide or, minimally, a cysteine derivative bearing any chemical moiety - can be ligated to a protein of interest inside living cells. Here, incorporation of the probe took place at the C-terminus of the protein, however, by interchanging the split intein components, it should be possible to introduce the probe at the N-terminus of a protein as well. Such a possibility increases the modularity of the approach. Future experiments will be performed in the laboratory to explore N-terminal ligations.

Semi-synthetic growth factor derivatives for whole animal studies.

This work was done in collaboration with Rob Flavell and Amersham Health.

Growth factors are extra-cellular polypeptide signal molecules that stimulate a cell to grow, proliferate, differentiate or survive.[85] They circulate in the blood stream and have specific receptors on the surface of target cells. Most animal cells require a precise combination of several growth factors to multiply, highlighting the combinatorial nature of the associated signaling pathways, and the need to study these polypeptides imbedded within their *in vivo* system. Furthermore, growth factors are tightly associated with the uncontrolled cell division seen in tumors and cancer.[85] These characteristics make growth factors an interesting and suitable proof-of-principle system for the investigation of protein semi-synthesis as a tool for whole animal studies. Initially, the vascular endothelial growth factor (VEGF) was chosen as a model protein.

VEGF is a critical regulator of angiogenesis, the formation of new blood vessels, and plays an important role in vascular development and wound healing.[55] VEGF-mediated angiogenesis is also implicated in a variety of pathological conditions, most notably the vascularization of solid tumors. Thus, there is great interest in understanding the role of VEGF in normal and aberrant vascular development and whether this function can be

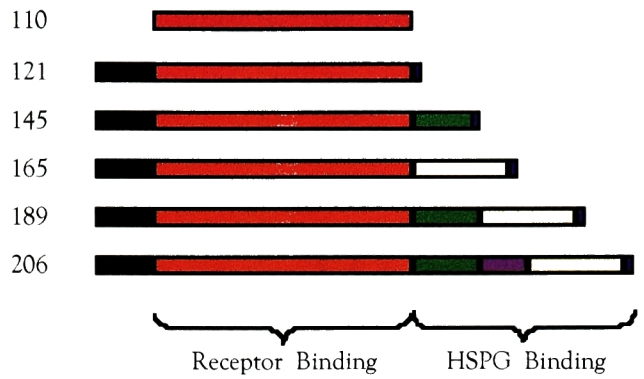


Figure 2.22. Splice variants of human VEGF. The human *VEGF* gene, through alternative splicing produces 5 iso-forms, which differ in the presence or absence of heparin binding domains. Black and red represent the pro-peptide sequence and the receptor-binding domain, respectively. Other colors represent heparan-sulfate proteoglycan (HSPG) binding domains encoded by different exons. $VEGF_{110}$ is the proteolytic product of $VEGF_{165}$ or $VEGF_{189}$.

exploited for the development of new therapies for disorders like cancer.[54] VEGF is a homo-dimeric, disulfide-linked glycoprotein that is secreted into the extra-cellular space. It has multiple splice variants (Figure 2.22), all of which contain a signal peptide sequence for secretion and heparin binding domains. In the case of $VEGF_{165}$ and $VEGF_{189}$, the heparin binding domains can be cleaved off by the protease plasmin, producing the freely circulating $VEGF_{110}$.[55]

VEGF binds to specific receptors, including VEGFR-1 and VEGFR-2 (Figure 2.23), with pico-molar affinities. These receptors are expressed predominantly in endothelial cells and can be up-regulated under hypoxic conditions or by the increase in ligand concentration that occurs during

angiogenesis. High concentrations of VEGFR-1 and VEGFR-2 serve as markers of active angiogenesis.[55] For these reasons, VEGF has been used in several studies to deliver specific cytotoxic proteins to cells expressing VEGF receptors as a potential cancer therapy.[5 8 9]

A circulating variant of VEGF, such as $VEGF_{110}$, coupled with an appropriate imaging tag, could be used as a probe to detect angiogenesis in whole animals. $VEGF_{110}$ lacks heparin-binding activity, potentially increasing its bio-availability,[80] but maintaining its high affinity for VEGFR-1 and VEGFR-2. As endothelial cells are among those exhibiting the lowest replication level in the body, (with only 0.01% cells engaged in cell division at any time[108]), site-specific localization of a radio-labeled semi-synthetic VEGF

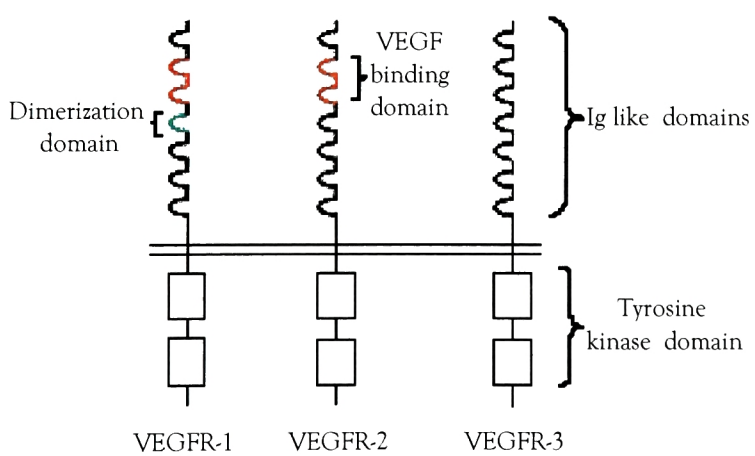


Figure 2.23. Main VEGF receptors. The three signaling tyrosine-kinase receptors of the VEGF family (VEGFR-1, VEGFR-2 and VEGFR-3) are shown. VEGF is also known to bind the accessory iso-form specific receptors neuropilin-1 and neuropilin-2, and heparan-sulfate proteoglycans.

onto VEGFR-1 and VEGFR-2 should concur with sites of angiogenesis (*e.g.* tumors).

Synthetic chelators of the radio-isotope ^{99m}Tc have been conjugated to a variety of peptides, and used to image pathological events such as blood clotting, inflammation, infection, and cancer.[26 122 158] In particular, epidermal growth factor has been labeled with ^{99m}Tc and used to image tumors in mice.[121] Antibodies and other proteins have also been labeled with ^{99m}Tc through standard conjugation methods, including alkylation of reduced cysteines, biotin-avidin interactions, or the addition of specific ^{99m}Tc chelating peptide sequences.[58 59 144]

^{99m}Tc has proven to be a versatile isotope because it is a pure γ emitter with optimal emission energy of 140 keV, ideal for use with γ cameras employed in nuclear medicine.[4] Additionally, its half-life of 6 hours and ease of production using a ^{99}Mo/^{99m}Tc generator make it well suited for

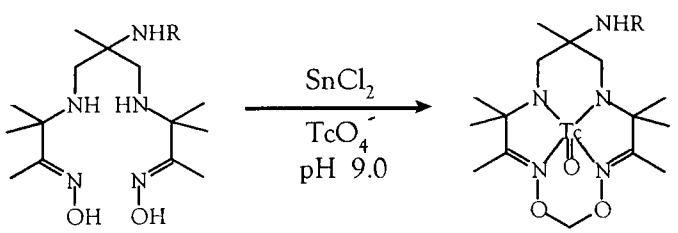


Figure 2.24. Diamine, dioxime chelates bind ^{99m}Tc in a pyramidal complex. Loading of ^{99m}Tc is achieved under reducing conditions at pH 9.0. The covalent interaction is stable under physiological conditions.

diagnostic use. A variety of high affinity synthetic chelators for ^{99m}Tc have been developed[12] (see below) and loading with ^{99m}Tc is a relatively simple procedure (Figure 2.24).

The methodology summarized in Figure 2.25 was envisioned to produce semi-synthetic VEGF with a small imaging tag that chelates ^{99m}Tc with high affinity. Following protein semi-synthesis, the ^{99m}Tc would be incorporated into the protein and allow imaging of the proteins' localization to vascularized tumors.

In this methodology, VEGF is first bacterially expressed fused to the

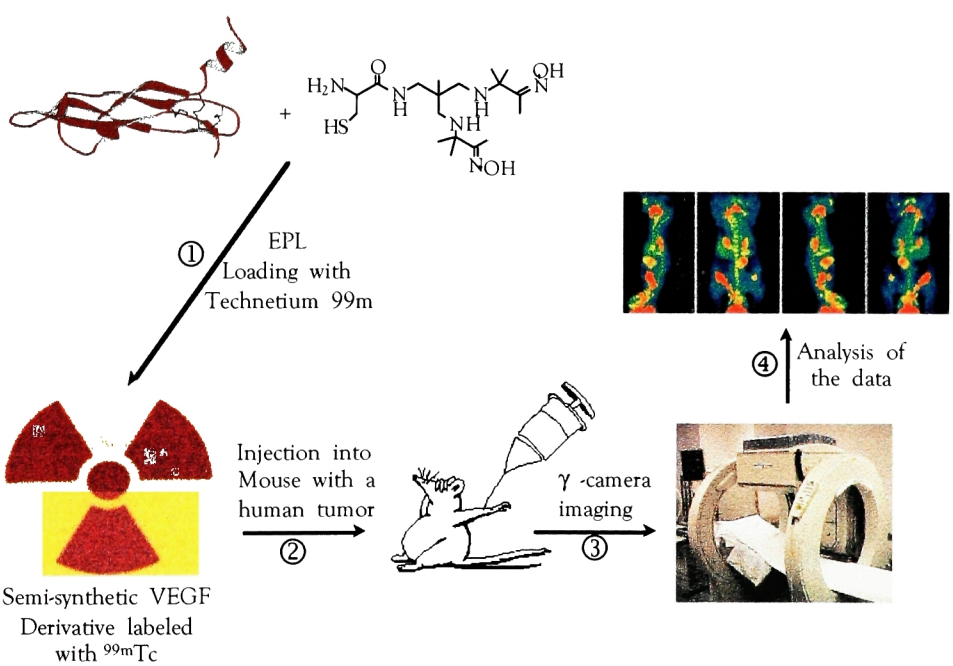


Figure 2.25. General scheme to investigate the use of semi-synthetic proteins produced by EPL in whole animal studies. VEGF was chosen as a model protein to be tagged with a ^{99m}Tc chelator. The semi-synthetic protein could allow imaging of tumors in animals.

modified GyrA intein. A synthetic cysteine derivative containing a high affinity chelator of ^{99m}Tc is then reacted with the purified protein, resulting in a semi-synthetic form of VEGF tagged with the chelator (Figure 2.25 ①). In the next step, the semi-synthetic protein is loaded with ^{99m}Tc and injected into mice bearing a human tumor (Figure 2.25 ① & ②). Imaging of the mice with a γ-camera would allow the localization of VEGF to be visualized (Figure 2.25 ③ & ④). This project was conceived in collaboration with Amersham Health, one of the world's leading biomedical imaging companies. In addition, VEGF, a disulfide linked dimeric protein, would further test the scope of the EPL methodology.

The crystal structure of $VEGF_{8\text{-}109}$ in complex with the VEGF binding domain of VEGFR-1 has been solved.[149] Analysis of the structural data shows that the C-terminus of $VEGF_{8\text{-}109}$ is not involved in receptor binding. Moreover, mutagenesis of this region has little effect on VEGFR binding.[81] Therefore, to minimize possible disruption of the interaction between $VEGF_{110}$ and its receptors due to an unnatural moiety, the ^{99m}Tc chelator was ligated onto the C-terminus of $VEGF_{110}$. An affinity handle was later attached to the chelator because in preliminary studies, hydrolyzed material could not be resolved from the ligation product (see below) (Figure 2.26).

$VEGF_{110}$ was expressed in *E.coli* N-terminally fused to the cleavable intein GyrA and a CBD. The majority of the fusion protein was expressed

in inclusion bodies. These were washed with a solution containing 2 M urea to remove loosely associated proteins, before being solubilized with a solution containing 8 M urea. Previous work in the Muir laboratory has shown that GyrA can be refolded from inclusion bodies and it regains protein splicing activity.[139] Such a characteristic makes GyrA a suitable intein for fusion with $VEGF_{110}$, a protein known to go into inclusion bodies when expressed in bacteria.[30] The solubilized $VEGF_{110}$-GyrA-CBD was refolded by

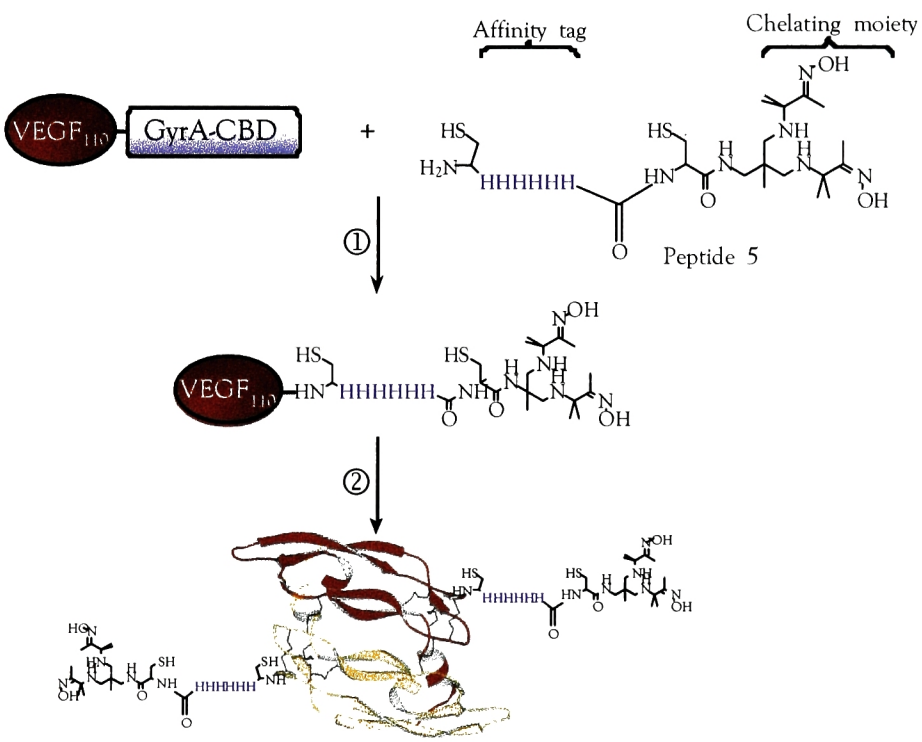


Figure 2.26. Preparation of semi-synthetic $VEGF_{110}$ by EPL. ① $VEGF_{110}$-GyrA-CBD is reacted with peptide 5 to produce the monomer form of semi-synthetic $VEGF_{110}$-Peptide 5, and is purified away from hydrolyzed $VEGF_{110}$ and un-reacted chelator. ② The chemically homogenous $VEGF_{110}$-Peptide 5 can then be refolded into the dimeric and biologically active form according to literature procedures.

dialysis using the reported procedure,[139] and affinity purified over chitin beads. An initial concern was the possibility that the eight cysteine residues in $VEGF_{110}$ could cleave the protein away from the GyrA-CBD. Previous experiments had shown that appropriately placed cysteines could induce *in vivo* cleavage of the intein domain.[20] However, as seen in figure 2.27, no *in vivo* cleavage was detected by SDS-PAGE analysis of the crude cell lysate or the affinity purified protein.

Amersham Health supplied five different ^{99m}Tc chelators (Figure 2.28), each were bi-functional, containing an N-terminal cysteine (necessary for NCL & EPL) and a C-terminal ^{99m}Tc chelating moiety. Scheme 2.1 shows the

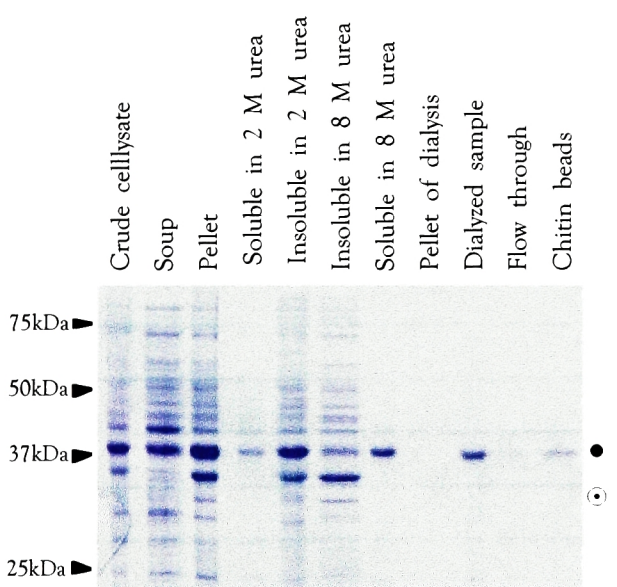


Figure 2.27. Expression and purification of $VEGF_{110}$-GyrA-CBD. $VEGF_{110}$-GyrA-CBD is denoted by •, and its expected mass is 40.5 kDa. ⊙ denotes where the GyrA-CBD would be if *in vivo* cleavage were taking place.

synthetic strategy used to prepare SN1400-01. All other chelators were synthesized in a similar fashion.

In preliminary experiments it was shown that all of the molecules were stable to the reaction conditions used in NCL and EPL. In particular, there was no decomposition of the oxime moieties in the presence of high concentrations of thiols (Scheme 2.2). Furthermore, all chelators could

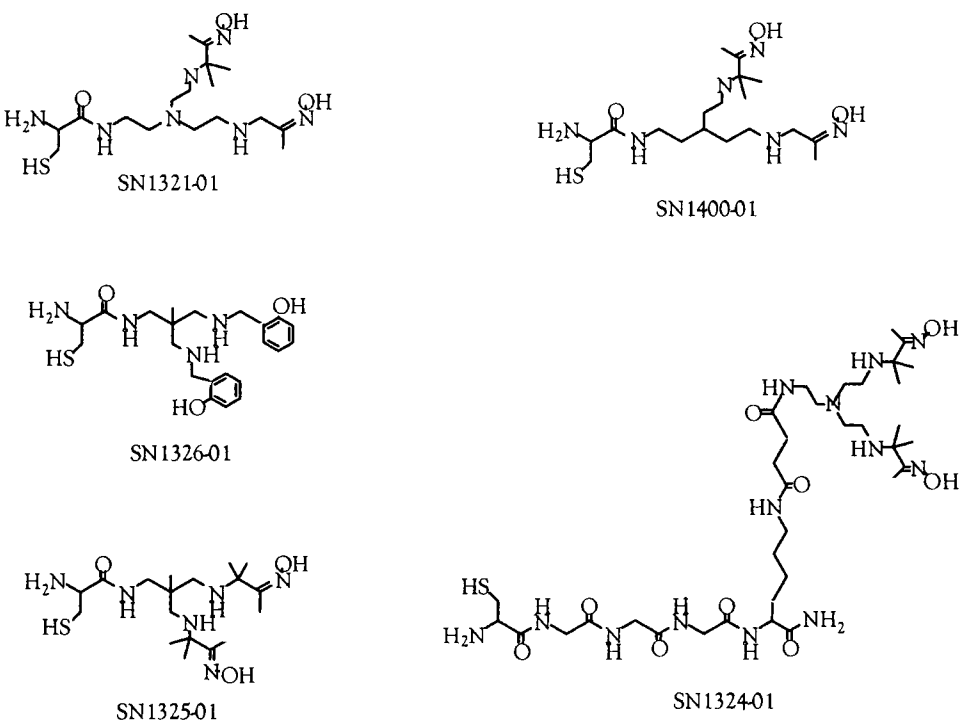
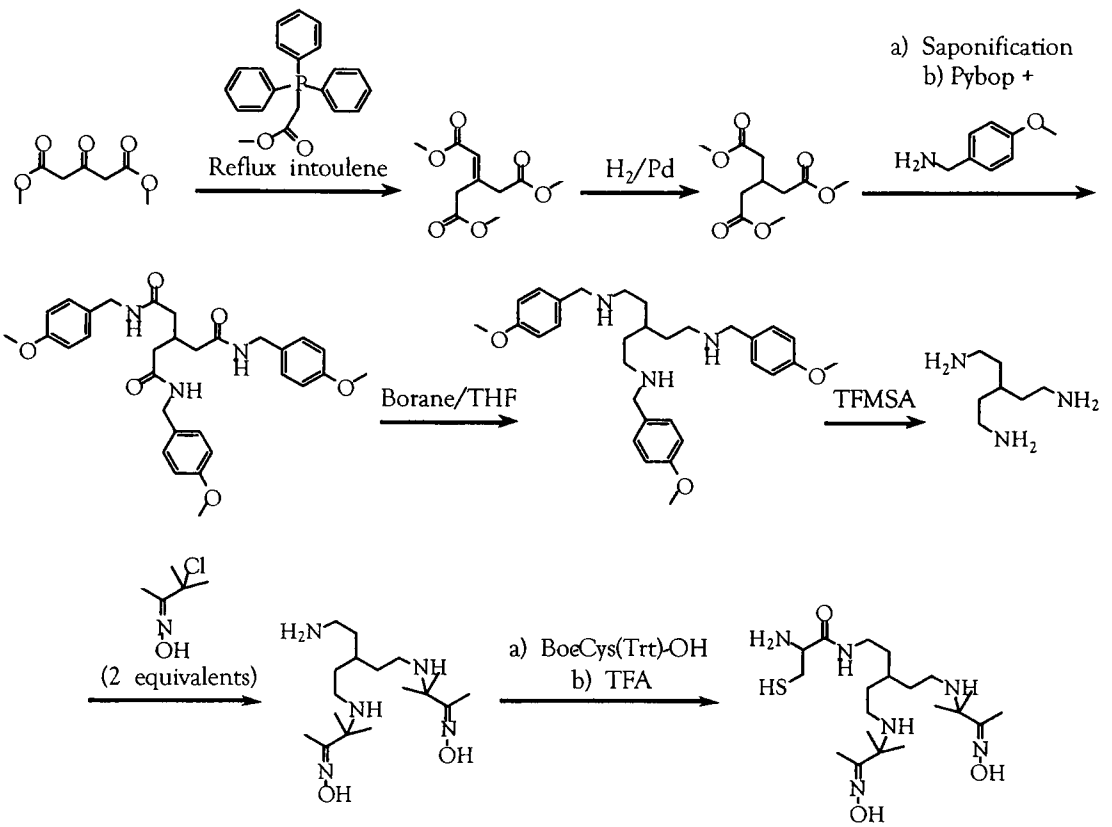
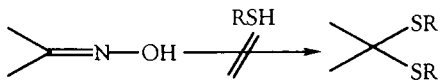


Figure 2.28. ^{99m}Tc chelators supplied by Amersham Health. Chelators are labeled with their identifying numbers. All of the molecules were stable to the reaction conditions used in NCL and EPL conditions. Furthermore, they could all ligate to a small α-thioester peptide and a protein tagged with an intein.



Scheme 2.1. Synthesis scheme of SN1400-01. All other chelators provided by Amersham Health were synthesized in a similar fashion.



Scheme 2.2. Possible decomposition of oxime moieties to a thio-ketal in the presence of high concentrations of thiols. This reaction was not observed with any of the chelators tested.

undergo efficient NCL with a model α-thioester peptide, and EPL with a model protein (Crk-SH3) (data not shown).

Initial EPL studies between $VEGF_{110}$-GyrA-CBD and SN1400-01 yielded the desired ligation product and a side product. Due to hydrolysis of the α-thioester during the ligation reaction, the free carboxylic acid of $VEGF_{110}$ was also generated (Figure 2.29 **B**). The two products could not be separated by chromatographic techniques, such as RP-HPLC, ion exchange chromatography, and affinity chromatography. Upon oxidative refolding,[30] three different dimers were produced: the two homo-dimers and the hetero-dimer (Figure 2.29 **C**). These three compounds were similarly inseparable by chromatography. The generation of a heterogeneous mixture using a one-step ligation protocol underlined the need for an alternative tactic for the preparation of the desired product in a chemically homogenous form.

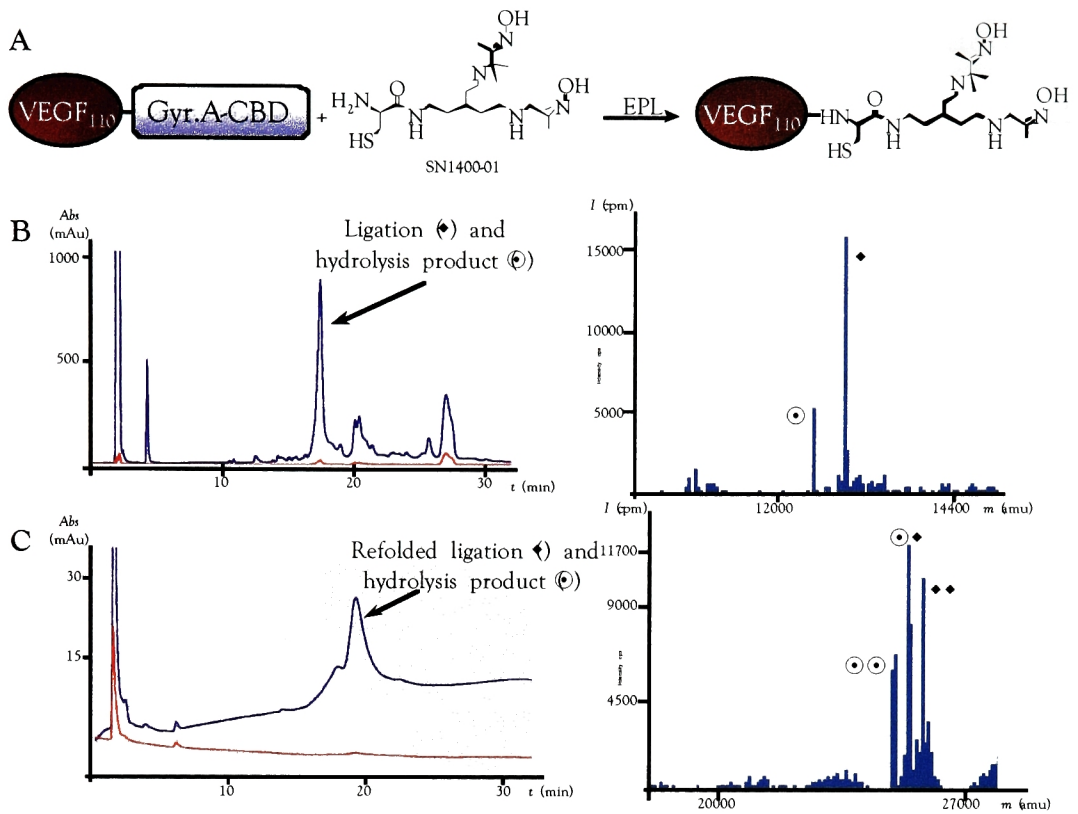


Figure 2.29. One-step ligation between $VEGF_{110}$-GyrA-CBD and SN1400-01. **(A)** Schematic representation of the reaction between $VEGF_{110}$-GyrA-CBD and SN1400-01. Purified protein and peptide were reacted in 100 mM MESNA and 2% ethanethiol at pH 7.2 for two days. **(B)** RP-HPLC (0-73% B) and ES-MS analysis of the ligation reaction. ◆ denotes the ligation product $VEGF_{110}$-SN1400-01 (OM = 12934.3 ± 1.4 Da; EM = 12930.6 Da) and ⊙ denotes the hydrolysis product $VEGF_{110}$ (OM = 12506.1 ± 1.4 Da; EM = 12502.3 Da). **(C)** RP-HPLC (0-73% B) and ES-MS analysis of the oxidative refolding reaction. ⊙⊙ denotes the hydrolysis homo-dimer (OM = 24994.0 ± 5.2; EM = 24988.6), ◆⊙ denotes the hetero-dimer (OM = 25417.9 ± 2.6; EM = 25416.7), and ◆◆ denotes the ligation homo-dimer (OM = 25845.2 ± 1.3; EM = 25845.2).

In principle, attachment of an affinity handle to the chelator should provide a strategy for purifying the final ligation product away from starting material or side-products (*e.g.* hyrolysis). Therefore a sequential ligation approach was developed to incorporate the secondary purification tag (Figure 2.30). The devised method permits modularity and bypasses the need to synthesize the chelators with the affinity handle (possibly a complicated synthesis).

Key to this strategy is the reversible protection of the N-terminal cysteine in the affinity tag. In the absence of such protection, cyclization or polymerization would occur in the first ligation step (Figure 2.30 ①). To conceal the N-terminal cysteine a thiazolidine masking strategy was chosen.[140] The encrypting moiety is removed upon treatment with methoxylamine, which is mild, easy and efficient (Figure 2.30). Although the thiazolidine masking methodology was well suited for this study, other strategies for the concealment of N-terminal cysteines, such as leader sequences that can be proteolytically removed[47] or chemical protecting groups, such as the 2-(methylsulfonyl) ethyloxycarbonyl (Msc) group,[131 156] are also available, and may be appropriate in different contexts.

While virtually any affinity tag could be used, a 6xHis tag was chosen due to its compatibility with denaturants and mildly reducing conditions necessary to maintain VEGF as an unfolded monomer during purification.

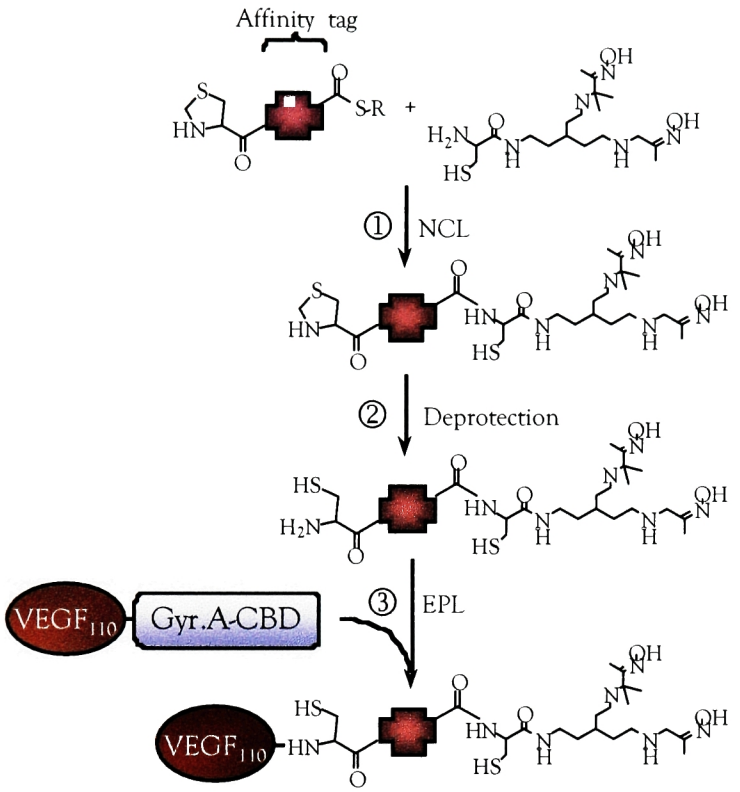


Figure 2.30. General strategy for tandem ligation of an affinity tag and a chelator onto $VEGF_{110}$. ① The chelator is ligated to an affinity tag containing a thiazolidine derivative that serves as a cryptic N-terminal cysteine. ② After ligation, the cysteine is unmasked by methoxylamine treatment, and can then undergo ligation with $VEGF_{110}$-GyrA-CBD ③.

The 6xHis tag (peptide 4) was synthesized by Boc SPPS on a 3-mercaptopropionamide derivatized PEGA resin, as outlined in figure 2.31 ①. The N-terminal cysteine of the peptide was incorporated as a thiazolidine derivative (thiaproline). The only protecting groups on the peptide were dinitrophenyl (DNP) groups on the histidine residues, which allowed for an easy and efficient de-protection and cleavage from the resin in a single step (overnight treatment with 10% aqueous ethanethiol). Peptide 4 was then

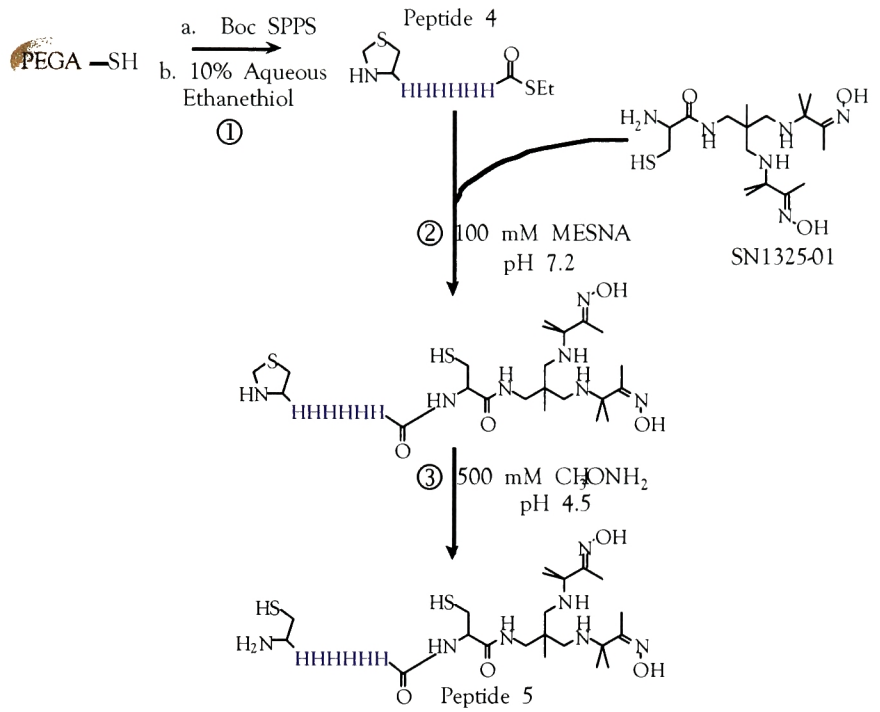


Figure 2.31. NCL between peptide 4 and chelator SN1325-01 to produce peptide 5. ① The 8 amino acid α-thioester peptide 4 was synthesized by Boc SPPS on a 3-mercaptopropionamide derivatized PEGA resin. The N-terminal cysteine of the peptide was incorporated as thiaproline. De-protection and cleavage from the resin in a single step (overnight treatment with 10% aqueous ethanethiol) resulted in the production of the α-thioester Peptide 4. ② Peptide 4 was then reacted with SN1325-01 to produce the protected peptide 5. ③ After ligation, the cryptic N-terminal cysteine in peptide 5 was unmasked by the addition of methoxylamine.

reacted with SN1325-01 using 100 mM MESNA as the thiol co-factor to produce the cysteine protected peptide 5. Unmasking of the cryptic N-terminal cysteine was accomplished by the addition of methoxylamine.[140] The latter reaction very clean with no side-products.

$VEGF_{110}$-GyrA-CBD was then reacted with purified peptide 5 generating the desired ligation product ($VEGF_{110}$-Peptide 5) and the free carboxylic acid of $VEGF_{110}$ previously described (see above). However, unlike initial experiments, $VEGF_{110}$-Peptide 5 could now be purified away from the side-product. Gel filtration chromatography, under denaturing and mildly reducing conditions, first removed un-reacted peptide 5 and thiol cofactors. The $VEGF_{110}$-Peptide 5 was then purified away from hydrolysis product by Ni-NTA chromatography. The chemically homogenous $VEGF_{110}$-Peptide 5 was refolded according to literature protocols,[30] and the dimeric product was purified by gel filtration chromatography. The final product was analyzed by non-reducing SDS-PAGE, RP-HPLC and ES-MS, and confirmed to be the homo-dimer of $VEGF_{110}$-Peptide 5 (Figure 2.32). Thus addition of the secondary affinity tag allowed recovery of the desired ligation product, which upon formation of the homo-dimer can be applied to studies in whole animals.

The purified semi-synthetic $VEGF_{110}$-Peptide 5 dimer was tested for biological activity using an endothelial cell proliferation assay.[31] Human umbilical vein endothelial cells (HUVEC) were stimulated with either the semi-synthetic protein or commercially available $VEGF_{121}$ (R & D Systems). Cell proliferation was measured by radiolabeled thymidine, and the EC_{50}s

were determined to be 591.6 pM (150.5 - 2325 pM; 95%CI) and 41.1 pM (11.9 - 141.7 pM; 95% CI) for $VEGF_{110}$-Peptide 5 and $VEGF_{121}$, respectively (Figure 2.33). The ten-fold discrepancy in the obtained EC_{50} values could be due to experimental errors (in particular in the concentrations of the stocks), or to unforeseen detrimental effects due to the presence of the chelator. Nonetheless, $VEGF_{110}$-Peptide 5 still binds with high affinity and has maximal potency.

The dimeric and biologically active semi-synthetic $VEGF_{110}$-Peptide 5

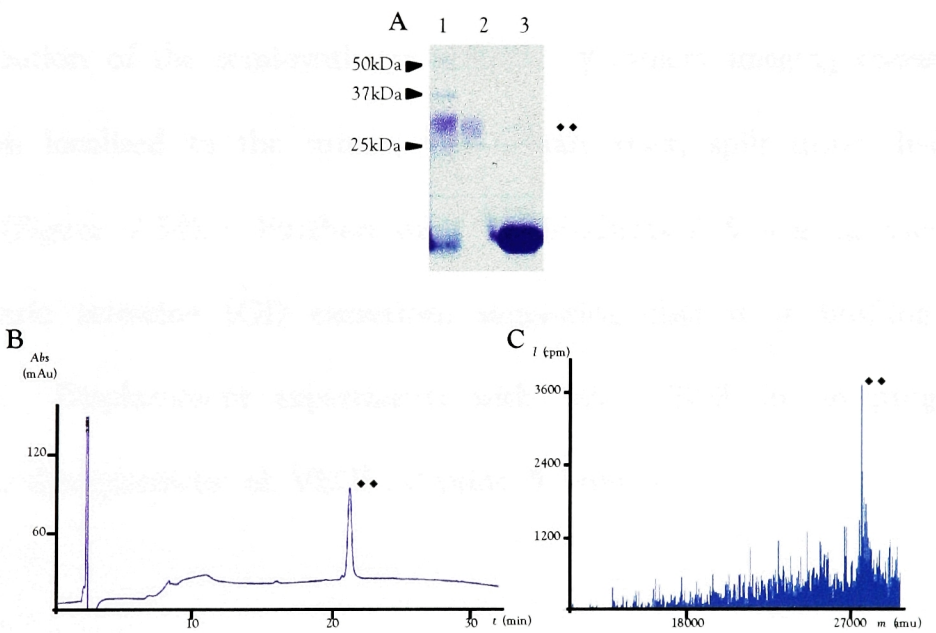


Figure 2.32. Analysis of the $VEGF_{110}$-Peptide 5 homo-dimer. (**A**) Non-reducing SDS-PAGE analysis. Crude refolding mixture (lane 1), gel filtration purification of dimer (lane 2) and reduction of the dimer with DTT (lane 3). (**B**) RP-HPLC (0-73% B) and (**C**) ES-MS analysis. $VEGF_{110}$-Peptide 5 homo-dimer is denoted by ◆◆ (OM = 27762.9 ± 13.0; EM = 27755.4).

sample was sent to Amersham Health for ^{99m}Tc loading and imaging in whole animals. Male C57BL/6 mice were injected sub-cutaneously in their inner right thigh with Lewis Lung carcinoma cells. Tumors were allowed to develop for 15 days, as previous model studies demonstrated that this time point showed the highest level of tumor angiogenesis. $VEGF_{110}$-Peptide 5 was radiolabeled with ^{99m}Tc with a low stoichiometry, under a nitrogen atmosphere, using a standard protocol.[93] Radiolabeled $VEGF_{110}$-Peptide5 was the injected into tumor bearing animals *via* their tail vein. Animals were killed and dissected at different time points after injection to determine the bio-distribution of the semi-synthetic protein. γ camera imaging revealed that γ-radiation localized to the tumor, the urinary tract, spilt urine, liver, and kidneys (Figure 2.34). Furthermore, $VEGF_{110}$-Peptide 5 was up-taken with little gastric intestine (GI) excretion, suggesting that it is binding to its receptors. Displacement experiments with cold VEGF are in progress to determine the specificity of $VEGF_{110}$-Peptide 5 binding.

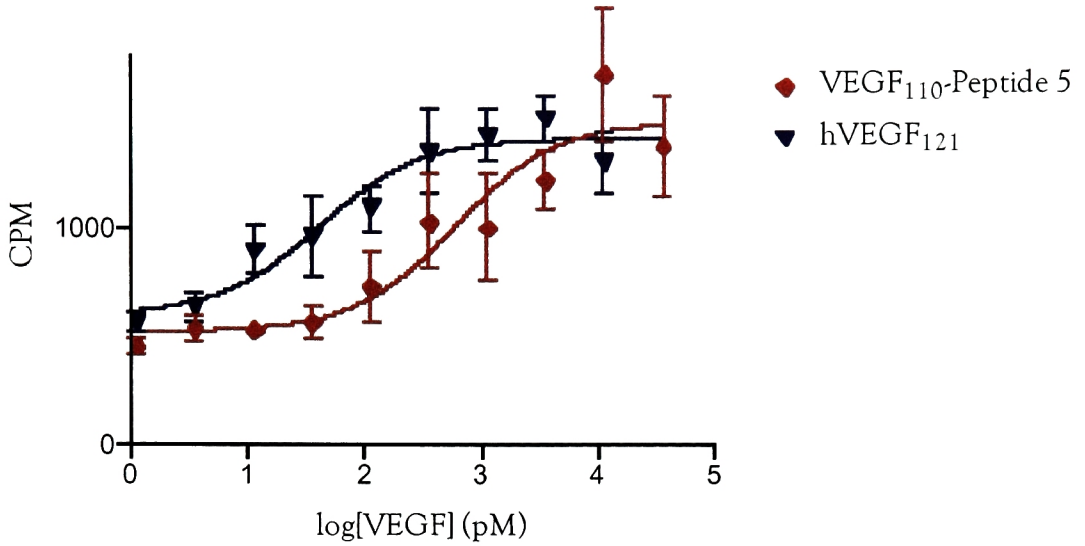


Figure 2.33. Endothelial cell proliferation assay. Incorporation of radiolabeled thymidine by HUVEC in response to $VEGF_{110}$-Peptide 5 (◆) or $VEGF_{121}$ (▼). The average count per minute (CPM) of untreated cells was 550. The experiments were performed in triplicates.

Although $VEGF_{110}$-Peptide 5 localized to the tumor, the negative control ($hVEGF_{110}$ without a chelator) had a similar γ emmission distribution in mice, however the level of contrast was inferior to that obtained with $VEGF_{110}$-Peptide 5. This observation suggests that VEGF may have an intrinsic affinity to ^{99m}Tc, or may bind ^{99m}Tc through an alternative mechanism. One such mechanism could involve the two GlyGlyCys sequences present within VEGF. These sequences have been reported to bind ^{99m}Tc within a variety of proteins, under the labeling conditions used.[17 58 129] Varying the ^{99m}Tc loading procedures or adding free chelators after labeling may help reduce background ^{99m}Tc labeling of control $VEGF_{110}$. Additionally, lengthening the time before imaging could allow better images

to be obtained. Such experiments are currently underway at Amersham Health.

It is important to note that in addition to the experiments underway at Amersham Health, the modular synthetic strategy developed can also be used to sidestep VEGF ^{99m}Tc labeling issues. For instance, the imaging tag can be easily changed to a near-infrared (NIR) fluorophore. NIR fluorescence imaging, with the advantage of minimal light absorption by hemoglobin and water,[147] could by-pass the problems observed, while still permitting imaging of angiogenesis. Alternatively, the protein may be exchanged to a tumor specific single chain antibodies. The modularity of protein, imaging tag, and affinity handle allows for the exploration of a number of proteins within a whole animal setting. Future experiments in the Muir laboratory, in collaboration with Amersham Health, will address such possibilities.

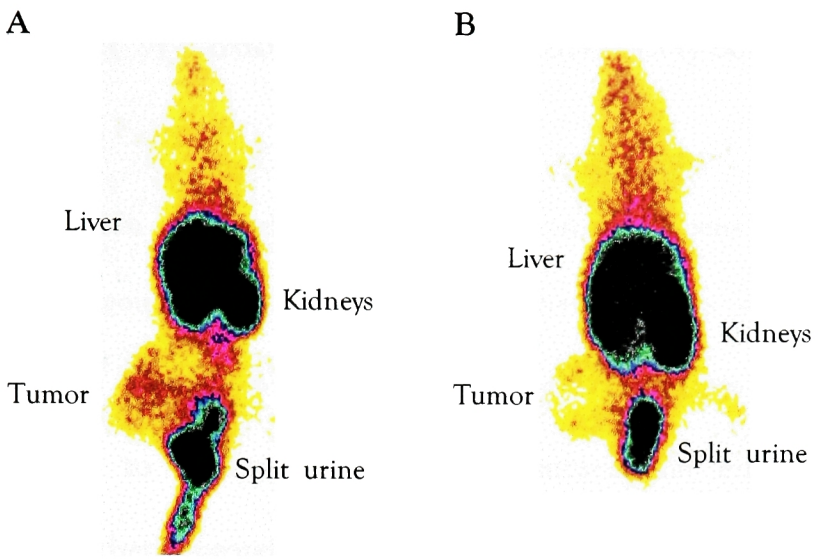


Figure 2.34. Bio-distribution of $VEGF_{110}$-Peptide 5 or $VEGF_{110}$ control observed by γ camera imaging at 120 minutes after injection (MAI). **(A)** Image of $VEGF_{110}$-Peptide 5. **(B)** Image of $VEGF_{110}$ (no chelator). The radiolabelled protein is localized to the tumor, urinary tract, liver and kidneys. Note the asymmetry in the image. No γ radiation emanates from the left thigh (tumor free) of the mouse.

Chapter 3: Conclusions and Future Perspectives

The introduction of unnatural moieties into proteins opens the possibility of exploiting the orthogonality and selectivity that chemistry has to offer to study polypeptides in their natural environment. This thesis describes the ligation of synthetic molecules to target proteins in an intracellular environment, and the use of semi-synthetic proteins for whole animal studies.

Protein semi-synthesis was first explored in a cellular setting. Specifically, a methodology was developed to site-specifically modify proteins in living cells. First, the ANTP PTD peptide was successfully evaluated as a means of introducing *in vitro* manipulated peptides into cultured eukaryotic cells. It was shown that ANTP could translocate up to 70 amino acids, and reliably deliver the Ic-FLAG polypeptide for reaction *in vivo*. A split intein system was then used to perform protein semi-synthesis *in vivo*. Protein *trans*-splicing was chosen because the split inteins are promiscuous with respect to their flanking sequences,[72] are inactive when expressed individually,[87 99 128 152] associate exclusively with their corresponding partner,[109] and, specifically for *Ssp* DnaE, the fragments have high affinity for each other.[96] Such characteristics allowed for the *in vivo* ligation of a synthetic FLAG tag onto four different cytosolic proteins (GFP, DHFR, GST and YFP), and one integral membrane protein (Opsin), thereby demonstrating that, in principle, any synthetic molecule - a polypeptide or, minimally, a cysteine derivative bearing any chemical moiety - can be ligated to a protein of interest inside living cells.

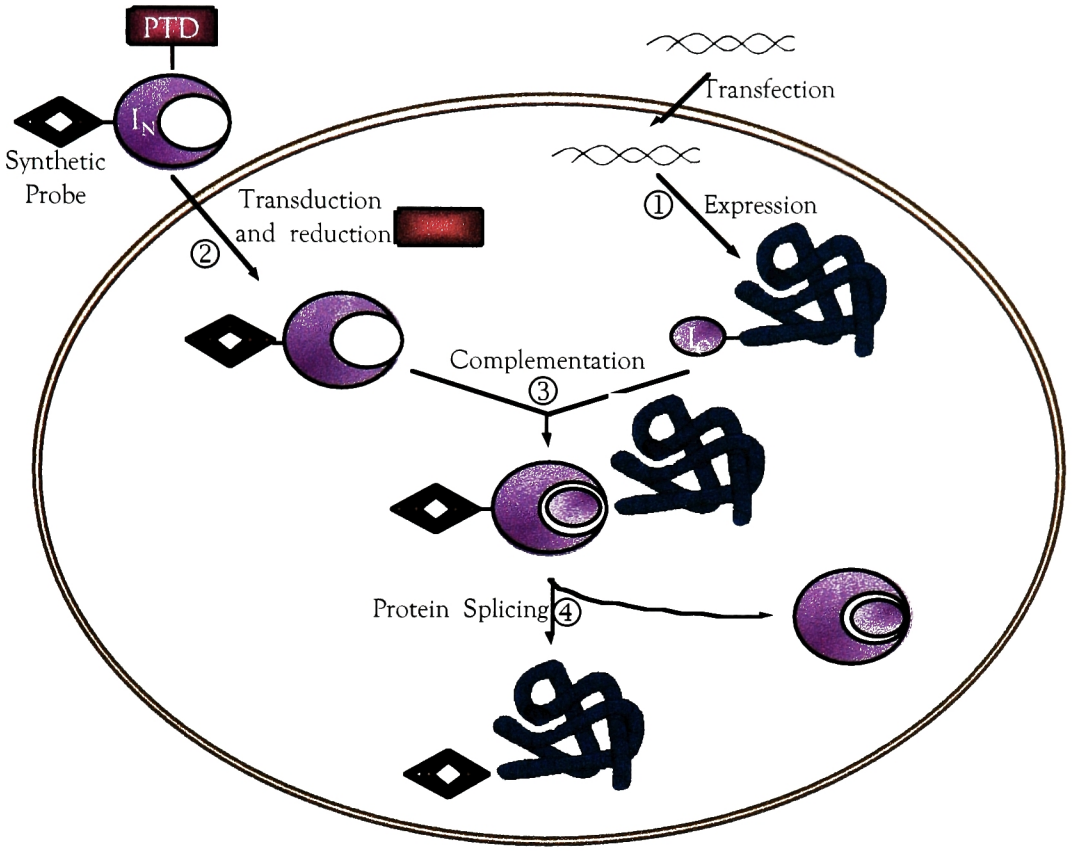


Figure 3.1. Principle of semi-synthetic protein trans-splicing at the N-terminus of proteins. The I_N and I_C components of the semi-synthetic protein *trans*-splicing methodology developed are interchanged to allow ligation of chemical moieties at the N-terminus of proteins.

Furthermore, since the reaction is triggered by the addition of one of the components onto the media of cells already expressing the other component, temporal control is inherent to the system. Lastly, protein *trans*-splicing could permit ligation at the C- or N-terminus of proteins by interchanging of the split intein components (compare Figure 2.5 with Figure 3.1).

This thesis explored protein *trans*-splicing as a strategy for the semi-synthesis of proteins *in vivo*, however, alternative methodologies could be developed. Specifically, NCL and EPL both have characteristics that make

them poteintially suitable for semi-synthesis in living cells. NCL is a robust reaction that can take place under physiological conditions and in the presence of all the functionalities commonly found in proteins, including free cysteine sulfhydryls.[36] Furthermore, the end product of NCL contains a native peptide backbone at the point of ligation.[36] EPL, with NCL as its last step,[103] inherits these same favorable characteristics. Additionally modified inteins have been shown to be active *in vivo*,[20] and cells are known to tolerate thiols needed as co-factors in both NCL and EPL.[65] Preliminary experiments have been conducted to test the feasibility of using NCL and EPL for the semi-synthesis of protein in living cells (data not shown).

Figure 3.2 illustrates the general methodology for semi-synthesis of proteins *in vivo* based on NCL. The protein of interest is expressed in cultured cells fused at its N-terminus to ubiquitin followed by a cysteine (Figure 3.2 ①). Co-translational removal of ubiquitin by ubiquitin hydrolase (naturally present in cells) unmasks the N-terminal cysteine, required for NCL, within the model protein (Figure 3.2 ②). Then, a synthetic molecule with an α-thioester is introduced into the cell (by use of PTDs), whereupon it can react with the available N-terminal cysteine (Figure 3.2 ③), resulting in the ligation of the probe to the selected protein through a normal peptide bond (Figure 3.2 ④). Key to this methodology will be the correct removal of ubiquitin from the protein of interest, and the stability of the resulting protein with an N-terminal

cysteine, although, removal of ubiquitin from a polypeptide in mammalian cells has been shown to work efficiently in the Muir laboratory (Hofmann and Muir, unplublished data). Another issue could be the stability of the synthetic α-thioester within the cell; however the latter issue could be by-passed if the reaction were to take place on the cell surface. This strategy would use a minute reactive tag on

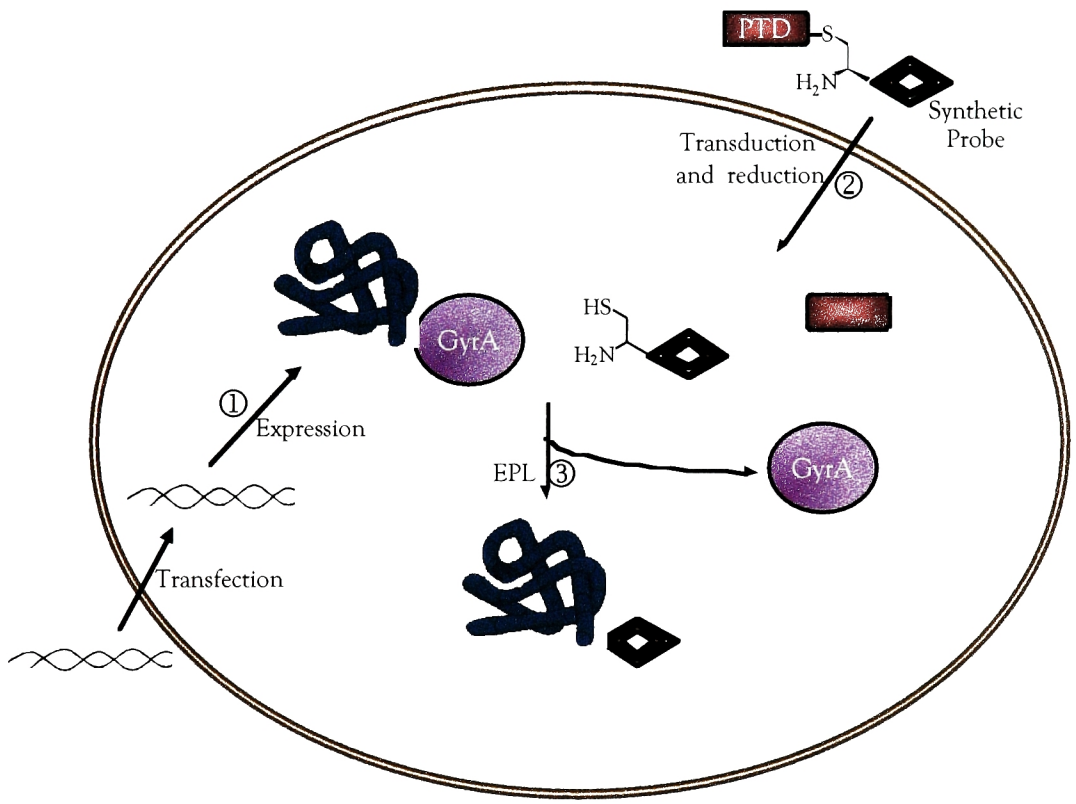


Figure 3.3. Principle of protein semi-synthesis based on EPL in living cells.

the protein of interest, and therefore have minimal effect on the function and localization of the protein. Furthermore, the chemical moiety would not have to be synthesized in tandem to an auxiliary polypeptide (*e.g.* the split intein halves) making the system potentially more straightforward to use.

Figure 3.3 illustrates the general methodology for semi-synthesis of protein *in vivo* based on EPL. The protein of interest is expressed in cultured cells with a modified GyrA intein fused to its C-terminus (Figure 3.3 ①). Then, a semi-synthetic cysteine derivative is introduced into the cell using PTDs (Figure 3.3 ②). The cysteine derivative would attack the α-thioester generated by the intein, displacing the intein and attaching itself to the protein through a native peptide bond (Figure 3.3 ③). Key to this

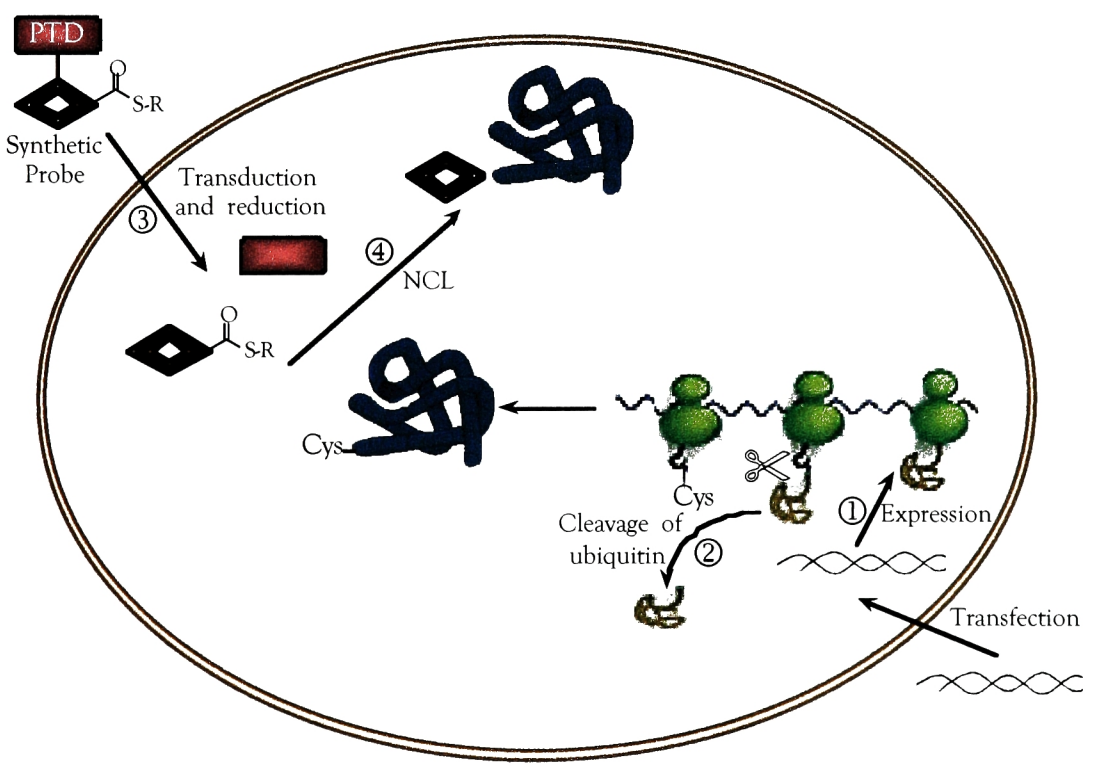


Figure 3.2. Principle of protein semi-synthesis based on NCL in living cells. Ubiquitin is in brown and the ubiquitin hydrolase is represented as a pair of scissors. When the ubiquitin is removed an N-terminal cysteine is unmasked.

methodology will be the specificity of the reaction in terms of the incoming peptide, and the minimization of *in vivo* cleavage. This strategy could by-pass some of the solubility and localization problems observed with protein *trans*-splicing, as the modified GyrA intein tends to express well. Furthermore, and as previously mentioned, the methodology would be more straightforward because the chemical moiety would not have to be synthesized in tandem to an auxiliary polypeptide (*e.g.* the split intein halves). It is important to note that both NCL and EPL could allow for semi-synthesis on membrane proteins, both intra and extra-cellularly.

Protein semi-synthesis *in vivo* could become a powerful tool for the analysis of proteins within their natural context. Probes, such as fluorophores, cross-linkers, unnatural amino acids and other sensors can now be added onto cellular proteins, using protein *trans*-splicing, in order to study function. The development of NCL and EPL for use in protein semi-synthesis *in vivo* could further increase the permutations for chemically modifying proteins, and open new routes for investigating and understanding polypeptides. Furthermore, all of the methodologies would have inherited temporal control. Table 3.1 presents some possible applications for semi-synthetic protein *trans*-splicing *in vivo*.

Ultimately, protein semi-synthesis could become an effective instrument for whole animal studies. To explore the utility of semi-synthetic proteins as

tools for biological questions in the context of whole animals, a modular tandem ligation method was developed. This approach combines three building blocks: the protein of interest, a second stage affinity tag, and a chemical moiety of choice.

Table 3.1. Possible applications for semi-synthetic protein trans-splicing in vivo.

Determination of protein turn-over rate	Protein turn-over could be determined independent of the half-life of the mRNA using this methodology. For example, to determine the turn-over rate of the membrane protein Opsin, a construct comprised of Opsin-I_N could be expressed in mammalian cells and "pulse" labeled with FLAG tag. The time that Opsin takes to completely turn-over would be determined by monitoring the disappearance of Opsin-FLAG.
Bio-sensor for membrane proteins	A sodium ion channel could be labeled with a small molecule that chelates Na^+ ions and, in doing so, becomes fluorescent. Such an approach would allow direct imaging of Na^+ transport through a specific channel.
Post-translational modifications	Post-translational modifications of proteins play a crucial role in cell biology. A common modification involves the addition or removal of phosphate groups. This methodology could allow the addition of non-hydrolysable phosphate groups to proteins of interest in an attempt to understand the role of

	these modifications. For example Smad-2, known to have two phosphorylation sites at its C-terminus, could be semi-synthetically modified to produce a unnatural version of Smad-2 containing two non-hydrolysable phosphate groups. This would allow to study the effects of a constitutevily active Smad-2.
Site- specific proteolysis	The breakage of a bond that takes place in protein splicing can be taken advantage of to create a site-specific proteolysis tool. Such a tool would allow to temporally control the separation of two fused poly-peptides inside the cell. For example, a kinase inhibitor peptide could be cloned onto the C-terminus a kinase of interest (PKA), with I_N present in between the two (PKA-I_N-Inhibitor). When this construct would be expressed in mammalian cells the PKA would be inactive due to the presence of the peptide inhibitor. Upon addition of I_C, semi-synthetic protein trans-splicing would remove the inhibition and PKA could become active. Such work in underway in the laboratory.

The first and last units are dictated by the experiment at hand, but the intervening unit can be varied to fit circumstantial needs, such as the conditions for purification. VEGF was used as a model protein in an attempt to develop a tumor-imaging agent. A ^{99m}Tc chelator was ligated to the C-terminus of $VEGF_{110}$ with a 6xHis second stage affinity handle, which

allowed recovery of a chemically homogenous monomeric protein. The refolded and dimeric semi-synthetic $VEGF_{110}$-Peptide 5 was shown to be biologically active and able to target tumors in mice.

This thesis has developed methodologies for the application of synthetic protein chemistry to *in vivo* systems. Effective approaches have been established to study proteins in both cellular and whole animal contexts in an attempt to provide tools to address the question of lifes' regulation by proteins.

Chapter 4: Materials and Methods.

General Methods.

All amino acid derivatives and resins were purchased from NovaBiochem (San Diego, CA), Peninsula Laboratories (Belmont, CA) and Bachem (Torrance, CA). All other chemical reagents were purchased from Sigma-Aldrich (Milwaukee, WI) or Fisher (Chicago, IL). Restriction enzymes were purchased from New England Biolabs (Boston, MA). Cell culture reagents were purchased from Gibco BRL. Fugene transfection reagent was bought from Boehringer Mannheim. Analytical gradient HPLC was performed on a Hewlett-Packard 1100 series instrument with detection at 214 and 280 nm. Analytical HPLC was performed on a Vydac C18 column (5 micron, 4.6 x 150 mm) at a flow rate of 1 mL/min. Preparative HPLC was routinely performed on a Waters DeltaPrep 4000 C18 column (15-20 micron, 50x250 mm) at a flow rate of 30 mL/min. All runs used linear gradients of 0.1% aqueous TFA (solvent A) *vs.* 90% acetonitrile plus 0.1% TFA (solvent B). Mass spectrometric analysis was routinely applied to all synthetic peptides and components of reaction mixtures. Electrospray mass spectrometry (ESMS) was performed on a Sciex API-100 single quadrupole electrospray mass spectrometer. Antibodies used were from Zymed, Sigma and Clontech. Primers were ordered from IDT or Invitrogen. HeLa, MCF-

7 and CHO cells were a generous gift from the Simon laboratory. HUVEC cells and their media were purchased from Cell Applications.

Peptide Synthesis.

All peptides were synthesized on a 0.5 mmol scale on a 4-methylbenzhydrylamine (MBHA) resin according to the *in-situ* neutralization/HBTU activation protocol for Boc SPPS, unless otherwise noted.[124] Following chain assembly, global de-protection and cleavage from the support was achieved by treatment with HF containing 4% v/v *p*-cresol, for 1 hour at 0°C. Following removal of the HF, the crude peptide products were precipitated and washed with anhydrous cold Et_2O before being dissolved in aqueous acetonitrile (50% B) and lyophilized.

Peptide 1:

H-MVKVIGRRSLGVQRIFDIGLPQDHNFLLANGAIAANCFGCGRQIKI

WFQNRRMKWKKGGDYKDDDDKGK(Rh)G-NH_2

The peptide was made by NCL between amino acids 1-39 (Peptide **a** - synthesized with an C-terminal α-thioester) and amino acids 40-70 (Peptide **b** - synthesized with an N-terminal cysteine).

The crude peptide **a** was purified by preparative HPLC using a linear gradient of 30-40% B over 60 minutes. The purified peptide was

characterized as the desired product by ESMS [observed mass (OM) = 4300.1 ± 1.2 Da; expected mass (average isotopic mass)(EM) = 4301.5 Da].

Rh was attached to the ε-NH_2 group of Lys^{30} of peptide **b** using an orthogonal (Fmoc/Boc) protection strategy. The crude peptide **b** was purified by preparative HPLC using a linear gradient of 20-40% B over 60 minutes. The purified peptide was characterized as the desired product by ESMS [OM = 4170.5 ± 1.4 Da; EM = 4172.6 Da].

Peptide **a** (10 mgr; 2.3 μmols) was mixed with peptide **b** (5 mgr; 1.2 μmols) in 50 mM Tris/HCl pH 7.5 with 100 mM MESNA and 20% (v/v) acetonitrile. The reaction was allowed to take place overnight at room temperature, and then was quenched with acid. The crude mixture was reduced with TCEP and peptide 1 was purified by preparative HPLC using a linear gradient of 35-45% B over 60 minutes. 11 mgr of peptide 1 was obtained. The purified peptide was characterized as the desired product by ESMS [OM = 8366.3 ± 1.1 Da; EM = 8366.9 Da].

Peptide 2 (L-FLAG):

H-MVKVIGRRSLGVQRIFDIGLPQDHNFLLANGAIAAN

CFNDYKDDDDKG-NH_2

The crude peptide was purified by preparative HPLC using a linear gradient of 25-45% B over 60 minutes. 41.7 mgr of peptide 2 was

obtained. The purified peptide was characterized as the desired product (Ic-FLAG) by ESMS [OM = 5319.2 ± 1.1 Da; EM = 5322.5 Da].

Ic-FLAG-Rh:

H-MVKVIGRRSLGVQRIFDIGLPQDHNFLLANGAIAAN

CFDYKDDDDK(Rh)G-NH_2

Rh was attached to the εNH_2 group of Lys^{47} using an orthogonal (Fmoc/Boc) protection strategy. The crude peptide was purified by preparative HPLC using a linear gradient of 30-50% B over 60 minutes. 10.3 mgr of Ic-FLAG-Rh was obtained. The purified peptide was characterized as the desired product (Ic-FLAG-Rh) by ESMS [OM = 5732 ± 1.6 Da; EM = 5732 Da].

Ic(SR)-FLAG:

Ac-CMVKVIGRRSLGVQRIFDIGLPQDHNFLLANGAIAAN

C(*S*-CH_2CONH_2)FNDYKDDDDKG-NH_2

The resin from the Ic-FLAG synthesis was used to couple an extra N-acetylated Cys residue protected on the thiol group with a 9-fluoronylmethyl (Fm) group. The crude peptide was purified by preparative HPLC using a linear gradient of 35-55% B over 60 minutes. The purified peptide was characterized as Ac-Cys(Fm)-Ic-FLAG by ESMS [OM = 5687.3 ± 1.7 Da; EM

= 5687.7 Da]. The peptide was then alkylated in solution by resuspending it in 5 mL of DMF with 5% DIEA with 10 mM iodoacetamide. The reaction was allowed to proceed for 15 minutes at room temperature and 500 μL of 1M DTT was then added to quench the un-reacted iodoacetamide. After 5 minutes at room temperature, piperidine was added to a final concentration of 20% to deprotect the N-terminal Cys. The reaction was then quenched with 5 volumes of 20% buffer B and was purified by preparative HPLC using a linear gradient of 0-70% B over 60 minutes. 3.2 mgr of I_c(S-R)-FLAG was obtained. The purified peptide was characterized as the desired product (I_c(S-R)-FLAG) by ESMS [OM = 5566.2 ± 1.7 Da; EM = 5567.6 Da].

ANTP:

H-C(NPyS)RQIKIWFNNRRMKWKKKG-NH_2

The crude peptide was purified by preparative HPLC using a linear gradient of 35-45% B over 60 minutes. 74.4 mgr of ANTP was obtained. The purified peptide was characterized as the desired product (ANTP) by ESMS [OM = 2730.0 ± 0.1 Da; EM = 2731.3 Da].

Rh-ANTP:

H-C(NPyS)RQIKIWFNNRRMKWKKK(Rh)G-NH_2

Rh was attached to the ε-NH_2 group of Lys^{18} using an orthogonal (Fmoc/Boc) protection strategy. The crude peptide was purified by preparative HPLC using a linear gradient of 20-40% B over 60 minutes. 6.6 mgr of Rh-ANTP was obtained. The purified peptide was characterized as the desired product (Rh-ANTP) by ESMS [OM = 2842.5 ± 0.5 Da; EM = 2842.9 Da].

Peptide 3:

H-CGRGK(Rh)G-NH_2

Rh was attached to the ε-NH_2 group of Lys^5 using an orthogonal (Fmoc/Boc) protection strategy. The crude peptide was purified by preparative HPLC using a linear gradient of 10-30% B over 60 minutes. 13 mgr of peptide 3 was obtained. The purified peptide was characterized as the desired product (peptide 3) by ESMS [OM = 987.9 ± 0.3 Da; EM = 988.2 Da].

Coupling of the I_c-probe to the PTD.

I_c-FLAG, I_c-FLAG-Rh or I_c(SR)-FLAG was dissolved together with ANTP or Rh-ANTP in degassed 50 mM Tris/HCl buffer pH 7.5 to a concentration of 1 mM in each. The peptides were allowed to react for 5 minutes at room temperature and the reaction was then quenched with 10

volumes of buffer A and purified over preparative HPLC using a liner gradient of 20-40% B over 60 minutes. The purified peptides were characterized as the desired products by ESMS [ANTP-S-S-Ic-FLAG, OM = 7895.8 ± 0.5 Da, EM = 7895.5 Da; ANTP-S-S-Ic-FLAG-Rh, OM = 8309.8 ± 1.5 Da, EM = 8309.8 Da; Rh-ANTP-S-S-Ic-FLAG, OM = 8267.2 ± 2.5 Da, EM = 8265.3 Da; ANTP-S-S-Ic(SR)-FLAG, OM = 5566.3 ± 1.7 Da, EM = 5568.7 Da].

Peptide 4:

Thial-PHHHHHH-SEt

The peptide was synthesized on a S-propionamide derivatized PEGA resin on a 0.15 mmol scale. The peptide was simultaneously cleaved from the resin and de-protected by overnight stirring in 10% aqueous ethanethiol. The ethanethiol was then removed *in vacuo*, and the resulting yellow solution was lyophilized. The crude peptide was purified by preparative HPLC using a linear gradient of 10-25% B over 60 minutes. The purified peptide was characterized as the desired product (Peptide 4) by ESMS [OM = 1057 ± 0.1 Da; EM = 1056.1 Da].

Peptide 5:

H-CHHHHHH-SN1325-01

SN1325-01 (12 mgr; 0.0287 mmol) and peptide 4 (47 mgr; 0.0431) were dissolved in 15 mL of buffer containing 200 mM sodium phosphate, 200 mM NaCl, 0.1% Triton X-100, 100 mM MESNA pH 7.2. The reaction was allowed to proceed for 3 days at room temperature. 5 mL of 1N NaOH was added to bring the pH to 12.5 and hydrolyze remaining α-thioesters. After 3 minutes, 7.5 mL of 1M HCl were added to bring the pH to 4.5. Methoxylamine-HCl (1.127 grams) was then added and the deprotection reaction was allowed to take place overnight at room temperature. The ligation product was purified by preparative HPLC using a linear gradient of 10-25% B over 60 minutes. The purified peptide was characterized as the desired product (Peptide 4) by ESMS [OM = 1400.3 ± 0.5 Da; EM = 1401.4 Da].

Expression plasmids.

All plasmids were constructed using standard molecular biology techniques,[123] and they were sequenced to verify correct coding.

pGFP-I_N: A PCR fragment amplified with IGpr63 and IGpr64, encoding I_N with 5 N-terminal extein residues (KFAEY) (using the plasmid reported in [151] as the template) was inserted at the C-terminus of GFP on pEGFPN1 (Clontech) between the BsRG I and Not I restriction sites.

p6xHis-GFP-I_N: A PCR fragment amplified with IGpr102 and IGpr103, encoding 6xhisGFP-I_N (using pGFP-I_N as the template) was inserted into pET11d (Novagene) between the Nhe I and BamH I restriction sites.

pGFP-I_N(Ala^{250}): Cys^{250} of pGFP-I_N was mutated to Ala using Quickchange Site-Directed Mutagenesis Kit (Stratagene) with IGpr111 and IGpr112.

pN1: The sequence of EGFP was removed from pEGFPN1 by digestion with Bsp 120 I and Not I, and the backbone was then re-ligated to make pN1.

pGST-I_N: A PCR fragment amplified with IGpr81 and IGpr83, encoding GST (using pGEX 2T (Amersham Bioscience) as the template) was inserted into pN1 between the Bgl II and EcoR I restriction sites. A PCR fragment amplified with IGpr84 and IGpr85, encoding I_N with 5 N-terminal extein, was then inserted at the C-terminus of GST between the Hind III and EcoR I restriction sites.

pDHFR-I_N: A PCR fragment amplified with IGpr86 and IGpr87, encoding 6xhisDHFR (using $pMT4^{90}$ as template) was inserted into pN1 between the Bgl II and Nhe I restriction sites. A PCR fragment amplified with IGpr90

and IGpr91, encoding I_N with 5 N-terminal extein, was then inserted at the C-terminus of 6xhisDHFR between the Xma I and Bgl II restriction sites.

pYFP-I_N-CFP: A PCR fragment amplified with IGpr99 and IGpr101, encoding I_N with 5 N-terminal extein, was inserted at the N-terminus of CFP on pCFP (Clontech) between the EcoR I and BamH I restriction sites. The resulting intermediate plasmid was digested with BsiW I and Not I to removed the I_N-CFP fragment and inserted it onto the C-terminus of YFP on pYFPN1 (Clontech) between the BsRG I and Not I restriction sites.

pOpsin-GFP-I_N: A PCR fragment amplified with IGpr75 and IGpr92, encoding Opsin, was inserted at the N-terminus of GFP on pEGFPN1 between the Sac I and EcoR I restriction sites. Then a PCR fragment amplified with IGpr63 and IGpr64, encoding I_N with 5 N-terminal extein residues (KFAEY) (using the plasmid reported in [151] as the template) was inserted at the C-terminus of Opsin-GFP between the BsRG I and Not I restriction sites.

pTnfR-GFP: A PCR fragment amplified with IGpr29 and IGpr30, encoding the trans-membrane domain of TnfR, was inserted at the N-terminus of GFP on pEGFPN1 between the Hind III and EcoR I restriction sites.

pTnfR-GFP-I_N: A PCR fragment amplified with IGpr63 and IGpr64, encoding I_N with 5 N-terminal extein residues (KFAEY) (using the plasmid reported in [151] as the template) was inserted at the C-terminus of TnfR-GFP between the BsRG I and Not I restriction sites.

pI_C-GyrA-CBD: A PCR fragment amplified with IGpr65 and IGpr66, encoding I_C with 3 C-terminal extein residues (CFN) (using the plasmid reported in [151] as template) was inserted at the N-terminus of Gyrase A on pTXB1 (NEB) between the Nde I and Sap I restriction sites.

p$VEGF_{110}$-GyrA-CBD: A PCR fragment amplified with RFpr09 and RFpr10, encoding $VEGF_{110}$ with four extra C-terminal amino acids (AAAAGG) (using a hVEGF coding plasmid from Research Genetics as template) was inserted at the N-terminus of Gyrase A on pTXB1 (NEB) between the Nde I and Sap I restriction sites.

Primers.

IGpr63: CGCGCGTACGGCGGCGGCGGCGGCGGCAAGTTTGCGGAAT ATTGC

IGpr64: CGCGCGGCGGCCGCTTATTTAATTGTCCCAGCG

IGpr102: CTAGCTAGCATGCACCACCACCACCACCACGGGGTGAGCA

AGGGCGAGG

IGpr103: CGGGATCCGCGGCCGCTTATTTAATTGTCCCAGCGTC

IGpr111: GCAAGTTTGCGGAATATGCCCTCAGTTTTGGCACC

IGpr112: GGTGCCAAAACTGAGGGCATATTCCGCAAACTTGC

IGpr81: GAAGATCTATGTCCCCTATACTAGG

IGpr83: CGAATTCTCATCAAAGCTTATCCGATTTTGGAGGATGG

IGpr84: GGGAAGCTTAAGTTTGCGGAATATTGCC

IGpr85: CGAATTCTCATTATTTAATTGTCCCAGCG

IGpr86: GGGGCTAGCATGCATCATCATCATCATCATGTTCGACCATTG

AACTCG

IGpr87: CCCGAGATCTGATTTCTTCTCGTAGACTTCAAAC

IGpr90: CCCCCCGGGAAGTTTGCGGAATATTGCC

IGpr91: GGGAGATCTTCATTATTTAATTGTCCCAGCGTC

IGpr99: GGAATTCCGTACGGGAAGTTTGCGGAATATTGCC

IGpr101:

CGGGATCCCCTCCTCCTCCTCCTCCTTTAATTGTCCCAGCGT

C

IGpr75: CCCGAGCTCATGAACGGTACCGAAGGCCC

IGpr92: GGAATTCGGGCAGGCGCCACTTGGCTGG

IGpr29:

CCCAAGCTTGCCGCCACCATGGCCATGGATCAAGCTAGATCA

GC

IGpr30: GGAATTCGTGCAGGGAAGTCCTCTCC

IGpr65: GGGAATTCCATATGGTTAAAGTTATCGG

IGpr66: GGGCCCGTACCGTTAAAACAATTGGCGGC

RFpr09: GGAATTCCATATGGCACCCATGGCAGAAGG

RFpr10: CCCAAGGCGCCGCCGGCGGCGGCGGCTGGTCTGCATTCAAC

ATTTG

Bacterial expression and purification.

6xHisGFP-I_N: *E. coli* BL21 (DE3) cells (Invitrogen) were transformed with p6xhisGFP-I_N. Expression was carried out in LB media at 37°C. Cells were grown to an OD_{590nm} of 0.8, and protein expression was induced by addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. After 5 hours the cells were harvested by centrifugation. The cell pellet was resuspended in bacterial lysis buffer (100 mM Tris/HCl pH 7.5, 500 mM NaCl, 20 mM β-mercaptoethanol and protease inhibitors (Roche)) and disrupted using a french press. Insoluble material was pelleted by centrifugation and the supernatant was added onto Ni-NTA agarose beads (Novagen) previously equilibrated in lysis buffer. After overnight incubation

at 4°C the beads were washed twice with lysis buffer and the protein was eluded with lysis buffer containing 500 mM imidazole. The eluant was dialyzed against storage buffer (100 mM Tris/HCl pH 7.5, 500 mM NaCl, 1 mM EDTA and 10 mM DTT), and stored at 4°C.

Ic-GyrA-CBD: *E. coli* BL21 (DE3) cells were transformed with pIc-GyrA-CBD. Expression was carried out in LB media at 25°C. Cells were grown to an OD_{590nm} of 0.8, and protein expression was induced addition of IPTG to a final concentration of 1 mM. After 5 hours the cells were harvested by centrifugation. The cell pellet was resuspended in bacterial lysis buffer (50 mM Tris/HCl pH 7.5, 200 mM NaCl, 5% glycerol, 0.1 mM EDTA and protease inhibitors) and disrupted using a french press. Insoluble material was pelleted by centrifugation and the supernatant was added onto chitin beads (NEB) previously equilibrated in lysis buffer. After overnight incubation at 4°C the beads were washed twice with lysis buffer and stored at 4°C.

$VEGF_{110}$-GyrA-CBD: *E. coli* BL21 (DE3) cells were transformed with $pVEGF_{110}$-GyrA-CBD. Expression was carried out in LB media at 37°C. Cells were grown to an OD_{590nm} of 0.8, and protein expression was induced by addition of IPTG to a final concentration of 1 mM. After 5 hours the cells were

harvested by centrifugation. The cell pellet was resuspended in bacterial lysis buffer (25 mM sodium phosphate pH 7.0, 250 mM NaCl, 0.1 mM EDTA, 5% glycerol, and protease inhibitors) and disrupted using a french press. Insoluble material was pelleted by centrifugation. The pellet was washed with 2 M urea in lysis buffer, and insoluble material was pelleted by centrifugation. The pellet was resuspended in 200 mM sodium phosphate pH 7.2 and 8 M urea and insoluble material removed by centrifugation. The supernatant was dialyzed for 24 hours against 50 mM Tris/HCl pH 7.5, 200 mM NaCl, 1 mM EDTA and 0.1% Triton X-100 in a Spectra Pore dialysis bag with a 10,000 Da molecular weight cutoff. The dialyzed solution was centrifuged to remove insoluble components, and was then loaded onto a chitin column (NEB) previously equilibrated with 200 mM sodium phosphate pH 7.2, 200 mM NaCl, 0.1% triton X-100. After overnight incubation at 4°C, the column was washed with buffer containing 200 mM sodium phosphate pH 7.2, 200 mM NaCl, 0.1% triton X-100 and stored at 4°C.

In vitro reactions.

Semi-synthetic protein trans-splicing: The bacterially expressed 6xHisGFP-I_N in 100 mM Tris/HCl pH 7.5, 500 mM NaCl, 1 mM EDTA and 10 mM DTT was added to I_C-FLAG to obtain a final concentration of the peptide of 100

μM. The mixture was allowed to react at room temperature for 5 hours and was then analyzed by SDS-PAGE gel (12%) and by an anti-FLAG blot.

EPL of Ic-GyrA-CBD with peptide 3: An aliquot of the bead slurry containing immobilized Ic-GyrA-CBD was incubated overnight with 1 mL of peptide 3 (1 mM) in PBS with 100 mM MESNA at room temperature. The reaction was monitored by SDS-PAGE. The supernatant was removed and the beads were washed with 8 M urea. The eluant was injected into the analytical HPLC to analyze the reaction, and to confirm product formation by ESMS [Ic-Peptide 3: OM = 5298.3 ± 2.0 Da; EM = 5299.7 Da].

EPL between $VEGF_{110}$-GyrA-CBD and SN1400-01: $VEGF_{110}$-GyrA-CBD chitin beads (30 mL) were treated in 50 mL ligation buffer (200 mM sodium phosphate pH 7.2, 200 mM NaCl, 0.1% triton X-100) with 100 mM MESNA and 2% (v/v) ethanethiol in the presence of SN1400-01 peptide (10 mgr; 22.4 μmols) and rocked for 2 days at room temperature. The column was drained, washed with ligation buffer, and the ligation product was eluted by washing the column with a buffer containing 8 M urea, 200 mM sodium phosphate, 20 mM DTT, pH 7.2. The urea solution was then dialyzed against 20 mM Tris/HCl pH 8.0, 200 mM NaCl. A mixture of the ligation products, $VEGF_{110}$-SN1400-01 and $VEGF_{110}$ (hydrolysis product)

were purified by preparative HPLC using a linear gradient of 32-47% B over 60 minutes. The purified proteins were characterized by ESMS [$VEGF_{110}$-SN1400-01 OM = 12934.3 ± 1.4 Da; EM = 12930.6 Da and $VEGF_{110}$ OM = 12506.1 ± 1.4 Da; EM = 12502.3 Da].

EPL between $VEGF_{110}$-GyrA-CBD and peptide 5: $VEGF_{110}$-GyrA-CBD chitin beads were treated with 2% ethanethiol in ligation buffer for 2 days. The $VEGF_{110}$ α-thioester was eluted by overnight treatment with 8 M Urea, 200 mM sodium phosphate, pH 7.2 with 2% ethanethiol, and purified by preparative HPLC using a linear gradient of 40-55% B over 60 minutes. $VEGF_{110}$ ethyl thioester (5 mgr; 0.4 μmols) were then dissolved in 6 M guanidine, 200 mM sodium phosphate pH 7.2, 200 mM sodium chloride, 2% ethanethiol and 100 mM MESNA in the presence of peptide 5 (5 mgr; 11.9 μmols). The reaction was allowed to take place overnight at room temperature. The ligation mixture was purified by size-exclusion chromatography on a Zorbax GF-450 column using a flow rate of 5 mL/min and a running buffer containing 6 M guanidine, 200 mM NaCl, 200 mM sodium phosphate pH 7.2, 20 mM 2-mercaptoethanol, and 0.1 mg/mL BSA. The fraction containing $VEGF_{110}$-Peptide 5 was then further purified by Ni-NTA chromatography over a Hi-Trap Affinity column (Pharmacia), on a Pharmacia FPLC system, at a flow rate of 5 mL/min using 6 M guanidine,

200 mM sodium phosphate, 500 mM NaCl, 20 mM 2-mercaptoethanol pH 7.4 as loading buffer, and an identical buffer with 500 mM imidazole as the eluting buffer. Purified $VEGF_{110}$-Peptide 5 was oxidatively refolded according to literature protocols[30] by dialyzing against 20 mM Tris/HCl pH 8.0, 400 mM sodium chloride, 1 mM cysteine. Dimeric $VEGF_{110}$-Peptide 5 was then separated from monomeric and other oligomeric species by gel filtration chromatography on a Zorbax GF-250 column using a flow rate of 5 mL/min and PBS as the running buffer.

Fluorescence Microscopy.

Live cell imaging: HeLa were grown in DMEM+ 10% fetal bovine serum (FBS) media at 37°C with 5% CO_2 in a microscopy chamber. Peptide 1 was added to a final concentration of 1 nM. After 30 minutes incubation, the cells were washed with Hanks' balanced salt solution (HBSS) and imaged live by fluorescence microscopy on an Axiovert 200 M inverted microscope. A TxRd filter (Chroma) was used to image the Rh signal, and the data was analyzed using MetaMorph (Universal Imaging).

Fixed cell imaging for peptide localization: HeLa cells were grown in DMEM + 10% FBS media at 37°C with 5% CO_2, on #1.5 glass cover slips (Fisher). Before adding the peptides (ANTP-S-S-Ic-FLAG-Rh or Rh-ANTP-S-S-Ic-FLAG),

the cells were washed twice with PBS, and then fresh media was added containing the peptides at a concentration of 2 nM. After 30 minutes of incubation, the cells were fixed with 3.7% formaldehyde, followed by BSA blocking and nucleic acid staining with Hoechst. The cover slips were inverted onto a glass slide and sealed with nail polish before fluorescence microscopy analysis. A DAPI or a TxRd filter (Chroma) was used to visualize the signals from Hoechst or Rh, respectively. The data was analyzed using MetaMorph.

Fixed cell imaging for protein localization: HeLa or MCF-7 cells were grown in DMEM + 10% FBS media at 37°C with 5% CO_2, on #1.5 glass cover slips, and transiently transfected with the desired plasmids using Fugene. After 24 hours of expression the cells were washed twice with PBS and fixed with 3.7% formaldehyde, followed by BSA blocking and nucleic acid staining with Hoechst. The cover slips were inverted onto a glass slide and sealed with nail polish before fluorescence microscopy analysis. A FITC filter (Chroma) was used to image the GFP signal, and the data was analyzed using MetaMorph.

Western Analysis of pGFP-I_N.

CHO cells, grown in F12-HAM + 10% FBS media at 37°C with 5% CO_2, were transiently transfected with pGFP-I_N using Fugene. After 24 hours of expression the cells were washed twice with PBS and harvested. The cell pellet was resuspended in SDS sample buffer and loaded onto an SDS-PAGE gel (12%) for blotting against GFP.

Semi-synthetic protein trans-splicing.

10 cm plates of CHO cells were transfected with a mammalian vector based on pEGFP-N1 (Clontech) encoding the I_N fusion protein (GFP-I_N, GFP-I_N(Ala^{250}), GST-I_N, DHFR-I_N, YFP-I_N-CFP or Opsin-GFP-I_N). After 24 hours of transient expression, the cells were washed twice with 10 mL of PBS and fresh media (2.5 mL) containing 0.5% DMSO or 0.5% DMSO with the peptide (ANTP-S-S-I_C-FLAG or ANTP-S-S-I_C(SR)-FLAG) to a final concentration of 100 μM was added. The cells were then incubated for three hours, washed twice with 10 mL of PBS and harvested. The cell pellet was resuspended in 1 mL of lysis buffer: PBS containing 1% (w/v) docecylmaltoside (DM), 10 mM iodoacetamide, 2 mM N-ethyl-maleamide and protease inhibitors (Roche). All subsequent manipulations were performed at 4°C. Lysis was carried out for 1 hour after which insoluble material was

removed by micro-centrifugation at 16,000g for 30 minutes. The supernatant was pre-cleared for immuno-precipitation by incubation with 50 µL of Protein-A immobilized on sepharose CL-4B (Sigma) or mouse IgG agarose (Sigma) for 1 hour. The beads were then removed by micro-centrifugation at 16,000g for 5 minutes and the supernatant added onto 50 µL of anti-FLAG-Protein-A sepharose, anti-GFP-Protein-A sepharose or anti-FLAG M2-agarose affinity gel (Sigma) pre-blocked with bovine albumin (BSA). Immuno-precipitation was allowed to take place overnight and the beads were then washed three times for 5 minutes with 1 mL of PBS before being resuspended in SDS sample buffer for analysis by western blotting.

Endothelial Cell Proliferation Assay.

A plate of HUVEC primary cultures (Cell Applications) was grown to confluence over 4 days using endothelial cell growth media (Cell Applications). VEGF samples were reconstituted by dissolving or diluting the samples to generate a 10 µg/mL stock in PBS with 1 mg/mL BSA (Sigma). These stocks were diluted into assay medium (medium 199 with 10 mM HEPES, 10% FBS, 100 units/mL penicillin, 100 µg/mL streptomycin) at various concentrations, and 75 µL of each dilution was plated in triplicate in a 96-well plate. 50 µL of a solution of 10^5 cells/mL in assay medium was then added to each well. The cells were allowed to grow for 48 hours

under 5% CO_2 at 37°C degrees. 10 µL of a 50 µCi/mL stock of ^{3}H-Thymidine (NEN) was added to each well, and the cells were allowed to grow for an additional 24 hours. The cells were dissociated with trypsin, and counted in a scintillation counter. The data were fitted to a sigmoidal dose response curve using Prism 3.0 (Graphpad Software), using the following equation:

$$E = Basal + E_{max} - Basal \Big/ 1 + 10^{(LogEC_{50} - Log[A]^{nH})}$$

where E denotes effect, [A] the VEGF concentration, n_H the midpoint slope, EC_{50} the midpoint location parameter, and E_{max} and $Basal$ the upper and lower asymptotes, respectively.

Academic Course Work

Transmission genetics.

Introduction to protein function.

Development of the nervous system.

Mammalian genetics.

Protein folding and structure prediction.

Gene expression.

Advanced organic chemistry.

Laboratory Rotations

September 1997 - December 1997: Dr. Ali Hemmati Brivalauo

January 1998 - April 1998: Dr. John Kuryain

May 1998 - March 1999: Dr. Titia de Lange

Publications

Smith, S., Giriat, I., Schmitt, A., de Lange, T., (1998) *Science*, 282 p.1484

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

Giriat, I., Muir, T.W., Perler, F.B., (2001) *Genetic Engineering*, 23 p.171.

Stanger, H.E., Syud, F.A., Espinosa, J.F., Giriat, I., Muir, T., Gellman, S.H., (2001) *PNAS*, 98 p.12015.

Giriat, I.; Muir, T.W. (2003) submitted to JACS.

Bibliography.

1. Adams, S.R., Campbell, R.E., Gross, L.A., Martin, B.R., Walkup, G.K., Yao, Y., Llopis, J., Tsien, R.Y., (2002) *JACS*, **124** p.6063.

2. Allinquant, B., Hantraye, P., Mailleux, P., Moya, K., Bouillot, C., Prochiantz, A., (1995) *JCB*, **128** p.919.

3. Anderson, D.C., Nichols, E., Manger, R., Woodle, D., Barry, M., Fritzberg, A.R., (1993) *Biochem. Biophys. Res. Comm.*, **194** p.876.

4. Arano, Y., (2002) *Annals of Nuclear Medicine*, **16** p.79.

5. Arora, N., Masood, R., Zheng, T., Cai, J., Smith, D.L., Gill, P.S., (1999) *Cancer Research*, **59** p.183.

6. Axelrod, D., Koppel, D.E., Schlessinger, J., Elson, E., Webb, W.W., (1976) *Biophys. J.*, **16** p.1055.

7. Ayers, B., Blaschke, U.K., Camarero, J.A., Cotton, G.J., Holford, M., Muir, T.W., (2000) *Biopolymers*, **51** p.343.

8. Backer, M.V., Aloise, R., Przekop, K., Stoletov, K., Backer, J.M., (2002) *Bioconjugate Chemistry*, **13** p.462.

9. Backer, M.V., Backer, J.M., (2001) *Bioconjugate Chemistry*, **12** p.1066.

10. Baird, G., Zacharias, D.A., Tsien, R.Y., (1999) *PNAS*, **96** p.11241.

11. Baird, G.S., Zacharias, D.A., Tsien, R.Y., (2000) *PNAS*, **97** p.11984.

12. Banerjee, S., Pillai, M.R., Ramamoorthy, N., (2001) *Seminars in Nuclear Medicine*, **31** p.260.

13. Barka, T., Van der Noen, H.M., (1996) *Hum. Gene Ther.*, **7** p.613.

14. Barka, T., van der Noen, H.M., (1997) *J. Histochem. Cytochem.*, **45** p.1533.

15. Becker-Hapak, M., McAllister, S.S., Dowdy, S.F., (2001) *Methods*, **24** p.247.

16. Berlose, J.P., Convert, O., Derossi, D., Brunissen, A., Chassaing, G., (1996) *Eur. J. Biochem.*, **242** p.372.

17. Bogdanov, A., Jr., Simonova, M., Weissleder, R., (1998) *Biochimica et Biophysica Acta*, **1397** p.56.

18. Budisa, N., Minks, C., Alefelder, S., Wenger, W., Dong, F., Moroder, L., Huber, R., (1999) *FASEB Journal*, **13** p.41.

19. Calnan, B.J., Biancalana, S., Hudson, D., Frankel, A.D., (1991) *Genes Dev.*, **5** p.201.

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

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

22. Camarero, J.A., Muir, T.W., (1999) *JACS*, **121** p.5597.

23. Caron, N.J., Torrente, Y., Camirand, G., Bujold, M., Chapdelaine, P., Leriche, K., Bresolin, N., Tremblay, J.P., (2001) *Molecular Therapy*, 3 p.310.

24. Chin, J.W., Martin, A.B., King, D.S., Wang, L., Schultz, P.G., (2002) *PNAS*, 99 p.11020.

25. Chin, J.W., Santoro, S.W., Martin, A.B., King, D.S., Wang, L., Schultz, P.G., (2002) *JACS*, 124 p.9026.

26. Choi, S.R., Yang, B., Plossl, K., Chumpradit, S., Wey, S.P., Acton, P.D., Wheeler, K., Mach, R.H., Kung, H.F., (2001) *Nuclear Medicine And Biology*, 28 p.657.

27. Chong, S., Mersha, F.B., Comb, D.G., *et al.*, (1997) *Gene*, 192 p.271.

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

29. Chong, S., Williams, K.S., Wotkowicz, C., Xu, M.-Q., (1998) *J. Biol. Chem.*, 273 p.10567.

30. Christinger, H.W., Muller, Y.A., Berleau, L.T., Keyt, B.A., Cunningham, B.C., Ferrara, N., de Vos, A.M., (1996) *Proteins*, 26 p.353.

31. Conn, G., Soderman, D.D., Schaeffer, M.T., Wile, M., Hatcher, V.B., Thomas, K.A., (1990) *PNAS*, 87 p.1323.

32. Cornish, V.W., Benson, D.R., Altenbach, C.A., Hideg, K., Hubbell, W.L., Schultz, P.G., (1994) *PNAS*, 91 p.2910.

33. Corringer, P.-J., Le Novere, N., Changeux, J.-P., (2000) *Annu. Rev. Pharmacol. Toxicol.*, **40** p.431.

34. Dang, H., England, P.M., Farivar, S.S., Dougherty, D.A., Lester, H.A., (2000) *Mol. Pharmacol.*, **57** p.1114.

35. Dawson, P.E., Kent, S.B.H., (2000) *Annu. Rev. Biochem.*, **69** p.923.

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

37. Derossi, D., Calvet, S., Trembleau, A., Brunissen, A., Chassaing, G., Prochiantz, A., (1996) *J. Biol. Chem.*, **271** p.18188.

38. Derossi, D., Chassaing, G., Prochiantz, A., (1998) *Trends Cell Biol.*, **8** p.84.

39. Derossi, D., Joliot, A.H., Chassaing, G., Prochiantz, A., (1994) *J. Biol. Chem.*, **269** p.10444.

40. Dockter, M.E., (1979) *J. Biol. Chem.*, **254** p.2161.

41. Dougherty, D.A., (1996) *Science*, **271** p.163.

42. Dougherty, D.A., (2000) *Curr. Opin. Chem. Bio.*, **4** p.645.

43. Dujon, B., Belfort, M., Butow, R.A., Jacq, C., Lemieux, C., Perlman, P.S., Vogt, V.M., (1989) *Gene*, **82** p.115.

44. Elliott, G., O'Hare, P., (1997) *Cell*, **88** p.223.

45. Ellis, R.J., (2000) *Trends Biochem. Sci.*, **25** p.210.

46. Embury, J., Klein, D., Pileggi, A., *et al.*, (2001) *Diabetes*, 50 p.1706.

47. Erlandson, D.A., Chytil, M., Verdine, G.L., (1996) *Chem. Biol.*, 3 p.981.

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

49. Evans, T.C., Martin, D., Kolly, R., *et al.*, (2000) *J. Bio. Chem.*, 275 p.9091.

50. Ezepchuk, Y.V., Fehringer, A.P., Harbeck, R., Freed, J.H., Leung, D.Y., (1999) *Mol Gen Mikrobiol Virusol*, 2 p.29.

51. Ezhevsky, S.A., Ho, A., Becker-Hapak, M., Davis, P.K., Dowdy, S.F., (2001) *MCB*, 21 p.4773.

52. Ezhevsky, S.A., Nagahara, H., Vocero-Akbani, A.M., Gius, D.R., Wei, M.C., Dowdy, S.F., (1997) *PNAS*, 94 p.10699.

53. Fawell, S., Seery, J., Daikh, Y., Moore, C., Chen, L.L., Pepinsky, B., Barsoum, J., (1994) *PNAS*, 91 p.664.

54. Ferrara, N., (2000) *Curr. Opin. Biotech.*, 11 p.617.

55. Ferrara, N., Davis-Smyth, T., (1997) *Endocrine Reviews*, 18 p.4.

56. Frankel, A.D., Pabo, C.O., (1988) *Cell*, 55 p.1189.

57. Futaki, S., Suzuki, T., Ohashi, W., Yagami, T., Tanaka, S., Ueda, K., Sugiura, Y., (2001) *J. Biol. Chem.*, 276 p.5836.

58. George, A.J., Jamar, F., Tai, M.S., *et al.*, (1995) *PNAS*, 92 p.8358.

59. Giblin, M.F., Wang, N., Hoffman, T.J., Jurisson, S.S., Quinn, T.P., (1998) *PNAS*, **95** p.12814.

60. Gimble, F.S., (2000) *FEMS Microbiol. Lett.*, **185** p.99.

61. Gimble, F.S., Thorner, J., (1992) *Nature*, **357** p.301.

62. Giriat, I., Muir, T.W., Perler, F.B., (2001) *Genetic Engineering*, **23** p.171.

63. Gonzalez, C., Bejarano, L.A., (2000) *Trends Cell Biol.*, **10** p.162.

64. Green, M., Loewenstein, P.M., (1988) *Cell*, **55** p.1179.

65. Griffin, B.A., Adams, S.R., Tsien, R.Y., (1998) *Science*, **281** p.269.

66. Griffiths, A.J.F.G., W.M.; Lewontin, R..C.; Miller, J.H., (2002) *Modern Genetic Analysis: Integrating Genes and Genomes*, .

67. Hackeng, T.M., Griffin, J.H., Dawson, P.E., (1999) *PNAS*, **96** p.10068.

68. Hall, H., Williams, E.J., Moore, S.E., Walsh, F.S., Prochiantz, A., Doherty, P., (1996) *Curr. Biol.*, **6** p.580.

69. Heim, R., Prasher, D.C., Tsien, R.Y., (1994) *PNAS*, **91** p.12501.

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

71. Hodges, R.A., Perler, F.B., Noren, C.J., Jack, W.E., (1992) *Nucleic Acids Res*, **20** p.6153.

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

73. Jayaraman, S., Haggie, P., Wachter, R.M., Remington, S.J., Verkman, A.S., (2000) *J. Biol. Chem.*, 275 p.6047.

74. Jo, D., Nashabi, A., Doxsee, C., Lin, Q., Unutmaz, D., Chen, J., Ruley, H.E., (2001) *Nature Biotech.*, 19 p.929.

75. Joliot, A., Pernelle, C., Deagostini-Bazin, H., Prochiantz, A., (1991) *PNAS*, 88 p.1864.

76. Kane, P.M., Yamashiro, C.T., Wolczyck, D.F., Neff, N., Goebl, M., Stevens, T.H., (1990) *Science*, 250 p.651.

77. Kanner, E.M., Klein, I.K., Friedlander, M., Simon, S.M., (2002) *Biochem.*, 41 p.7707.

78. Karginov, A.V., Lodder, M., Hecht, S.M., (1999) *Nucleic Acids Res.*, 27 p.3283.

79. Keppler, A., Gendreizig, S., Gronemeyer, T., Pick, H., Vogel, H., Johnsson, K., (2003) *Nature Biotech.*, 21 p.86.

80. Keyt, B.A., Nguyen, F.H., Ferrara, N.,. 1998, (Genentech, Inc., USA; Keyt, Bruce A.; Nguyen, Francis Hung; Ferrara, Napoleone). p. 66 pp.

81. Keyt, B.A., Nguyen, H.V., Berleau, L.T., Duarte, C.M., Park, J., Chen, H., Ferrara, N., (1996) *J. Biol. Chem.*, 271 p.5638.

82. Kneen, M., Farinas, J., Li, Y., Verkman, A.S., (1998) *Biophys. J.*, 74 p.1591.

83. Koh, J.T., Cornish, V.W., Schultz, P.G., (1997) *Biochem.*, 36 p.11314.

84. Kwon, H.Y., Eum, W.S., Jang, H.W., Kang, J.H., Ryu, J., Ryong Lee, B., Jin, L.H., Park, J., Choi, S.Y., (2000) *FEBS Lett.*, 485 p.163.

85. Lehninger, A.L., Nelson, D.L., Cox, M.M., (1993) *Principles of Biochemistry, Pt. 1. 2nd Ed,* .

86. Levin, V.A., (1980) *J. Med. Chem.*, 23 p.682.

87. Lew, B.M., Mills, K.V., Paulus, H., (1998) *J. Biol. Chem.*, 273 p.15887.

88. Lissy, N.A., Davis, P.K., Irwin, M., Kaelin, W.G., Dowdy, S.F., (2000) *Nature*, 407 p.642.

89. Llopis, J., McCaffery, J.M., Miyawaki, A., Farquhar, M.G., Tsien, R.Y., (1998) *PNAS*, 95 p.6803.

90. Lo, B.a.S., M., (1997) *J. Biol. Chem.*, 273 p.903.

91. Loret, E.P., Vives, E., Ho, P.S., Rochat, H., Van Rietschoten, J., Johnson, W.C., Jr., (1991) *Biochem.*, 30 p.6013.

92. Mabrouk, K., Van Rietschoten, J., Vives, E., Darbon, H., Rochat, H., Sabatier, J.M., (1991) *FEBS Lett.*, 289 p.13.

93. MacWhorter, S.E., Wu, T.-L., White, D.S.,. 1997, (Amersham International PLC, UK). p. 19 pp.

94. Mahajan, N.P.H.-S., D.C.; Michaux, J. and Herman B., (1999) *Chem. Biol.*, 6 p.401.

95. Mann, D.A., Frankel, A.D., (1991) *EMBO J.*, **10** p.1733.

96. Martin, D.D., Xu, M.Q., Evans, T.C., Jr., (2001) *Biochem.*, **40** p.1393.

97. Mathys, S., Evans Jr, T.C., Chute, C.I., Wu, H., Chong, S., Benner, J., Liu, X.-Q., Xu, M.-Q., (1999) *Gene*, **231** p.1.

98. Matz, M.V., Fradkov, A.F., Labas, Y.A., Savitsky, A.P., Zaraisky, A.G., Markelov, M.L., Lukyanov, S.A., (1999) *Nature Biotech.*, **17** p.969.

99. Mills, K.V., Lew, B.M., Jiang, S.-Q., Paulus, H., (1998) *PNAS*, **95** p.3343.

100. Mitchell, D.J., Kim, D.T., Steinman, L., Fathman, C.G., Rothbard, J.B., (2000) *J. Pep. Res.*, **56** p.318.

101. Miyawaki, A., Llopis, J., Heim, R., McCaffery, J.M., Adams, J.A., Ikura, M., Tsien, R.Y., (1997) *Nature*, **388** p.882.

102. Mootz, H.D., Muir, T.W., (2002) *JACS*, **124** p.9044.

103. Muir, T.W., Sondhi, D., Cole, P.A., (1998) *PNAS*, **95** p.6705.

104. Nagahara, H., Vocero-Akbani, A.M., Snyder, E.L., Ho, A., Latham, D.G., Lissy, N.A., Becker-Hapak, M., Ezhevsky, S.A., Dowdy, S.F., (1998) *Nature Med.*, **4** p.1449.

105. Nakanishi, J., Nakajima, T., Sato, M., Ozawa, T., Tohda, K., Umezawa, Y., (2001) *Anal. Chem.*, **73** p.2920.

106. Noren, C.J., Anthony-Cahill, S.J., Griffith, M.C., Schultz, P.G., (1989) *Science*, **244** p.182.

107. Nowak, M.W., Kearney, P.C., Sampson, J.R., *et al.*, (1995) *Science*, 268 p.439.

108. Ortega, N., Hutchings, H., Plouet, J., (1999) *Frontiers in Bioscience*, 4 p.D141.

109. Otomo, T., Ito, N., Kyogoku, Y., Yamazaki, T., (1999) *Biochem.*, 38 p.16040.

110. Ozawa, T., Kaihara, A., Sato, M., Tachihara, K., Umezawa, Y., (2001) *Anal. Chem.*, 73 p.2516.

111. Ozawa, T., Nogami, S., Sato, M., Ohya, Y., Umezawa, Y., (2000) *Anal. Chem.*, 72 p.5151.

112. Ozawa, T., Umezawa, Y., (2001) *Curr. Opin. Chem. Biol.*, 5 p.578.

113. Perler, F.B., (1999) *Trends Biochem Sci*, 24 p.209.

114. Perler, F.B., (2000) *Nucleic Acids Res*, 28 p.344.

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

116. Pietrokovski, S., (1998) *Protein Sci.*, 7 p.64.

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

118. Porter, J.A., Ekker, S.C., Park, W.-J., *et al.*, (1996) *Cell*, 86 p.21.

119. Prochiantz, A., (2000) *Curr. Opin. Cell Biol.*, 12 p.400.

120. Rizzuto, R., Brini, M., Pizzo, P., Murgia, M., Pozzan, T., (1995) *Curr. Biol.*, 5 p.635.

121. Rusckowski, M., Qu, T., Chang, F., Hnatowich, D.J., (1997) *J. Pept. Res.*, 50 p.393.

122. Rusckowski, M., Qu, T., Pullman, J., Marcel, R., Ley, A.C., Ladner, R.C., Hnatowich, D.J., (2000) *Journal Of Nuclear Medicine*, 41 p.363.

123. Sambrook, J.a.R., David W., (2001) *Molecular Cloning: A Laboratory Manual.*, .

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

125. Schwarze, S.R., Ho, A., Vocero-Akbani, A., Dowdy, S.F., (1999) *Science*, 285 p.1569.

126. Scott, C.P., Abel-Santos, E., Wall, M., Wahnon, D.C., Benkovic, S.J., (1999) *PNAS*, 96 p.13638.

127. Shih, C.K., Wagner, R., Feinstein, S., Kanik-Ennulat, C., Neff, N., (1988) *MCB*, 8 p.3094.

128. Southworth, M.W., Adam, E., Panne, D., Byer, R., Kautz, R., Perler, F.B., (1998) *EMBO J.*, 17 p.918.

129. Tait, J.F., Brown, D.S., Gibson, D.F., Blankenberg, F.G., Strauss, H.W., (2000) *Bioconjugate Chemistry*, **11** p.918.

130. Terskikh, A., Fradkov, A., Ermakova, G., *et al.*, (2000) *Science*, **290** p.1585.

131. Tesser, G.I., Balvert-Geers, I.C., **(1975)** *Int. J. Pept. Prot. Res.*, **7** p.295.

132. Thorn, K.S., Naber, N., Matuska, M., Vale, R.D., Cooke, R., (2000) *Protein Sci.*, **9** p.213.

133. Thorner, J., Emr, S.D., Abelson, J.N., eds. *Methods in Enzymology*. Vol. 328. 2000, Academic Press.

134. Tropsha, A., Hermans, J., **(1992)** *Prot. Eng.*, **5** p.29.

135. Troy, C.M., Derossi, D., Prochiantz, A., Greene, L.A., Shelanski, M.L., **(1996)** *J. Neuroscien.*, **16** p.253.

136. Troy, C.M., Stefanis, L., Prochiantz, A., Greene, L.A., Shelanski, M.L., **(1996)** *PNAS*, **93** p.5635.

137. Tsien, R.Y., **(1998)** *Annual Review in Biochemistry*, **67** p.509.

138. Tyagi, M., Rusnati, M., Presta, M., Giacca, M., (2001) *J. Biol. Chem.*, **276** p.3254.

139. Valiyaveetil, F.I., MacKinnon, R., Muir, T.W., (2002) JACS, **124** p.9113.

140. Villain, M., Vizzavona, J., Rose, K., (2001) *Chem Biol*, **8** p.673.

141. Vives, E., Brodin, P., Lebleu, B., (1997) *J. Biol. Chem.*, 272 p.16010.

142. Vocero-Akbani, A.M., Vander Heyden, N., Lissy, N.A., Ratner, L., Dowdy, S.F., (1999) *Nature Med.*, 5 p.29.

143. Wadia, J.S., Dowdy, S.F., (2002) *Curr. Opin. Biotech.*, 13 p.52.

144. Waibel, R., Alberto, R., Willuda, J., *et al.*, (1999) *Nat. Biotechnol.*, 17 p.897.

145. Wang, L., Brock, A., Herberich, B., Schultz, P.G., (2001) *Science*, 292 p.498.

146. Wang, L., Brock, A., Schultz, P.G., (2002) *JACS*, 124 p.1836.

147. Weissleder, R., (2001) *Nature Biotech.*, 19 p.316.

148. Wender, P.A., Mitchell, D.J., Pattabiraman, K., Pelkey, E.T., Steinman, L., Rothbard, J.B., (2000) *PNAS*, 97 p.13003.

149. Wiesmann, C., Fuh, G., Christinger, H.W., Eigenbrot, C., Wells, J.A., de Vos, A.M., (1997) *Cell*, 91 p.695.

150. Wills, K.N., Atencio, I.A., Avanzini, J.B., *et al.*, (2001) *J. Virol.*, 75 p.8733.

151. Wu, H., Hu, Z., Liu, X.-Q., (1998) *PNAS*, 95 p.9226.

152. Wu, H., Xu, M.-Q., Liu, X.-Q., (1998) *Biochim. Biophys. Acta.*, 1387 p.422.

153. Xu, M.-Q., Comb, D.G., Paulus, H., Noren, C.J., Shao, Y., Perler, F.B., (1994) *EMBO J.*, **13** p.5517.

154. Xu, M.Q., Southworth, M.W., Mersha, F.B., Hornstra, L.J., Perler, F.B., (1993) *Cell*, **75** p.1371.

155. Xu, R., Ayers, B., Cowburn, D., Muir, T.W., (1999) *PNAS*, **96** p.388.

156. Yamashiro, D., Li, C.H., (1988) *Int. J. Pept. Protein. Res.*, **31** p.322.

157. Yamazaki, T., Otomo, T., Oda, N., Kyogoku, Y., Uegaki, K., Ito, N., Ishino, Y., Nakamura, H., (1998) JACS, **120** p.5591.

158. Zhang, J., Wang, X., Lu, G., Tang, Z., (2001) *Applied Radiation And Isotopes*, **54** p.745.

159. Zhang, Z., Wang, L., Brock, A., Schultz, P.G., (2002) *Angew. Chem. Int. Ed.*, **41** p.2840.