technische universität
dortmund

# Modification of Cell Surface Proteins by Protein *Trans*-Splicing using the *Npu* DnaE Intein

Tulika Dhar

2011

# Modification of Cell Surface Proteins by Protein *Trans*-Splicing using the *Npu* DnaE Intein

Dissertation

zur

Erlangung des Doktorgrades

der Naturwissenschaften

(Dr. rer. nat.)

der Fakultät Chemie

der Technischen Universität Dortmund

vorgelegt von

Tulika Dhar

aus West Bengalen, Indien

Dortmund, 2011

This work was supervised by Professor Dr. Henning D. Mootz from September 2007 to August 2011 in the Fakultät Chemie - Chemische Biologie at the Technischen Universität Dortmund.

Erstgutachter: Prof. Dr. Henning D. Mootz
Zweitgutachter: Prof. Dr. Martin Engelhard
Tag der Abgabe:
Tag der Disputation:

*....dedicated to my loving family, fiancé and friends*

## ACKNOWLEDGEMENT

My Ph.D work is the result of the support and encouragement of several people. Without the help of these people the successful completion of this thesis was not possible. I hereby take this opportunity to personally thank all those wonderful people who have contributed to the fulfillment of this work.

First and foremost I would like to thank my supervisor Prof. Henning D. Mootz for his constant guidance, support, encouragement and valuable ideas and for giving me the great opportunity of being a part of his research group. Even in the midst of a busy schedule he always had time to discuss science. His assistance both in scientific and non-scientific matters has been invaluable.

Further I would like to thank Prof. Martin Engelhard and Dr. Leif Dehmelt for being my mentors and for their kind assistance.

I am heartily grateful to all the lab members of the AK Mootz group (both present and past) in Dortmund and in Muenster who provided a lively working atmosphere. I must specially thank Dr. Thomas Kurpiers, Dr. Christina Ludwig, Dr. Joachim Zettler, Dr. Tim Sonntag and Dr. Steffen Brenzel from whom I learnt a lot and who patiently answered all my queries. The work performed by Vivien Schütz was important for this project and I am truly grateful to her for her cooperation during the entire duration of this project. Here I must also thank all the members of the Dehmelt group (especially Silke and Julia) whose company, help and good humor were vital to my work in the cell culture laboratory.

I am profoundly grateful to Dr. Leif Dehmelt for assistance with miscroscopy work and for his valuable discussions. Dr. Sven Müller has provided constant support for all the confocal microscopy measurements. I am grateful to Jian Hou for helping me with the Olympus microscope. Assistance with the FCCS measurements was provided by Dr. Eli Zamir and Dr. Pete Lommerse. I am also thankful to Prof. Philippe Bastiaens and to the MPI Dortmund for allowing me the use of their equipment.

For the GPI-anchor project I am thankful to Dr. Miria Schumacher, Dr. Ralph Seidel and Prof. Martin Engelhard for their helpful disscussions.

I acknowledge here the immense help provided by Dr. Tim Sonntag both during my Ph.D and for his corrections and critical comments on this thesis. I am also grateful to Alexandra Wasmuth, Ilka Thiel and Julia Arens for helping with german translations.

I must thank here also all my friends (Anchal, Shyamal, Shobhna, Nachiket, Pipola, Malay, Saonli and Anupam) for providing a friendly living environment in Germany and for their constant help and support. I also thank Anchal here for assisting me on the Leica confocal microscope and for her friendship and company during the writing of this thesis.

My family (parents Samir Kumar Dhar and Kum Kum Dhar and sister Tuhina Dhar) has provided me with love, care and encouragement during the entire course of my Ph.D work for which I will always be grateful to them. I am especially indebted to my father for his immense belief in me and for his guidance throughout my life. Here I also show deep appreciation for my uncle Subir Dhar and aunt Bairnhildt Möller Dhar for treating me as a part of their own family and making my stay in Germany a truly memorable one.

A most heartfelt thanks and a deep sense of gratitude is felt for my fiancé, Dr. Souvik Rakshit without whose understanding, love, help, encouragement and timely support this work was not possible.

I am also thankful to all the members of the IMPRS team, especially Waltraud Goody and Christa Hornemann for their continuous assistance. In the end I would like to thank the IMPRS Dortmund and the DFG for providing financial support.

## SUMMARY

Cell surface proteins are of ongoing interest in several areas of research and development. This can be attributed to the fact that in addition to the dynamic behavior of the membrane which affects the nature and properties of these proteins, cell surface proteins are involved in important cellular activities such as signal transduction, immune response, cell adhesion and communication. To elucidate their structure and function several cell surface protein labeling strategies have been developed in recent years. These strategies in conjunction with fluorescence microcopy techniques provide a powerful tool for studying membrane proteins.

Split inteins have established themselves in the field of biochemistry and molecular biology as influential workhorses for the manipulation of proteins. This is evident from the numerous biotechnological applications where they have been utilized. Split inteins (both artificially created in the laboratory as well as naturally occurring) are expressed as separate fusion proteins and undergo protein splicing in *trans*. This autocatalytic protein *trans*-splicing (PTS) event requires the intein halves to associate with each other, resulting in the formation of an active complex. Consequent removal of the split intein halves joins the flanking extein sequences with a native peptide bond.Naturally split inteins confer certain advantages in terms of solubility and folding to the intein fusion proteins as they have evolved to splice two protein halves.

In this work, the ability of the naturally split DnaE intein from *Nostoc punctiforme* for labeling cell surface proteins has been demonstrated. The superior splicing properties of this intein make it an ideal candidate for this purpose. Using confocal microscopy and western blots it could be shown that intein halves fused to non-native exteins can be expressed on the extracellular surface of mammalian cells and efficiently labeled. The PTS reaction could be carried out on the surface of live cells at the physiological temperature of 37 °C in serum free cell culture media within minutes. The reaction was independent of the presence of a reducing agent and the choice of membrane anchor (transmembrane domain or GPI). It was however affected by the intein half being expressed on the cell surface as well as the length of the linker region between the intein half and the transmembrane anchor. The extein sequences were also shown to play an important role in the splicing reaction on the cell surface. In conclusion, the conducted studies establish this strategy as a fast and efficient means for labeling cell surface proteins. It is expected to be beneficial for the manipulation of membrane proteins as well as preparation of GPI-anchored proteins.

In addition, an attempt has been made in this work for labeling cell surface proteins using the PCP-domain and performing PTS on the surface of quantum dots for labeling any POI.

# ZUSAMMENFASSUNG

Oberflächenproteine stellen ein Forschungs- und Entwicklungsgebiet von anhaltendem Interesse dar. Dies lässt sich auf die Tatsache zurückgeführten, dass zusätzlich zu dem dynamischen Verhalten der Plasmamembran, die das Verhalten und die Eigenschaften von Proteinen beeinflusst, Zelloberflächenproteine in wichtige zelluläre Vorgänge involviert sind. Zu diesen zählen u.a. die Signaltransduktion, Immunantwort, Zelladhäsion und Zellkommunikation. Zur Aufklärung von Stuktur und Funktion von Oberflächenproteinen wurden in jüngster Zeit einige Markierungsstrategien entwickelt. In Verbindung mit der Technik der Fluoreszenzmikroskopie stellen diese Strategien ein leistungsstarkes Werkzeug zur Erforschung von Membranproteinen dar.

Gespaltene Inteine haben sich in den Gebieten der Biochemie und Molekularbiologie als effizient Werkzueg zur Manipulation von Proteinen etabliert, was anhand ihrer vielfältigen biotechnologischen Anwendungen ersichtlich wird. In diesen Fällen werden entweder die artifiziell im Labor erzeugten oder natürlich vorkommenden gespaltenen Inteine als separate Fusionsgene exprimiert und anschließend die Proteinspleißreaktion *in trans* durchgeführt. Dieser autokatalytische Vorgang des Protein-*trans*-Spleißens (PTS) erfordert zunächst die Assoziation der Inteinhälften und die Bildung des aktiven Komplexes. Anschließend werden während der Entfernung der Inteinhälften die flankierenden Exteinsequenzen in einer nativen Peptidbindung miteinander verknüpft. In Bezug auf Löslichkeit und Faltung der Inteinfusionsproteine besitzen natürlich gespaltene Inteine einige Vorteile, da sie sich zur PTS-Reaktion von gespaltenen Proteinen evolviert haben.

In dieser Arbeit wurde des natürlich gespaltene DnaE-Intein aus *Nostoc punctiforme* erfolgreich zur Markierung von Oberflächenproteinen verwendet. Die ausgezeichneten Spleißeigenschaften dieses Inteins machen es zu einem idealen Kandidaten für diesen Zweck. Unter Verwendung von Konfokalmikroskopie und Western Blot Analyse konnte gezeigt werden, dass die mit nicht-nativen Exteinen fusionierten Inteinhälften exprimierten und auch effizient auf der extrazellulären Oberfläche von Säugerzellen markiert werden konnten. Bei physiologischer Temperatur von 37 °C und in serumfreiem Zellkulturmedium lief die PTS-Reaktion auf der Öberfläche von lebenden Zellen innerhalb von Minuten ab. Dabei war die Reaktion weder von der Anwesenheit reduzierender Agenzien noch von der Wahl des Membranankers (Transmembrandomäne oder GPI) abhängig. Allerdings wurden sie einerseits von der Art der auf der Zelloberfläche exprimierten Inteinhälfte beeinflusst und andererseits von der Länge der Linkerregion zwischen Inteinhälfte und Transmembrandomänenanker. Des Weiteren nahmen die verwendeten Exteinsequenzen eine besondere Rolle in der Spleißreaktion auf der Zelloberfläche ein. Aufgrund der durchgeführten Studien konnten die entwickelte Strategie als eine schnelle und effiziente Methode zur Markierung von Oberflächenproteinen etabliert werden. Es ist davon auszugehen, dass diese bei der Manipulation von Membranproteinen und zur Herstellung von GPI-verankerten Proteinen vorteilhaft sein wird.

In einem weitern Teil der Arbeit wurde versucht Zelloberflächenproteine mit Hilfe der PCP-Domäne zu markieren. Weiterhin erfolgten Experimente zum Protein *trans*-Spleißen auf der Oberfläche von Quantum dots um beliebige Proteine von Interesse zu markieren.

# PUBLICATIONS

**Tulika Dhar and Henning D. Mootz (2011)**

Modification of transmembrane and GPI-anchored proteins on living cells by efficient protein trans-splicing using the *Npu* DnaE intein

*Chemical Communications (Cambridge)*, **47**, 3063-3065

**Tulika Dhar, Thomas Kurpiers and Henning D. Mootz (2011)**

Extending the Scope of Site-Specific Cysteine Bioconjugation by Appending a Prelabeled Cysteine Tag to Proteins Using Protein Trans-Splicing.

*Methods in Molecular Biology*, **751**, 131-142

# 1 INTRODUCTION

Integral membrane proteins carry out several vital functions across cell membranes such as signal transduction, material transport and intercellular communication.[1] Investigations of such membrane proteins in living cell membranes is important because biomembranes are heterogeneous and dynamic which affects their behavior. Visualization of membrane proteins in living cell membranes is essential for observing the dynamic behavior of proteins such as endocytosis and also allows detection of conformational changes and monitoring protein-protein interactions by FRET which detects changes in distance and/or orientation of different fluorophores having a spectral overlap. Site-specific labeling of proteins is a key technique for the detection of a target protein by fluorescence microscopy. Surfaces of cultured cells are freely accessible to chemical treatment and therefore labeling cell surface proteins with fluorescent molecules appears as an attractive strategy to equip them with probes that allow for their functional characterization.[2]

In this work, a **split intein** mediated approach (**protein *trans*-splicing/PTS**) for **site-specific labeling** of **cell surface proteins** using the ***Npu* DnaE** intein is presented. Two genes are known to code for the α-subunit of the DNA polymerase III (DnaE) protein in *Nostoc punctiforme*. One gene codes for the N-terminal half of the protein ($DnaE^N$/$Ext^N$) which is succeeded by the N-terminal half of an in-frame intein sequence ($Int^N$). The second gene codes for the C-terminal half of the intein ($Int^C$) followed by the C-terminal half of the protein ($DnaE^C$/$Ext^C$). The "split" intein sequence is not a part of the functional DnaE protein and is removed by a post-translational autocatalytic mechanism inherent to the intein sequence which ligates the two halves of the DnaE protein with a native peptide bond. Iwai and coworkers demonstrated that the intein retained its catalytic character when the DnaE sequences were replaced with non-native extein sequences.[3]

Theoretically, if a transmembrane protein is fused to one half of the split intein with the intein part facing the extracellular environment and is expressed on the surface of mammalian cells, then the exogenous addition of a recombinantly expressed and purified protein containing a desired modification fused to the other split intein half will lead to the intein fragments recognizing and associating with each other. Autocatalytic removal of the intein sequences will then result in a modified/labeled cell surface protein (**Figure 1**).[4]

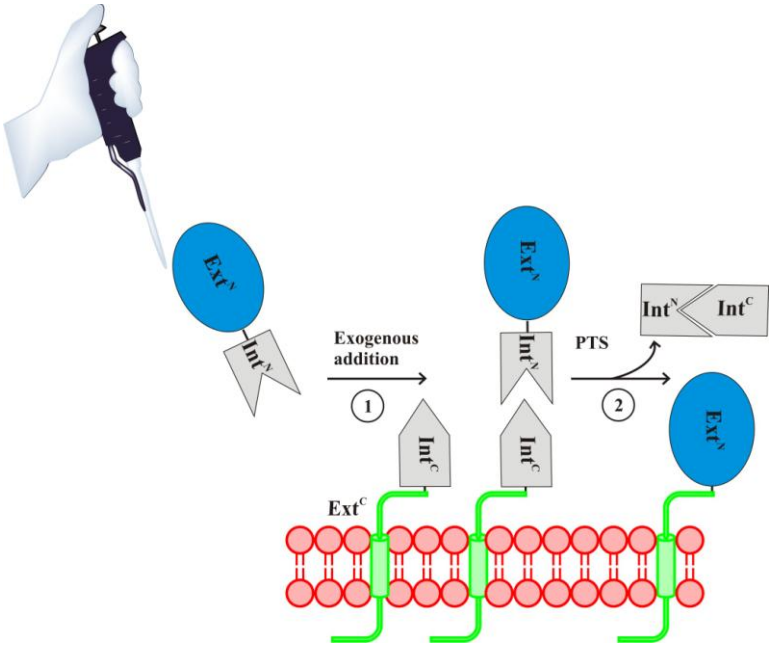



**Figure 1 Schematic representation of cell surface protein labeling using the split-intein mediated approach.** The $Int^C$ is fused to the N-terminus of a membrane protein ($Ext^C$) and expressed on the surface of mammalian cells. The $Int^N$ is fused to the C-terminus of the desired protein of interest or a peptide sequence ($Ext^N$) and is recombinantly expressed in *E.coli*. In the first step, the purified $Int^N$ fusion protein is added to the surface of cells expressing the $Int^C$ fusion protein. In the second step, the intein halves are expected to recognize each other and associate non-covalently, form an active intein conformation and catalyze a protein *trans*-splicing reaction. $Ext^N$= N-Extein sequence, $Ext^C$= C-extein sequence, $Int^N$= Intein N-terminal half, $Int^C$= Intein C-terminal half, PTS = Protein *trans*-splicing

Our view of the simple 'central dogma' of biology (**Figure 2**) is constantly challenged by rapidly evolving evidence for editing at the protein level including proteolytic cleavage of polyproteins, protein autoprocessing, non-ribosomal addition of moieties and proteasome mediated peptide ligation.[5] A recent member to join this bandwagon of post translational modifications is 'PROTEIN SPLICING'.

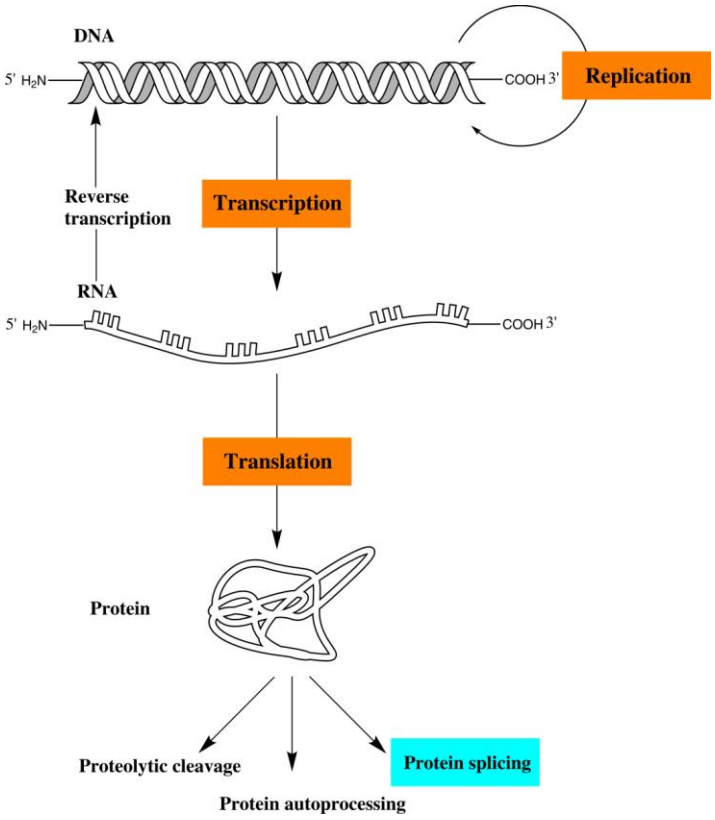



**Figure 2 The central dogma of biology.** The central dogma of biology deals with the sequential transfer of information from DNA to RNA to proteins. It states that information cannot be transferred back from proteins to nucleic acids.

## 1.1 Inteins and Protein Splicing

In the early 1990's, it was discovered that the VMA1 gene from *Saccharomyces cervevisiae* encodes a 120 kDa translational product from which a 50 kDa endonuclease domain is excised out, producing the final mature product of 70 kDa.[6] This deleted 50 kDa sequence present as an in-frame insert in the gene was later termed **INTEIN** (INTernal protEIN).

Inteins are protein sequences present in the host protein sequence and must be removed in order to form a functional protein. In contrast to analogous introns, inteins are transcribed as well as translated together with their host protein.[7] The two parts of the host protein separated by the disrupting intein sequence are called exteins (EXTernal protEIN).[8] During post-translational processing this intein sequence is excised out by a process called '**protein splicing**' concomitantly joining its flanking extein sequences with a native peptide bond.[9] Two stable proteins, the mature protein and the intein are the products of the protein splicing process.

Since the discovery of the VMA1 intein, inteins have been identified in all three domains of life: eucaryotes, archaebacteria and eubacteria as well as in phages and viruses and range from 134-608 amino acids. They have also been found in proteins with diverse functions, including metabolic enzymes, proteases, DNA and RNA polymerases, helicases, gyrases and vacuolar type ATPase.[10] The intein regristry InBase at http:// www.neb.com/inteins.html has listed more than 450 inteins along with their sequences, conserved motifs, host organisms and host proteins.[11] Intein names include a genus and species designation, abbreviated with three letters, and a host gene designation according to the currently accepted nomenclature. For example, the *S.cerevisiae* VMA1 intein is known as *Sce* VMA1. Most known genes encode only one intein and inteins found at the same insertion site in homologous proteins in different organisms are called intein alleles.[12] In very rare cases, such as the ribonucleotide reductase gene of the oceanic $N_2$-fixing cyanobacterium *Trichodesmium erythraeumsome*, four inteins are encoded.[13]

Inteins are considered as parasitic selfish genetic elements as they contribute no known advantage to their host. However, due to their apparent negligible disruption of their host gene or the protein product formed, they are not weeded out. In addition, endonucleases integrated into some inteins results in transfer of the inteins into intein-less alleles. Endonucleases prefer inteins as integration sites because in these locations the functions encoded by the surrounding DNA would not be disrupted.[14] Also, since inteins are inserted in highly conserved points in genes coding for essential proteins, their removal can be detrimental.[15]

### 1.1.1 Intein Classification

Inteins are classified into four groups, large (maxi), minimal (mini), *trans*-splicing and Ala inteins[16]:

Large or maxi inteins (**Figure 3**) consist of two domains: a self-splicing domain and an endonuclease domain. They were the first to be discovered. The homing endonucleases are site-specific, double strand DNA endonucleases that promote intein spread between genomes in a recombination dependent process known as ‘homing’. Usually, homing endonuclease are encoded by an open reading frame within an intein.[17] The N- and C-terminal regions of the large inteins contain the elements necessary for splicing. The *Sce* VMA1 intein belongs to this class.

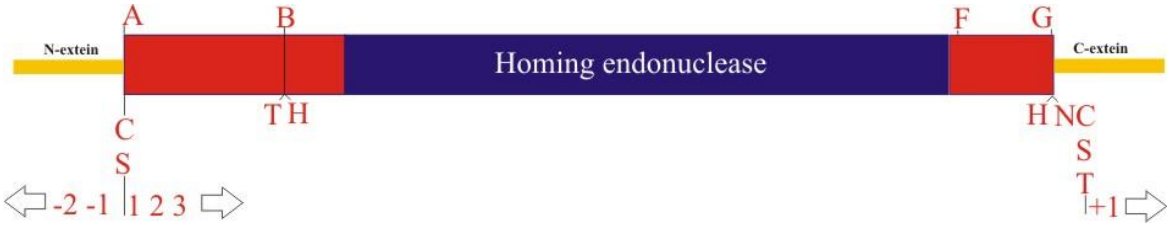



**Figure 3 Schematic representation of a maxi-intein along with conserved intein motifs and numbering.** Blocks A and B are present in the N-terminal splicing domain, while F and G are in the C-terminal splicing domain. The endonuclease domain contains the blocks C, D, E and H. The first amino acid residue upstream of the intein is numbered -1 with numbering continuing upstream. The intein is numbered sequentially beginning with the N-terminal amino acid residue. The first C-extein residue is the +1 position with numbering proceeding toward the C-terminus.

Mini-inteins (**Figure 4**) have the typical N- and C-terminal splicing domains; however, the endonuclease domain is not present. The earliest mini-inteins were created by deleting the endonuclease domain at the gene level from the *Sce* VMA or the *Mycobacterium tuberculosis* RecA inteins.[18] Splicing in the absence of this endonuclease domain demonstrated the sufficiency of the protein splicing domains for activity. Shortly thereafter, the naturally occurring mini-intein was reported in the *Mycobacterium xenopi* gyrA gene.[19] The smallest known mini-intein *Mth* RIR1 intein is only 134 amino acids long and contains all the information necessary for splicing.[20]

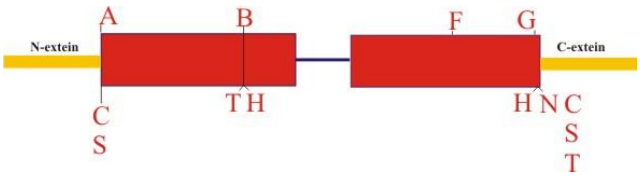



**Figure 4 Schematic representation of a mini-intein.** The mini inteins lack an endonuclease domain but the remaining conserved residues are the same as a maxi intein. The endonuclease region has been replaced by a short linker region.

*Trans*-splicing inteins or split inteins (**Figure 5**) lack a covalent linkage between their N- and C-terminal splicing domains. *Trans*-splicing was demonstrated using artificially split inteins before a naturally split intein was discovered.[21] The naturally split *Synechocystis sp* DnaE intein is characteristic of this class of inteins.

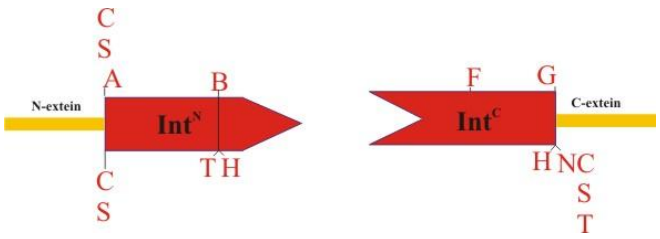



**Figure 5 Schematic representation of a split intein.** These inteins lack a covalent linkage between their N- and C-terminal splicing domains.

Alanine inteins (**Figure 6**) contain a naturally occurring N-terminal alanine residue instead of the conserved cysteine or serine reported for the earlier inteins. The *Methanococcus jannaschii* KlbA intein belongs to this class of inteins.

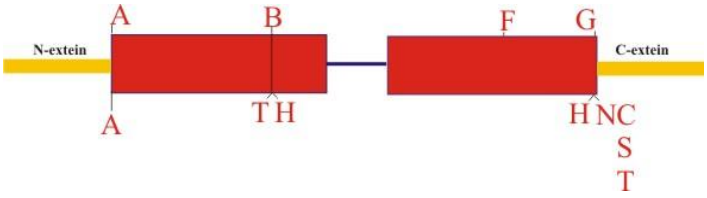



**Figure 6 Schematic representation of an Ala intein.** These inteins begin with an Ala instead of the conserved C, S or T. They belong to the Class II inteins and splicing proceeds via an alternative splicing mechanism where the first C-extein amino acid directly attacks the amide bond at the N-terminal splice junction.

### 1.1.2 Conserved Motifs

Intein sequence analysis has revealed conserved motifs, blocks A, B, F, G, N2 and N4 which are found in all inteins. Groups of related amino acids populate these conserved motifs.[22] Blocks A and B are present at the intein N-terminus, blocks F and G are at the C-terminus and blocks N2 and N4 are close to the N-terminal.[11-12, 22b, 23] Maxi inteins with an endonuclease domain have another four conserved motifs (blocks C, D, E, and H) (**Figure 3**).

Known inteins share a low degree of sequence similarity with most inteins beginning with Ser or Cys and ending in His-Asn. However, several putative inteins are known to begin with Ala (see Alanine inteins; section 1.1.1), Gln and Pro or end in Asp and Gln.[24] The first amino acid of the C-extein is an invariant Ser, Thr or Cys. Block B contains a highly conserved His

and a less conserved Thr residue. However, residues proximal to the intein-splicing junction at both the N- and C- terminal exteins were recently found to influence protein splicing.[25]

Protein splicing has so far not been observed in multicellular organisms. However, the hedgehog autoprocessing domain in the hedgehog family of embryonic signaling protein from *Drosophila melanogaster* and the splicing domain of inteins have a common structural fold (the HINT module) which share two motifs (blocks A and B) and perform similar chemical reactions.[26] The HINT (hedgehog intein) module is composed largely of β-strands and forms a horseshoe-shaped structure with two large β-strands forming the horseshoe core. The N- and C-terminal splice junction residues reside in the central cleft of this horseshoe-like structure in the form of antiparallel β-sheet. The splice junction residues are tightly associated with numerous hydrogen bonding contacts and hydrophobic packing interactions.

### 1.1.3 Mechanism of Protein Splicing

The widely accepted mechanism for protein splicing pathway is known to proceed via 4 intra-molecular steps[14, 27] outlined in **Figure 7**

I. N→X acyl shift (X is either S if the first intein residue is Cys or O if the residue is Ser)

Protein splicing is initiated by the side chain attack of the first intein residue on the preceding carbonyl (C=O) group, resulting in a shift of the N-extein to the side chain of the first intein residue and forming a thioester bond.

II. Transesterification

The hydroxyl or thiol group of the first C-extein residue (Ser, Thr, Cys) attacks the thioester linkage resulting in the transfer of the N-extein to the side chain of the +1 residue forming a branched thioester intermediate.

III. Asparagine Cyclization

Cyclization of the C-terminal Asn residue of the intein results in the removal of the intein due to the cleavage of the amide linkage yielding exteins linked by a thioester bond.

IV. S→N or O→N acyl rearrangement

A spontaneous acyl shift results in a stable peptide bond being formed between the exteins.

Most known inteins fall into the category of **Class I** inteins which carry out the standard protein splicing steps mentioned above. The importance of the splice junction residues mentioned in section 1.1.2 was clearly explained by the splicing mechanism outlined above. However, it did not account for the penultimate His and the block B Thr and His residues. It was later demonstrated by mutation and structural studies that the block B His assists in activating the N-terminal splice junction.[27b, 28] The less conserved Thr was also shown to assist in N-terminal reactions. The penultimate His (His-Asn) activates the C-terminal splice junction for Asn cyclization. Mutating this residue in various inteins led to either no effect or total inhibition of the splicing reaction.

In addition, an alternative splicing process is carried out by the alanine inteins (see section 1.1.1) also known as the **Class II** inteins. In this case, the first residue of the C-extein (Cys) directly attacks the amide bond at the N-terminal splice junction leading to the formation of the branched intermediate thereby bypassing the first acyl shift initiation step.[29]

In the **Class III** intein of the mycobacteriophage *Bethlehem* DnaB protein, a Cys residue present in the conserved block F attacks the peptide bond at the N-terminal splice junction forming a branched intermediate with a labile thioester linkage. In the next transesterification step, the N-extein is transferred to the +1 residue of the C-extein, a Thr, resulting in the branched intermediate as in **Class I** inteins.[30]

Inteins with the conserved C-terminal Asn replaced by Asp or Gln have also been discovered as mentioned in section 1.1.2.[31] These residues are capable of similar cyclization reactions as Asn, although other mechanisms have been proposed.

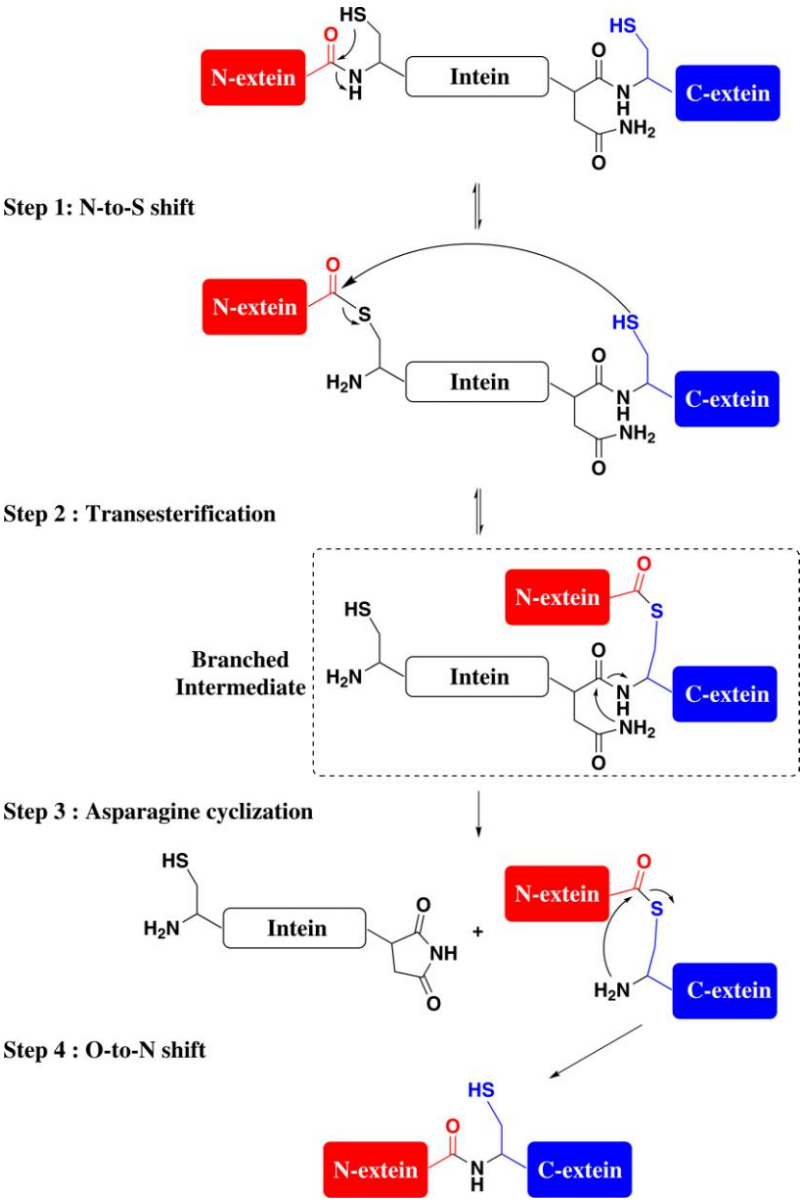



**Figure 7 Protein splicing mechanism.** The side chain sulfhydryl or hydroxyl of the intein N-terminal residue initiates an N-S or N-O acyl shift. This forms a (thio) ester linkage between the N-extein and the intein. The downstream nucleophile, side chain of Cys, Ser or Thr attacks the N-terminal (thio) ester in a trans (thio) esterification reaction to generate a branched intermediate. Cyclization of the intein C-terminal Asn coupled to peptide bond breakage releases the intein. The (thio) ester bond between the N-and C-exteins is converted to a native peptide bond by a spontaneous S-N or O-N acyl rearrangement.

### 1.1.4 Side Reactions

An excised intein and a ligated extein are the products expected after the completion of the above mentioned splicing process. However, products from side reactions are also obtained in

addition. N-terminal cleavage and C-terminal cleavage reactions only occur when inteins are inserted into foreign contexts or made to function under non-physiological conditions, for example, expressing splicing precursors in a heterologous host. These products arise from cleavage at either one or both splice junctions. Hydrolysis of the linear thioester intermediate or the branched intermediate prior to Asn cyclization results in cleavage at the N-terminal splice junction. On the other hand, cleavage at the C-terminal splice junction arises due to the Asn cyclization without the N-extein transesterification step. Mutating catalytic residues at one junction generally increases the cleavage reactions at the other junctions.[32] For example, mutation of the Asn residue at the intein C-terminus abolishes steps 3 and 4 of the splicing reaction resulting in N-terminal cleavage but step 1 can still occur. Spontaneous hydrolysis of the linear thioester bond separates the N-extein from the intein/C-extein part. Replacing the conserved first N-terminal residue of the intein abolishes steps 1, 2 and 4 of the splicing reaction leading to C-terminal cleavage. In such a case, step 3 still occurs separating the C-extein from the N-extein/intein portion.

### 1.1.5 Effect of the Proximal Extein Amino Acids

Inteins in their natural context have co-evolved with their extein sequences for maximum splicing efficiency. Therefore, placing them in heterologous insertion sites leads to a ~10-fold lower splicing efficiency and/or yields inactive precursor or cleavage byproducts.[27b, 28a, b, 33] Also, different inteins show different substrate specificity. Amino acids that are most similar to the native residue (for example amino acids of comparable size, shape and polarity) work best under these conditions. It has also been observed that splicing with inteins in a non-native context becomes temperature dependent suggesting that folding of the intein is impaired.[18b, 19, 34] All three nucleophiles involved in the protein splicing pathway are splice-junction residues; therefore, suboptimal proximal extein sequences may lead to destabilization of the active site and finally inhibit splicing.

Data from the past few years have demonstrated the effect of the adjacent N-extein (N-1) and two C-extein residues (C+1, C+2) on splicing activity.[3, 32a, 32c, 35] Amitai *et al*. recently reported that a bias also exists for the N-2, C+2 and C+3 extein sequences.[25] They identified a number of N-1 residues that preferentially enhance or decrease the rate of N-terminal cleavage. The C+3 amino acid was shown to enhance N-terminal cleavage activity. It is therefore generally accepted that splicing is improved when 1-5 native extein sequences flank

the intein on either or both sides. Also, the further away the amino acid is from the intein the less likely it will have a significant impact on the splicing reaction.

### 1.1.6 Effect of Divalent Cation

The $Zn^{2+}$ ion has been shown to affect splicing. It has been reported that splicing is inhibited when $Zn^{2+}$ was added to the reaction mixture of the two intein halves for the *Mtu* RecA and the *Ssp* DnaE intein.[36] $Zn^{2+}$ was also shown to be present in the active site of the *Sce* VMA1 intein. Although not involved in large structural perturbations, it was in contact with the penultimate His and the +1 Cys residues. However, till date it is still speculative whether the $Zn^{2+}$ binds to and interferes with a folded intein. $Zn^{2+}$ ions interference with the activity of an intein *in vivo* also remains to be determined. $Cu^{2+}$[37] and $Cd^{2+}$[36a] are also known inhibitors of protein splicing activity. It was therefore hypothesized that metal ion co-ordinations have negligible influence on protein structure. Mobility restriction of key residues from metal coordination is likely the key cause of metal inhibition of intein splicing.[38]

## 1.2 Split Inteins and Protein *Trans*-Splicing

Split inteins/*Trans*-splicing inteins (section 1.1.1) can be artificially engineered from contiguous inteins by dividing the intein into two parts and further reassembled to yield an active protein splicing element. This phenomenon was demonstrated both *in vivo* and *in vitro* in the 1900's by several independent groups.[18a, 21] These so called split inteins self-associate and catalyze protein splicing activity in *trans* termed 'Protein *trans*-splicing (PTS)'.

PTS proceeds similarly as protein splicing; however, the first step involved is association of the intein fragments and their folding into an active conformation (**Figure 8**). A protein of interest (POI) is reconstituted from two fragments and the two polypeptide chains get linked by a native peptide bond. The intein fragments are excised making the linkage virtually traceless.

Inteins can be split both within the splicing domain as well as within the endonuclease domain. Splicing was found to be more efficient using a split mini-intein with the split at the endonuclease insertion position. Recently, several other split positions have also been

identified.[39] However, fusion proteins with artificially split inteins show reduced solubility and are only soluble if the complementary pairs are co-expressed or if fused to a solubility enhancing domain. Therefore, this often necessitates an undesirable denaturation/renaturation step prior to splicing. This suggests that artificially split inteins often have exposed hydrophobic domains which lead to protein aggregation due to misfolding in the absence of the complementary fragment. The tendency to aggregate may contribute to reduced splicing efficiency.[21a, b, 40] The *Ssp* DnaB[41] and the *Mxe* GyrA[42] artificially split inteins have been reported to retain solubility and splicing activity under native conditions using purified proteins.

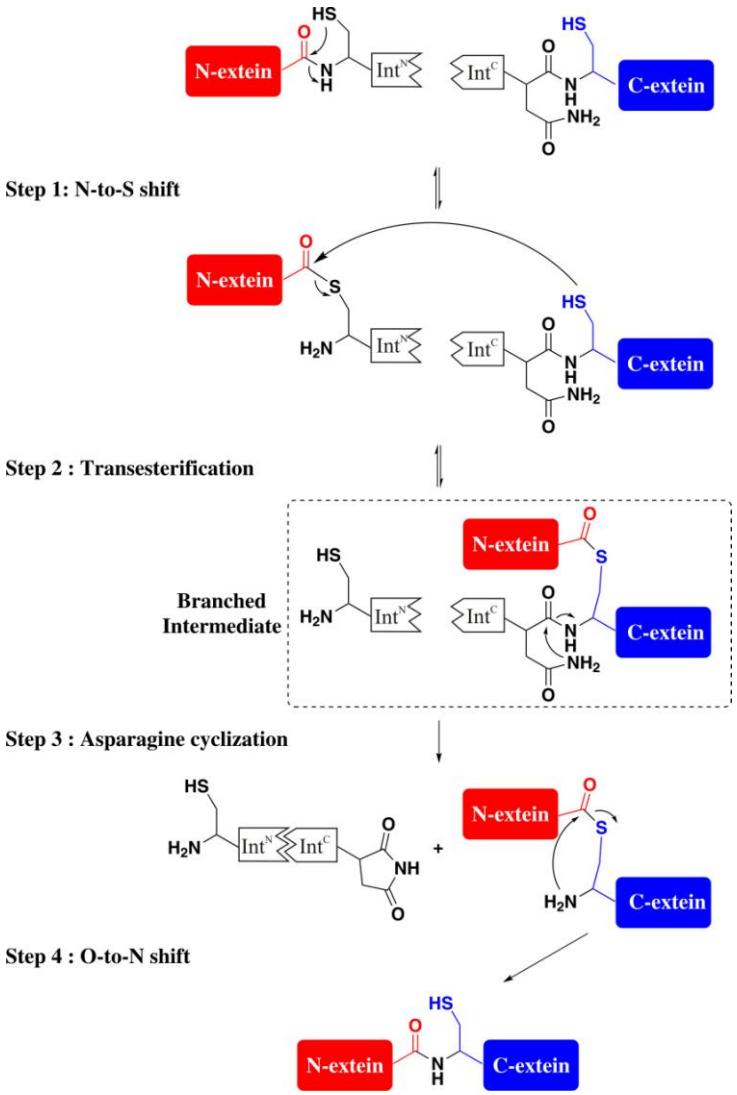



**Figure 8 Protein *trans*-splicing mechanism.** Both artificially and naturally split inteins undergo proteins *trans*-splicing via the mechanism outlined above. Intein association is the first step in PTS and the following steps are the same as for contiguous inteins. For a comparison with the protein splicing mechanism of contiguous inteins see **Figure 7**.

Naturally occurring split inteins lack an endonuclease domain and the split site corresponds to the location of this missing domain. This region often coincides with a loop region within the intein structure.[43] They are believed to be descendants of a continuous mini-intein which later lost its sequence continuity due to one or more genomic rearrangements. Naturally split inteins are generally believed to exhibit better solubility due to higher affinity among intein fragments and because they are naturally expressed as separate fragments compared with artificially split inteins.[44] One of the best characterized naturally split inteins (DnaE intein) is found in the dnaE gene of a large group of cyanobacteria.

Split intein engineering of naturally split inteins have recently revealed that new split inteins can be created by introducing split sites at locations differing from the canonical split site of the intein. These split sites must however correspond to surface loops.[45] Similarly, new split sites have also been engineered for artificially split inteins. The *Ssp* DnaB intein split at four new positions was found to show protein splicing activity.[39] PTS was carried out with an N-intein fragment that is only 11 amino acids long.[46] Split *Ssp* GyrB intein with a C-intein fragment of 6 aa has also been reported.[47]

### 1.2.1 The DnaE Intein

Cyanobacteria are a large and diverse group of bacteria and it was recently reported that the split intein of the *dnaE* gene is quite common in several groups of these cyanobacteria.[48] The catalytic α subunit of the DNA polymerase III (DnaE protein) was found to be encoded by two genes in several cyanobacteria- *Synechocystis* species PCC 6803 (*Ssp*), *Synechococcus* species PCC 7002, *Nostoc* species PCC 7120, *Nostoc punctiforme* (*Npu*), *Thermosynechococcus elongates* (*Tel*) and *Trichodesmium erythaeum* (*Ter*). The dnaE gene coding sequences are split in the exactly same extremely conserved point into two genes. The N-terminal part of the dnaE gene is followed by a N-terminal part of the split intein. Similarly, the C-terminal half of the dnaE gene is preceded by a C-terminal half of the split intein. The DnaE split intein amino acid sequence is conserved along their entire sequence except for the 3' end of the N-terminal half which differ from each other in sequence and in length. The length of the $DnaE^N$ part is between 101 and 123 aa and the $DnaE^C$ is usually 35-36 aa is length. All the DnaE inteins studied so far have the six conserved sequence motifs that define the intein protein splicing fold and active site and none of them has an endonuclease domain. dnaE genes present in other cyanobacteria like *Prochlorococcus*

*marinus* are contiguous with no inteins and are flanked by the same genes. However, split dnaE genes are flanked by different genes. DnaE split intein amino acid sequences are more similar to each other than other inteins. Split intein genes are unlikely to be transferred by homing as both need to be copied at the same time and also because of the lack of the endonuclease domain.

Although the *Ssp* DnaE from *Synechocystis* species PCC 6803 (*Ssp*) is one of the better studied and broadly used split inteins, the present work employs the *Npu* DnaE intein from *Nostoc punctiforme*. In order to appreciate the superior properties of this intein and understand the reasons for its use in this work, a comparison with the *Ssp* DnaE intein is however necessary.

### 1.2.2 The *Ssp* DnaE and the *Npu* DnaE Split Inteins

The *Ssp* DnaE and the *Npu* DnaE inteins are members of a family of naturally split DnaE inteins all of which use the canonical "CFN" tripeptide splice junction but whose splicing activities differ from each other.[3, 49]

The DnaE protein of *Synechocystis* species PCC 6803 is coded by a split gene interrupted by a split intein. The N- and C-terminal halves of the catalytic subunit α of DNA pol III are encoded by two separate genes located ~750 kb apart in the genome and on opposite DNA strands.[43a] The dnaE-n product consists of a N-extein followed by a 123 aa intein N-terminal sequence ($Int^N$) whereas the dnaE-c product consists of a 36 aa intein C-terminal sequence ($Int^C$) followed by the C-extein sequence. Co-expression of the intein halves in *E.coli* produced an intact DNA pol III.[43a]

The intein halves have been shown to perform spontaneous *trans*-splicing both *in vitro* and *in vivo* in a foreign context.[49b, 50] It is known from literature that *trans*-splicing with this intein also generates cleavage products, together with stable complexes of noncovalently, tightly interacting intein intermediates.[51] Kinetic analysis suggests extreme rapidity of association and a slow dissociation. Protein *trans*-splicing is extremely rapid ($6.6 \pm 1.3 \times 10^{-5}$ $s^{-1}$) and occurs at a low nanomolar affinity.[44, 51-52] The crystal structure obtained with artificially joined split intein halves revealed a high structural similarity to the characteristic compact intein fold and its active site.[43d]

The *Npu* DnaE intein present in the α subunit of the DNA pol III gene of *Nostoc punctiforme* came into the limelight because of a recent discovery that this DnaE intein besides having the obvious advantages of naturally split inteins, is more robust in protein *trans*-splicing than the *Ssp* DnaE intein.[3, 49b]

The naturally split *Npu* DnaE intein is homologous to the *Ssp* DnaE with a sequence identity of 67 % (68/102) for the N-terminal intein and 53 % (19/36) for the C-terminal intein half. The C-terminal fragment of the *Ssp* DnaE contains 21 aa lacking in the *Npu* DnaE intein, which is not involved in the intein structural fold[43d] and is dispensable for the splicing activity. Furthermore, all known conserved amino acids essential for catalytic activity are identical in both cases. Therefore, one can only speculate upon the origins of the unique properties of this intein.

The *Npu* DnaE can perform spontaneous *trans*-splicing *in vivo* in a foreign context.[49b] In comparison to the *Ssp* DnaE amount of unprocessed precursor is 20 fold less with a splice product yield more than 10-fold higher in comparison to the *Ssp*. It is known to display by far the highest rate constant in the protein *trans*-splicing reaction (with a $t_{1/2}$ of about 1 min at 37 °C using purified proteins) while many split inteins including the *Ssp* DnaE intein exhibit reduced yields and increased formation of hydrolysis side products at this temperature.[49b] Also, in contrast to the *Ssp* $DnaE^N$, the *Npu* $DnaE^N$ can accommodate many substitutions at the second residue of the C-terminal extein (Phe+2) with no or modest reduction in the *trans*-splicing activity thereby increasing the applications of *trans*-splicing as it has often been limited by the inevitable insertion of the native extein sequence at the ligation junction.[3] The two split inteins were also found to have cross-splicing activity[3] but the rate of the reaction was compromised suggesting that only a native combination of the *Npu* intein fragments was capable of performing the superior *trans*-splicing activity reported.

Thus, it is evident that the *Npu* DnaE intein holds great promise and has potential for becoming an invaluable tool in various areas of biotechnology with drastically shortened reaction times and high yields being the major selling points. Its robustness under high urea conditions and ability to splice efficiently even at lower temperatures will see its applications where poorly soluble or temperature sensitive proteins are required. Moreover, an engineered split *Npu* DnaE intein with a shortened C-terminal fragment[43b] will allow labeling of proteins via protein semi-synthesis.

## 1.3 The PCP domain

Non-ribosomal peptide synthetases (NRPSs) are large modular enzymes which are responsible for the synthesis of a variety of microbial bioactive peptides.[53] An NRPS comprises of a set of distinct modules, with each module consisting of defined domains.[54] These modules each recognize and incorporate one specific amino acid into the peptide product. Each amino acid of the peptide product undergoes adenylation, thioester and peptide bond formation as well as substrate modifications catalyzed by these domains. An amino acid is activated by the adenylation domain (A) to form the aminoacyl adenylate.[55] The peptidyl carrier protein (PCP) is the thioester domain of these modules.[56] This domain is subjected to phosphopantetheinylation at a Ser residue within a conserved sequence motif of the PCP domain shown in **Figure 9** which converts the apo-PCP (inactive) to the holo (active) form. This reaction, catalyzed by phosphopantetheine transferases (PPTases) occurs by tethering the phosphopantetheinyl moiety of cosubstrate coenzyme A (CoA) in phosphodiester linkage to the hydroxymethyl side chain of the conserved serine residue in the PCP domain.[57] The PCP domain, under native conditions, tethers the activated amino acylated product to the csyteamine thiol group of 4'-PP. The activated amino acid fixed to the PCP can undergo further optional modifications.[56, 58] The downstream condensation domain (C) then catalyzes peptide bond formation between the amino acids bound to the 4'-PP cofactors of the PCP domains in two adjacent modules. This results in a peptide product tethered to the second PCP domain and release of the free –SH group of the first 4'-PP.[59] In this way the reaction is continued until all amino acids are incorporated into the peptide chain which is then released from the last module by hydrolysis or cyclisation.[60]

MDKMPLTPND KIDRKALPEP DLTANQSQAA YHPPRTETES ILVSIWQNVL
GIEKIGIRDN FYSLGGDSIQ AIQVAARLHA YQRKLDTKDL LNYPTIEQVA
LFVKKSTTRRS

**Figure 9 Ribbon model of the solution structure of the PCP domain.** Sequence of the PCP domainof the TycA gene from *B.brevis*. The 4'-PP binding site is underlined with the CoA attachment site (Ser) marked with a blue arrow. PDB ID: 1DNY

## 1.4 Fluorescence cross-correlation spectroscopy

For understanding the function and regulation of a protein it is crucial to understand the repertoire of interactions it undergoes with other molecules. Several biochemical and biophysical techniques have been developed for investigating such interactions and each of these approaches have their own strengths and weaknesses especially with regard to the sensitivity and specificity of the method.[61] Biochemical approaches such as co-immunoprecipitation (CoIP) and Yeast two-hybrid (Y2H) have been extensively applied for studying protein-protein interactions between various proteins.[61b, 62] These techniques have been further extended in real time in living cells with the aid of various forms of fluorescence microscopy. However, pairs of interacting proteins are too small to be resolved by conventional fluorescence microscopy. Fluorescence resonance energy transfer (FRET)[63] and fluorescence cross correlation spectroscopy (FCCS)[64] have overcome this barrier and it is therefore now possible to investigate protein-protein interactions using fluorescence microscopy.

FCCS, an extension of the popular fluorescence correlation spectroscopy (FCS) technique[65] is a method that measures the temporal fluctuations emanating from two differently labeled molecules diffusing through a small sample volume.[64, 66] The basic concept of FCS is to

make the number of observed molecules low enough so that each of them contributes substantially to the measured signal. The parameter of primary interest in contrast to other fluorescence techniques is not the intensity of the fluorophore itself, but rather the spontaneous intensity fluctuations caused by the minute deviations of the small system from thermal equilibrium. The fluctuations can be quantified in their strength and duration by temporally autocorrelating (A mathematical representation of the degree of similarity between a given time series and a lagged version of itself over successive time interval) the recorded intensity signal.[67]

In dual colour FCCS interacting molecules can be tagged with spectrally different fluorophores and their interaction studied by following intensity fluctuations of both molecules (**Figure 10)**. By separately labeling the reactants with differently emitting fluorophores, the probes can be simultaneously excited with two different laser lines and detected in separate channels. The signals from both channels are cross-correlated in time and the doubly labeled products can be easily distinguished from the singly labeled reactants, independent of their mass. This was earlier not feasible with FCS as the sensitivity of this technique to detect binding of two or more components depends on the relative change in mass upon binding which is not pronounced in a multi-component system where the mass of the product should differ from the reactants by at least a factor of 4.[68]

The basic characteristics of an FCS and FCCS setups are the choice of laser lines, dichroics and filters. FCCS with dual laser excitation involves the use of two spectrally distinct fluorophores each of them excited by one of the lasers.[64] This offers the highest signal to noise ratio since different fluorophores can be excited with lasers having excitation wavelengths that match the fluorophores' absorption maxima. Fluorophores can also be chosen to have widely separated emission wavelength to minimize cross-talk. This requires matching the two laser beams to the same focal spot to maximize the two excitation volumes and detection volumes making it experimentally challenging.[69] Single wavelength FCCS (SW-FCCS) overcame this limitation as it uses a single laser line for the excitation of two fluorophores with similar excitation but different emission spectra due to distinct Stokes shifts.[70] This had the advantage that it eliminated the problems of chromatic aberrations which caused displacement shifts in the observation volume in the previous case and the requirement for focal volume overlap. Resolution could be improved by increasing the S/N ratio through maximizing counts/min and minimizing detector cross-talk. This is influenced

by several factors, importantly fluorophore brightness as there is currently a lack of bright fluorophores with a wide range of Stokes shifts excited by a single laser line.

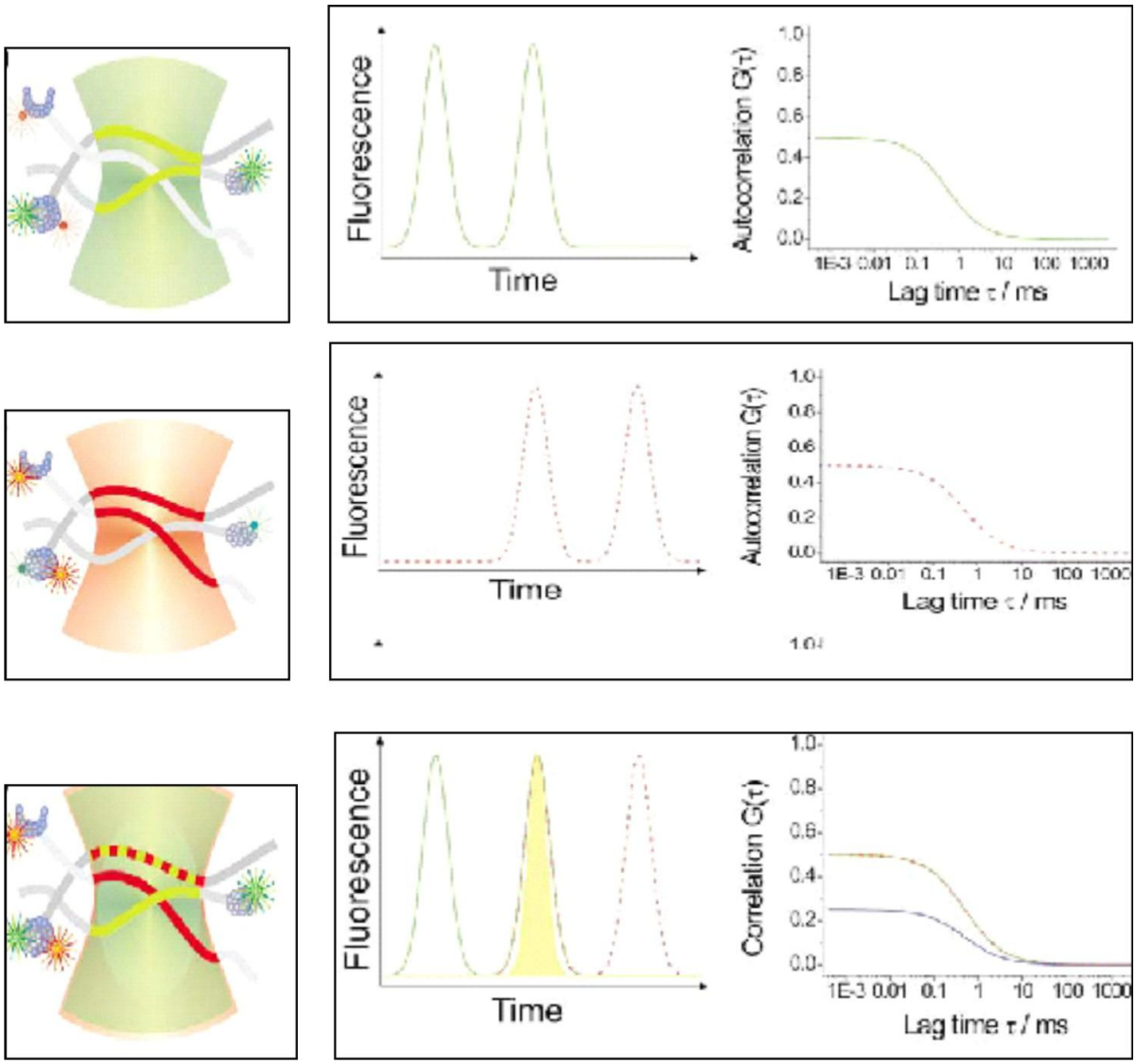



**Figure 10 Schematic representation of the FCCS phenomenon.** A green fluorescently labeled molecule on passing through the observation volume is detected as a fluctuation in the green channel and an auto-correlation curve is obtained for the molecule. Similarly, a red fluorescently labeled molecule will result in a fluctuation being observed in the red channel. An interaction between these two molecules will result in simultaneous fluctuations being observed in both the red and green channels and the cross-correlation curve thus obtained provides characteristic information about the interacting molecules.

FCCS applications require probes with high quantum yields and long-term photostability. It is also advantageous for dyes to have narrow emission spectra for minimal cross talk as red-

shifted dyes usually tend to have broader emission spectra. Organic dyes, quantum dots (QD's) and fluorescent proteins have been used for various FCS applications.[71] Quantum dots in general are superior to organic fluorophores and genetically encoded fluorescent proteins in various aspects such as long-term photostability, high quantum yield, multiple labeling with several colors and single wavelength excitation for all colors. Quantum dots are semiconductor nanocrystals that are coated with a polymer shell or other ligands that allow the materials to be conjugated to biological molecules.[72] They have the unique property of size dependent emission wavelengths.[73] However, quantum dots have limitations which occur due to their blinking characteristics (random switching between bright and dark states), aggregation tendency and large size which affects the mobility of the target molecule.[74] Nevertheless, they hold great promise as fluorescent probes due to their intense brightness, low photobleaching rate and tunable emission wavelengths with broad absorption spectra.

## 1.5 Application of Inteins in Biotechnology

Since their discovery in the nineties, much has been learned about inteins. Crystal structures, kinetics and mutagenesis studies have further added to the understanding of inteins and the splicing reaction. The potential of these inteins as powerful tools for protein engineering has been recognized in the last decade and a number of intein-based technologies have been developed.

Intein mediated protein ligation has been used for ligating peptides and proteins.[75] Intein applications have included segmental labeling of proteins for NMR analysis,[43b, 76] cyclization of proteins,[50a] controlled expression of toxic proteins,[77] quantum dot conjugation to proteins[78] and incorporation of unnatural amino acids.[79] Furthermore, assembly of modular proteins from fragments has been accomplished by PTS resulting in a biologically active full-length protein. This has allowed monitoring of *in vivo* protein-protein interactions[80] and spatio-temproal expression of proteins using reporter genes like enhanced green fluorescent protein (EGFP)[81] and luciferase[82] for the production of transgenic plants by reconstitution of herbicide resistance proteins.[83] PTS has also been useful for translocating proteins into cellular organelles,[17a, 84] for gene therapy[85] and recently for reconstitution of a membrane protein.[86] A few of the important applications of inteins are discussed below.

## I. Intein mediated protein ligation:

(IPL) or expressed protein ligation (EPL) is well established in molecular biology.[28b, 87] It is an *in vitro* technique that ligates polypeptides having an intein generated C-terminal α-thioester with polypeptides starting with Cys and a peptide bond is formed after the ligation (**Figure 11**). Various molecules can be ligated with this methodology, however only if it has a Cys with a free amino group. PNA[88] or DNA[89] tags have been linked to proteins using EPL.

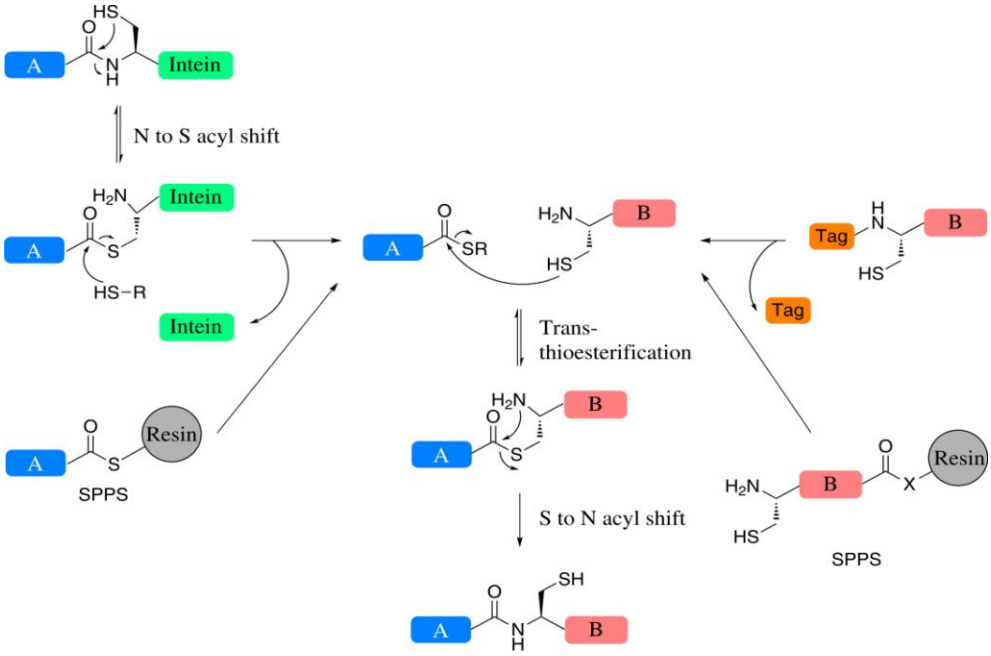



**Figure 11 Expressed protein ligation.** *Trans*-thioesterification of the protein α-thioester by the N-terminal Cys polypeptide is followed by an S to N acyl shift to generate a new peptide bond linking the two polypeptides. α-thioesters can be obtained recombinantly, using engineered inteins, or by chemical synthesis. N-terminal Cys polypeptides can also be produced recombinantly or made using standard SPPS.

## II. Protein semi-synthesis:

Split inteins have been used as versatile tools for protein semi-synthesis (generating a polypeptide from an expressed recombinant protein and a segment obtained by organic peptide synthesis). Short $Int^C$ or $Int^N$ fragments can be chemically synthesized with a small C-extein or N-extein sequence by solid phase peptide synthesis and the remaining protein is fused to the complementary intein fragment by recombinant gene expression. Split inteins with extremely short N- and C-terminal halves are of particular interest in protein labeling because they can provide a facile means for the site-specific incorporation of unnatural amino

acids, fluorescent labels or other biophysical probes into a protein in combination with chemical synthesis (**Figure 12**). PTS between a target protein fused to the remainder of the split intein and the engineered peptide results in a labeled POI.[46, 90] Giriat and Muir reported the semisynthesis of proteins in living cells.[91]

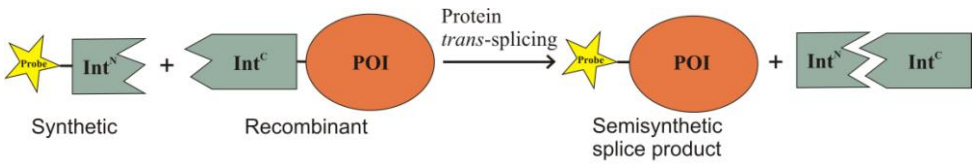



**Figure 12 Protein semi-synthesis.** The scheme depicts semisynthesis of N-terminally modified proteins using protein splicing in *trans*. In a similar manner, proteins can be modified at their C-terminus by chemically synthesizing a short $Int^C$-fragment and reacting with the recombinantly produced complementary half.

Regioselective fluorescent labeling of a protein of interest containing more than one cysteine can also be achieved by combining PTS with cysteine bioconjugation.[42, 86, 92] Split inteins lacking a cysteine at the splice junction and also free of cysteines in the remaining sequence are well suited for this purpose. The $Int^C$ fragments of the artificially split *Ssp* DnaB[92a] and *Mxe* GyrA[42] inteins are cysteine free and serine and threonine occupy the +1 position, respectively. A short peptide sequence with a single cysteine, termed the Cys tag was expressed in fusion with these intein fragments and labeled. Ligation by PTS to proteins with multiple cysteines resulted in a chemically modified protein of interest (**Figure 13**). An internal cysteine residue within a POI is addressed by splitting the POI itself and reconstituting via PTS.[86]

The end result of both IPL and PTS is the formation of a native peptide bond and both techniques have been used in numerous types of modifications.[27b, 28a, b, 88-89, 91, 93] However, both differ from each other in the following points: PTS is driven by strong attractive forces whereas EPL is driven by mass action. The intein generated α-thioester allows ligation of numerous types of molecules besides polypeptides whereas PTS is limited by the substrate preferences of the split intein. Nevertheless, in favor of PTS, the intein directly forms a peptide bond between the two exteins and a chemical reaction is not required. This implies that synthesis of a thioester or any other functional group is not a prerequisite and high concentration of thiols is unnecessary. Furthermore, the high affinity and specificity between

the split intein fragments permits the reaction to be performed at low micromolar to nanomolar concentrations and in the presence of other proteins, for example, in a cell lysate.

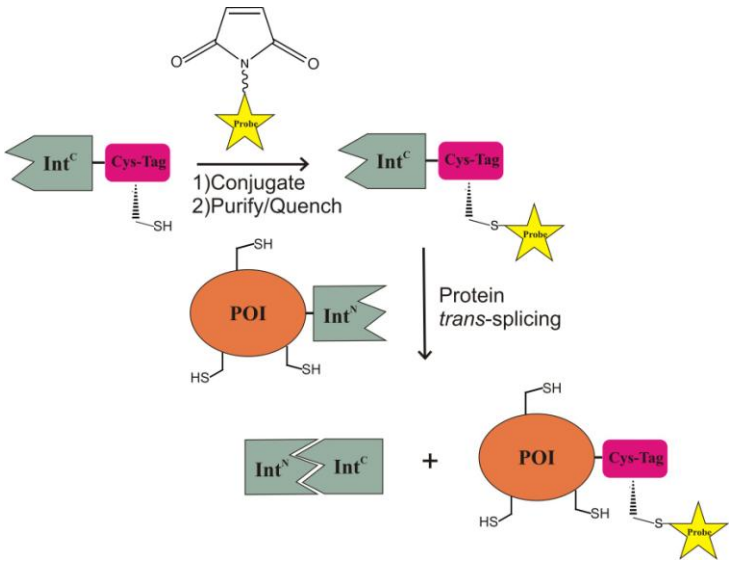



**Figure 13 Combining PTS with cysteine bioconjugation.** A short peptide tag with a single cysteine is chemically modified with a labeling reagent. In a second step, ligation by PTS results in a modified protein. This approach circumvents the problem of regioselective labeling of proteins containing more than one essential cysteine. The split intein fragment fused to the Cys-tag should be free of any catalytic cysteines.

## III. Segmental isotopic labeling

Conventional NMR techniques are generally limited to low molecular weight proteins of upto 30 kDa. Segmental isotope labeling of N- and C-terminal protein segments has been established both *in vitro* and *in vivo* using the PI-*Pfu*I intein and PTS which has allowed for the labeling of larger proteins.[40, 76, 94] Multi-fragment ligation was also successfully performed with orthogonal inteins for labeling at an internal segment of a protein with a heavy atom in NMR spectroscopy.[93, 95]

## IV. Protein-purification

Purification of a variety of proteins has been facilitated by inteins.[96] The principle involves an exchange of the exteins of an engineered intein between the purification tag and the target protein (**Figure 14**). Splice junction residues are mutated to abolish splicing. After immobilization of the POI either the N-terminal cleavage reaction is induced by DTT or a C-

terminal cleavage by a pH or temperature shift resulting in release of the POI. The intein mediated purification with affinity chitin-binding tag (IMPACT) is commercially available.

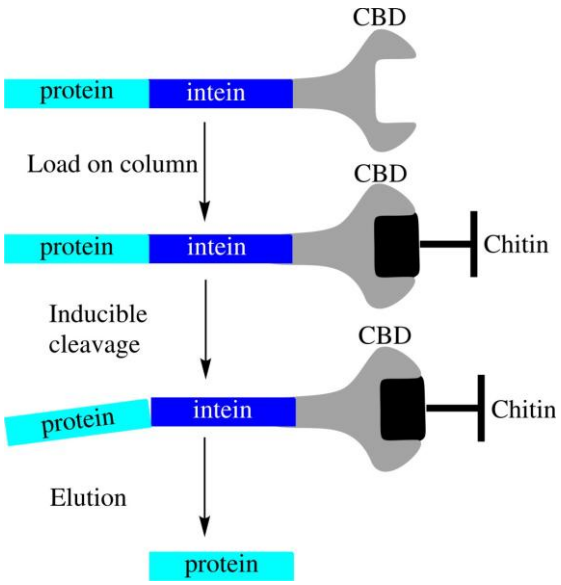



**Figure 14 Intein mediated protein purification.** A mutated intein variant is used for the purification of a protein fused to its N-terminus. A chitin binding domain (CBD) is fused to the C-terminus of the intein and purification is performed by utilizing a chitin column. The intein is mutated at the C-terminus Asparagine so that cleavage only occurs at the N-terminus which results in the release of the target protein from the column.

## V. Conditional Protein Splicing

Inteins can be used as *in vivo* molecular switches by conditional protein splicing. The engineered split fragments of *Sce* VMA1 intein were each fused to either a FKBP or FRB domain. The FKBP-FRB dimerize in response to rapamycin, triggering *trans*-splicing.[97] Similarly, different receptors were inserted into the *Mtu* RecA intein and splicing was activated in response to the respective ligands.[98] Gene function in yeast and drosophila could be spatially and temporally controlled by inserting temperature sensitive splice mutants into transcription regulators.[99]

## VI. Cyclic peptide synthesis

Split intein mediated circular ligation of peptides and proteins (SICLOPPS) has been demonstrated *in vivo* using the *Ssp* DnaE intein. The desired POI is inserted between a

circularly permuted split intein and PTS will result in the N- and C-termini of the target protein being joined together by a peptide bond. The order of the intein fragments was reversed in the precursor i.e. $Int^C$-POI-$Int^N$ and splicing resulted in a fully functional circular protein (**Figure 15**).[50a, 100]

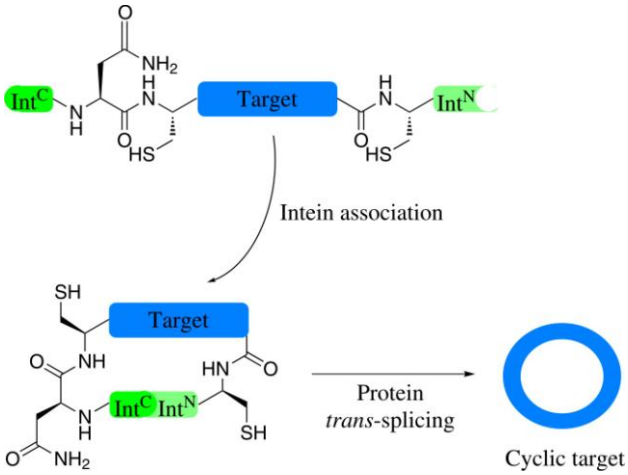



**Figure 15 Schematic representation of intein mediated peptide/protein cyclization.** The target polypeptide is flanked at its N- and C- termini by the $Int^C$ and $Int^N$ fragments, respectively. PTS ligates the N and C termini by a peptide bond generating a circularized peptide or protein.

Thus, it is well documented in literature that modification of inteins both *in vitro* and *in vivo* is of biotechnological significance. Protein *trans*-splicing promises to provide further opportunities in the rapidly growing field of molecular and structural biology.

## 1.6 Cell Surface Protein Labeling Strategies

Visualization of proteins in living systems in real time have been key to understanding the spatiotemporal distribution, trafficking, binding interactions, stability, fate and function of either an individual or multiple proteins. Genetically encoded fluorescent proteins have been the dominant rulers in cellular imaging due to their remarkable stability and capacity to label proteins with absolute specificity. GFP and its variants have indeed revolutionized the ability to visualize and monitor proteins *in vivo*.[101] However, this system also suffers from potential disadvantages. First, these proteins are large enough to interfere with the localization, structure or activity of the protein to which they are fused. The large size also greatly restricts

use for stie-specific labeling within a protein. Secondly, the barrel like structure isolates the chromophore from the cellular environment making fluorescent proteins poor probes for intracellular changes in pH or ion concentrations.[102] There is only a finite potential for modifying the spectral and biochemical properties of fluorescent proteins.[103] Fluorescent proteins are not very bright or photostable thereby making them unsuitable for single molecule microscopy. Precise control of the labeling ratio for multicolor FRET measurements is cumbersome with fluorescent proteins. These disadvantages have led to the development of alternate chemical labeling approaches.

**Site-specific labeling of proteins *in vivo***

Site-specific labeling techniques in combination with fluorescence microscopy have been widely used to detect proteins in living cells. Labeling methods developed so far, rely on protein-ligand interaction, peptide-peptide interaction, peptide fluorophore interaction, metal chelation and enzymatic reactions.[2, 104] Specificity in labeling integral membrane proteins can only occur if the protein of interest harbors a unique reactivity that distinguishes it from other biomolecules present on the cell surface. However, only in very few cases the intrinsic properties of a protein can provide such a unique functionality and therefore in majority of the cases, molecular tricks need to be employed. Several post-translational labeling methods are currently known which use a genetically encodable tag and synthetic probes which are specific for those tags. A good labeling strategy should possess high specificity for the target protein, does not interfere with the biochemical functions or cellular localization of the labeled protein and minimally perturbs the normal cellular processes. Criteria's such as cell permeability, non-toxicity, specific reactivity and good fluorescent properties should be carefully considered during the design and tailoring of small molecule probes.

Some of the popular techniques currently utilized for site specific labeling of cell surface proteins have been discussed below.

### 1.6.1 Non-covalent Methods

#### I. Biarsenical Dyes

This method developed by Tsien *et al.* exploits the well-known specificity of organoarsenicals with pairs of thiols (**Figure 16**).[105] F1AsH, a cell permeable fluorescein derivative or ReAsH, a resorufin derivative was used to label a short peptide sequence CCXXCC (X= non-

Cys) fused to a protein of interest. F1AsH contains two As (III) substituents conjugated to ethanedithiol (EDT). F1AsH bound to EDT is virtually non-fluorescent but upon binding to the CCXXCC motif, a marked increase in fluorescence occurs.[106]

Used for studying the role of conformational changes in GPCR-mediated cell signaling[107] and activation.[108]

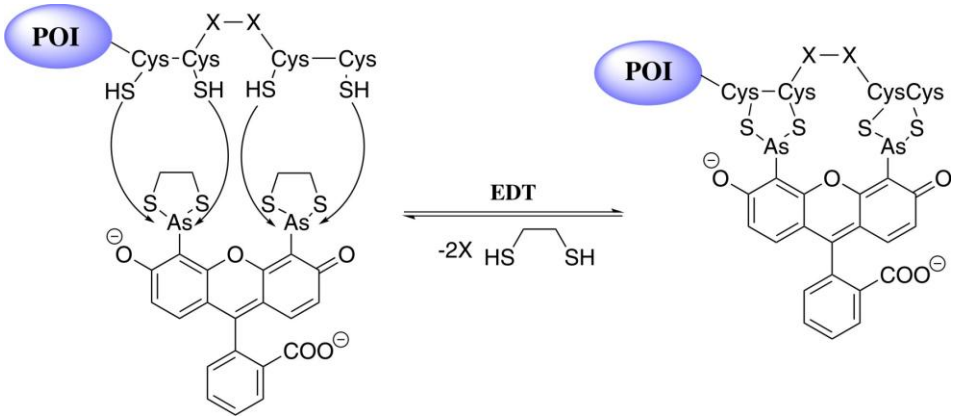



**Figure 16 Site-specific labeling of tetracysteine motif by F1AsH.** A short peptide sequence CCXXCC (X is a non-cysteine amino acid) was genetically fused to the protein of interest which was subsequently recognized and labeled by F1AsH, a cell permeable fluorescein derivative. POI = Protein of interest.

## II. Labeling using DHFR fusion

This non-covalent technique developed by *Cornish et al.* uses the well-described interaction between methotrexate (Mtx) and the enzyme dihydrofolate reductase (DHFR) to label DHFR fusion proteins with fluorescent derivatives of Mtx.[109] Mtx is a small molecule which binds with subnanomolar affinity to the monomeric protein, DHFR. An improved system using the bacterial DHFR and trimethoprim (TMP) has been recently described.[109b]

*E.coli* dihydrofolate reductase (eDHFR) at the plasma membrane has been labeled by ligand TMP by fusing the N-terminus of eDHFR to the myristoylation/palmitoylation sequence MGCKISKGKD.[109b]

### III. Reversible labeling with NTA probes

Vogel *et al*. reported a site-selective and reversible method for labeling membrane proteins containing a polyhistidine sequence which was based on the known interaction between polyhistidine sequences and the $Ni^{2+}$-NTA moiety.[110]

Structure and plasma membrane distribution of the 5-hydroxytryptamine serotonin receptor ($5HT_3R$) was studied by this approach in HEK 293 cells.[110]

### IV. Non-covalent labeling with FKBP12

Nolan *et al.* devised a technique that made use of the tight interaction between a specific mutant of FK506-binding protein 12 (FKBP12) and the synthetic ligand SLF' that was derivatized with fluorescein. SLF' has sub-nanomolar affinity and an extremely high selectivity for FKBP12 and lacks the immunosuppressive effects of FK506 (original FKBP12 ligand).[111]

This approach was used to label a variety of proteins including caveolin, rac and rho in HeLa cells although only two types of fluorophores (TMR and fluorescein) were suitable for effective labeling.[111-112]

### V. Oligo-Asp/$Zn^{2+}$complex labeling

Hamachi *et al.* devised an elegant technique for labeling cell surface proteins based on metal-ligand interaction. A short oligo- aspartate tag was fused to the protein of interest based on the fact that several carboxylate groups will react favorably with a cationic metal complex. Binuclear Zn(II)-DpaTyr fluorescein or cyanine derivatives were used as the metal complex.[113]

CHO cells expressing the muscarinic acetylcholine receptor (mAChR) was successfully labeled with FITC using this technique without affecting its signaling activity.[113]

### 1.6.2 Covalent Methods

#### I. Labeling using hAGT fusion

Keppler *et al.* developed a strategy whereby specific covalent labeling was achieved using human $O^6$-alkylguanine-DNA alkyltransferase (hAGT) (**Figure 17**). Function of hAGT in the cellular environment involves the repair of alkylated DNA where it irreversibly transfers the alkyl group from $O^6$-alkylguanine DNA to one of its reactive cysteines and following this transfer the AGT is typically degraded. However, $O^6$-benzylguanine (BG) and BG-derivatives substituted at the 4-position of the benzyl ring can also act as substrates. BG-derivatives containing biotin and fluorescein were successfully used for labeling hAGT fusion proteins [(G160W) mutant of hAGT which possess higher activity with BG] expressed in live cells.

SNAP-tag- Efforts to decrease non-specific binding of BG derivatives to endogenous AGT lead to the identification of an inhibitor against *wt* AGT together with the generation of a mutant $^M$AGT which is refractory to this inhibitor. This SNAP-tag has reduced affinity for alkylated DNA and exhibits a 50-fold faster labeling kinetics.[114] It was fused to the GPCR $NK_1$ with an exoplasmic N-terminus.[115]

Recently, this technique was used to specifically label cell-surface G-protein coupled receptors (GPCR'S).[116]

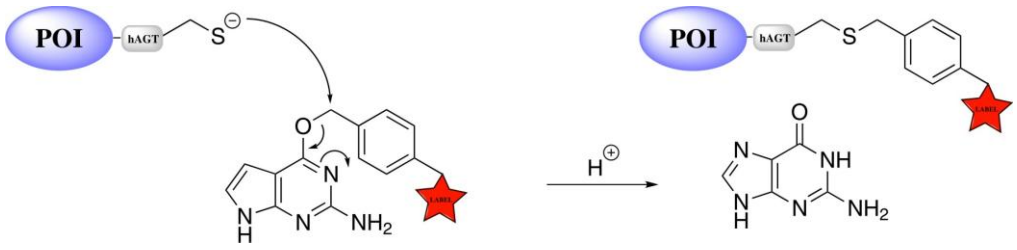



**Figure 17 hAGT mediated protein labeling.** Covalent labeling of a protein of interest fused to hAGT using $O^6$-benzylguanine derivatives. The native activity of hAGT is to repair alkylated DNA by transferring the alkyl group from $O^6$-alkylguanine DNA to an active site Cys residue of the hAGT enzyme. Here, the hAGT catalyzes the transfer of small molecule probes from $O^6$-benzylguanine conjugates to the active site Cys. POI = Protein of interest, Red star = probe

#### II. Halo-Tag mediated labeling

Halo Tag protein is a ~33 kDa monomeric engineered haloalkane dehalogenase from *Rhodococcus rhodochrous* devised by researchers at Promega that cleaves carbon-halogen

bonds in aliphatic halogenated compounds. Nucleophilic attack by the haloalkane to Asp106 in the enzyme results in the formation of an ester bond between the ligand and protein followed by ester hydrolysis to yield alcohol as the final product (**Figure 18**). A His272Phe mutant of the enzyme, which prevents ester hydrolysis, fused to the protein of interest would allow the covalent attachment of alkyl halide conjugated chemical probes to the Halo-Tag for protein labeling by forming a stable bond with the substrate haloalkane.[117]

Organic fluorophores like TMR and fluorescein as well as inorganic quantum dots have been attached to the extracellular domain of platelet derived growth factor receptors for long-term imaging.[118]

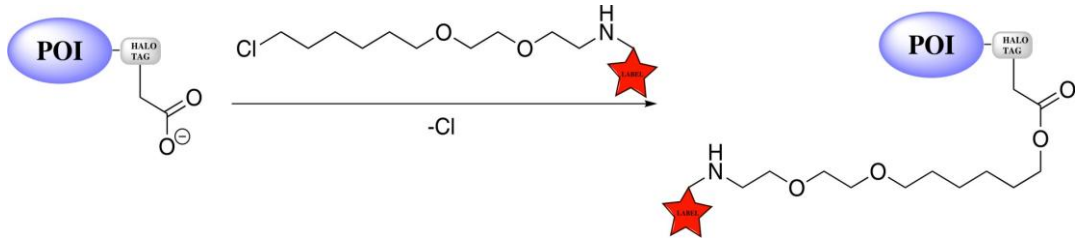



**Figure 18 Halo tag mediated protein labeling.** Haloalkane dehalogenases under native conditions hydrolyze alkyl halides. In protein labeling reactions, chemical probes linked to an alkyl halide can be covalently conjugated to an active site Asp residue in the Halo tag. POI = Protein of interest, Red star = probe

## III. Labeling by enzymatic posttranslational modification

**Cutinase:** This enzyme is a ~22 kDa globular, fungal serine esterase which under native conditions hydrolyzes cutin and is known to also hydrolyze fatty acid esters and triacyl glycerol.[119] Fusion proteins can be targeted using probes based on the *p*-nitrophenyl phosphonate (pNPP) scaffold which is a suicide inhibitor for cutinase and forms a covalent adduct with the enzyme by conjugation with a specific Ser residue in the enzyme active site (**Figure 19**).[120] In protein labeling catalyzed by cutinase, pNPP-SH derivatized with maleimide-functionalized fluorophores are transferred onto the cutinase fused to the target protein.

Cell surface protein integrin LFA-1 expressed on the surface of BAF cells has been labeled with quantum dots or Alexa Fluor 488 to study the dynamic redistribution of the receptor.[121]

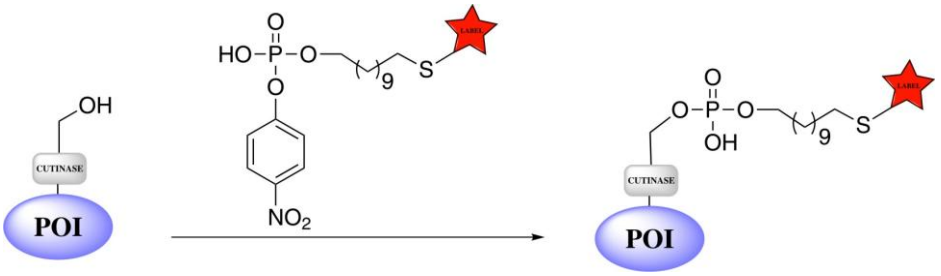



**Figure 19 Cutinase catalyzed protein labeling.** Native activity of cutinase is to hydrolyze the polyester linkages in cutin. For labeling proteins, small molecule probes are transferred from their phosphonate derivatives to a specific Ser residue in the cutinase tag. POI = Protein of interest, Red star = probe

**<u>Biotin Ligase (BirA)</u>:** In nature, BirA accepts only biotin and its derivatives as substrates for covalently attaching a biotin to a specific Lys side chain of the biotin carboxyl carrier protein (BCCP) in the multi-enzyme complex of Acetyl carboxylase A with an amide linkage. A two-step labeling strategy was established by the group of Ting *et al.* where they demonstrated the acceptance of a ketone isostere of biotin by an *E.coli* BirA and transfer to the lysine side chain within a 15 amino acid acceptor peptide (AP) sequence (**Figure 20**). Incorporation of the ketone into the AP is followed by bio-orthogonal ligation with hydrazide or hydroxylamine bearing probes.[104a]

Epidermal growth factor receptor (EGFR) expressed on the surface of HeLa cells was labeled by this procedure.[104a]

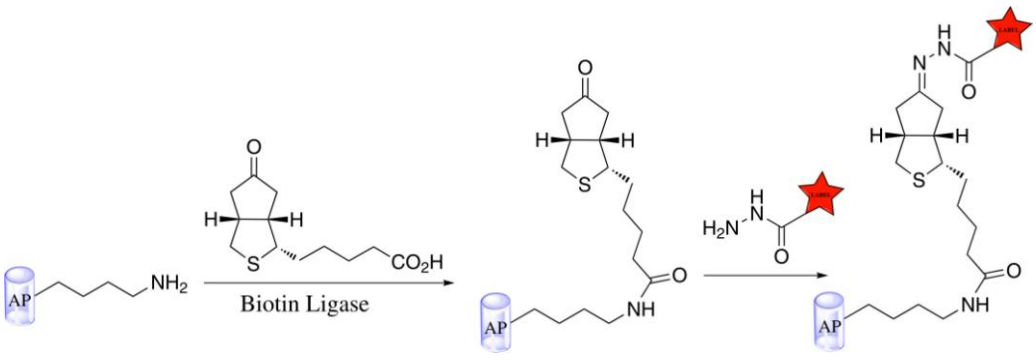



**Figure 20 General method for labeling acceptor peptide (AP)-tagged protein of interest using biotin ligase.** Biotin ligase catalyzes the ligation of ketone to the AP. A subsequent bio-orthogonal ligation between ketone and hydrazide (or hydroxylamine) introduces the probe. POI = Protein of interest, Red star = probe

**Lipoic acid ligase mediated labeling:** Similar to the above mentioned biotin ligase strategy, a two-step labeling can be carried out by using lipoic acid ligase (LplA) which uses an alkyl azide as substrate instead of the native lipoic acid.[122] An LplA acceptor peptide (LAP) was identified that could be appended to the target protein and the LplA ligates the alkyl azide to the ε-$NH_2$ of the lysine side chain in the LAP sequence (**Figure 21**). This is followed by its derivatisation with a cyclo-octyne conjugate modified with dyes, QD's or biotin (Huisgen cycloaddition).

Labeling of LAP-low density lipoprotein receptor (LDLR) fusion protein expressed on the cell surface was carried out which did not impair protein activity or trafficking.

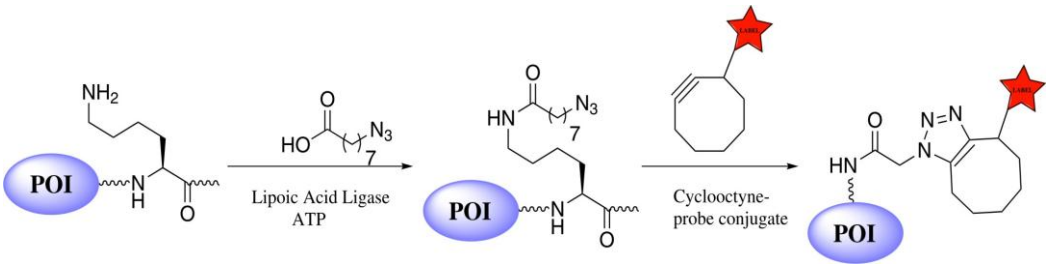



**Figure 21 Scheme for lipoic acid catalyzed fluorescent tagging.** Lipoic acid ligase ligates an alkyl azide to a lysine side chain within a peptide recognition sequence. The azide is selectively functionalized with a cyclo-octyne conjugate to a probe of interest to give a triazole adduct. POI = Protein of interest, Red star = probe

**Transglutaminase (TGase):** TGases catalyze amide bond formation in multicellular organisms between glutamine and lysine side chains and are active during apoptosis, wound healing and migration events.[123] Acyl transfer from the γ-carboxamide group of a Gln residue in one protein to the ε-amino group of a Lys residue in another protein establishes an isopeptide bond between the proteins (**Figure 22**). The TGase from guinea pig liver (gpTgase) exhibits high specificity for Gln containing proteins which act as acyl donors and low specificity for the Lys containing proteins which end up as acyl acceptors.[124] Amines as diverse as fluorescein cadaverine and biotin cadaverine can be utilized by gpTGase instead of lysine.

Ting *et al.* labeled cell surface proteins fused to a "Q-tag" (PKPQQFM) with biotin or Alexa-568 derivatives of cadaverine using guinea pig liver TGase. The TGase specifically recognized the Q-tag of a cadaverine conjugated fluorophore to a glutamine side chain in Q-rich tag sequences expressed in HeLa cells.[125]

EGFR conjugated to a Q-tag was successfully labeled with Biotin cadaverine with no impairment in receptor response to EGF.[125]

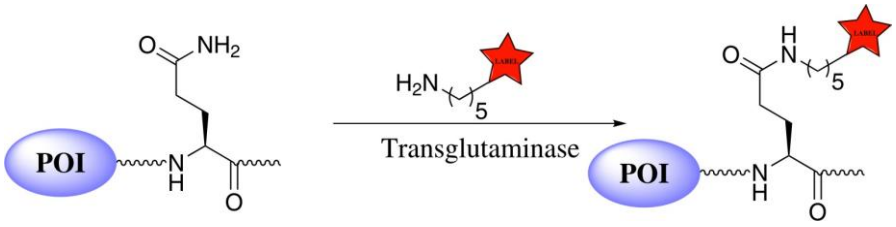



**Figure 22 Transglutaminase catalyzed protein modification.** Under native conditions TGase catalyzes the transamination reaction between Gln and Lys residues in the modified proteins. In TGase catalyzed protein labeling, cadaverine functionalized small molecule probe is transferred to a Gln residue in the Q tag fused to a protein of interest. POI = Protein of interest, Red star = probe

**<u>Phosphopantetheine Transferase (PPTase)</u>:** The native activity of PPTases is to transfer the 4'-phosphopantetheinyl group (Ppant) from CoA to a conserved Ser residue of acyl carrier protein (ACP) or peptidyl carrier protein (PCP).[126] Fusion proteins to ACP/PCP can be modified with CoA derivatized with a fluorophore or other probes using either AcpS (for ACP only)[126a] or Sfp (for both ACP and PCP) (**Figure 23**).[127]

ACP fused to the N-terminus of neurokinin-1 ($NK_1$) receptor allowed mobility studies down to single receptor sensitivity by FRET on cell membranes of HEK293 using Cy3 and Cy5 fluorophores.[128]

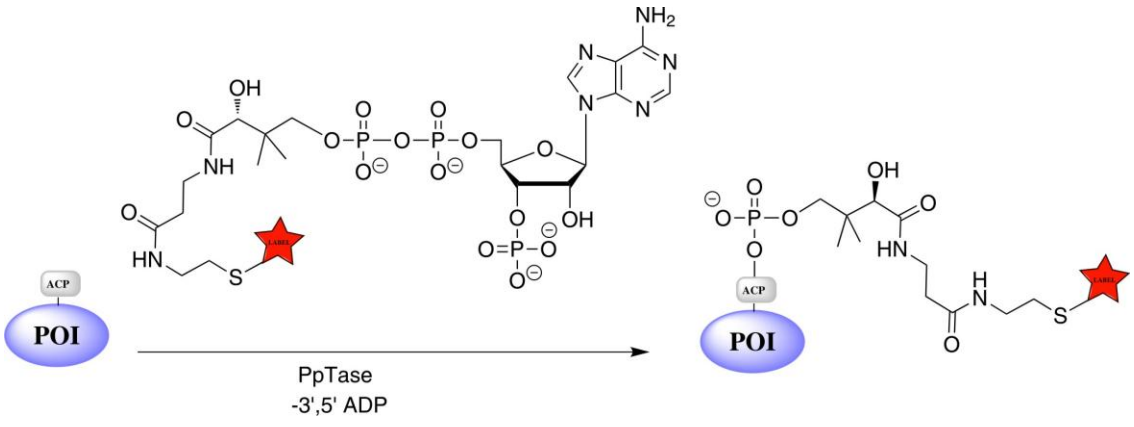



**Figure 23 Phosphopantetheinyl transferase (PPTase) catalyzed protein modification.** PPTases use CoA as the source for Ppant and attach it as a phosphodiester to an invariant Ser residue of the acyl carrier protein (ACP) or peptidyl carrier protein (PCP). Fusion proteins to ACP/PCP are labeled by CoA derivatized with desired probe. POI = Protein of interest, Red star = probe

**Sortase mediated site specific labeling:** Sortase A (SrtA) from *Staphylococcus aureus* natively catalyzes the hydrolysis of the peptide bond between Thr and Gly in a LPXTG recognition motif in the cell surface protein and subsequently forms a new peptide bond between the carboxyl group of the exposed Thr and the amino group of the pentaglycine cross bridge on the cell surface peptidolglycan thereby anchoring the protein to the cell wall (**Figure 24**).[129] In 'Sortagging' the SrtA uses biotin and fluorescence probes conjugated to peptides with oligo Gly as substrates for modifying the protein of interest fused to the LPETG peptide tag.[130] The tag should be located close to the C-terminus of the target protein in a flexible region.

Membrane proteins CD40[131] and osteoclast differentiation factor (ODF)[130c] were labeled in intact HEK293T cells.

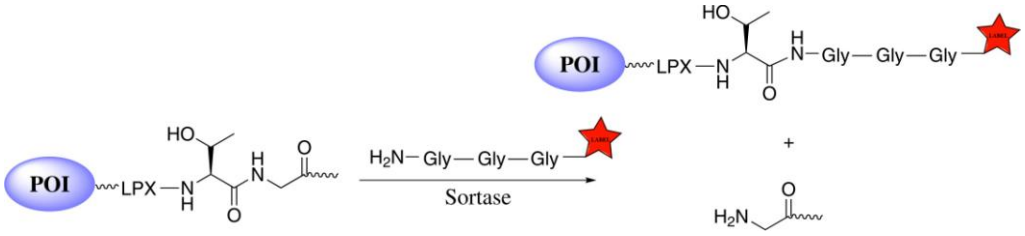



**Figure 24 Sortase mediated ligation of triglycine derivatives to LPXTG tagged proteins.** Sortase recognizes the LPXTG sequence, cleaves between the Thr and Gly residues and subsequently links the carboxyl group of threonine to an amino group of N-terminal glycine oligomers by a native peptide bond. POI = Protein of interest, Red star = probe

## IV. Small molecule labeling using unnatural amino acids

Pioneered by Schultz *et al.*, the idea behind the strategy involves the amber suppressor tRNA which is chemically modified with an unnatural amino acid by a unique aminoacyl tRNA-synthetase corresponding to a unique tRNA-codon pair.[132] Several of these unnatural amino acids have novel properties which are useful for studying biochemical and cellular functions of protein.

m-acetyl-phenylalanine was site-specifically incorporated into the LamB membrane protein and subsequent labeling with hydrazide dyes was carried out.[133]

## V. Miscellaneous methods

The HisZiFit, histidine-zinc fluorescent tag is useful for labeling surface-exposed hexahistidine-tagged proteins. Surface exposure of a membrane protein, stromal interaction molecule (STIM1) from the ER in HEK293 cells was successfully detected with this method.[134]

A 13-amino acid peptide that binds to α-bungarotoxin has been successfully used as a tag for labeling extracellular domains of proteins of interest, for e.g., membrane trafficking of AMPA receptors and a vesicle-associated protein VAMP2 was observed by this method.[135]

Peptide-peptide interaction between the $E_3$ tag $(EIAALKE)_3$ and probe $K_3$ $(KIEELEK)_3$ or $K_4$ $(KIEELEK)_4$ has been used for specific labeling of cell-surface receptors such as prostaglandin EP3β, EGF and β2- adrenergic receptors.[136]

## VI. PTS for labeling cell surface proteins

In protein *trans*-splicing (see section 1.2), a polypeptide is reconstituted from split halves (**Figure 25**). The semi-synthetic PTS approach (see section 1.5) has earlier been successfully applied for labeling cell surface proteins at either their N- or C-terminus using artificially split and chemically synthesized intein fragements.[90a, 90c] The present work takes advantage of the superior properties of the fastest known naturally split *Npu* DnaE intein for labeling cell surface proteins.[4]

The PTS approach in contrast to other techniques mentioned above allows for the direct manipulation of the polypeptide backbone of the protein of interest (POI). Such backbone engineering would also be interesting for protein reconstitution or the creation of nanostructures in artificial systems. Majority of the earlier methods for cell surface labeling only resulted in the attachment of a modified prosthetic group to a particular side chain in the tag and are thus not suitable. A tabular comparison of the various advantages and disadvantages of each technique is given in

**Table** 1**.** However, such backbone engineering can also be achieved by the Sortase mediated ligation.[130c] This reaction is however hampered by the requirement for high concentrations of both the substrates and the bond-forming enzyme sortase A. The split intein approach does not necessitate an extra enzyme as the PTS reaction is a self-processing reaction. Moreover,

the reaction can be virtually traceless because the split intein fragments are removed during the process.

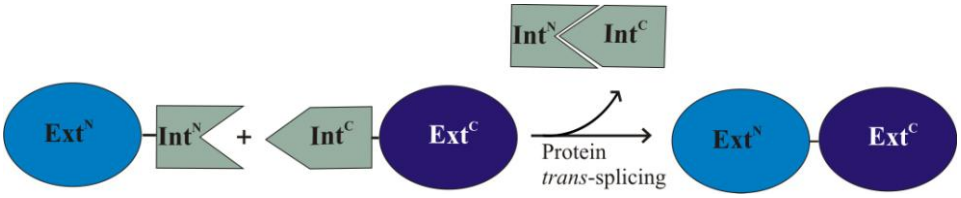



**Figure 25 Schematic representation of protein *trans*-splicing.** The intein domain is split into two pieces that are connected to the flanking extein sequences. The first step comprises of a fast intein association whereby the inteins fold into an active conformation. The rate limiting second PTS step then excises the intein joining the two extein sequences with a peptide bond.

**Table 1 Comparison of various cell surface labeling techniques**

| Method | Advantage | Limitation |
|---|---|---|
| **Biarsenical dyes** | Small size, stable self-labeling | Arsenic toxicity, Cys reduced, dithiols required, high background |
| **DHFR fusion** | High Specificity | DHFR deficient cell lines required |
| **NTA-His** | Site-specific, fast and reversible, does not require cell fixation | Low affinity and stability, quenching, $Ni^{2+}$ toxicity |
| **FKBP12** | Cell permeable, non-toxic, high affinity and selectivity | Only two types of fluorophores are suitable, protein-ligand dissociation |
| **Oligo-Asp/$Zn^{2+}$** | Low Affinity | Zn requirement, large size |
| **SNAP-tag fusion** | High specificity, minimal background, inert,non-cytotoxic | Large size of hAGT, AGT-deficient cell lines required, degradation of fusion protein |
| **Halo-Tag fusion** | Endogenous haloalkanes absent, easy labeling with both organic dyes and inorganic QD's | Very large tag |
| **Cutinase** | Low background, easy synthesis | Suicide inhibitors may react with endogenous proteins |
| **BirA Labeling** | Endogenous keto-groups absent, selective, small size, easy synthesis of hydrazides | Millimolar concentrations of probe required |

| | | |
|---|---|---|
| **LpIA labeling** | High specificity, minimal background | High probe concentrations required |
| **TGase labeling** | Applicable to wide range of substrates | Poor specificity |
| **ACP/PCP-Tag** | Tolerates various chemical substitutents, specific | Large size of tag |
| **Suppressor tRNA** | Specificity, versatile, minimal structural perturbation, small molecule label | tRNA/synthetase pair for each amino acid, endogenous amber codons, low suppression efficiency |
| **HisZiFit** | Tight affinity | Suboptimal dye characteristics |
| **Sortagging** | Protein backbone modification | Necessity of sortase A and high concentration of substrate. |
| **Protein-*trans* splicing** | Protein backbone modification, traceless, autocatalytic | Size of intein, effect of extein sequence |

Therefore, it is apparent that no one technique is applicable for labeling the large repertoire of cell surface proteins. Each technique has its unique benefit and also suffers from certain drawbacks and limitations. An appropriate choice of a technique is possible depending on its intended use.

The aim of this work was to modify proteins expressed on the cell surface and facing the extracellular environment in a fast and efficient manner. Protein *trans*-splicing using the naturally split *Npu* DnaE intein from *Nostoc punctiforme* was the method of choice for this purpose because it is currently the fastest known intein with a $t_{1/2}$ of ~60s.[49b] The split inteins remove themselves due to their autocatalytic character which makes ligation of even entire protein domains feasible. This technique will also allow the manipulation of the primary sequence of the protein sequence in a post-translational manner. The PTS reaction for labeling membrane proteins will allow for the rapid and almost traceless modification and the additional superior properties conferred by the *Npu* DnaE intein will allow for its use in various applications.

# 2 MATERIAL

## 2.1 Apparatus

**Table 2 Apparatus list**

| Apparatus | Manufacturer |
|---|---|
| -80 °C freezer | *New Brunswick scientific* ultra low temperature freezer |
| -152 °C deep freezer | *SANYO* Ultra low temperature freezer, MDF-1156 |
| Agarose gel station | *PEQLAB* Perfect Blue Gelsystem M and L |
| Weighing balance | *Mettler* PM400; *Sartorius* Basic |
| Autoclave | *H+P Laboratory technology* Varioklav 135 S |
| Chromatographie system | *GE Healthcare* ÄKTA prime plus |
| $CO_2$ incubator | *Thermo Scientific Heraeus* Heracell 150 |
| Agarose and SDS-PAGE gel documentation | *Canon* PowerShot G5 |
| Electroporator | *Bio-rad* MicroPulser |
| Emulsifier | *Avestin* Emulsiflex-C5 |
| Heating block | *Eppendorf* Thermostat Plus; *Eppendorf* Thermomixer comfort |
| Horizontal Shaker | *Heidolph* Promax 1020 |
| Incubator | *Memmert* INE 800 |
| Incubator Shaker | *Infors* HT Aerotron; *New Brunswick Scientific* I26,TC-7 |
| Microscope | Olympus IX-81; Olympus FV100 confocal |
| PCR | *Biometra* TPersonal Combi |
| pH-meter | *HANNA* instruments pH211 |
| Pipettes | *Eppendorf* Reference |
| Double distilled water | *Millipore* Milli-Q |
| Scanner | *Cannon* CanoScan 8400F |
| SDS-PAGE Apparatus | *Bio-Rad* Mini-PROTEAN 3 Cell; *Bio-Rad* PowerPac Basic |
| Safety hood | *Nuaire* NU 440-500E |
| UV-screen | *H.Saur* Transilluminator IL-200M |
| UV/Vis-Spectrophotometer | *GE Healthcare* Ultrospec 3100 pro; Varian Cary 100Bio |

| | |
|---|---|
| Vortex | *LMS* VTX 300 |
| Waterbath | *P-D Industriegesellschaft* WB20 |
| Western blot apparatur | *Bio-Rad* Mini Tans-Blot Cell |
| Cell density counter | *GE Healthcare* Ultrospec 10 |
| Centrifuge | *Eppendorf* Centrifuge 5415 D; *Eppendorf* Centrifuge 5804 R; *Sorvall* Evolution |

## 2.2 Chemicals, Enzymes and Other Instruments

A comprehensive list of the chemicals used in this work is given in **Table 3**.

**Table 3 Chemical list**

| **Supplier** | **Product(s)** |
|---|---|
| *AppliChem* GmbH (Darmstadt) | Glycine, IPTG,Tris |
| *Becton, Dickinson and Company* (New Jersey, USA) | Yeast extract |
| *Biolegio* (Nijmegen, Netherlands) | Oligonucleotide |
| *Carl Roth* (Karlsruhe) | Acrylamide solution for SDS-PAGE, Agar Nr.1, Ampicillin, APS, L-Arabinose, Bromophenol blue, BSA, Coomassie Brilliant blue R 250, Chloramphenicol, DMSO, Ethanol, Ethidium-bromide, Glucose, β-mercaptoethanol, Milk powder, SDS, TEMED; Triton X-100. Tween-20, Visking-dialysis bag |
| *Covance* (Princeton, USA) | Anti-GFP, Anti-HA antibodies |
| *Eurogentech* (Seraing,Belgium) | Agarose, Electroporation cuvette |
| *Eurofins MWG Operon* (Ebersberg) | Oligonucleotides |
| *Fermentas* (Burlington, Canada) | dATP, dTTP, dGTP, dCTP, DTT, DNA and Protein marker, T4 DNA-ligse, Restriction enzymes |
| *Finnzymes* (Espoo, Finnland) | Phusion$^{TM}$ DNA polymerase |
| *GE Healthcare* (Little Chalfont, Great Britain) | ECL and ECL advanced western blotting kit, HyperfilmTM ECL, Whatman GB500 Blotting paper, Anti-cmyc antibody |
| *IBA* (Göttingen) | Strep-tactin sepharose, D-desthiobiotin |
| *Invitrogen* (Darmstadt) | Quantum dot 655 starter kit (streptavidin coated) |
| *MatTek Corporation* (Ashland, USA) | 35-mm-glassbottom dish |

| | |
|---|---|
| *Millipore* (Billerica, USA) | Sterile filter (45 nm) |
| *New England Biolabs* (Ipswich, USA) | Phusion high fidelity DNA-polymerase, Restriction enzymes |
| *PAN Biotech GmbH* (Aidenbach) | Media, Buffer and Solutions for cell culture |
| *Qiagen* (Hilden) | QIAquick PCR-Purification and Gel-Extraction kit, Ni2+-NTA-Superflow |
| *Sarstedt* (Nümbrecht) | 100-mm-cell culture dish, 6 well plates, filtration units |
| *Sigma-Aldrich* (St-Louis, USA) | EDTA, PEG 3350 |

## 2.3 Vectors

### 2.3.1 pRSFDuet

The pRSFDuet vector (3.8 kb) from *Novagen* was developed for the co-expression of two target proteins. It encodes two MCS each of which is preceded by a T7 promoter, lac operator, and a ribosomal binding site. In addition, it also includes the Kanamycin-resistance gene, the lacI gene and the replication origin RSF derived from the RSF1030 replicon. In this work, this vector was not used for co-expression but for the expression of a single target gene cloned into the first MCS.

### 2.3.2 pBAD

The pBAD vectors from Invitrogen allow the expression of a protein of interest by the addition of L- arabinose. The target gene is under the control of the araBAD promoter at its 5'-end which in turn is regulated by the AraC-repressor protein. In addition, for maximal expression, the binding of L- arabinose to the AraC-Repressor requires the complex of cAMP and cAMP activator protein (CAP).

### 2.3.3 pmCherry-N1

The pmCherry-N1 vector (4.7 kb) is derived from the Discosoma sp. red fluorescent protein, DsRed. This mammalian expression vector can be used for the localization determination of proteins fused in frame to the N-terminus of mCherry-N1 by fluorescence microscopy. The inserted gene should contain an ATG-initiation codon with no in-frame stop codons in between. The excitation and emission maxima of mCherry is 587 nm and 610 nm,

respectively. A CMV promoter upstream of the MCS allows for its expression in mammalian cells. The pmCherry-N1 vector can also be used for simply expressing the mcherry protein in a cell line and can be easily transfected using any given transfection protocol. It contains the gene for conferring kanamycin resistance to *E.coli* cells.

### 2.3.4 pEGFP-N1

The pEGFP-N1 (4.7 kb) vector from Clontech was derived from the wt-GFP containing a double amino-acid mutation of Phe-64 to Leu and Ser-65 to Thr. This mammalian expression vector can be used for the localization determination of proteins fused in frame to the N-terminus of pEGFP-N1 by fluorescence microscopy. The inserted gene should contain an ATG-initiation codon with no in-frame stop codons in between. The excitation and emission maxima of pEGFP is 488 nm and 507 nm, respectively. A CMV promoter upstream of the MCS allows for its expression in mammalian cells. The pEGFP-N1 vector can also be used for simply expressing the EGFP protein in a cell line and can be easily transfected using any given transfection protocol. It contains the gene for conferring kanamycin resistance in *E.coli* cells. Genbank: U55762.1

### 2.3.5 pCDNA3

The pcDNA3 vector (5.4 kb) from Invitrogen is a shuttle vector for the expression of proteins in mammalian cells. After the cloning of the gene of interest in the multiple cloning site (MCS), the expression can occur under the control of the CMV promoter. The vector contains an Ampicillin-resistance gene (bla) that serves as a selectivity marker in *E. coli*.

### 2.3.6 pDisplay

The pDisplay mammalian expression vector (5.3 kb) allows for the expression of proteins on the cell surface. The gene of interest is fused to the N-terminus of the Ig-κ chain leader sequence, directing the protein to the secretory pathway. At the C-terminus is the transmembrane domain of the platelet derived growth factor receptor (PDGFR) which anchors the expressed protein to the extracellular side of the plasma membrane. The vector backbone also contains the *hemagglutinin A* (HA*)* and the *myc* epitopes for detection by western blot and immunofluorescence.

## 2.4 Micro-Organisms and Cell Lines

### 2.4.1 *E.coli* Strains

**I. DH5α**

This heat competent *E.coli* strain (Subcloning efficiency ™, DH5α ™) was purchased from *Invitrogen.*

Genotype: F-Φ80*lacZ*Δ*M15*Δ(*lacZYA-argF*)U169 *deoR recA*1 *endA*1 *hsdR17*($rk^-$, $mk^+$) *phoA supE44 thi*-1 *gyrA96 relA1* λ-

**II. XL-10 Gold**

This strain was used for preparing electro-competent cells for the purpose of cloning and was purchased from *Stratagene*.

Genotype: *Tetr Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr) endA1 supE44 thi-1 recA1gyrA96 relA1 lac Hte [F´ proAB lacIqZΔM15 Tn10 (Tetr) Amy Camr]*

**III. BL21-Gold (DE3)**

This strain was used for the preparation of heat competent cells for the purpose of protein expression and was purchased from *Stratagene*.

Genotype: F- *ompT hsdS* (rB-mB-) *dcm*+ Tetr *gal* λ(DE3) *endA* Hte

### 2.4.2 Eucaryotic Cell Lines

**I. Neuro2a :** This cell line was derived from the neuroblast cells from mouse brain. They are amoeboid like stem cells and grow as an adherent monolayer.

**II.** **COS-7 :** This cell line was derived from the kidney of the African green monkey. They are fibroblast-like cells and grow as an adherent monolayer.

**III.** **CHO :** This cell line was derived from the ovary of an adult Chinese Hamster. They are epithelial-like cells and grow as an adherent monolayer.

## 2.5 Growth Media

### 2.5.1 *E.coli* Growth Media

LB-media was used as full growth media for *E.coli* cells. The media was sterilized by autoclaving for 30 mins at 121ºC and 1,5 bar. For preparing Agar plates, 1,5 % (w/v) agar-agar was added to the LB-media before autoclaving.

| | | |
|---|---|---|
| **LB medium:** | 10 | g/L Tryptone |
| | 5 | g/L Yeast Extract |
| | 5 | g/L Nacl |

The relevant antibiotic (heat sensitive) was added to the media after autoclaving as a sterile filtered solution in the following concentration: Ampicillin (100 µg/mL), Kanamycin (50 µg/mL).

### 2.5.2 Mammalian Cell Culture Growth Media

**I.** **Neuro2a:** MEM (Minimum Essential Media) from PAN-Biotech was used for the growth of the Neuro2a mammalian cells.

| | | |
|---|---|---|
| **MEM:** | EBSS (Earle's balanced salt solution) | |
| **(including)** | 2,2 | g/L $NaHCO_3$ |

In order to prepare the full growth media, MEM was supplemented with the following components and sterile filtered and then stored at 4 ºC.

| | | |
|---|---|---|
| **MEM:** | 10 % | (v/v) Fetal Bovine Serum (FBS) |
| **(full growth media)** | 2 | mM L-Glutamine |
| | 100 | U/mL Penicillin |
| | 100 | µg/mL Streptomycin |
| | 1 mM | Sodium Pyruvate |

**II.** **COS-7:** DMEM (Dulbecco's Modified Eagle's Media) from PAN-Biotech was used for the growth of COS-7 cells

| | | |
|---|---|---|
| **DMEM:** | 4,5 | g/L Glucose |
| **(including)** | 4 mM | L-Glutamine |
| | 1 mM | Sodium Pyruvate |
| | 3,7 | g/L $NaHCO_3$ |

In order to prepare the full growth media, DMEM was supplemented with the following components and sterile filetered and then stored at 4 °C.

| | | |
|---|---|---|
| **DMEM:** | 10 % | (v/v) FBS |
| **(full growth media)** | 2 | mM L-Glutamine |
| | 100 | U/mL Penicillin |
| | 100 | µg/mL Streptomycin |

**III.** **CHO:** Ham's F-12 media from PAN-Biotech was used for the growth of CHO cells.

| | | |
|---|---|---|
| **Hams' F-12:** | 1,176 | g/L $NaHCO_3$ |
| **(including)** | 1 | mM Sodium Pyruvate |

In order to prepare the full growth media, HAM's –F12 media was supplemented with the following components and sterile filtered and then stored at 4ºC.

| | | |
|---|---|---|
| **Hams' F-1:** | 10 % | (v/v) FBS |
| **(full growth media)** | 2 | mM L-Glutamine |
| | 100 | U/mL Penicillin |
| | 100 | µg/mL Streptomycin |

## 2.6 Buffers and Solutions

| | | | |
|---|---|---|---|
| **I.** | **Coomassie Staining Solution:** | 40 % | (v/v) Ethanol |
| | | 10 % | (v/v) Acetic Acid |
| | | 2,5 | g/L Coomassie Brilliant Blue R 250 |
| **II.** | **Coomassie Destaining Solution:** | 40 % | (v/v) Ethanol |
| | | 10 % | (v/v) Acetic Acid |
| **III.** | **Dulbecco's Phosphate-Buffered saline without $Ca^{2+}$ and $Mg^{2+}$ (PBS):** | 0,2 | g/L KCl |
| | | 0,2 | g/L $KH_2PO4$ |
| | | 8 | g/L NaCl |
| | | 2,2 | g/L $Na_2HPO4 \cdot 7H2O$ |
| **IV.** | **EB-Buffer:** | 10 | mM Tris-HCl, pH 8,5 |
| **V.** | **$Ni^{2+}$-NTA-Buffer A:** | 50 | mM Tris/HCl, pH 8,0 |
| | | 300 | mM NaCl |
| **VI.** | **$Ni^{2+}$-NTA-Buffer B:** | 50 | mM Tris/HCl, pH 8,0 |
| | | 300 | mM NaCl |
| **VII.** | **SDS-Gel Running Buffer:** | 25 | mM Tris |
| | | 250 | mM Gylcine |
| | | 0,1 % | (w/v) SDS |
| **VIII.** | **4x SDS- Loading Buffer:** | 50 | mM Tris/HCl, pH 6,8 |
| | | 8 % | (w/v) SDS |
| | | 40 % | (v/v) Glycerol |
| | | 20 % | (v/v) β-Mercaptoethanol |
| | | 5 | mg/L Bromphenolblau |
| **IX.** | **Strep-Tag Buffer W:** | 100 | mM Tris/HCl, pH 8,0 |
| | | 150 | mM NaCl |
| | | 1 | mM EDTA |

| | | | |
|---|---|---|---|
| **X.** | **Strep-Tag Buffer E:** | 100 | mM Tris/HCl, pH 8,0 |
| | | 150 | mM NaCl |
| | | 1 | mM EDTA |
| | | 2,5 | mM D-Desthiobiotin |
| **XI.** | **Western blot** | 3 | g/L Tris |
| | **Running Buffer:** | 14,4 | g/L Glycine |
| | | 150 | mL Methanol (for 1L) |
| **XII.** | **TBST-Buffer:** | 50 | mM Tris/HCl, pH 7,6 |
| | | 150 | mM NaCl |
| | | 0,1 % | (v/v) Tween-20 |
| **XIII.** | **TE-Buffer:** | 10 | mM Tris/HCl pH 8,0 |
| | | 1 | mM EDTA |
| **XIV.** | **Amylose re-suspension** | 20 | mM Tris/HCl, pH 7,4 |
| | **Buffer:** | 200 | mM NaCl |
| | | 1 | mM EDTA |
| **XV.** | **Amylose Elution** | 20 | mM Tris/HCl, pH 7,4 |
| | **Buffer:** | 200 | mM NaCl |
| | | 1 | mM EDTA |
| | | 10 | mM Maltose |
| **XVI.** | **Agarose gel** | 89 | mM Tris |
| | **Running buffer** | 89 | mM Boric Acid |
| | **(TBE):** | 0,02 | mM EDTA (0,5 M) |
| **XVII.** | **Bradford's solution:** | 0,1 | mM Coomassie Brilliant Blue G 250 |
| | | 25 | mL Ethanol |
| | | 50 | mL Phosphoric Acid (85 %) |
| | | For 500 mL solution | |

# 3 METHODS

## 3.1 General Molecular Biology Methods

Standard molecular biology techniques were used for this work. These methods include: plasmid preparation from *E.coli*, DNA-amplification with polymerase chain reaction (PCR), DNA-fragment and plasmid purification with QIAquick-spin colums from Qiagen, ligation of DNA fragments using T4-DNA ligase as well as sequencing of all plasmid constructs from GATC Biotech AG (Constanz). A detailed account of these methods including the transformation of competent *E.coli* cells with plasmid-DNA can be found in the following book of protocols.[137] A list of all of the constructs used in this work is given in **Table 4**.

### 3.1.1 Bacterial Expression Plasmids

The source of various bacterial plasmids used in this work is given below in the order in which they appear in the text.

Construct **2** (pVS07; ST-EGFP-*Npu*DnaE$^{N}$), Construct **4** (pVS18; ST-EGFP-*Npu*$_{C1S}$DnaE$^{N}$) and Construct **5** (pVS01, *Npu*DnaE$^{C}$-Trx-His$_6$) were kindly provided by Vivien Schütz and have been previously reported.[49b]

Construct **8** (pTD060, MBP-*Npu*DnaE$^{N}$-His$_6$): The *Npu*DnaE$^{N}$-His$_6$ sequence was amplified from plasmid pVS24 (kindly provided by Vivien Schütz), ST-Ald Tag-*Npu*DnaE$^{N}$-His$_6$ using primers **oTD048 FP** 5'- ATA GAA TTC TGT TTA AGC TAT GAA ACG GAA ATA TTG -3'and **oTD048 RP** 5'- TAT AAG CTT AGT GGT GGT GGT GGT GGT GCT C -3'. The product obtained was double digested with EcoRI / HindIII and cloned downstream of the MBP sequence of the pSTDuet vector of plasmid pVS22 (kindly provided by Vivien Schütz) encoding MBP-TEV-*Npu*DnaE$^{N}$ (Arg mutant)-His$_6$, resulting in plasmid pTD060.

Construct **9** (pAU07, *Npu*DnaE$^{C}$-EGFP-His$_6$) and Construct **10** (pAU08, ST-gpD-*Npu* DnaE$^{N}$) were obtained from a previous work of Annika Urbanek and have been previously reported.[49b]

Construct **12** (pJZ68, *Npu*DnaE$^{C}$-gpD-His$_6$) was provided by Dr. Joachim Zettler.

Construct **14** (pTD043, ST-PCP-*Npu*DnaE$^{N}$) and Construct **15 (**pTD040, *Npu*DnaE$^{C}$-PCP-$His_6$) were prepared by bachelor student Christina Geraldine Kantzer.[138]

### 3.1.2 Mammalian Expression Plasmids

Construct **1** (pTD038; Igκ-HA-*Npu*DnaE$^{C}$-Trx-$His_6$-myc-TMD-mCherry): The mCherry-N1 encoding gene was PCR amplified from plasmid mCherry-N1 using the following primers: forward primer **oTD032 FP** 5'- ATA GCT AGC ATG GTG AGC AAG GGC GAG GAG G -3'; reverse primer **oTD032 RP** 5'- TAT GCG GCC GCT ACT TGT ACA GCT CGT CC -3'. The product obtained was double digested with NheI /NotI and cloned downstream of the PDGFR trans membrane domain (TMD) encoding sequence of the p-Display vector of plasmid pTD036 (obtained from Dr.Leif Dehmelt) encoding Igκ-HA-SNAP tag-myc-TMD-EGFP-N1, resulting in plasmid pTD037 coding for Igκ-HA-SNAP tag- myc-TMD-mCherry-N1. In a second step, the gene fragment encoding *Npu*DnaE$^{C}$-Trx-$His_6$ was amplified using PCR from bacterial plasmid *Npu*DnaE$^{C}$-Trx-$His_6$ (pVS01) using the following primers: forward primer **oTD018 FP** 5'- ATA GGG CCC AGA TGA TCA AAA TAG CCA CAC GTA AAT -3'; reverse primer **oTD018 RP** 5'- TAT GTC GAC GTA GTG ATG GTG ATG GTG ATG AGA -3'. The PCR product obtained and plasmid pTD037 were digested with ApaI/SalI and ligated to yield the mammalian expression plasmid pTD038.

Construct **3** (pTD056; Igκ-HA-*Npu*DnaE$^{C}_{(C+1S)}$-Trx- $His_6$-myc-TMD-mCherry)**:** The C+1S mutant gene fragment was PCR amplified from bacterial plasmid *Npu*DnaE$^{C}_{(C+1S)}$-Trx-$His_6$ (pVS08) using the following primers: **oTD018 FP** forward primer 5'- ATA GGG CCC AGA TGA TCA AAA TAG CCA CAC GTA AAT -3'; reverse primer **oTD018 RP** 5'- TAT GTC GAC GTA GTG ATG GTG ATG GTG ATG AGA -3'. The resulting sequence and vector plasmid pTD038 were double digested with Apa1/Sal1 and ligated to obtain the construct pTD056.

Construct **6** (pTD058; Igκ-HA-*Npu*DnaE$^{C}$-myc-TMD-mCherry)**:** The gene fragment coding for *Npu*DnaE$^{C}$ was amplified from plasmid pVS01 using primers **oTD047 FP** forward primer 5'- ATA GGG CCC TGA TGA TCA AAA TAG CCA CAC GTA AAT ATT TAG -3' and **oTD047 RP** reverse primer 5'- TAT GTC GAC ATT GAA ACA ATT AGA AGC TAT G -3' and double digested with ApaI/SalI and further ligated into plasmid pTD038 which was also digested with the same enzymes.

Construct **7** (pTD061; Igκ-HA-gpD-*Npu*DnaE$^{C}$-myc-TMD-mCherry)**:** The gene fragment coding for gpD-*Npu*DnaE$^{C}$ was amplified from plasmid pVS05 (ST-gpD-*Npu*DnaE$^{C}$-His$_6$) using primers **oTD050 FP** forward primer 5'- ATA GGG CCC AGG CGA GCA AAG AAA CCT TTA C -3' and **oTD050 RP** reverse primer 5'- TAT GTC GAC ATT GAA ACA ATT AGA AGC TAT GAA GCC -3' and double digested with ApaI/SalI along with vector pTD038 and was ligated into the vector to yield plasmid pTD061.

Construct **11** (pTD020; Igκ-HA-ST-EGFP-*Npu*DnaE$^{N}$-myc-TMD): The gene fragment coding for ST-EGFP-N1-*Npu*DnaE$^{N}$ was amplified from plasmid pVS07 using primers **oTD017 FP** forward primer 5'- ATA GGG CCC AGA TGG CCA GTT GGA GC -3' and **oTD017 RP** reverse primer 5'- TAT GTC GAC TTA ATT CGG CAA ATT ATC AAC CCG -3'. The PCR amplified product obtained and the pDisplay vector were double digested with ApaI/SalI and the insert was ligated in frame into the vector between the HA-tag and the myc-tag sequences resulting in plasmid pTD020.

Construct **13** (pTD057; Igκ-HA-*Npu*DnaE$^{C}$-Trx-His$_6$-myc-Gas1p): The gene fragment coding for the Gas1p sequence was amplified using PCR from mammalian expression plasmid pCDNA 3.1/Zeo Kre-Prp-Gas1p (kindly provided by Prof. Martin Engelhard and Dr. Miria Schumacher), using the following primers: **oTD045 FP** forward primer 5'- ATA GTC GAC TCT TCT TCT TCT TCT TCA GCT TCA TC -3'; **oTD045 RP** reverse primer 5'- TAT GCG GCC GCT TAA ACC AAA GCA AAA CCG AC -3'. The resulting PCR product and plasmid pTD038 were digested with SalI/NotI and ligated to yield the plasmid pTD057.

**Table 4 List of constructs**

| Construct | Sequence |
|---|---|
| 1 | Igκ-HA-Int$^{C}$-Trx-Myc-TM-mCherry |
| 2 | ST-EGFP-Int$^{N}$ |
| 3 | Igκ-HA-Int$^{C}_{(C+1S)}$-Trx-Myc-TM-mCherry |
| 4 | ST-EGFP$_{(C1S)}$-Int$^{N}$ |
| 5 | Int$^{C}$-Trx-His$_6$ |
| 6 | Igκ-HA-Int$^{C}$-Myc-TM-mCherry |
| 7 | Igκ-HA-gpD-Int$^{C}$-Myc-TM-mCherry |
| 8 | MBP-Int$^{N}$-His$_6$ |
| 9 | Int$^{C}$-EGFP-His$_6$ |
| 10 | ST-gpD-Int$^{N}$ |
| 11 | Igκ-HA-ST-EGFP-Int$^{N}$-Myc-TM |
| 12 | Int$^{C}$-gpD-His$_6$ |
| 13 | Igκ-HA-Int$^{C}$-Trx-Myc-GPI |
| 14 | ST-PCP-Int$^{N}$ |
| 15 | Int$^{C}$-PCP-His$_6$ |

## 3.2 Protein Expression and Purification

### 3.2.1 Recombinant Protein Expression in *E.coli*

Recombinant protein production was carried out by transforming the bacterial strain *E.coli* BL21 (DE3) with the desired plasmid. Test expressions were carried out in 4 mL of LB-medium. Samples were collected before and after induction and analyzed on an SDS-Page.

For carrying out preparative scale protein production, 300-800 mL of LB-medium containing the relevant antibiotic in the required concentration (50 µg/mL Kanamycin,100 µg/mL Ampicillin), to maintain selection pressure, was inoculated with an overnight *E.coli* expression culture at a ratio of 1:100. The cells were grown at 37 ºC under 250 rpm conditions to an O.D.$_{600}$ of 0.5-0.7. At this point the temperature was lowered to 28 ºC and protein expression was induced by adding either 0.4 mM IPTG or 0.02-0.2 % Arabinose and

the cells were grown for an additional 4-5 hours. After expression, the cells were harvested by centrifugation (8000 rpm, 20 min, 4 ºC) and the pellet was re-suspended in a small volume of the appropriate buffer and stored at -20 ºC.

### 3.2.2 Cell Lysis

Cell lysis is the release of the expressed proteins from *E.coli.* An EmulsiFlex ®-C5 high pressure homogenizer was used for this purpose. The device was pre-cooled with ice in order to prevent the proteins from being digested. The resuspended cells were thawed and then passed under high pressure through a small opening till a pressure difference of 6.9 MPa was reached. The resulting shear forces thereby destroyed the cell walls and released the proteins present in the cytosol. In order to achieve the most complete disruption, the cell suspension was passed twice through the homogenizer. To separate the cell lysate from the cell debris and insoluble proteins, the suspension was centrifuged for 30 min at 17,000 rpm and 4 °C. The clear supernatant was used directly for the specific affinity chromatography.

### 3.2.3 Protein Purification

The purification of the proteins was carried out by utilizing the N-and/or C-terminal fusion affinity tags. In this work, the affinity tags used were the hexahistidine sequence ($H_6$), the Strep-tag II (ST) and the Maltose Binding protein (MBP). The exact steps were performed based on the information provided with the column material by the manufacturer. For the affinity chromatography on immobilized Nickel ion for proteins fused with a $H_6$-tag, material from Qiagen ($Ni^{2+}$-NTA Superflow) was used. Proteins fused with a Strep-tag II were purified with the Strep-Tactin Sepharose material obtained from iBA. The Amylose resin for purifying proteins fused with a MBP-tag was obtained from NEB.

### 3.2.4 Protein Purification under Denaturing Conditions

Recombinant $Int^C$-fusion proteins from construct **3** (*Npu*$DnaE^C$-Trx-$His_6$) and construct **9**, (*Npu*$DnaE^C$-EGFP-$His_6$) and construct **14** (*Npu*$DnaE^C$-PCP-$His_6$) were known to undergo proteolysis and therefore cell lysis was performed in the presence of buffer containing 8 M urea and subsequent $Ni^{2+}$-NTA purification under denaturing conditions was performed. After protein expression, the cells were harvested by centrifugation (8000 rpm, 20 min, 4 ºC) and the pellet was frozen at -80 °C. The cells were later thawed and re-suspended in $Ni^{2+}$-NTA buffer A containing 8 M urea. Post addition of 8 M urea containing buffer, all steps were

performed under room temperature conditions. Cell lysis was performed and the cells were further harvested by centrifugation at 10,000 rpm, 30 min, R.T. The supernatant was directly used for protein purification using the $His_6$-tag at the C-terminus of the protein. Purified protein was dialysed against PBS buffer containing 10 % glycerol. 2 mM DTT was added to the dialysis buffer unless otherwise mentioned in order to keep the protein in the reduced form.

2 mL of $Ni^{2+}$-NTA agarose material was added to the column and was equilibrated with $Ni^{2+}$-NTA buffer A containing 8 M urea. The clear supernatant was passed through the column once and the flow through was collected. Washing was performed stepwise by initially washing with 5 column volumes of $Ni^{2+}$-NTA buffer A containing 20 mM imidazole and 8 M urea and further a second washing step was performed using 5 column volumes of $Ni^{2+}$-NTA buffer A containing 40 mM imidazole and 8 M urea. Elution was carried out by the addition of 5 column volumes of $Ni^{2+}$-NTA buffer A containing 250 mM imidazole and 8 M urea. Elution fractions were collected and dialysed against PBS buffer containing 10 % glycerol. 2 mM DTT was added to the dialysis buffer unless otherwise mentioned in order to keep the protein in the reduced form.

### 3.2.5 Calculation of Protein Concentration

The concentration of a purified protein was calculated by measuring its absorbance at $\lambda$ = 280 nm in solution. The extinction coefficient was calculated by using the software program Protean (DNAstar). This photo metric calculation was performed in a quartz cuvette using a Ultrospec $^{TM}$ 3100 pro UV/Vis spectrophotometer. The purified protein was further aliquoted and flash frozen under liquid $N_2$ conditions. The aliquots were stored at -80 °C.

### 3.2.6 PCP-Labeling by Sfp PPTase

Labeling of the PCP-domain was carried out in the following manner. To 5-20 µM of the PCP fusion construct dissolved in PBS buffer, 10 mM $MgCl_2$ was added along with 40 equivalents of CoA-derivatized with either tetramethylrhodamine (TMR) or Dylight 488. To this, 0,02 equivalents of the Phosphopantetheinyl transferase (PPTase) Sfp purified from *Bacillus subtilis* was added and the reaction mixture was incubated at 4 °C overnight. Excess of CoA was removed via dialysis against PBS buffer.

### 3.2.7 *In vitro Trans*-Splicing Reaction

Purified *Npu* DnaE intein fusion proteins were added in equimolar concentration ranging from 10-20 µM in PBS buffer (DMEM wherever mentioned) and incubated at 37 °C. To each reaction mixture 2 mM DTT was added from a freshly made stock. At various time points aliquots were collected and the reaction was stopped by adding 4x SDS-loading buffer. For the 0 h sample, 4x SDS- loading buffer was added to the first construct and the mixture was boiled for 5 min and the partner protein was then further added to this sample. The reaction was followed on an SDS-PAGE gel. All reactions were performed at least in duplicate.

In case of reactions under non-reduced conditions DTT was either completely omitted or one partner protein was first reduced with DTT and the excess DTT was removed via dialysis against PBS buffer. In the latter case, protein *trans*-splicing was performed with the second partner construct being non-reduced and in the absence of DTT.

## 3.3 Mammalian Cell Culture Techniques

### 3.3.1 Passaging of Cells

Mouse Neuro-2a cells were cultured in 12 mL of full growth MEM at 37 ºC under 5 % $CO_2$ in a 100 mm cell culture dish and passaged after every 48 h. At this point, when the cells were 70-80 % confluent, the old media was removed and the cells were washed once with 5 mL of prewarmed 1x sterile DPBS. 1 mL of pre-warmed Trypsin/EDTA solution was added drop-wise to the cells and the cells were incubated for 5 min at 37 ºC, 5 % $CO_2$. 6 mL of fresh pre-warmed full growth MEM was added to the cells to de-activate trypsin and the cells were washed thoroughly to detach them from the culture dish. 1 mL from this solution was then plated on 12 mL of full growth MEM on a fresh culture dish and the cells were incubated for a further 48 h.

CHO cells were cultured as described above with HAM's F-12 full growth media and COS-7 cells were cultured in DMEM full growth medium.

### 3.3.2 Transfection

Transfection was carried out using Gene Juice transfection reagent according to the following protocol. $1x10^5$ cells were plated on either a 35 mm glass bottom mat-tek dish or a 6 well-plate one day before transfection. 24 hours post cell plating, a 100 µL transfection mixture was prepared/dish, in an eppendorf cup, containing 1µg DNA, 4 µL gene juice transfection reagent and the remaining of serum free DMEM media. A master mix was prepared in case of several transfections and volumes were adjusted accordingly.

For expression plasmids encoding for proteins which could not be observed by simple fluorescence microscopy, 0,1 µg of mCherry-N1/EGFP-N1 was co-transfected along with 0,9 µg of the plasmid DNA. mCherry-N1/EGFP-N1 was also used in a concentration of 0,1 µg as transfection control and the transfection mixture volume was adjusted accordingly.

### 3.3.3 Fixation

Fixation of cells was either carried out directly after 24 h of transfection with the required construct in case of cells expressing a non-fluorescent fusion protein in order to observe its expression and targeting or post *trans*-splicing with an exogenously added recombinant protein. Fixation was performed by completely removing the growth media and then washing the cells twice with pre-warmed 1x DPBS. 1mL of fixing solution (1 mL 37 % HCHO + 8 ML 1x DPBS) was added to the cells and incubation was carried out for 20 min at 37 °C. Cells were then again washed 3 times with 1x DPBS to remove all traces of formaldehyde and 1mL of 1x DPBS was finally added. Cells were either directly observed with confocal microscopy or stored at 4 °C for later use.

### 3.3.4 Immunostaining

The immunostaining reactions were carried out on cells plated on glass bottom Mat-tek dishes for carrying out confocal microscopy. Post transfection, cells were fixed with 1 mL of fixing solution for 20 min at 37 °C. Washing was carried out 3 times with DPBS and the cells were either permeablized with 1 mL permeablization solution or taken for further processing. The permeablization reaction was carried out for 15 min at room temperature. The permeabilized cells were then washed 3 times with DPBS. In the next step, blocking was carried out using 150 µL of blocking solution and incubating for 1 h at 37 ºC. Blocking solution was replaced with either 250 µL of anti-HA or anti-MBP monoclonal primary antibody, used at a dilution

of 1:500 in antibody dilution solution. Cells were incubated at for 1 h, 37 ºC and then washed 3 times with DPBS. 250 µL of goat anti-mouse secondary antibody tagged with either Alexa-fluor 568 or Alexa- fluor 488 was used at a dilution of 1:1000 in antibody dilution solution and added to the cells and again an incubation of 1h at 37 ºC was carried out. The cell layer was again washed with DPBS 3 times and finally 1 mL of DPBS was added to prevent the cell layer from drying. The cells were either stored at 4 ºC for later use or directly observed using an Olympus fluoview (FV) 1000 confocal laser microscope. GFP excitation maxima occurs at ~508 nm and therefore the 488 nm line of argon laser was chosen for excitation. The 568 nm line of krypton laser was used for exciting mCherry-N1 showing an excitation maxima at ~610 nm. The cells were viewed using the 63x oil immersion objective lens.

### 3.3.5 Protein *Trans*-Splicing on Living Cells

1 x $10^5$ Neuro-2a cells were plated on 35 mm glass bottom mat-tek dishes or 6 well-plates, as required, and incubated at 37 ºC and 5 % $CO_2$. 24 h post plating, when the cells were 70-80 % confluent, they were transfected with the relevant construct and incubated for another 24 h. mCherry-N1 was transfected separately as a negative control for the experiments. For experiments with construct **11**, EGFP-N1 was used as a negative control. Expression and targeting of fluorescently tagged fusion proteins was directly observed by fluorescence microscopy. In case of cells expressing a non-fluorescent fusion protein, (construct **13**) mCherry-N1 was co-transfected as a transfection control.

For the *trans*-splicing experiment a mixture of 5 µM protein, 2 mM DTT and the remaining of serum-free DMEM was made in an eppondorf cup to a final volume of 300 µL/dish or /well. Volumes were adjusted accordingly. 24 h post-transfection, the cells were washed once with

1 mL of serum-free DMEM. The splicing reaction was carried out by incubating the cells with 300 µL of the reaction mixture at 37 ºC for 1 h, unless otherwise stated. After the incubation the cells were washed 2-3 times with serum-free DMEM to remove excess of protein. Fixation was carried out by adding 1 mL of fixing solution as mentioned above under caption 3.3.3**.** The cells were further washed 3 times with DPBS and finally 1 mL of DPBS was added to the cells to prevent the cell layer from drying. The cells were either immediately analysed by confocal microscopy or stored at 4 °C for later analysis.

For *trans*-splicing on the surface of COS-7 cells serum free DMEM was used but for CHO cells the DMEM media was replaced with serum-free HAM's-F12 media.

### 3.3.6 Western Blot Analysis

$1x10^5$ Neuro-2a cells were plated on each well of a 6 well-plate. The cells were incubated for 24 h at 37 ºC and 5 % $CO_2$. After 24 h, they were transfected with the relevant expression plasmid and again incubated for 24 h to allow for protein expression. Transfection and protein expression were confirmed by fluorescence microscopy. Further, *trans*-splicing on the cell surface was performed as mentioned above in section **3.3.5**. Post splicing, the cells were washed 3 times with serum-free medium. Following washing, 100 µL of 2x SDS-loading buffer was added to the cells drop-wise. The cells were thoroughly scraped off from the plates and boiled for 10-15 min. The protein samples were separated on a SDS-PAGE gel and transferred onto a PVDF membrane for western blotting. Western blotting was carried out using the Amersham ECL™ advanced western blotting kit. Detection of protein bands was carried out using relevant antibodies-anti-GFP (1 mg/mL; mouse IgG1, Covance, 1:10,000 dilution), anti-MBP (1 mg/mL; murine IgG2a, Fermentas, 1:25,000 dilution), anti-HA (1 mg/mL; mouse IgG1, Covance, 1:10,000 dilution), anti-cmyc (1:10,000). Anti-mouse antibody (Ig-HRP conjugate, GE healthcare, 1:100.000) was used as the secondary antibody for the western blot procedure in all cases.

# 4 RESULTS

Proteins present on the surfaces of cells are involved in several crucial cellular processes such as signalling, differentiation, growth, adhesion and cell-cell communication. A number of strategies have been recently developed (see introduction 1.6) for the investigation of the structure, function and cellular localization as well as protein-protein interactions of such proteins by enabling protein manipulations without loss of protein function. As the surfaces of cells are freely accessible, labeling them with synthetic probes is an attractive strategy for their functional characterization. In this work, protein *trans*-splicing (PTS) using the naturally split *Npu* DnaE intein was used for modifying cell surface proteins.

A protein of interest (POI) on the cell surface can be modified with virtually any fluorophore/molecule if a PCP-domain is fused as the extein sequence. The PCP-domain can be loaded with a Co-A derivatized molecule and the PTS system will result in a labeled POI.

In addition, in this work a strategy for performing PTS on the surface of quantum dots has also been presented. This would allow for the labeling of any POI with these inorganic nanocrystals.

## 4.1 Proof of Principle Experiments (expression and targeting to the plasma membrane)

For splicing to be performed on the cell surface, the *Npu* DnaE$^{C}$ fragment was required to be presented on the extracellular cell surface. Therefore, expression and secretion of construct **1** to the plasma membrane was the required first experiment.

Proteins are targeted to the various compartments of the cell (e.g. nucleus, plasma membrane, mitochondria etc.) by a mechanism called protein sorting.[1] This targeting information is contained in the protein sequence itself as a signal peptide or a signal sequence usually present at the N-terminus of the protein. In case of the plasma membrane, this signal peptide, which consists of a few amino acids (3-60 AA) acts as a recognition sequence and targets the protein to the secretory pathway. After being synthesized in the cytosol, it is transported into the endoplasmic reticulum where the signal sequence is cleaved off. This protein then gets further processed in the golgi apparatus and finally ends up in vesicles which fuse to the plasma membrane, thereby exposing the protein to the extracellular surface.

In this work, the murine Ig-κ chain leader sequence[139] is fused to the N-terminus of the protein of interest (POI), targeting the protein for secretion. The transmembrane domain of the Platelet Derived Growth Factor receptor (PDGFR-TM)[140] anchors the secreted protein to the plasma membrane for display. Additionally, mCherry-N1 was fused to the C-terminus of the PDGFR-TM in construct **1** (pTD038, Igκ-HA-*Npu*DnaE$^{C}$-Trx-His$_6$-myc-TMD-mCherry), enabling the direct visualization of the expressed protein at the plasma membrane by fluorescence microscopy.

Immunostaining (section 3.3.4) against the HA-tag was additionally performed for confirmation of protein expression and secretion. Alexa fluor 488 tagged secondary antibody was used for detection of the HA-tag. It was clear from the microscopy data obtained (**Figure 26)** that targeting of construct **1** to the cell membrane of transiently transfected mouse N2a cells was successful. Although most of the fluorescence was observed at the membrane, high fluorescence intensity was also detected in the intracellular compartments.

It was noteworthy that cells transfected with construct **1** and permeabilized with Triton X-100 did not show significant amounts of intracellular staining with Alexa-488. This suggests that maybe the hemeagglutinin sequence is not being fully recognized by the anti-HA antibody. It could also result from a cleavage of the independently folding mCherry domain. Another reason could be the degradation of the N-terminal HA-Int$^{C}$ sequence. Nevertheless, the recognition of the Alexa fluor 488 tagged HA-antibody on the plasma membrane was an indication of significant amounts of protein undergoing proper folding in the secretory pathway and being displayed on the plasma membrane.

**A**


**B** **C**

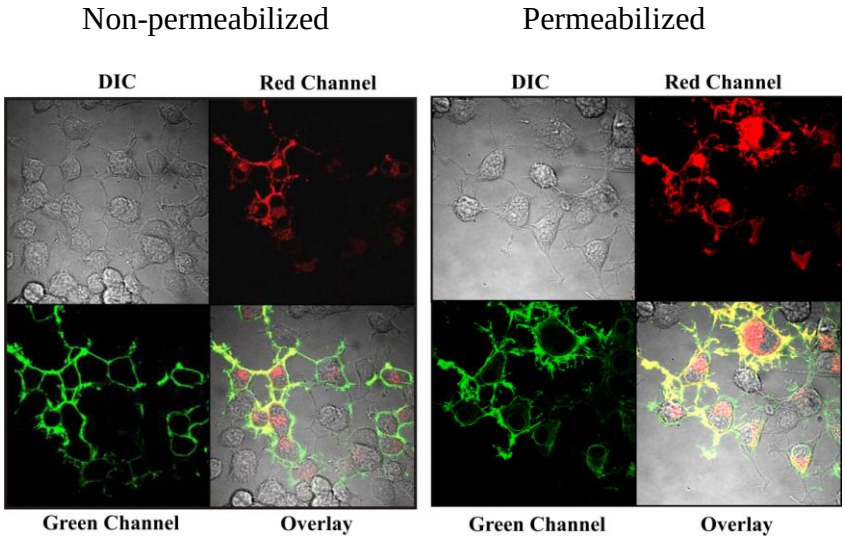



**Figure 26 Sequence and confocal microscopy images of construct 1 (pTD038).** A) DNA sequence representing constructing **1** is shown. Igκ = Murine chain leader sequence, HA = hemagglutinin, $Int^C$ = Intein C-terminal half, Trx = Thioredoxin, Myc = myc epitope, TM = Platelet Derived Growth Factor receptor transmembrane domain, mCh = mCherry. **1** was transfected into neuro-2a cells and the cells were fixed 24 h later. Membrane targeting was confirmed both by mCherry-N1 fluorescence observed at the plasma membrane (red channel) as well as immunostaining against the HA-tag with Alexa-fluor 488 (green channel). B) non-permeabilized cells. C) Cells permeabilized post fixation with 0.25 % Triton X-100.

The next step would be splicing a recombinantly produced *Npu* $DnaE^N$ fusion protein to the above expressed protein.

## 4.2 Protein *Trans*-Splicing on Cell Surface

Immunostaining experiments confirmed the display of construct **1** on the cell surface. Test for splicing activitity would require splicing a target protein/ peptide onto the membrane. EGFP was chosen as the protein of interest due to its stability and its autofluorescent nature which would allow direct observation of the splicing reaction by fluorescence microscopy. This was based on the hypothesis that only cells transfected and expressing construct **1** would also get labeled with EGFP after PTS. The scheme of this reaction is shown in **Figure 27A** and **B**.

Cells were transfected with expression plasmid **1** and the splicing reaction was carried out by adding protein **2** (VS07; ST-EGFP-*Npu*DnaE$^{N}$) under the reaction conditions mentioned earlier. The reaction was carried out in serum free media in order to avoid interference from the components present in FBS. Post washing, cells were observed for splicing activity by confocal fluorescence microscopy. For a complete list of reaction conditions see methods section 3.3.5

EGFP fluorescence was clearly observed on the surface of cells transfected with construct **1** as seen in (**Figure 27C**). The red (mCherry-N1) and the green (EGFP-N1) fluorescence overlap each other to produce the yellow colour. In contrast to this, untransfected cells and cells expressing lower levels of the Int$^{C}$ fusion protein did not show any significant green fluorescence. On cells transfected with only the control plasmid (cytoplasmic mCherry-N1), no green fluorescence was detected (**Figure 27D**). Furthermore, this control experiment indicated a very low unspecific binding of the Int$^{N}$ fragment to the cell surface.

This data indicated a clear interaction between the split intein halves and hinted at PTS. However, split intein fragment association is known to be extremely stable and therefore it was concievable that this interaction could be only associated intein fragments. To rule out this possibility the logical next experiment was immunoblotting to definitely prove splice product formation.

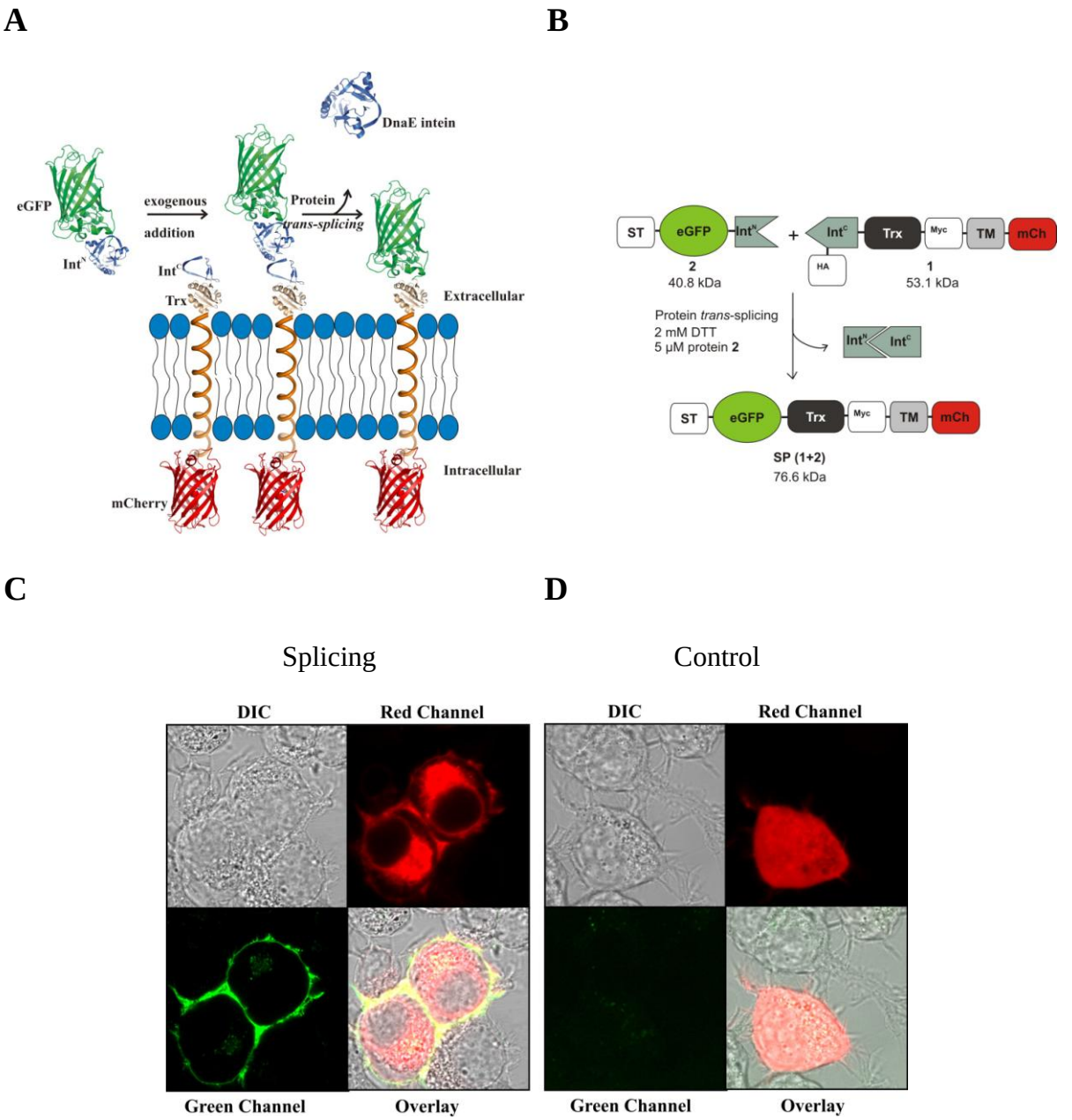



**Figure 27 Protein *trans*-splicing on cell surface.** A) Schematic representation of PTS on cell surface. The $Int^C$ fusion protein is targeted and anchored to the plasma membrane with a transmembrane domain (shown in orange) with the $Int^C$-fragment facing the extracellular environment. The $Int^N$ fusion protein is recombinantly expressed and purified from *E-coli* cells and exogenously added onto the cell surface. Splicing results in the C-extein and N-extein being ligated together by a peptide bond and the intein getting removed. B) Fusion constructs involved in the splicing reaction along with the molecular weight of the constructs and splice product formed. C) Neuro-2a cells were transfected with the relevant expression constructs. 24 h later the cells were washed and protein **2** was added exogenously and the reaction was allowed to continue for 1 h. Post washing, the cells were fixed and visualized by confocal microscopy. Cells transfected with and expressing construct **1** at the plasma membrane and reacted with protein **2**. D) Cells transfected with control construct cytoplasmic mCherry-N1 and reacted with protein **2**. SP = splice product

## 4.3 Confirmation of Splice Product Formation

From the microscopy data obtained in the previous experiment, it was evident that the $Int^N$ and the $Int^C$ fragments indeed associate with each other. This complex has been described as very stable (survives SDS treatment) and therefore could account for the observed co-localization. However, this non-covalent complex dissociates under 5 min boiling conditions.[51] Taking advantage of this fact, a western blot analysis of a cell lysate using the anti-GFP antibody was performed to rule out the possibility of intein association.

Neuro-2a cells were transfected with construct **1** and reacted with protein **2**. Post splicing and washing, the cells were lysed by treating with SDS and boiled for 10 min. The samples were further processed for immunoblot analysis. The immunoblot showed a new band at ~77 kDa corresponding to the calculated weight of the splice product (**SP 1+2**) **(Figure 28).** This was clearly indicative of a covalent bond formation between the extein sequences. This band was absent in the control experiment where the EGFP-$Int^N$ half was reacted with cells expressing only cytoplasmic mCherry-N1. It was also absent from the lysate obtained from cells expressing construct **1** but not reacted with protein **2**.

A second, high intensity band was observed at the height of ~40 kDa. It was believed to be the excess EGFP-$Int^N$ fragment by molecular weight analysis (40.6 kDa) and was only observed in those lanes where the cells had been exposed to the $Int^N$ fragment.

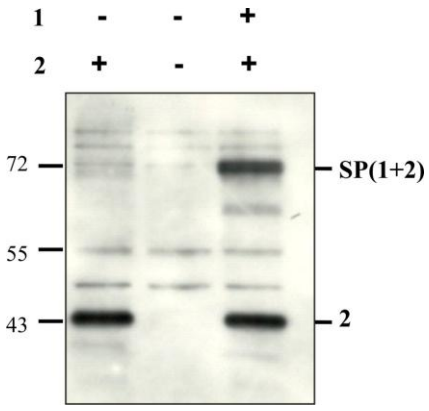



**Figure 28 Anti-GFP western blot for confirmation of splice product formation.** Neuro-2a cells were transfected with construct **1** and after 24 h these cells were reacted with protein **2**. After washing, 2x SDS containing buffer was added for lysis and the cells were further processed for immunoblotting. Control experiments were performed with cells transfected with mCherry-N1, SP = splice product

Thus, formation of splice product indeed occurred on the cell surface indicative of a successful PTS reaction. Further characterizations with splice mutants were required to fully demonstrate PTS on the membrane surface.

## 4.4 Effect of Mutations on Protein *Trans*-Splicing

To fully authenticate the occurrence of a reaction, it is imperative to show absence of or anomalous behaviour with a mutant. This proves that the given reaction is specific and occurs only under wild type conditions. Therefore, to confirm that the data obtained was indeed due to *trans*-splicing, similar splicing experiments were performed on the cell surface with splice mutants.

The 'CFN' tripeptide at the C-terminal splice junction of the DnaE intein corresponds to the first 3 amino acids of the C-extein sequence.[49] In construct **3** the catalytic +1Cys was replaced by Ser. For the *Npu* DnaE intein, this Cys residue is responsible for the nucleophilic attack on the linear thioester to form the branched intermediate (see introduction section 1.2). *In vitro* experiments with this mutant construct had resulted in only marginal splice product formation.[49b]

N2a cells were transfected with the splice mutant **3** (pTD056; Igκ-HA-*Npu*DnaE$^{C}_{(C+1S)}$-Trx-$His_6$-myc-TMD-mCherry) and *trans*-splicing was performed with protein **2** on the cell surface. The scheme of this reaction is shown in **Figure 29A**. Post washing, confocal fluorescence microscopy was performed and the results obtained with this splice mutant are shown in **Figure 29B**.

The expression level of construct **3** in the cells was similar to that obtained for construct **1** and it was processed via the secretory pathway and targeted to the plasma membrane. EGFP green fluorescence after addition of protein **2** was visible on the surface of the transfected cells.

**A** **B**

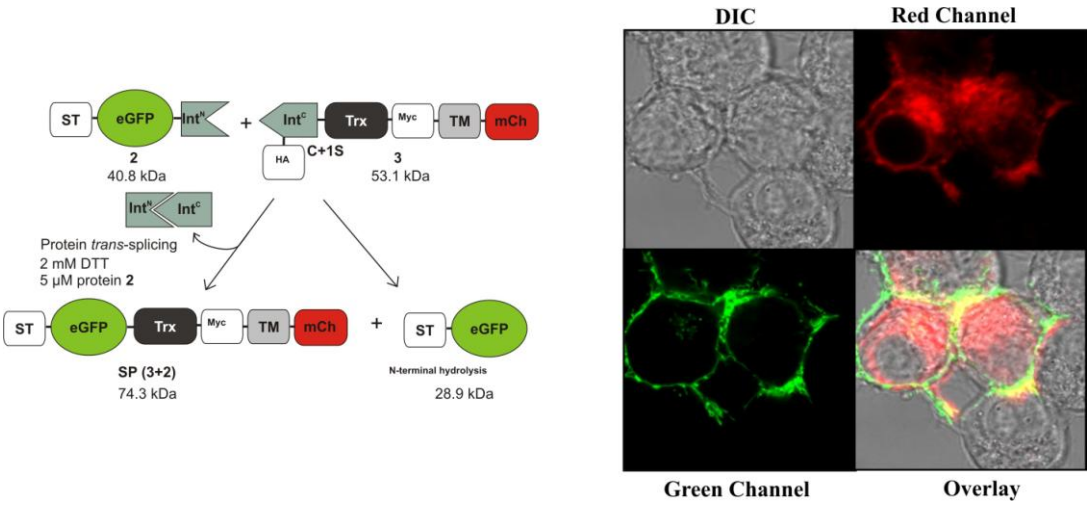



**Figure 29 Protein *trans*-splicing using splice mutant 3.** A) Schematic representation of the protein *trans-splicing* reaction on the cell surface using splice mutant **3**. In this construct the first amino acid of the C-extein, Cys, was mutated to a Ser by site directed mutagenesis. B) Confocal microscopy images of the splice mutant (red channel) and the reaction with protein **2** (green channel). SP = splice product

In another construct, a point mutation was introduced at the N-terminal splice junction where the Cys1 was mutated to Ser (**Figure 30A**). This Cys is the first amino acid of the *Npu* DnaE intein and acts as a nucleophile in the first step of the splicing mechanism by attacking the peptide bond at the $Ext^N$-$Int^N$ junction (see introduction 1.2). It was known from earlier *in vitro* experiments with this intein that a Cys1Ala mutation completely abolished splicing.[49b] Also, similar results were obtained with the *Ssp* DnaE intein where the Cys1 was mutated to Ala.[52b] In this work, Martin *et al*. had also suggested that a non-covalent interaction occurred between the intein halves.[52b]

For cell surface splicing with mutant construct **4**, N2a cells were transfected with expression construct **1**. 24 h later, protein *trans*-splicing was performed on the cell surface with protein **4** (VS18; ST-EGFP-$Npu_{C1S}DnaE^N$) which contained the Cys1Ser mutation at the N-terminal splice junction. The splicing reaction was performed under the reaction conditions described earlier in section 3.3.5. The cells were visualized by confocal microscopy. Green fluorescence was detected on the surface of the transfected cells as seen in **Figure 30B.** Thus, no difference was detected in the fluorescence pattern of the wild type construct and the mutants indicative of intein association in all three cases.

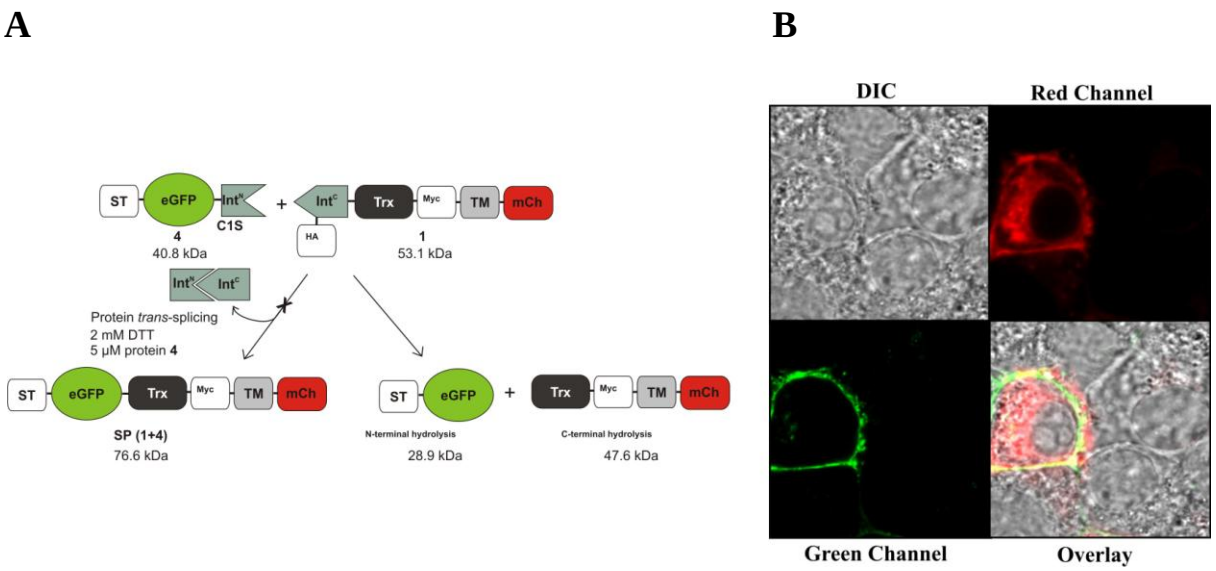



**Figure 30 Protein *trans*-splicing using splice mutant 4.** A) Schematic representation of the protein *trans-splicing* reaction on the cell surface using splice mutant **4**. In this construct the first amino acid at the N-terminus of the intein, Cys, was mutated to Ser by site directed mutagenesis. B) Confocal microscopy images of the splice mutant (red channel) and the reaction with protein **2** (green channel). SP = splice product

Validation of splice product formation however necessitates an immunoblot analysis (**Figure 31**). The immunoblot performed with the anti-GFP antibody showed low yields of splice product formation in the case of the Cys+1Ser splice mutant in comparison to the wild type plasmid and was in accordance with the data obtained for the *in vitro* assay.[49b] This could be due to the lower rate of *trans*-splicing as was reported earlier.[141] An increase in N-terminal hydrolysis does occur, however, since this hydrolysis product (ST-EGFP) would be in the growth media, it was not observed on the western blot of the cell lysates. In contrast to the Cys+1Ser mutant, which showed splice product formation, no product formation was detected for the Cys1Ser mutant. This led to the conclusion that similar to the Cys1Ala mutant, the intein fragments only associated non-covalently but this complex fell apart when treated with SDS containing buffer and boiled at 95 ºC.

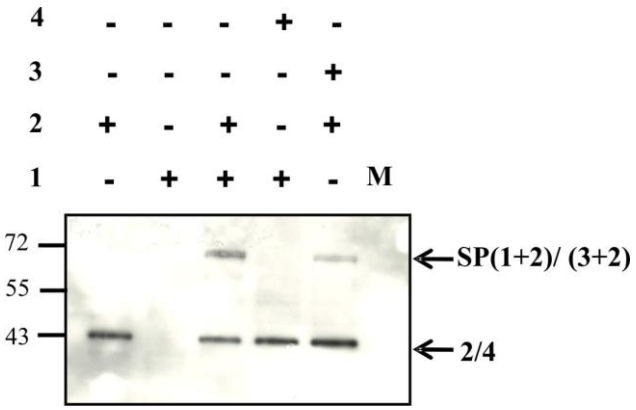



**Figure 31 Anti-GFP western blot showing the effect of various mutations on the splicing reaction.** Cells were transfected with the relevant plasmid and *trans*-splicing was carried out. They were lysed after washing and western blot analysis was performed with the cell lysate. Lane 1: Cells transfected with cytoplasmic mCherry-N1 and reacted with protein **2**, Lane 2: Cells transfected with construct **1** but no splicing reaction was carried out, Lane 3: Positive control with cells transfected with construct **1** and reacted with protein **2**, Lane 4: Cells were transfected with mutant construct **3** and reacted with protein **2**, Lane 5: Cells transfected with construct **1** and reacted with protein **4**. SP = splice product; M = marker

Thus, the data generated above validated the specificity of the PTS reaction on the cell surface. For further characterization of the PTS reaction on the cell surface and for advanced applications of this technique, a series of experiments were additionally carried out as described from section 4.5 onwards.

## 4.5 Monitoring the Cell Surface Splicing over a Time Course

A useful labeling strategy should ideally work with extremely short handling times because it reduces the stress on the living sample as well as increases the time for subsequent experiments. The studies so far using an intein for the extracellular modification of a membrane protein involved artificially split semi-synthetic inteins and reported reaction times of several hours which is unsuitable for most labeling reactions.[90c] The *Npu* DnaE intein was shown to possess the highest reported rate for the PTS reaction with a $t_{1/2}$ of 60 sec.[49b] However, this was demonstrated *in vitro* with purified proteins and might significantly vary when the same reaction is carried out on the surface of living cells in the presence of numerous other cell surface molecules (for, e.g. glycoproteins). Therefore, to assess the speed

of the PTS reaction on cell surface, splice product formation was monitored overt 60 min after the addition of protein **2**. Samples were collected by lysing cells at various intervals with SDS-loading buffer and immunoblot analysis was performed. The cell lysate was blotted both against anti-GFP (**Figure 32A**) and anti-HA (**Figure 32B**) antibodies.

Splice product (**SP 1+2**) formation was observed after an incubation of only 5 min and increased over time. This observation was supported by the anti-HA western blot which showed a gradual decrease in the amount of **1** due to the cleavage of HA-Int$^{C}$ by PTS. These results suggested that a nearly complete conversion of **1** into the splice product takes place.

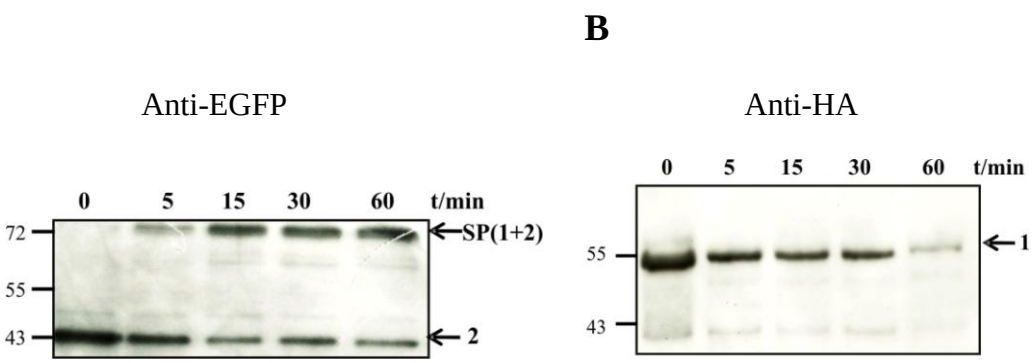



**Figure 32 Time dependent splice product formation on Neuro-2a cell surface.** Neuro-2a cells transfected with construct **1** were reacted with protein **2** for varying lengths of time. After each time point, the cells were washed and lysed by adding 2x SDS containing buffer and further processed for immunblotting. A) Anti-GFP western blot showing increase in splice product formation over time and corresponding decrease of starting material, protein **2**. B) Anti-HA western blot showing loss of the HA-tag during the time course due to the PTS reaction. SP = splice product

Thus, it can be concluded from the above mentioned data that PTS on the cell surface proceeds at a considerably fast rate (slower than with purified proteins) making this technique suitable for labeling of membrane proteins requiring short handling times.

## 4.6 Effect of Reducing Agents on the Splicing Reaction

The first amino acid of the Int$^{N}$-fragment of the *Npu* DnaE intein is a Cys which is essential for the initiation of the splicing reaction and for N-terminal cleavage.[32d] The C-extein Cys+1 acts as a nucleophile during the *trans*-splicing reaction, attacking the linear thioester to form the branched intermediate (see introduction 1.2.2). These catalytically important cysteines are

exposed to the extracellular side of the cell membrane which could lead to oxidation of either or both these cysteines. In order to ensure reduced intein fragments, DTT was added at a concentration of 2 mM to the reaction mixture routinely. However, this may lead to reduction of disulfide bridges of other cell surface proteins, resulting in structural and functional changes. Therefore, it was important to characterize the *trans*-splicing reaction in the absence of this additional DTT.

It was known from *in vitro* assays using purified proteins in splice buffer that the absence of DTT completely abolished protein splicing.[49b] In order to rule out any possible effects of a change in buffer (proteins present as a solution in PBS) or the growth media (used for cell surface splicing), similar assays were performed with purified proteins using either PBS (**Figure 33A** and **B**) or serum free DMEM (**Figure 33C** and **D**). Protein **5** (VS01; *Npu*DnaE$^{C}$-Trx-His$_6$) and protein **2** were added to the reaction mixture at a concentration of 15 µM, either in the presence or absence of 2 mM DTT and the splice reactions were analysed via SDS-PAGE.

In a separate assay, non-reduced protein **5** was reacted with reduced protein **2** in the absence of DTT (**Figure 33E**). It should be noted here that the Int$^{C}$-fusion protein underwent proteolysis within the Int$^{C}$ part upon cell lysis and therefore cell lysis was performed in the presence of buffer containing 8 M urea and followed by $Ni^{2+}$-NTA purification under denaturing conditions.[49b] The protein could be refolded by a simple dialysis step against PBS buffer.

Similar results were obtained for the splice reaction in PBS and DMEM as in the presence of splice buffer dismissing any effects due to the media. Efficient *trans*-splicing was observed in both cases with 2 mM DTT and no *trans*-splicing was detected in the total absence of DTT. The rate at which the splice product formation occurred was slightly lower in the case where DMEM was used as the buffer. However, splice product was observed in the case where protein **2** was reduced and the excess DTT was removed by dialysis and the reaction was carried out in the absence of DTT and non-reduced protein **5**.

Encouraged by the above mentioned results the *trans*-splicing reaction was further performed under the same conditions on the surface of the cells. Neuro-2a cells were transfected with construct **1** and the splicing reaction was carried out as described above.

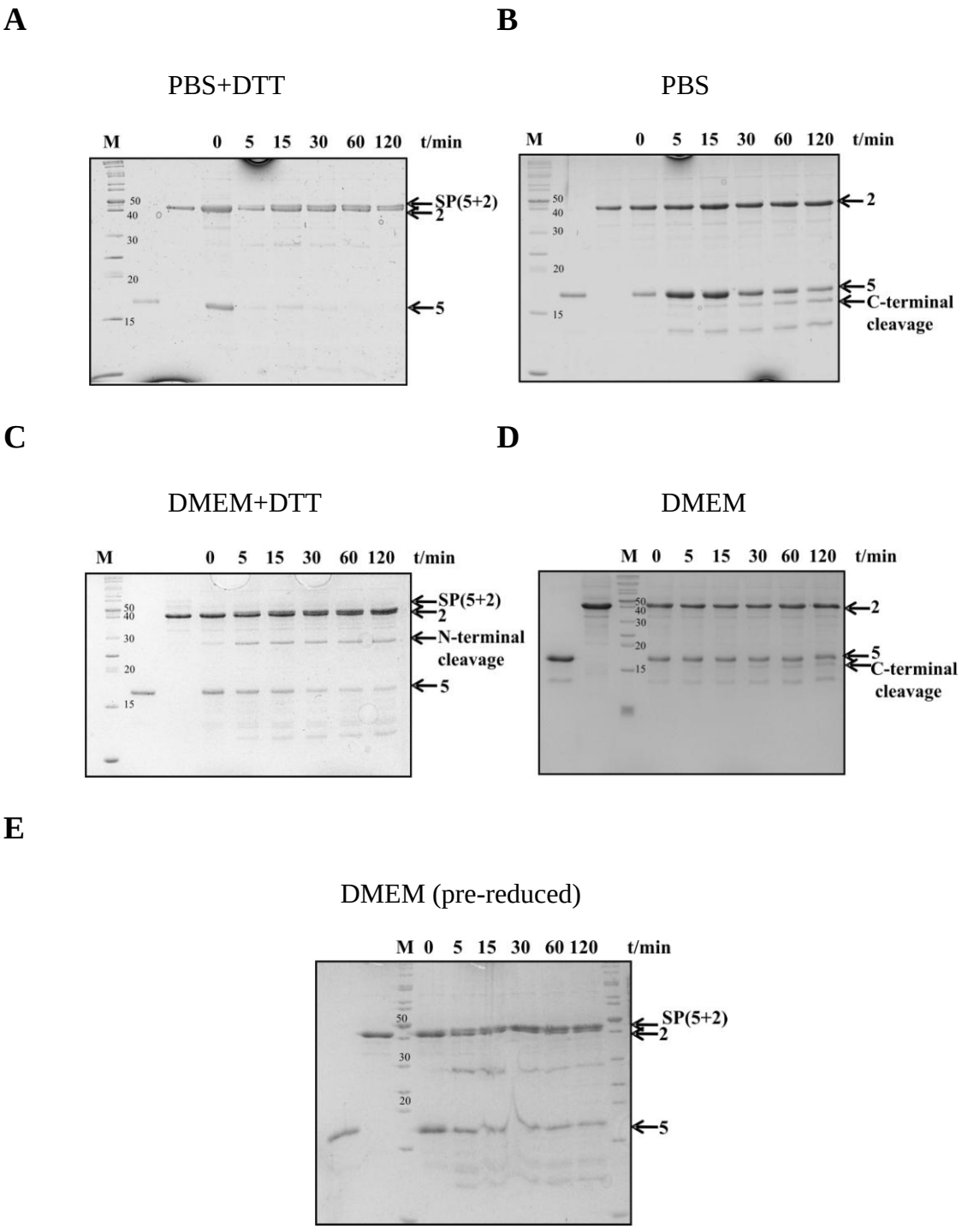
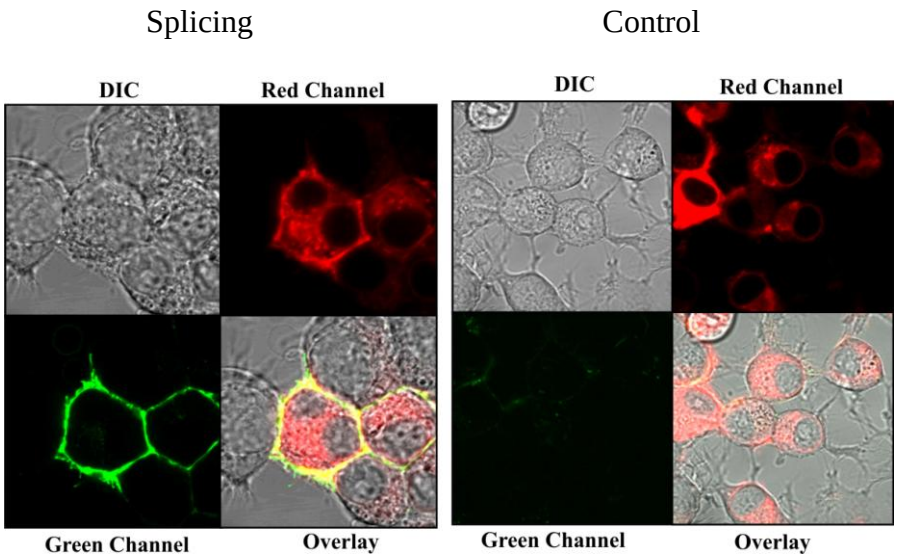



**Figure 33 Effect of various buffer conditions on the splicing reaction.** Protein **2** and Protein **5** were reacted at equimolar concentrations of 15 µM *in vitro* over a time span of 2 h at 37 °C. The reaction was followed by SDS-gel analysis. Samples were taken at various time points by adding 2x SDS sample buffer to stop the reaction. A) PTS in PBS buffer with 2 mM DTT. B) PTS under non-reduced condition with PBS. C) Splicing reaction in DMEM buffer with 2 mM DTT. D) Splicing reaction under non-reduced condition in DMEM. E) Splicing carried out in the absence of DTT but with reduced protein **2**. M = marker; SP = splice product

Microscopy data (**Figure 34A** and **B**) and immunoblot (**Figure 34C**) were consistent with the *in vitro* results obtained with purified proteins. Green fluorescence from the EGFP was clearly visible on the surface of transfected cells reacted with the reduced protein **2** as seen earlier with cells reacted in the presence of DTT. Partly background fluorescence was observed for cells reacted in the total absence of DTT.

**A** **B**




**C**

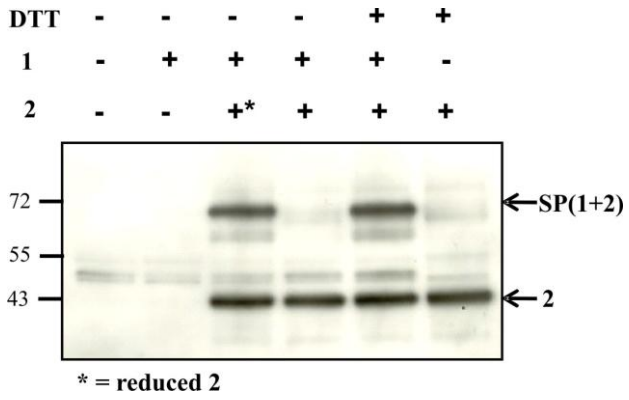



**Figure 34 Effect of DTT on the splicing reaction.** The splicing reaction was carried out on the surface of cells in the absence of DTT. Protein **2** was reduced and the excess DTT was removed by dialysis against PBS buffer and the reaction was performed as earlier. A) Confocal image of neuro-2a cells transfected with **1** and reacted with reduced **2**. B) Cells transfected with **1** and reacted with non-reduced protein **2** in the complete absence of DTT. C) Anti-GFP western blot of the *trans*-splicing reaction. SP = splice product

For the western blot, the band corresponding to the splice product formation was seen in cell lysates where the reaction had either occurred in the presence of 2 mM DTT or with reduced protein **2**. This band was not seen in lysates where DTT was completely removed.

These results indicated that no exposure of the cells to reducing agents was necessary and that the splicing reaction can be performed if exogenously added intein fusion proteins were reduced prior to the assay. This is advantageous as it will prevent reduction of disulfide bonds of other cell surface proteins present. The minimum requirement could be the presence of one partner protein in its reduced form.

## 4.7 Effect of the C-Extein Sequence on Cell Surface Splicing

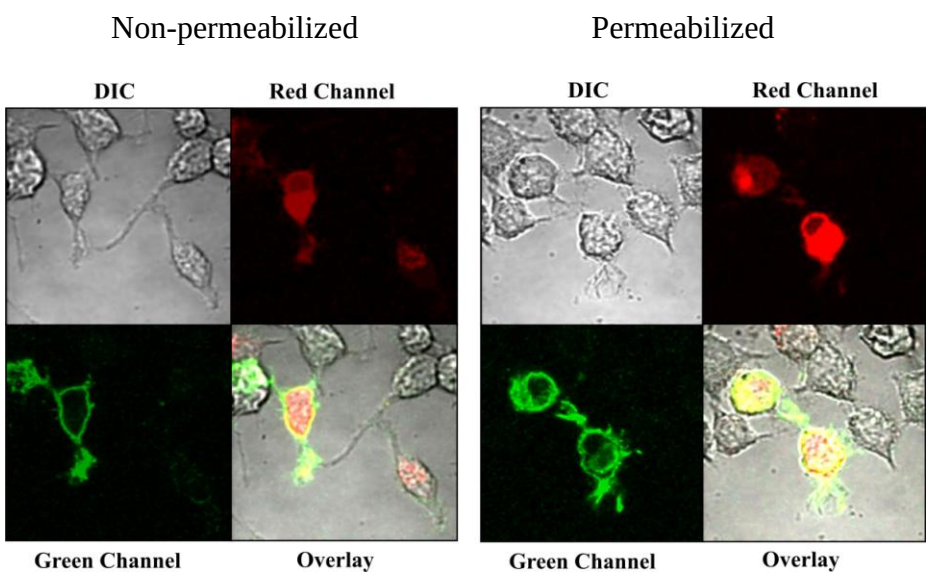
The present work was targeted at directly labeling the backbone of a cell surface protein. In order to achieve that, one needs to be able to go as close to the membrane as possible. A Trx domain in the $Int^C$ fusion construct was introduced as part of an earlier work involving *in vitro* study of the *Npu* DnaE intein.[49b] It enhances protein folding and solubility. Nonetheless, being a part of the C-extein sequence it is incorporated along with the N-extein modification onto the membrane protein. The splicing efficiency was extremely high in this scenario; however the extra Trx was not desirable.

In order to overcome this drawback for future applications, an expression plasmid was designed where the Trx sequence was entirely removed. Construct **6** (pTD058; Igκ-HA-*Npu*$DnaE^C$-myc-TMD-mCherry) was transfected and the expression pattern was observed. Fluorescence microscopy revealed that in contrast to the wild type plasmid, this protein was highly misfolded and subcellular mislocalization occurred **Figure 35A** and **B**. Only small amounts of the expressed protein was localized to the plasma membrane as can be seen from the immunofluorescence images.

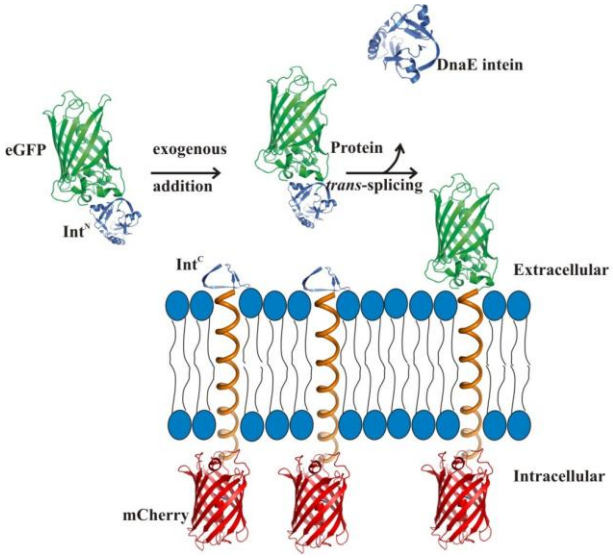
PTS was performed on the surface of neuro-2a cells expressing construct **6**. The scheme of protein *trans*-splicing along with the calculated weights of the reactants and products is shown in **Figure 35C** and **D**. Intensity of the EGFP fluorescence observed after *trans*-splicing with protein **2** on the surface of cells transfected with construct **6** was very low and required an increased laser power (**Figure 35E**).

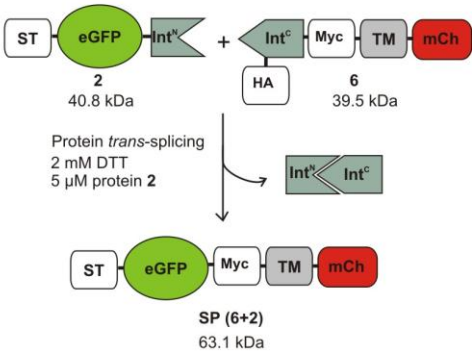
The above results pointed towards the fact that the highly soluble Trx domain might somehow be responsible for the correct expression and targeting of the entire fusion protein and it assisted the $Int^C$ terminal half in proper folding. Also, the distance of the $Int^C$ half from the transmembrane domain/ plasma membrane might be a crucial factor in the protein attaining a properly folded conformation.

**A** **B**




**C** **D**



**E**

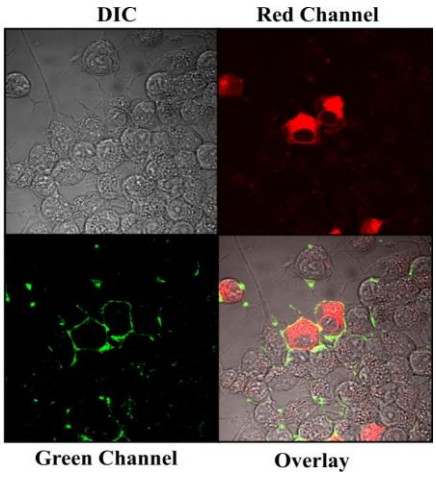


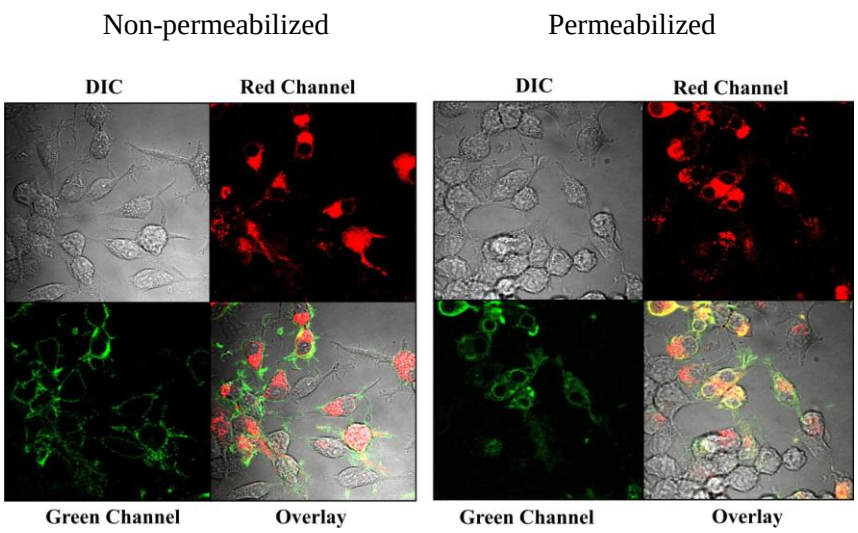

**Figure 35 Protein *Trans*-splicing with the Trx domain removed from the C-terminal fusion protein.** Immunostaining was performed against the HA-tag using Alexa-fluor 488 labeled secondary antibody to observe expression and targeting of construct **6**. Neuro-2a cells were transfected with construct **6** and fixed 24 h later. Cells were either A) non-permeabilized or B) permeabilized with Triton X-100. C) Schematic representation of the *trans*-splicing reaction on cell surface with fusion proteins **6** and **2**. D) Calculated molecular weights of the reactants and products of the PTS reaction. E) Confocal image of cells transfected with construct **6** and reacted with purified protein **2** observed at ~5x higher laser power than earlier in both the red and green channel. SP = splice product

In order to circumvent the solubility issue another small and soluble protein, gpD, (bacteriophage lambda head protein D) was fused to the N-terminus of $Int^C$ in construct **6** to generate expression plasmid **7** (pTD061; Igκ-HA-gpD-*Npu*DnaE$^C$-myc-TMD-mCherry). gpD has been used as a fusion partner for high level expression of soluble heterologous proteins in *E.coli*.[142] Several mammalian proteins have been produced by fusion to its C-terminus. It is thought to mediate optimal translation initiation while reducing inclusion body formation and protein degradation. Being fused N-terminally to the $Int^C$ terminal half in the fusion protein, it will be fully removed during the reaction and will not be included as part of the cell surface display.

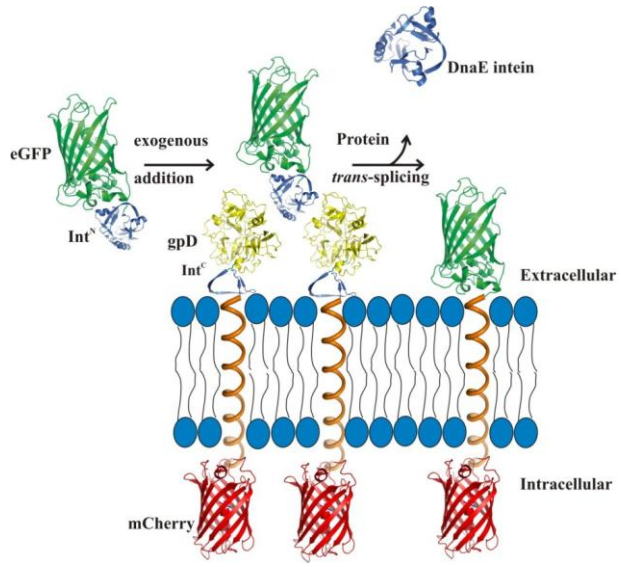
Neuro-2a cells were transfected with expression plasmid **7** and the expression pattern was followed by immunofluorescence and confocal microscopy (**Figure 36A** and **B**). Indeed, fusion of gpD rescued to some extent the misfolding and the mislocalization of the fusion protein. However, in comparison to construct **1**, the intensity of the fluorescence observed on the cell surface was much lower. The majority of the expressed protein was still misfolded

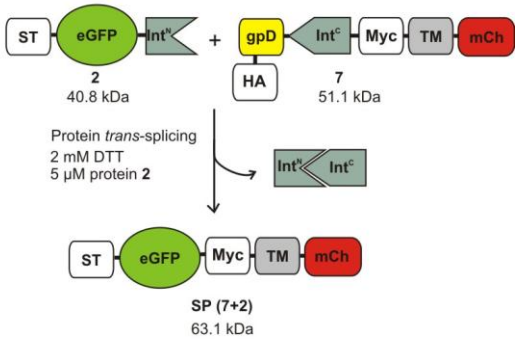
and was correlating to the subcellular compartments involved in the secretory pathway. The schematic representation and the fusion constructs involved in the PTS reaction along with the molecular weights of the reactants and products is illustrated in **Figure 36C** and **D**. Confocal microscopy analysis of the cell surface PTS with protein **2** revealed low intensity EGFP fluorescence requiring high laser power (**Figure 36E**).

**A** **B**




**C** **D**



**E**

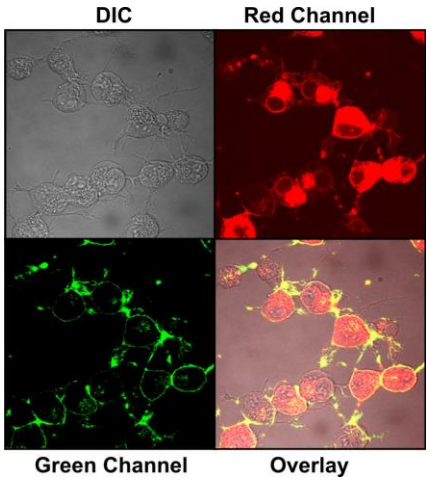
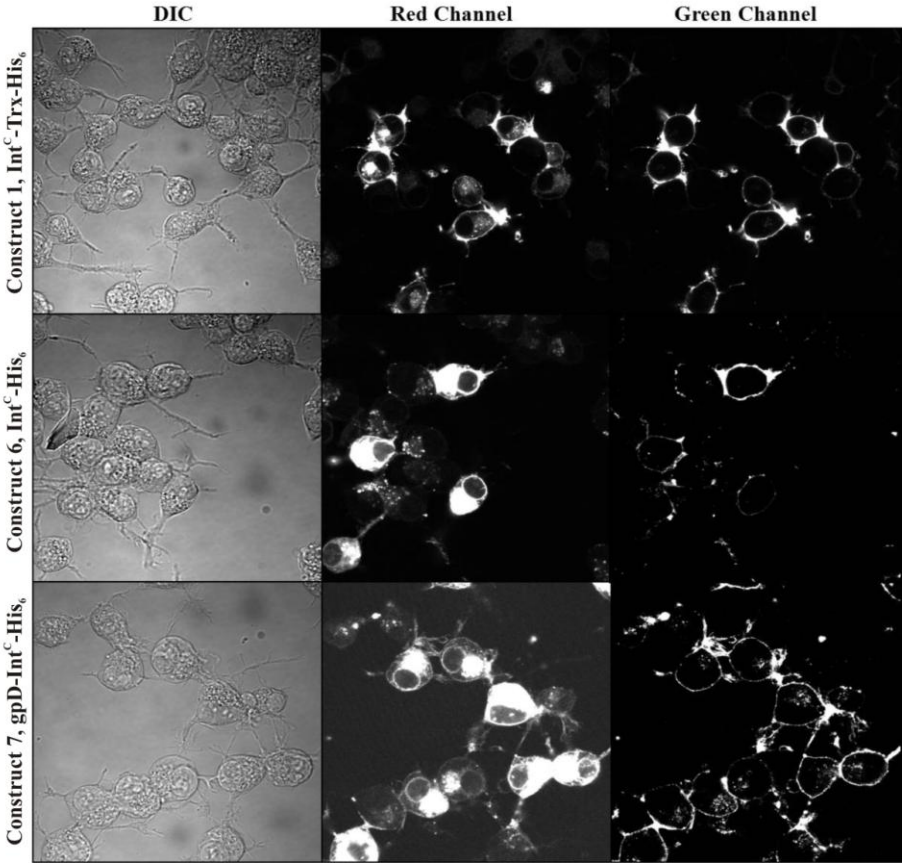


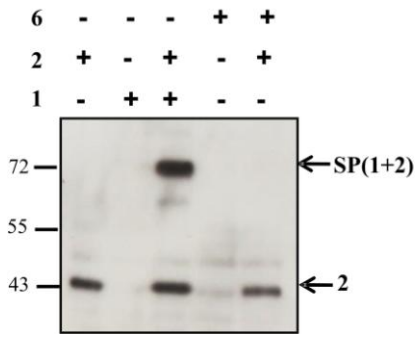

**Figure 36 Protein *Trans*-splicing with gpD fused at the N-terminus of the Int$^{C}$ sequence.** Cells were transfected with construct **7** and fixed 24 h later. Immunostaining was performed against the HA-tag using Alexa-fluor 488 labeled secondary antibody. Cells were either A) non- permeabilized or B) permeabilized with Triton X-100. C) Schematic representation of the *trans*-splicing reaction with construct **7** and protein **2**. D) Calculated molecular weights of the reactants and products of the PTS reaction. E) Confocal image of cells transfected with construct **7** and reacted with purified protein **2** observed at ~5x higher laser power than with wild type construct. SP = splice product

A comparison between the expression patterns of construct **1**, construct **6** and construct **7** is shown in **Figure 37A**. Majority of protein **1** appears to be targeted to the plasma membrane although minor amounts were also visualized in the subcellular compartments. In contrast to this, bulk of protein **6** was aggregated in the intracellular compartments with negligible amounts present on the membrane. The expression pattern of protein **7** indicates a rescue of the protein **6** phenotype. This protein was observed to be targeted to the plasma membrane; however, this rescue was only partial as can be observed from the western blot data in **Figure 37 B** and **C**.

**A**




**B**




**C**

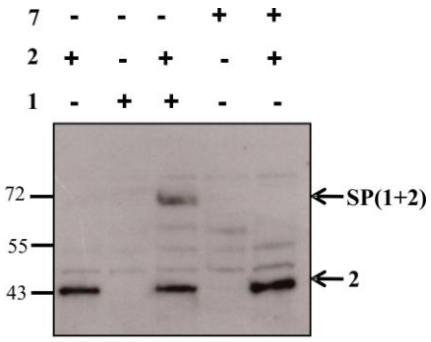



**Figure 37 Comparison of PTS performed with varying Int$^C$-fusion proteins.** A) Confocal images after exogenous addition of protein **2** to cells expressing various Int$^C$-fusion proteins. B) Anti-GFP western blot showing a comparison of the PTS reaction and product formation for construct **1** and construct **6** with protein **2**. C) Anti-GFP western blot showing a comparison of the PTS reaction and product formation for construct **1** and construct **7** with protein **2**. SP = splice product

Western blot analysis was performed with cells transfected with either construct **1**, **6** or **7** and the splicing reaction was carried out with protein **2**. Cells lysates were prepared after addition

of SDS loading buffer for stopping the reaction. In accordance with the above mentioned fluorescence microscopy pattern observed, the western blot analysis showed no splice product formation in case of PTS performed with construct **6** and protein **2**. Also, no splice product band was observed when the PTS was performed with construct **7** and protein **2** even though fluorescence microscopy indicated the targeting of protein **7** to the plasma membrane. This could be due to the low amount of this protein available on the extracellular side of the membrane. Thus, splice product formation might be occurring in the latter case but it may be below detection limit.

## 4.8 Protein Splicing with Decreased Protein Concentration

The $Int^N$-terminal half fusion protein **2** was used at a concentration of 5 µM in our experimental strategy. Inspite of most of the protein being involved in protein splicing, a comparable amount of residual protein was observed in the western blot analysis after protein *trans*-splicing and after washing of the cells as mentioned in section 4.1.3. This residual protein could be accounted for by unspecific binding and was undesirable.

In order to resolve this issue, an attempt at decreasing this residual protein band was performed by improvising upon the wash conditions. PTS was performed by transfecting constructing **1** and adding protein **2** as described earlier. Post splicing, either harder washes were performed with normal wash media or 50 U/ml heparin was added to the wash media before washing steps were performed. Washing cells harder resulted in the cells getting detached from the plate. The cells were then subsequently pelleted by centrifugation and the supernatant was removed. The cells were again re-suspended in wash buffer and a second centrifugation and supernatant removal steps were performed. Finally, the cells were suspended in 100 µl of 2x SDS containing buffer.

A second attempt at reducing this residual protein band was made by lowering the concentration of protein **2** in the splicing reaction. No detectable difference in fluorescence intensity was observed in confocal microscopy analysis when the concentration of protein **2** was lowered from 5 µM to 2 µM (**Figure 38A** and **B**). Green fluorescence was clearly visualized on the cell membrane of neuro-2a cells transfected with construct **1** and reacted with 2 µM protein **2**.

Western blot analysis from the experiments performed above indicated that lower protein concentrations of protein **2** and harder washing indeed decreased the residual band of protein **2** (**Figure 39**). This residual protein band was also observed even in the case where no cells were plated but the $Int^N$ fusion protein was added. This pointed towards the fact that at least a fraction of the residual protein **2** was un-specifically attached to the cells or to the plates. It is thus possible that all of the background observed was a consequence of this unspecific binding, although the possibility that some protein **2** was taken up by the cells by endocytosis could not be ruled out.

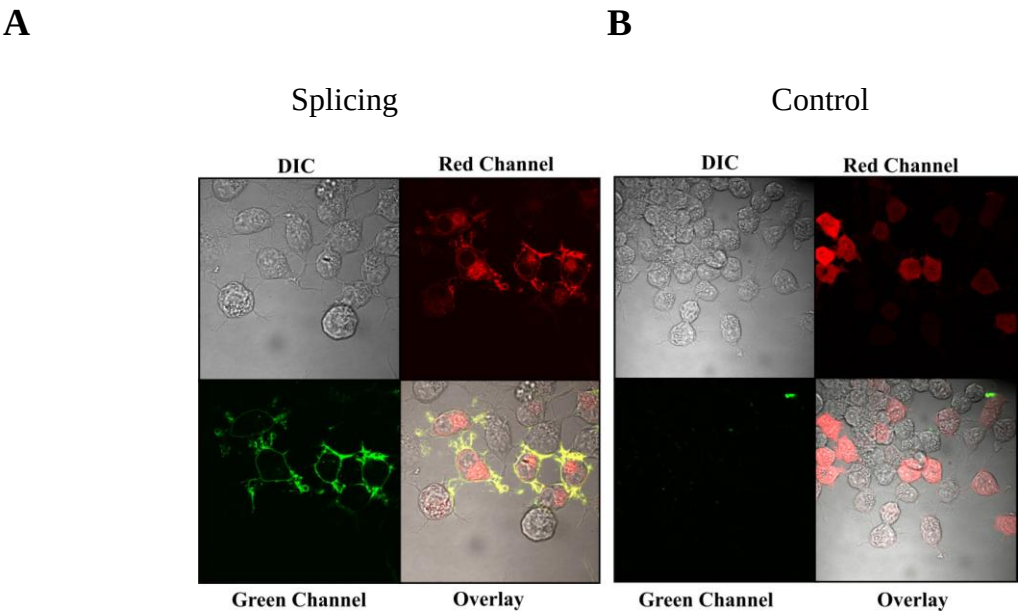



**Figure 38 Confocal images of splicing with a lowered protein concentration.** Cells were transfected with construct **1** and PTS was performed with 2 µM of protein **2** as described earlier in section 3.3.5. Cells were washed and fixed and microscopy analysis was performed. A) Microscopy image of cells transfected with construct **1** and reacted with protein **2**. B) Controls cells transfected with plasmid mCherry-N1 and reacted with protein **2**.

The important result which was obtained from these experiments was that the splicing reaction could be performed with protein **2** at only 2 µM concentration with no observable loss of splice product formation. This is indeed beneficial where yield of the recombinant $Int^N$ fusion proteins is low.

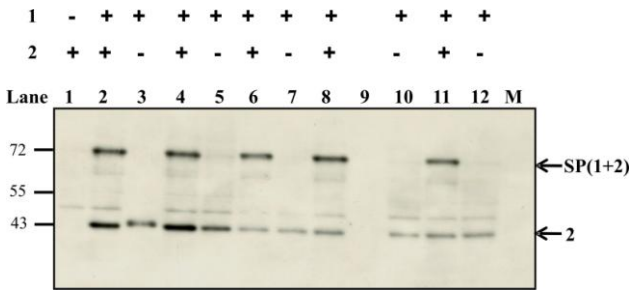



**Figure 39 Anti-GFP western blot showing control experiments for removal of excess protein 2 after protein *trans*-splicing.** These experiments were performed to address the cause for the residual protein **2** that can be seen in the western blot analysis after protein *trans*-splicing and after washing of the cells. For heparin wash conditions, heparin (50 U/ml) was added to the wash media. For harder wash conditions, cells were washed during the second wash until they detached from the plate; cells were then subsequently pelleted by centrifugation and the supernatant removed. Lanes 1 & 2: protein **2** at 5 µM with normal wash conditions; lane 3: no cells with protein **2** at 5 µM with normal wash conditions; lanes 4 & 5: protein **2** at 5 µM with heparin wash conditions; lanes 6 & 7: protein **2** at 5 µM with harder wash conditions; lanes 8 & 10: protein **2** at 2 µM with normal wash conditions; lane 9: blank; lanes 11 & 12: protein 2 at 2 µM with heparin wash conditions. SP = splice product

## 4.9 Protein Splicing on the Surface of COS-7 and CHO Cells

Protein *trans*-splicing could be efficiently performed on the surface of Neuro-2a cells. However, to show the generality of our approach, it was important to show that this labeling technique could be universally performed and was not limited to a particular cell type. For this reason, splice reactions under conditions similar to those performed on Neuro-2a cells mentioned earlier in section 3.5.5 were carried out on the surface of Chinese hamster ovary (CHO) and COS-7 cells, two commonly used laboratory cell lines.

Results obtained from confocal microscopy are shown in **Figure 40A** for CHO cells and **Figure 40B** for COS-7 cells. EGFP green fluorescence was clearly visualized on the surface of cells transfected and expressing protein **1**. Thus, the PTS reaction was not dependent and is not partial to a given cell type and can be efficiently performed on the surface of different cell lines. As the culture conditions for CHO and COS-7 requires a different growth media, the PTS reaction was performed in the serum free media of the specific cell type, i.e. HAM's-F12 for CHO and DMEM for COS-7. Similar results were obtained in each case.

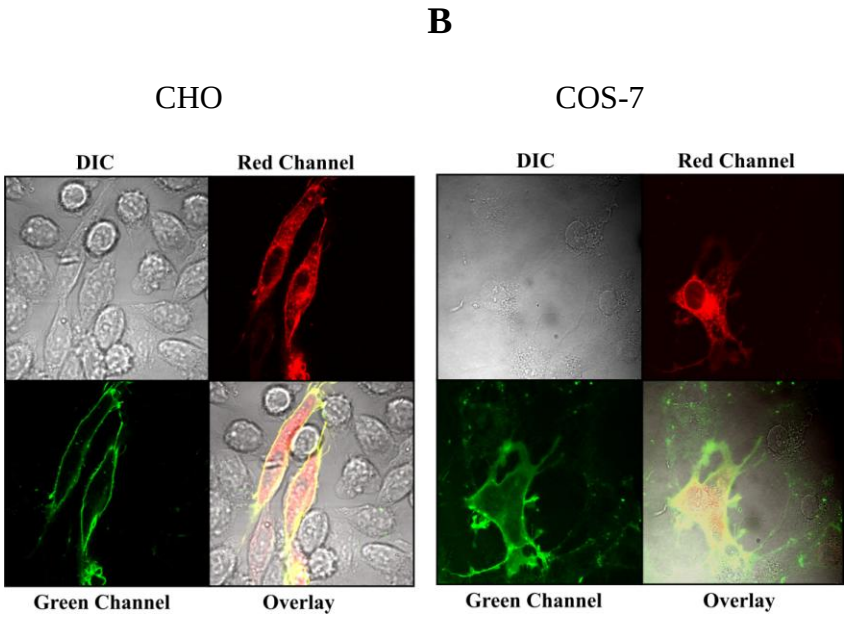



**Figure 40 Protein *trans*-splicing on the surface of various mammalian cells.** Cells were transfected with construct **1** and PTS was performed by addition 5 µM protein **2** as described earlier in section 3.3.5. Confocal microscopy images were taken post washing and fixation. A) Microscopy image of PTS performed on the surface of Chinese hamster ovary (CHO) cells. B) Microscopy image of PTS performed on the surface of COS-7 cells.

Thus, this technique is advantageous as it can be applied to various cell lines and under various media and buffer conditions as per experimental requirement.

## 4.10 Protein Splicing Using Various N-exteins

A good labeling/modification strategy should be amenable to labeling by various functional moieties. This would allow for its diverse applications. The PTS reaction in the experimental strategy mentioned in this work so far involved direct observation of fluorescence post splicing with protein **2** due to EGFP being the N-extein sequence. Therefore, it was imperative to study the PTS on the cell surface by replacing the EGFP sequence with a different POI.

To this end, we constructed and expressed a non-fluorescent recombinant fusion protein replacing the N-extein EGFP with the maltose binding protein (MBP). MBP is used to increase the solubility of recombinant proteins expressed in *E.coli* which assists in purification of the fusion protein. MBP can itself be used as an affinity tag for purification of

recombinant proteins. This fusion protein binds to amylose columns while all other proteins flow through. The MBP protein fusion can be purified by eluting the column with maltose.

Construct **8** (TD060; MBP-*Npu*DnaE$^{N}$-His$_6$) was expressed recombinantly in *E.coli*.and purified via affinity column purification using MBP as the affinity tag. Purified protein **8** (~56 kDa) was obtained at a very high concentration and was diluted to the appropriate concentration before use (**Figure 41A** and **B**). *In vitro* splicing was performed with construct **9** (AU07; *Npu*DnaE$^{C}$-EGFP-His$_6$) as the Int$^{C}$ fusion protein in order to show splicing activity with purified proteins.

Construct **9** has EGFP as its C-extein; however, this recombinant protein is very unstable and undergoes degradation during protein expression in *E.coli.* and could not be obtained as a single purified band even when purified under denaturing conditions. Therefore, degradation bands from this protein purification were also observed on the SDS-PAGE analysis of the *in vitro* splice reaction (**Figure 41C**). Protein splicing was carried out using 5 µM of protein **8** and **10** µm of protein **9** and the splicing proceeded efficiently with the splice product band (~70 kDa) observed after 5 min of incubation at 37 ºC.

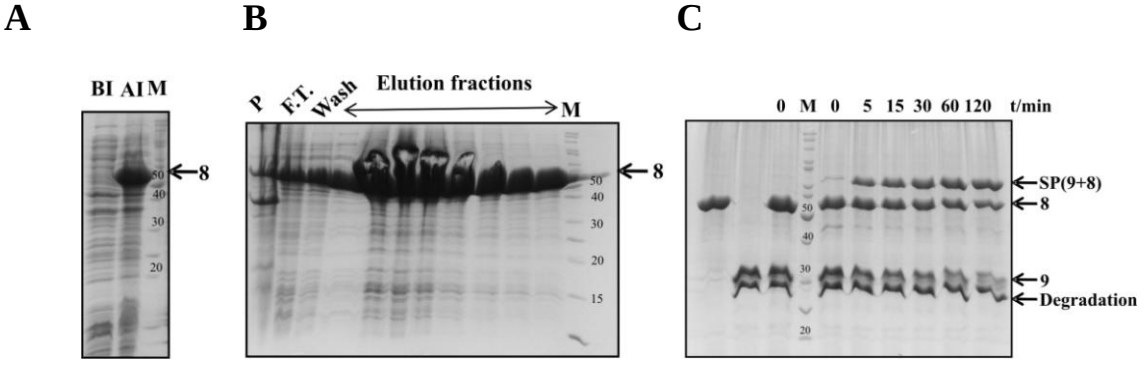



**Figure 41 Characterization of protein 8 with MBP as the N-extein.** A) SDS-PAGE analysis of recombinant protein expression in *E.coli.* Protein expression was induced with 0.4 mM IPTG. B) Purification of protein TD060 via amylose affinity chromatography. Protein was eluted with amylose elution buffer containing 10 mM Maltose. The protein was obtained in pure form and dialysed against PBS buffer. C) *In vitro trans*-splicing of protein **8** with protein **9** followed on an SDS-PAGE gel. Splice product formation at the calculated height of 71.4 kDa was observed after 5 min of incubation. Protein **9** = *Npu*DnaE$^{C}$-EGFP-His$_6$; BI = before induction; AI = after induction; P = pellet; F.T. = flow through; SP = splice product; M = marker

As protein **8** was splice active, PTS was further performed on the cell surface (see **Figure 42A**). Scheme of the PTS reaction between protein **1** and protein **8** is shown in **Figure 42B**. Neuro-2a cells were transfected with expression plasmid **1** and incubated with 5 µM of protein **8** under the same reaction conditions mentioned earlier in section 3.5.5. However, following incubation with the $Int^N$-fusion protein, washing and fixation, direct fluorescence was not possible due to MBP being a non-fluorescent protein. In order to overcome this limitation and based on the understanding that the protein *trans*-splicing reaction would result in the MBP being covalently ligated to the C-extein sequence on the cell surface of Neuro-2a cells, immunofluorescence was performed using an anti-MBP primary antibody as described in the methods section 3.3.4. Alexa-fluor 488 labeled secondary antibody was used for detecting the ligated MBP and confocal microscopy was performed.

It was clear from the microcopy data obtained that MBP was present on the membrane of neuro-2a cells (**Figure 42C**). To confirm splice product formation, western blot analysis was carried out with anti-MBP antibody. The splice product band of the correct size of 90.6 kDa was observed on the blot (**Figure 42D**). However, a second high intensity band was also observed which could be a degradation band from the splice product.

An additional experiment was performed using gpD as the N-extein. Protein **10** (AU08; ST-gpD-*Npu*$DnaE^N$) was recombinantly expressed in *E.coli* purified via strep-tag II affinity chromatography and dialyzed against PBS buffer. The schematic representation of the reaction carried out the cell surface is shown in **Figure 43A**. Neuro-2a cells were transfected with construct **1** and reacted with 5 µM of protein **10**. The reaction scheme with calculated molecular weights of the proteins and splice product is shown in **Figure 43B**.

For a direct comparison of the three constructs **2**, **8** and **10** with EGFP, MBP and gpD as the N-exteins respectively, a western blot analysis was performed using the c-myc antibody. The c-myc tag is present just after the Trx domain in construct **1** and will be retained in the splice product. The results from this western blot are shown in **Figure 44**.

The band observed at ~55 kDa corresponded to expressed protein **1**. The splice product bands were observed at their calculated molecular weights in each case. However, the difference in the intensity of the **SP** band could either be due to the difference in the efficiency of the reaction as a consequence of the N-extein used or due to improper loading of the sample. In support of the former argument, it is known from earlier published data that the overall splice

product yield reduces to 50 % when EGFP was used as the N-extein in comparison to gpD. A second low intensity band was observed below the **SP** band in the reaction with MBP as N-extein hinting at **SP** degradation.

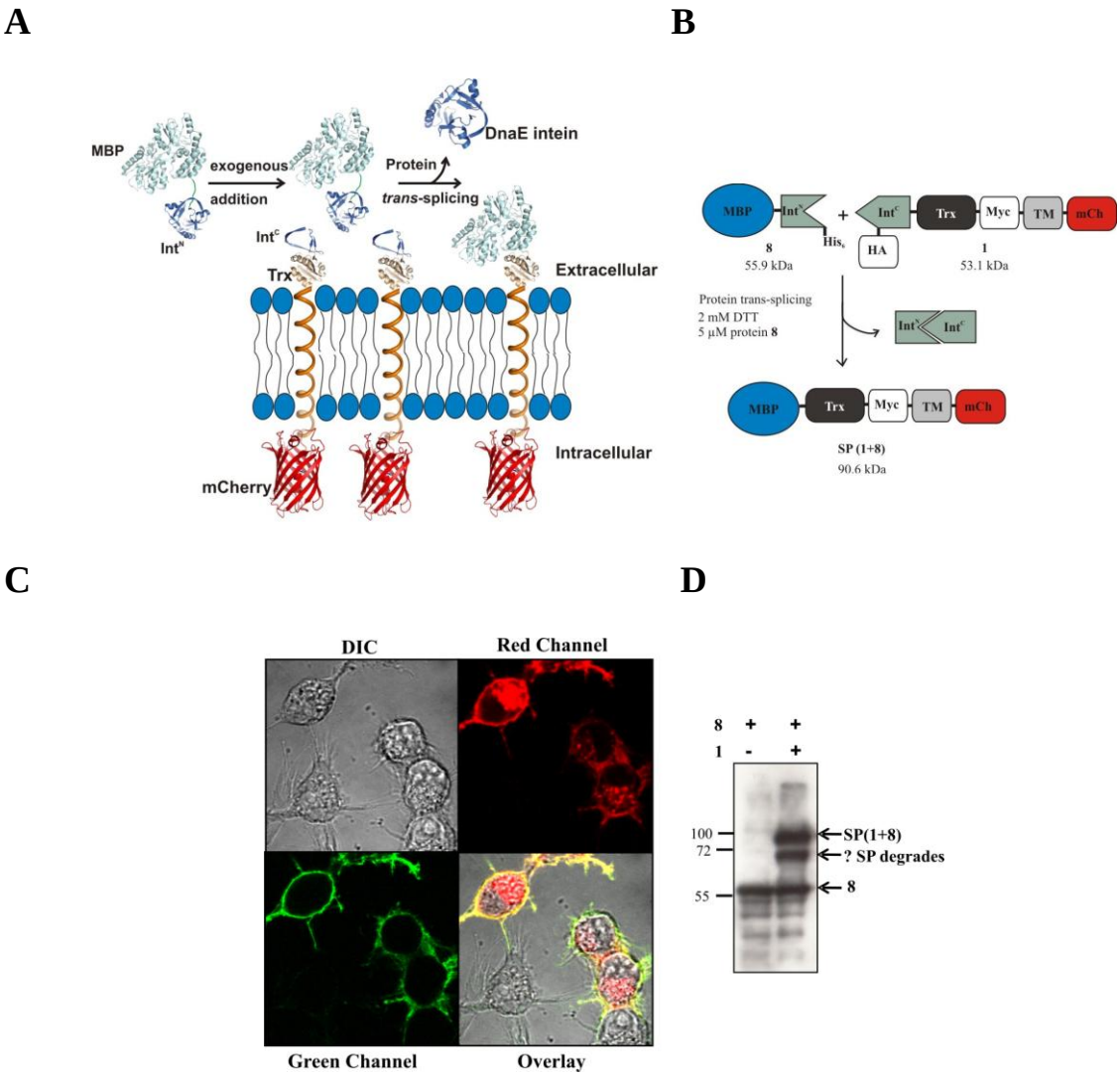



**Figure 42 PTS on the surface of neuro-2a cells with MBP as the N-extein.** A) Schematic representation of the protein *trans*-splicing reaction performed on the surface of neuro-2a cells. B) PTS reaction of protein **1** with protein **8** and splice product formation with their calculated molecular weights. C) Confocal microscopy image of the PTS reaction carried out on the surface of neuro-2a cells. Following PTS, the cells were washed and fixed. Immunofluorescence was performed using mouse anti-MBP antibody. Detection was carried out by Alexa-488 labeled goat-anti mouse antibody. D) Anti-MBP western blot showing bands at the calculated molecular weight of protein **8** at~56 kDa and **SP** at ~91 kDa. The additional band observed might be the result of **SP** degradation. SP = splice product

**A** **B**

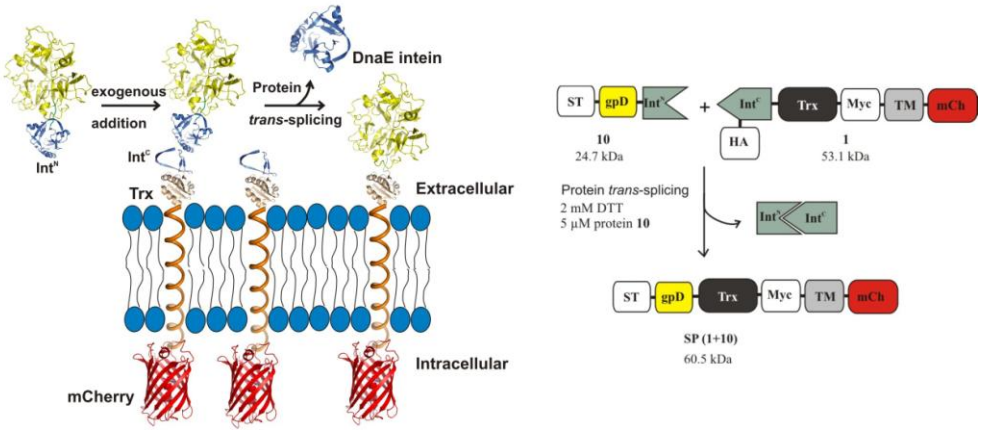



**Figure 43 Schematic representation of the PTS with gpD as the N-extein.** A) Scheme of the reaction on cell surface with protein **10** being exogenously added. B) PTS reaction with the calculated molecular weights of the fusion proteins involved and the expected splice product.

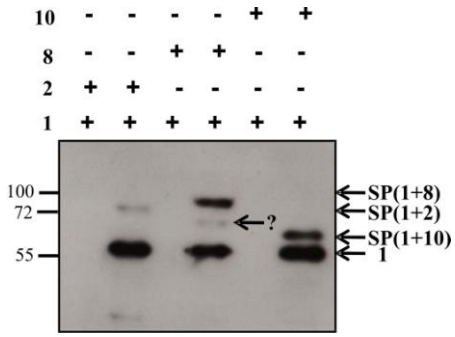



**Figure 44 *Trans*-splicing using various N-exteins.** Neuro-2a cells were transfected with construct **1** and the PTS reaction was carried out using recombinant fusion proteins with varying N-exteins. Post splicing the cells were washed and lysed by the addition of 2x SDS-loading buffer and further processed for analysis by western blot using anti-myc antibody. SP = splice product

Thus, PTS on the cell surface can be efficiently performed with various N-extein sequences thereby allowing a variety of modifications to be performed on membrane proteins.

## 4.11 Expression and Targeting of the Int$^N$ Fusion Protein

In the PTS technique described in this work, the Int$^C$-terminal half of the intein is fused to a POI and expressed in mammalian cells. This Int$^C$-fusion protein is then amenable to protein modification by various N-exteins fused to the Int$^N$-terminal half of the *Npu* DnaE intein.

However, this limits the application of this technique to membrane proteins with their N-terminal sequence on the extracellular surface. In order to widen its usage to also include membrane proteins with their C-terminal sequence on the outside of the plasma membrane, the PTS reaction required a switching of the intein partners.

As proof of principle for the above mentioned approach, EGFP fused to the N-terminus of the $Int^N$-terminal half of the *Npu* DnaE intein was constructed as a cell surface protein (construct **11**; pTD020; Igκ-HA-ST-EGFP-*Npu*$DnaE^N$-myc-TMD). gpD fused to the C-terminus of the $Int^C$ terminal half of the intein when recombinantly expressed in *E.coli* and exogenously added to the cell surface would result in EGFP being removed due to the PTS reaction. However, in this case the splice product will be found in the reaction media.

To this end, construct **11** was transfected and expressed on the surface of neuro-2a cells. The expression pattern of this protein could be directly followed by fluorescence microscopy due to the presence of the EGFP at the N-terminus of the cell surface fusion protein. Immunostaining was performed with a primary antibody against the HA-tag and an Alexa-568 tagged secondary antibody as performed earlier in section 3.3.4 to show targeting to the plasma membrane and the results are shown in **Figure 45A** and **B.** A combination of the direct EGFP fluorescence observed and immunostaining microscopy images point towards the fact that majority of this protein is present in the subcellular compartments and very small amount is actually available on the cell surface. This might be due to an improper folding of the $Int^N$-fusion protein due to the presence of the 102 aa long $Int^N$ domain.

Nevertheless, in the experimental setup for PTS, construct **11** was transfected into neuro-2a cells. 24h post transfection, cells were reacted with protein **12**, washed and fixed for image acquisition. The scheme of this reaction is depicted in **Figure 45C** and **D**. Confocal microscopy analysis did not indicate any significant decrease in the EGFP fluorescence due to the loss of EGFP in the PTS reaction (data not shown). For detecting splice product formation a western blot analysis was also performed as shown in **Figure 45E**. A sample from the supernatant post PTS was taken for identifying the removed EGFP. Corresponding to this removal, a decrease in the intensity of the detected protein **11** band should occur in comparison to the negative control. However, no splice product band was detected on the anti-EGFP western blot. Moreover, no significant decrease of the initial protein **11** band was observed either.

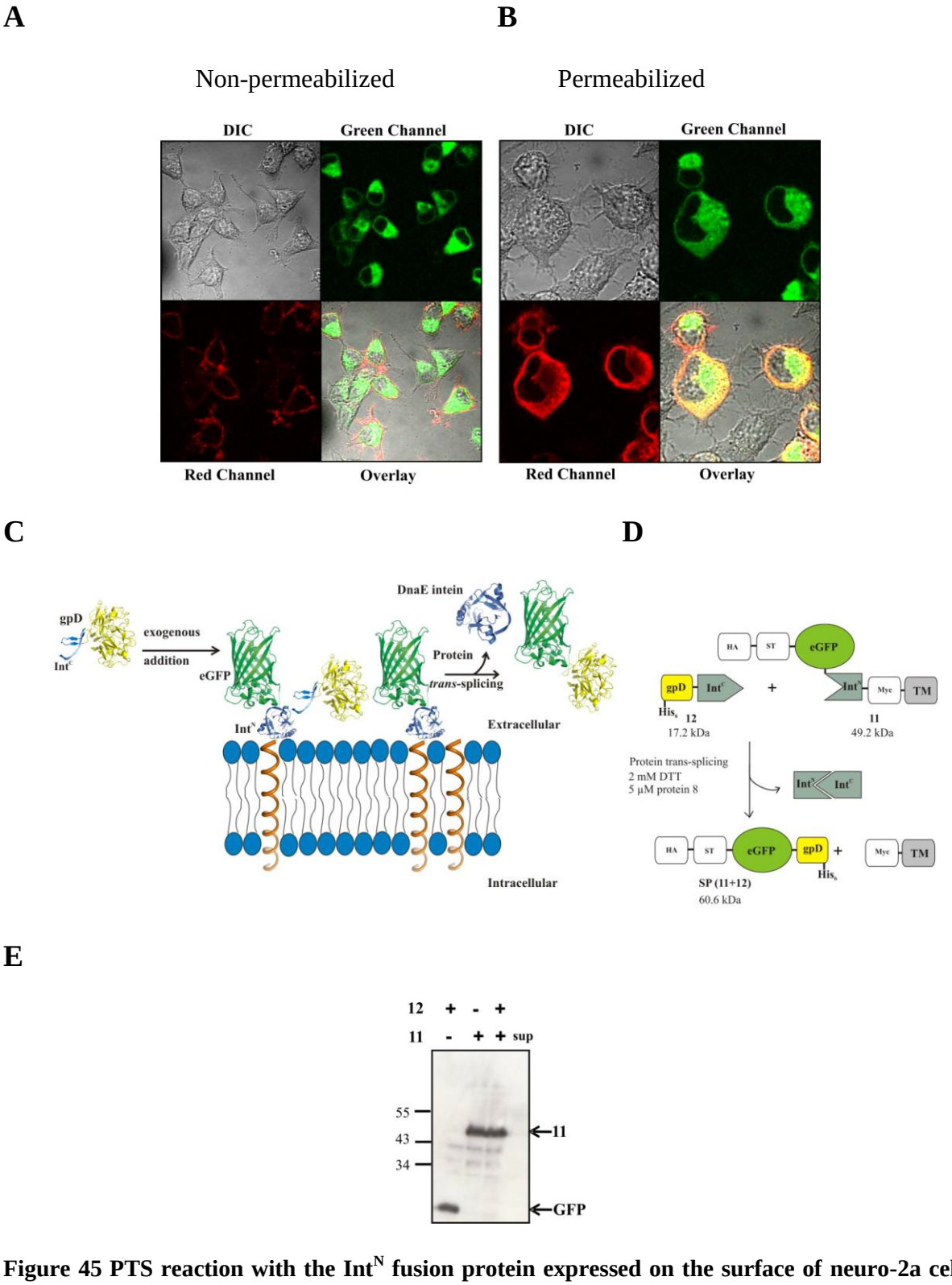



**Figure 45 PTS reaction with the Int$^N$ fusion protein expressed on the surface of neuro-2a cells.** Cells expressing protein **11** were fixed and immunostained with Alexa-fluor 568 tagged antibody against the surface exposed HA-tag. Cells were either A) non-permeabilized or B) permeabilized with Triton X-100. C) Schematic representation of the PTS reaction on the surface of cells. D) Schematic representation of the constructs involved in the PTS reaction and the splice product formed. Calculated molecular weights of the proteins are given. E) Anti-EGFP western blot for the PTS reaction between protein **11** and **12**. SP = splice product

Therefore, the results obtained from these experiments indicate that only a negligible amount of the $Int^N$ fusion protein is actually available at the cell surface for a successful PTS reaction to take place possibly because of the incomplete fold of the $Int^N$ fragment resulting in proteolysis in the secretory pathway.

## 4.12 Protein *Trans*-Splicing Using GPI Anchored Fusion Proteins

GPI anchored proteins are localized to the plasma membrane of eucaryotic cells and carry out several important functions.[143] A GPI anchor is a post translational modification that is added to the C-terminus of proteins and anchors the modified protein in the outer leaflet of the cell membrane. A GPI transamidase catalyzes the transfer of a preformed glycolipid anchor to substrate proteins that possess an appropriate C-terminal GPI signal sequence. This modification occurs in the endoplasmic reticulum and subsequent translocation of the protein to the plasma membrane takes place, where the proteins are attached via the anchor.[144] GPI biosynthesis pathway is conserved from mammals to protozoa and is carried out in the ER. Post translational modification with transamidase complex loads the proteins with synthesized GPI precursors. These proteins bearing the GPI precursors in their carboxyl terminals are transported to the Golgi complex where the GPI moiety is further subjected to modifications and the protein is then finally inserted into the plasma membrane as mature GPI-anchored proteins.[144-145]

The importance of these GPI anchored proteins cannot be underestimated judging by the fact that approximately 0.5 % of proteins are modified with a GPI-anchor and loss or absence of this anchor is usually lethal. Diminished biosynthesis of the GPI anchor due to a genetic mutation leads to haemolytic anemia in humans.[146] Loss of GPI anchoring is embryonic lethal in mammals[143b] and creates conditional lethality in yeast.[147]

Therefore, going by the significance of these GPI-anchored proteins, their modification by protein *trans*-splicing will be a powerful tool for several biotechnological applications. To this end, the transmembrane domain and mCherry sequences from construct **1** were replaced with the GPI anchor-recognition motif of the Gas1p sequence from yeast to yield construct **13** (pTD057; Igκ-HA-*Npu*DnaE$^C$-Trx-$His_6$-myc-Gas1p). The Gas1p is a glycoprotein anchored to the outer leaflet of the plasma membrane through a glycosylphosphatidylinositol and plays a key role in cell wall assembly.[148] This Gas1p protein from *Saccharomyces cerevisiae* is

synthesized as a precursor with a hydrophobic extension at the carboxyl terminus. It is removed and replaced with an inositol containing glycolipid that anchors the protein to the plasma membrane.[149]

Proof of principle experiments required the expression and targeting of construct **13** to the plasma membrane and its precise anchoring with a GPI anchor. The recombinantly expressed intein fusion partner is added exogenously on the cell surface during the PTS reaction. This would result in modification of the POI. As the fusion protein **13** was non-fluorescent, direct observation by fluorescence microscopy was not possible. For this purpose, immunostaining was performed against the HA-tag present C-terminally to the Igκ leader sequence. Cells transfected with construct **13** were fixed 24 h post transfection and reacted with mouse anti-HA primary antibody and detected with Alexa fluor-488 labeled goat anti-mouse secondary antibody. Cytoplasmic mCherry-N1 was co-transfected as a control. Microscopy analysis shown in **Figure 46A** and **Figure 46B** confirmed the expression of protein **13** and assumption of its correct processing and attachment to the membrane with a GPI-anchor was made due to the presence of significant amounts of this protein at the plasma membrane.

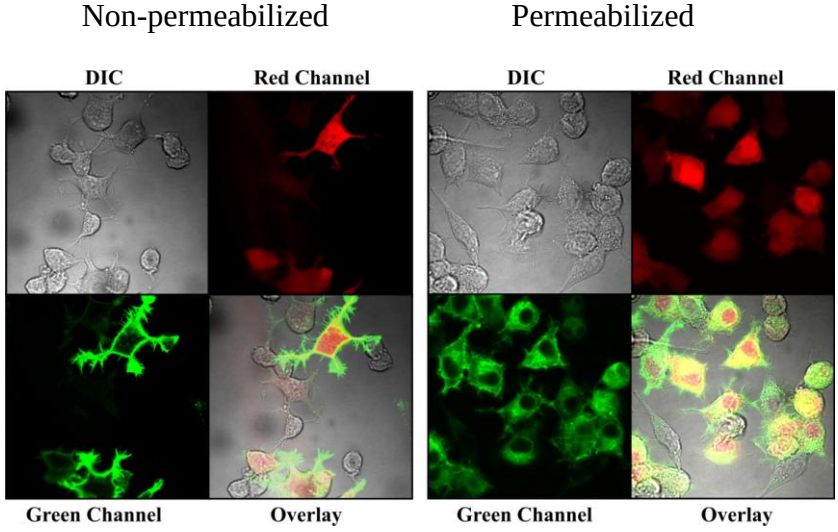
Modification of this GPI-anchored protein was carried out by performing PTS on the surface of neuro-2a cells. Schematic representation of this reaction is shown in **Figure 46C** and **D**. Construct **13** was co-transfected with cytoplasmic mCherry into neuro-2a cells. 24 h later, PTS was performed on the cell surface by adding the $Int^N$ recombinant fusion protein **2**. Visualization of green fluorescence on the surface of transfected cells post washing and fixation indicated the successful completion of the PTS reaction as shown in **Figure 46E**.The splice product formation was further confirmed by western blot analysis performed against the anti-GFP antibody. Splice product band was clearly observed at the calculated height of ~48 kDa as seen in **Figure 46F**.

Presently only a few techniques are known to modify a POI with a natural GPI anchor and most GPI anchor mimics only represent significantly simplified analogs. To get access to a native GPI anchor, Schumacher *et al.* utilized the yeast post-translational protein modification machinery for the production of a protein, which can be used for the C-terminal GPI-modification of any protein of interest.[150] TEV protease cleavage of the expressed and purified protein from yeast cells resulted in a peptide comprising of the C-terminal GPI anchor and N-terminal Cys. They used a covalent capture strategy for purifiying and enriching the GPI-anchored peptide. This modified peptide can further be ligated with a POI

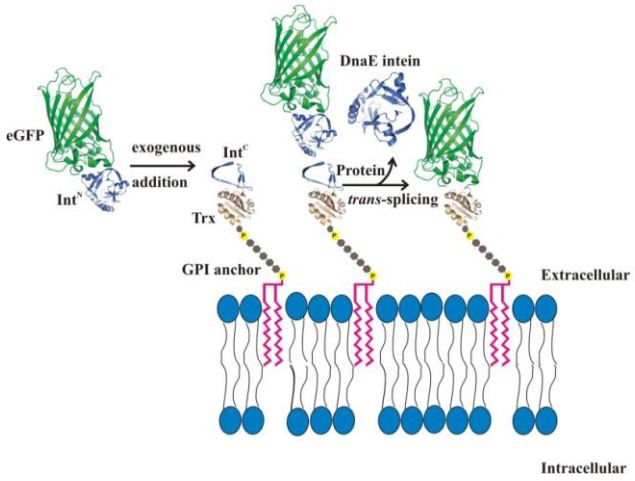
consisting of a C-terminal thioester resulting in a GPI anchored modified protein by using the NCL strategy.

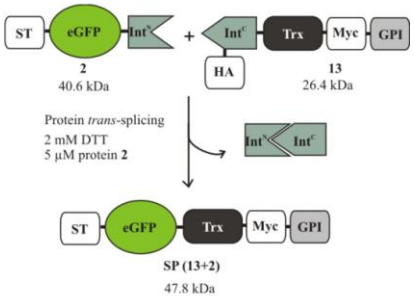
Using the technique presented in this work, one can modify any POI with a native GPI anchor and study its properties. For this puropose, the prion protein was used as the model protein because so far the expression yield and the isolation of the GPI-anchored protein from cells is not efficient enough to get enough protein to study its biochemical behavior. To this end, in collaboration with Dr.Miria Schumacher and Prof.Martin Engelhard, EGFP was replaced with the prion protein sequence Prp and the recombinant protein was isolated from yeast. However, the addition of this protein on the surface of mammalian cell expressing construct **13** resulted in cell death (data not shown). The apparent reason could be the toxicity of the prion protein. Therefore, a different model protein is necessary.

**A** **B**




**C** **D**



**E** **F**

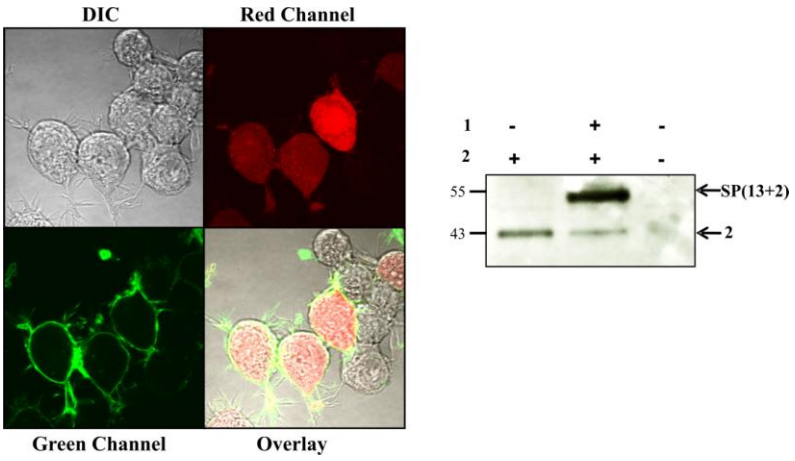



**Figure 46 Expression and targeting of GPI-anchored fusion protein.** Cells expressing protein **13** were fixed and immunostained with Alexa-fluor 488 tagged antibody against the surface exposed HA-tag. Cells were either A) non-permeabilized or B) permeabilized with Triton X-100. Cytoplasmic mCherry was co-transfected as transfection control.Schematic representation of the PTS reaction on the surface of cells with a GPI-anchored POI. C&D) Schematic representation of the PTS reaction between protein **13** and protein **2** with calculated molecular weights of the starting material and splice product. E) Neuro-2a cells were co-transfected with construct **13** and cytoplasmic mCherry-N1. Protein **2** was exogenously added on the cell surface. After 1h cells were washed, fixed and observed via confocal microscopy. F) Anti-EGFP western blot for confirmation of splice product formation at the calculated molecular weight. SP = splice product

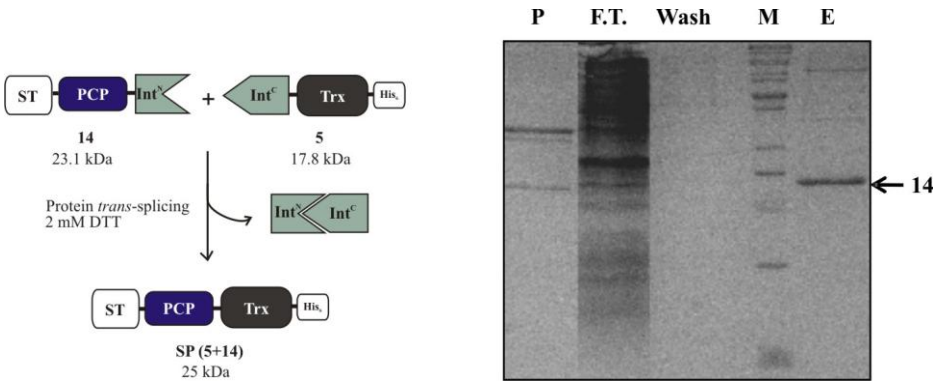
Nevertheless, these results demonstrate that a GPI-anchored protein can be site-specifically and efficiently modified via the protein *trans*-splicing reaction.

## 4.13 PCP-Labeling of Cell Surface Proteins with PTS

Fusion of a PCP domain would enable modification with a variety of organic dyes and small molecules. Labeling of a POI fused to the PCP domain has been earlier reported by various groups.[126] This was based on the promiscuity of the *Bacillus subtilis* phosphopantetheinyl transferase Sfp[151] for loading chemically synthesized peptidyl-CoA substrates onto apo-PCP instead of coenzyme A(CoA-SH).[152]

For the current work, the PCP domain was cloned from the TycA gene from *Bacillus brevis* ATCC 8185.[56] TycA is one of three peptide synthetases of a large enzyme complex involved in the production of the cyclic decapeptide antibiotic tyrocidine produced by via non-ribosomal peptide synthesis (NRPS).[153] The PCP domain was fused to the N-terminus of the $Int^N$-terminal fragment to form construct **14**. This construct was expressed in *E.coli* and purified via affinity chromatography as shown in **Figure 47A**. Prior to performing PTS, the PCP domain was labeled either with CoA-maleimide derivatized Dylight 488 or TMR. The labeling procedure has been described in the methods section 3.2.6. For confirming splice activity, *in vitro* splicing was performed with protein **5** and the reaction was followed on an SDS-PAGE and the results obtained are shown below (**Figure 47B**). Before addition of the coomassie staining solution, the gel was visualized under a UV-transilluminator.

**A**




**B**

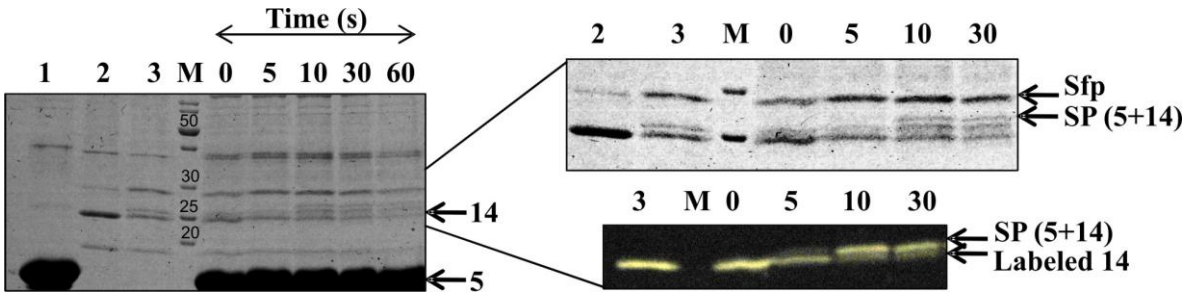



**Figure 47 PCP-labeling strategy for labeling cell surface proteins via PTS.** A) Schematic representation of the *in vitro* splicing reaction with the relevant molecular weights of the reactants and the splice product. Protein purification of construct **14**. B*) In vitro* splicing reaction between protein **5** and labeled protein **14** followed on an SDS-PAGE over time. Lane 1= protein **5**; Lane 2 = protein **14**; Lane 3 = Labeled protein **14**. The right panel shows the magnified form of a part of the gel. Upper half of left panel: Grayscale false color of the gel, Lower half: Fluorescence image of the SDS gel. P = Pellet; F.T. = Flow Through; M = Marker; E = Elution; SP = splice product

Protein yields of construct **14** were very low. Bacterial expression levels and solubility of this fusion protein seemed unaffected by the general conditions used in the lab for protein expression as mentioned earlier in the methods section 3.2.6. Poor quality of the strep tag material was therefore the reason for the low yields of the protein obtained. Furthermore, the labeling reaction did not go to completion as can be seen from the occurrence of two bands at ~25 kDa in lane 3. Nevertheless, splice product formation was observed after 5 min incubation as seen under UV-light. This band was also observed in the coomassie stained gel. Due to the extremely low concentration of protein **14**, the yield of the splice product was also minute. An extra unaccounted for band of ~30 kDa was observed throughout the reaction. Molecular weight analysis revealed this to be the 27 kDa Sfp PPTase. This was expected as the molecular mass of Sfp is greater that the cut of limit of the dialysis tube used.

This experimental section proves that a labeled PCP domain is active in PTS and can further be used for labeling of cell surface proteins.

## 4.14 Detecting Protein-Protein Interactions Using FCCS and PTS

Zamir *et al.* have devised an elegant approach for studying protein-protein interactions in living cells using the FCCS approach.[154] They have used quantum dots as small bait recruiting surfaces that can be microinjected into cells expressing fluorescently tagged prey proteins. Their strategy involved coupling of the quantum dot to recruiting agents such as antibodies or peptides that bind endogenous bait proteins with high specificity and high affinity. However, this would require protein specific expensive antibodies or synthesis of specific peptides. One way of overcoming this drawback would be via the intein based method earlier discussed.

This intriguing concept has been outlined in the figure given below (**Figure 48**). The $Int^C$ terminal fragment is fused at the N-terminus of either a PCP domain[56] or a Cys tag[92a] (C-extein). PCP can further be labeled with biotin-CoA via the Sfp-catalyzed reaction.[155] Biotin can also be conjugated to the Cys tag via maleimide chemistry.[156] This biotinylated fusion protein when incubated with streptavidin coated quantum dots gets labeled with these nanoparticles via the high affinity streptavidin-biotin interaction.[157] The bait protein (N-extein) fused to the $Int^N$ terminal fragment when reacted with the $Int^C$ fusion protein thus gets

labeled with fluorescent quantum dots by the protein *trans*-splicing mechanism. This would allow the *in vitro* coupling of any protein of interest to quantum dots.

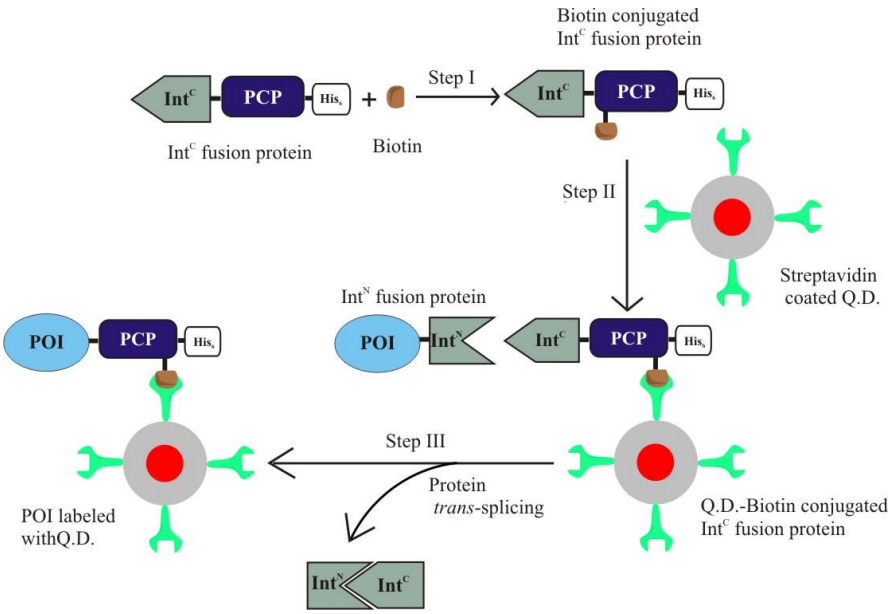
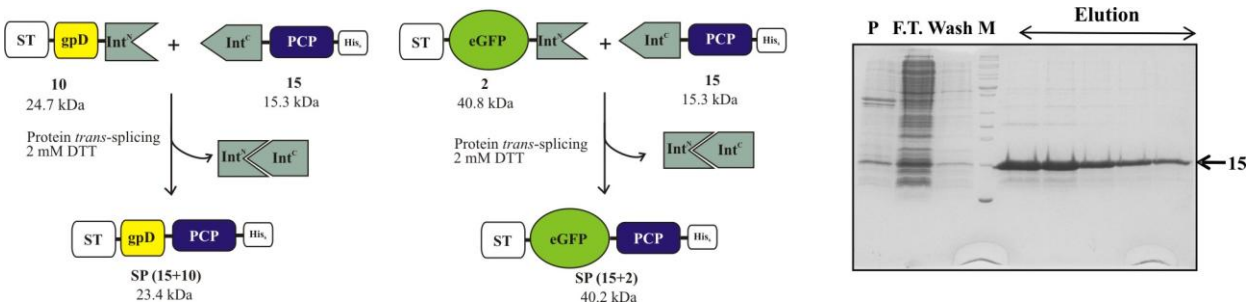



**Figure 48 Schematic representation for labeling a POI with quantum dots for FCCS.** In the first step the $Int^C$-PCP fusion protein is biotinylated at a specific serine residue via the Sfp PPTase method. In the second step streptavidin coated quantum dots are conjugated to the fusion protein by the biotin-streptavidin mediated interaction. The last step involves PTS with a POI fused to the N-terminus of the $Int^N$ terminal half of the split intein. Removal of the intein halves results in a POI labeled with a Q.D.

For proof of principle experiments, PCP fused to the C-terminus of the intein C-terminal fragment (construct **15**) was recombinantly expressed in bacteria and purified. The protein underwent proteolysis within the $Int^C$ part upon cell lysis and therefore cell lysis was performed by the addition of buffer containing 8 M urea (**Figure 49A**). The protein was finally obtained in a pure form in high concentrations and dialyzed against PBS buffer. Test for splicing activity was carried out *in vitro* both with protein **2** and protein **10**. Prior to splicing protein **15** was labeled either with Co-A derivatized TMR or Dylight 488 via the Sfp labeling[155] method described earlier in the methods section 3.2.6. Equimolar concentrations of each protein was added to the reaction mixture and the reaction was followed on an SDS-PAGE. The reaction was carried out in PBS buffer under 2 mM DTT conditions and the results obtained are shown in **Figure 49B**. This work was carried out together with diploma student Christina Kantzer.[138]

**A**




**B**

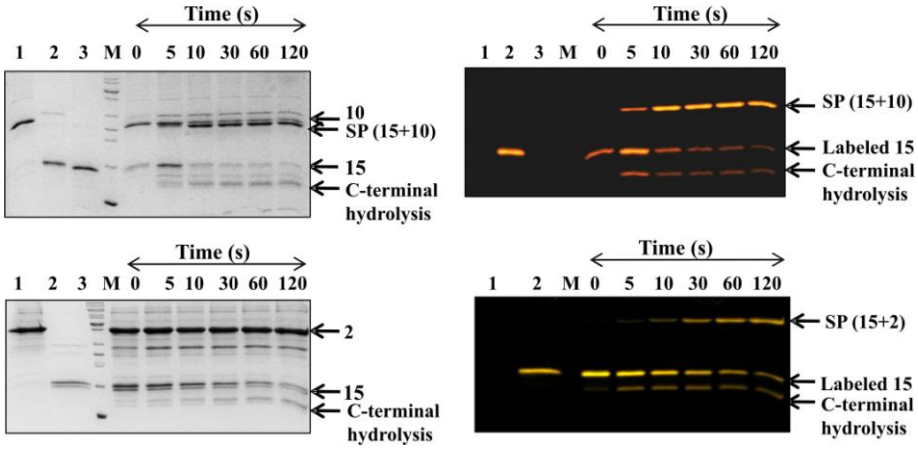



**Figure 49** ***In vitro*** **splice activity test for Int$^{C}$-PCP fusion construct.** A) Schematic representation of the splicing reaction performed *in vitro* using two different N-extein fusion proteins along with the relevant molecular weights of the reactants and splice products. The extreme right image depicts the protein purification of protein **15** under denaturing conditions. B) Upper panel: SDS-PAGE analysis of the splice product reaction. Grayscale and UV images of the PTS reaction of protein **15** with protein **10**.Lane 1: Protein **10**; Lane 2: Labeled protein **15**; Lane 3: Protein **15.** Lower panel: Grayscale and UV images of the PTS reaction of protein **15** with protein **2**. Lane 1: Protein **2**; Lane 2: Labeled protein **15**. Equimolar concentrations of the purified proteins were used in each case.

Protein **15** was obtained in a purified form in high yields. Splice product formation was observed after an incubation period of 5 min with the two different N-extein sequences. However, the yield of the splice product was higher in the case of gpD as the N-extein in comparison to the EGFP. This was in accordance with earlier reported data.[49b] C-terminal cleavage as a side product of the PTS reaction was also observed in each case. The splice product was not clearly visible in the coomassie stained gel due to the small mass difference between protein **2** (40.8 kDa) and the **SP** (40.2 kDa). However, it was clearly observed in the

UV-illuminated gel as only the **SP** will be labeled in this case. An extra band at ~30 kDa was observed in each case which corresponded to the mass of the Sfp PPTase. Thus, a successful PTS reaction could be performed with a PCP labeled protein as the C-extein and this protein can further be used for labeling with other CoA-derivatives. According to the earlier designed strategy (**Figure 48**), the next step would involve labeling of this PCP fusion protein with CoA derivatized with biotin.

Prior to modification of the $Int^C$ construct with biotin, control experiments were required with streptavidin quantum dots QD-655 (red emission) to rule out any unspecific binding of the N-terminal intein fragment. EGFP was used as the bait protein in this scenario as a partner fluorescent protein. 500 nM EGFP-$Int^N$ was incubated with 20 nM of Q.D's at 37 °C for 15 min to 30 min and washes were carried out with PBS with a centrifugal filter column (YM-100) to remove excess/unbound EGFP-$Int^N$ protein and the interaction was followed by FCCS analysis. In a separate experiment, incubation was carried out for the same time period at R.T.

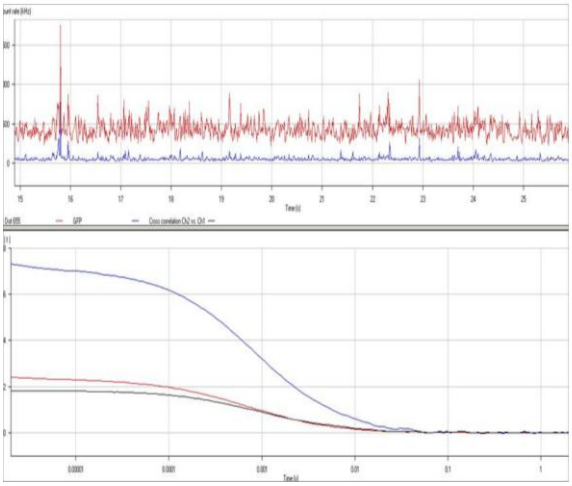
FCCS measurements were carried out on a LSM-510 ConfoCor 3 (Zeiss) using a 40x water immersion objective. Samples were excited with a 488 nm laser line (SW-FCCS) and fluorescence emission was split by a dichroic mirror NFT 565 and detected using a 505-610 nm band-pass and a 655 nm long-pass filter for the green and red channel detection. The line trace of the FCCS data and the correlation curves obtained are shown below (**Figure 50** and **Figure 51**).

A smooth cross-correlation curve with an average cross-correlation value $G(t) > 1$ is considered significant A significant cross-correlation curve was obtained when incubation was performed at 37 °C between $Int^N$-terminal fusion protein and QD 655 [$G(t) < 2$] (**Figure 50B**). The preliminary data obtained showed no cross-correlation [$G(t) \leq 1$] where incubation was carried out at R.T. (**Figure 51**). Only a noise signal was obtained in this scenario. This assumption was made based on the lack of simultaneous fluctuations in the red and the green channel. Only individual auto-correlation curves were obtained in each case. This led us to believe that the reaction maybe temperature dependent and at higher temperatures there is greater unspecific binding taking place between the quantum dots and the split intein fusion protein. A further control experiment with only EGFP reacted with QD's yielded a similar result. A high cross-correlation value [$G(t) > 1$] was obtained when EGFP was incubated with the quantum dots at 37 °C. This was further proof of unspecific binding at higher temperatures

The intensity of the QD is much higher than EGFP as is clearly seen. Quantum dots were incubated at R.T. with increasing concentrations of protein EGFP-$Int^N$ (500nM-2μM) in order to further verify the absence of cross-correlation between the two molecules. The number of EGFP molecules was extremely low as was expected. However, with increasing concentrations of the fusion protein, the number of molecules in the GFP channel was also observed to increase. This could be due to the fact that as similar numbers of washes were performed in each case, larger number of GFP molecules were retained in the column when a higher concentration of the fusion protein was used. Nevertheless, no cross-correlation was observed with the fusion protein even at 2 μM concentration. Further experiments with the $Int^C$ fusion protein will have to be performed. The strategy outlined above can in principle be performed with both artificially and naturally split inteins. However, artificially split inteins aggregate or misfold and may also require higher concentrations of the fusion protein giving rise to false cross-correlation signal. This was apparent from the FCCS analysis performed under similar conditions as described above using the *Mxe* GyrA intein (**Figure 50A**). The *Mxe* GyrA artificially split intein showed high cross-correlation value [G(t) ~ 1.2] at a temperature of 25 °C.

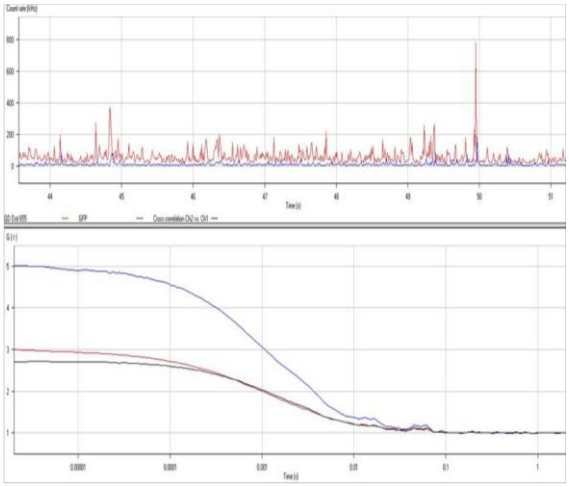
**A**

Q.D. + Mxe *GyrA*$^{N}$ (25 °C)



**B**

Q.D. + *Npu* DnaE$^{N}$ (37 °C)



**C**

Q.D. + GFP (37 °C)

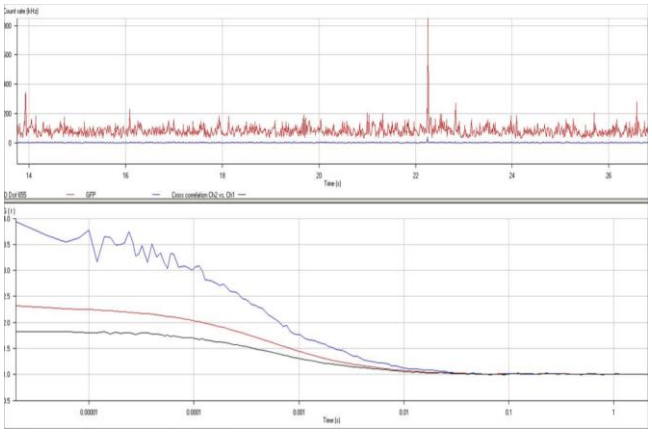
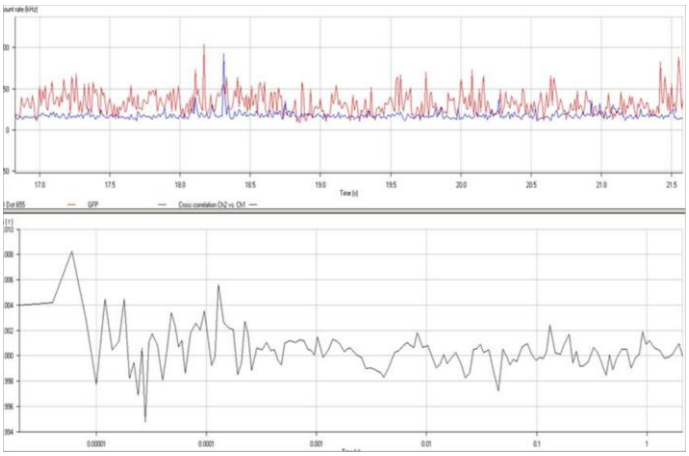


**Figure 50 Control experiment for studying non-specific binding of Q.D.655 to EGFP-Int$^{N}$.** Control experiments were performed to rule out non-specific binding of the EGFP-Int$^{N}$ fragment to Q.D's lacking the Int$^{C}$ fragment. The graph shows the average cross correlation between the two molecules over time. The height of the peak corresponds to the intensity of the fluctuations. Fluctuations arise when a fluorescent molecule passes the small observation volume. Simultaneous fluctuations of both fluorescent molecules determines the degree of binding. Blue curve: EGFP auto-correlation; Red curve: Q.D auto-correlation; Black curve: cross-correlation

**A**

Q.D. + *Npu* DnaE$^{N}$ (R.T.)

Cross-correlation curve



**B**

Auto-correlation curve

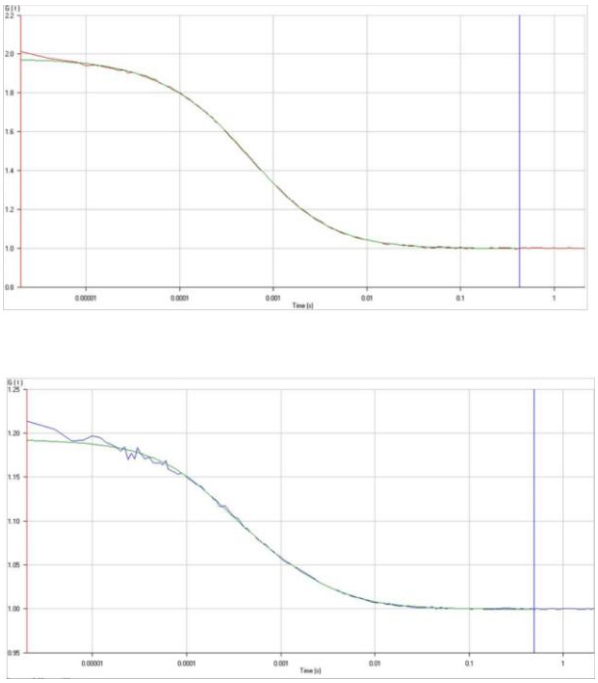


**Figure 51 Control experiment for studying interaction of Q.D. 655 with EGFP-Int$^{N}$ at R.T.** 20 nM Q.D. were reacted with 500 nM of protein EGFP-Int$^{N}$. Washing was performed to remove excess protein. FCCS was performed to assess the degree of bound GFP molecules. A) Cross-correlation curve between Q.D and EGFP-Int$^{N}$. B) Auto-correlation curves of Q.D (Right upper panel) and EGFP-Int$^{N}$ (Right lower panel). Blue curve: EGFP auto-correlation; Red curve: Q.D auto-correlation; Black curve: cross-correlation

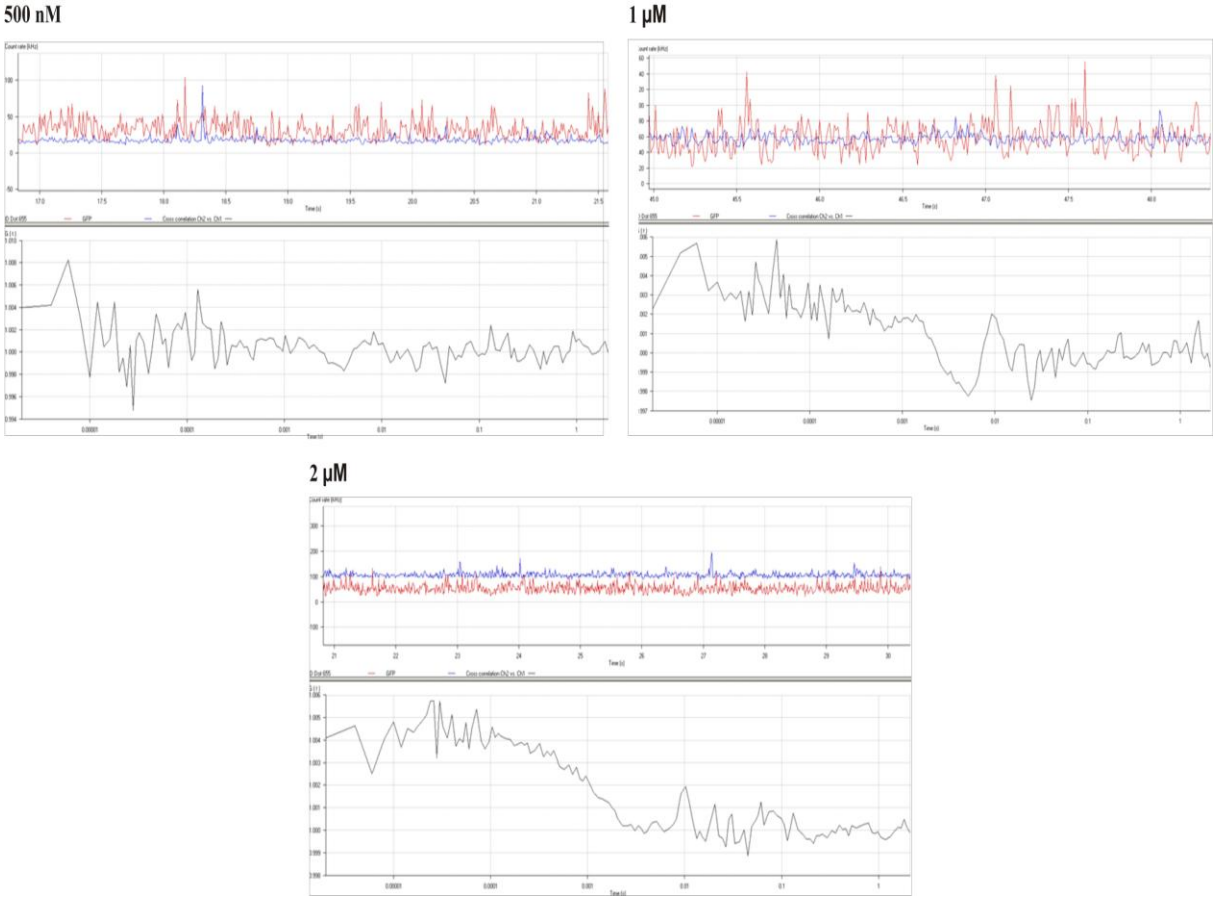



**Figure 52 Effect of increasing EGFP-Int$^N$ concentration.** Q.D. 655 were incubated with increasing concentrations of EGFP-Int$^N$ at R.T. FCCS measurements were carried out using the Zeiss confocor. Count rates *vs* time graphs are plotted here. With increasing concentrations the number of GFP molecules were also found to increase in the focal volume but there was no significant cross-correlation curve was obtained. Blue curve: EGFP auto-correlation; Red curve: Q.D auto-correlation; Black curve: cross-correlation

The technique reported here will not only be useful for studying protein protein interactions but will also enable PTS on the surface of quantum dots resulting in labeling of any POI with these inorganice nanocrystals. Intein mediated site-specific conjugation of QD's to proteins has been reported earlier.[78] The authors synthesized the Int$^C$ terminal half of a split intein via SPPS and performed *in vitro* conjugation of Int$^C$-biotin to streptavidin coated quantum dots. In another case quantum dot nanosensors for detecting protease activity were created by intein-mediated specific bioconjugation and detected by bioluminescence resonance energy transfer (BRET).[158] Quantum dot conjugation to proteins via the split intein mediated approach will additionally add to the repertoire of fluorescent techniques and widen biotechnological applications.

# 5 DISCUSSION and FUTURE OUTLOOK

## 5.1 On Various Labeling Strategies for Cell Surface Protein Modification

Chemical modifications of proteins can be accomplished by various techniques such as total chemical synthesis[159], enzyme mediated reactions[2a], nonsense suppression mutagenesis[160] and a variety of bioconjugation[161] and protein ligation methods.[162] These modifications provide the ability to site-specifically incorporate a range of chemical probes such as fluorescent molecules, photocrosslinkers, unnatural amino acids and are invaluable tools for studying cellular living systems. Consequently, development of strategies that enable the specific modification of proteins in cellular contexts is an area of immense interest in chemical biology. Proteins on the surface of cells have been of particular interest for this purpose because not only do they play a significant role in relaying extracellular signals, in transport etc. but also because their structure and function is affected markedly due to the dynamic nature of the plasma membrane. They play a central role in various biological processes and are important drug targets. Numerous cell surface protein labeling strategies have been developed (see section 1.3) with each approach having its own advantage and limitations and no one approach is suitable for every POI. Therefore, the demand for more tools is ever increasing.

Most of the labeling methods developed so far rely on the development of a tag for the specific labeling by a probe which equips the protein of interest with a new functionality.[2b] They require several other features like high specificity, small tag, no toxicity, no perturbation of the target protein, versatility in the choice of available fluorophores, a short labeling time etc. These features are often incompatible with each other and there is often a trade-off between specificity and size. For e.g. labeling using a hAGT fusion involves the use of a ~20 kDa protein that may compromise biological function of the protein of interest due to its size.[163] In contrast, PTS is traceless and results in the autocatalytic removal of the intein fragments. Chemical modification of the smaller ACP/PCP fusions[126a] or tags for biotin ligase[164] and lipoic acid ligase[122] require the addition of the conjugation enzyme next to the actual synthetic molecule. Furthermore, these methods result only in the attachment of a modified prosthetic group to a particular side chain in the tag and are thus not suitable to directly manipulate the peptide backbone structure of the POI. PTS, however, can be used for

manipulating the structure and primary sequence of membrane proteins. The sortase mediated ligation[130c] method does allow for the backbone engineering of proteins but the reaction requires high concentrations of both the substrates and the enzyme Sortase A. PTS, in contrast, does not necessitate an extra enzyme as the protein *trans*-splicing is a self-processing reaction.[27b, 165]

Most of the techniques for cell surface labeling developed so far are very specific for labeling membrane proteins which is desirable in most instances. Their applicability to internal proteins is however restricted due to several reasons. One major cause for this is the difference in the extra-(oxidising) and intra-(reducing) cellular environment. The tetracysteine system[105], for e.g., requires a reducing environment. PTS with known naturally split inteins progresses efficiently in an intracellular environment due to the presence of catalytic cysteines.[166] This might be a drawback for labeling proteins in the extracellular region as it will lead to oxidation of the conserved cysteines. This might be overcome by using a different intein lacking a catalytic cysteine residue at the +1 position, e.g. the *Mxe* GyrA[42] and *Tli*-Pol2 inteins[167] have a +1Thr at this position and the *Ssp* GyrB intein[5b] and Psp GBD Pol[86] have a Ser+1. However, applications of these to the cell surface labeling method discussed in this work would require artificial splitting which might further lead to insolubility problems. Moreover, splicing activity may be considerably affected.

PTS in an intracellular environment would be unlikely in the absence of both the intein halves. However, techniques like the tetracysteine system[105], the Texas Red binding peptide system[168] suffer from background labeling. Other techniques like the biotin ligase/lipoic acid ligase, ACP/PCP labeling are restricted to labeling proteins present only on the cell surface due to the requirements of the labeling strategy. Moreover, the biotin ligase method requires millimolar concentrations of the probes to achieve significant labeling. PTS on the other hand has been demonstrated to occur at low micromolar concentrations for labeling both intra-[166] and extra-cellular[4] proteins.

Since the majority of the techniques mentioned here require genetic fusions to peptide/proteins, their expression, folding, localization or functions may alter. Large fusion partners, that possess localization signals and highly charged fusion partners often alter the function of the target protein. A flexible linker region may reduce chances of altering the functions of a POI. The tetracysteine and biotin ligase system are preferred in this regard.

Therefore, it is evident that different strategies are required for labeling different cell surface proteins and no unique system is applicable to them.

## 5.2 On PTS as a Suitable Method for Labeling Cell Surface Proteins

In the present work we have succeeded in labeling cell surface proteins by the split intein mediated method of protein *trans*-splicing. The strategy for labeling cell surface proteins using PTS, although described earlier, has never been reported using a naturally split intein. Earlier work by Ando *et al.*[90c] and Volkmann *et al.*[90a] demonstrated cell surface protein labeling utilizing the semisynthetic approach (see section 1.3b).[165a]

Ando *et al.* performed extracellular modification of a membrane protein on the surface of CHO cells using the artificially split semi synthetic *Ssp* DnaB intein where a reaction time of 8 h was reported. Moreover, affinity between the split partners was very low (requiring over 10 μM reactant concentrations) and an integration of the receptor-ligand interaction between eDHFR and its ligand trimethoprim was required for conferring high affinity in order to perform the labeling reaction. The artificially split *Ssp* DnaB intein consists of a small 11aa N-intein and a large C-intein[39, 46] and thus was not suitable for labeling proteins with their C-terminus facing the extracellular surface. Volkmann *et al.* took advantage of the *Ssp* GyrB split intein consisting of a very small C-intein (6aa) and a large 150 aa N-intein[47] for labeling the transferrin receptor on the surface of CHO cells. This method was however carried out over 18 h and was applicable for labeling only the C-terminus of cell surface proteins. In contrast, incubation time of only 5 min was sufficient in this work for modifying a cell surface protein using the *Npu* DnaE intein and within 1 h the reaction was almost complete. It was clear from the time lapse experiment performed (**Figure 32**) that the splice product formation was very fast. However, in comparison to the *in vitro* reaction with purified proteins where the proteins are well separated, freely accessible and not influenced by other biological factors[49b], the reaction on the cell surface proceeds at a slower rate. Different experimental conditions (DMEM media instead of the standardized splice buffer) and the presence of other extracellular proteins on the cell surface may further impede the reaction. Therefore, it is evident that the major advantage conferred by PTS using a naturally split intein method over semi-synthetic *trans*-splicing reaction is the speed of the modification. However, this may be greatly affected by the choice of the intein. In an earlier report, the $Int^C$ terminal fragment of the *Ssp* DnaE intein was synthesized semi-synthetically and used to label

an intracellular POI.[166] However, an incubation of 3 h was reported here. Zettler *et al.* reported a $t_{1/2}$ of ~60s for the *Npu* DnaE intein using purified proteins which is 33 to 170 fold higher than the rate reported for the *Ssp* DnaE intein.[49b] A second advantage would be the concentrations of the reactants required in each case. PTS demonstrated so far with artificially split inteins require peptide concentrations in the higher micromolar range. The present study was able to show successful labeling with a protein concentration of only 2 µM which is much lower in comparison to the existing methods.

Since a naturally split intein has evolved to splice together two halves, the intein fragments are expected to behave better in terms of folding and solubility. Artificially split inteins often suffer from protein aggregation problems due to the exposure of hydrophobic patches on the surface and often require a protein denaturation and renaturation step. This might prove to be detrimental for sensitive proteins.[165a]

Shorter labeling times is advantageous for membrane protein labeling as it will allow for a longer visualization time before the labeled protein is internalized by endocytosis. Additionally, use of photolabile moieties requires short handling and data analysis time.

## 5.3 On Cell Surface Protein Modification Using the *Npu* DnaE Intein

### I. Choice of the intein

In this work the naturally split *Npu* DnaE intein from *Nostoc punctiforme* has been used for labeling cell surface proteins. This intein was chosen based on various criteria discussed below. Iwai and co-workers had demonstrated the robust *trans*-splicing activity of this intein with non-native exteins in *E.coli*.[3] It was shown to possess > 98 % ligation efficiency and 20-fold less residual precursor proteins in comparison to the *Ssp* DnaE intein. Further, Zettler *et al.* had determined an apparent first order rate constant of (1.1±0.2) $x10^{-2}$ $s^{-1}$ at 37 ºC for this intein using purified proteins[49b] which is 5.5 fold higher than the rate reported for the *Sce* VMA intein and 33 to 170 fold higher than the *Ssp* DnaE intein and is by far the highest rate reported for the PTS reaction. Thus, it was a very suitable choice for fast and efficient modification of proteins expressed on living mammalian cells (see also section 1.2.2).

The reasons for the superior properties of this intein is as of yet not clearly understood as it is highly homologous to the *Ssp* DnaE intein but still displays a remarkable difference in

splicing activity. All conserved amino acid known to be essential for the catalytic function are identical in both cases. The *Ssp* DnaE$^{N}$ fragment has a C-terminal extension of 21 aa lacking in *Npu* DnaE intein which is not part of the intein structural fold.[43d] The *Npu* DnaE and the *Ssp* DnaE inteins have also been shown to cross-react, however the rate reported for the *Npu* DnaE intein was only possible in the native combination of the inteins.

One could possibly think of extending the applications of the *Npu* DnaE naturally split intein by further performing semi-synthetic *trans*-splicing. The Int$^{C}$ (36 aa) has been synthesized by solid phase peptide synthesis for covalently modifying target proteins with lipid anchor peptides.[169] A shortened functional C-intein that still retains its robust splicing activity was designed by Aranko *et al.* with the engineered intein bearing the C-terminal 15aa of the C-intein and 123 aa long Int$^{N}$.[45] The ligation efficiency of this new intein was ~96 % and it was shown to improve protein ligation by accelerating the PTS reaction. The kinetic constants for the *trans*-splicing reaction were estimated to be $4.8 \pm 0.3 \times 10^{-5}$ $s^{-1}$ in the presence of 0.5 mM TCEP. The split site of the engineered intein is located in the loop between β12 and β13 of the *Npu* DnaE intein structure. Oeemig *et al.* solved the NMR structure of a single chain variant of *Npu* DnaE intein (see **Figure 51**).[43b] They further engineered a C-intein half consisting of only 6 aa which was efficient in PTS.

**Figure 53 NMR structure of a single chain variant of *Npu* DnaE intein.** Ribbon diagram of the crystal structure of the *Npu* DnaE intein solved by Oeemig *et al.*[43b] PDB ID = 2KEQ

A second problem that could restrict the application of this intein is the C-terminal junction containing the canonical 'CFN' tripeptide sequence which remains in the product after splicing thereby resulting in a mutant protein of interest.[49] It is currently unclear why this sequence is required for efficient PTS. The critical catalytic activity of this C-extein cysteine residue also limits its usage in the cysteine bioconjugation reaction (see section 1.3b) although

the $Int^C$- half in itself is free of any cysteine residues. However, PTS still occurs although at a compromised rate when this Cys is mutated to a Ser. The yield of the reaction is unaffected.

## II. Specificity

A strong and highly specific interaction between the split intein halves is suggested from the demonstrations (**Figure 39**) that the split inteins were able to splice in the presence of several other extracellular proteins even at low micromolar concentrations. This is because split inteins are under evolutionary pressure to associate rapidly with high fidelity and dissociate slowly post-splicing as they splice together essential gene fragments in an endogenous environment, for e.g., proteins involved in DNA replication and formation of unproductive complexes would be deleterious to the survival of the cell.

Despite the broad utility of the PTS reaction, little is known about what drives this highly efficient association of split intein fragments.[49a] Ionic interactions have been postulated to play a role in this split intein assembly.[44, 49a] These studies revealed segregation of charges at specific residues of the intein fragments. Acidic residues were concentrated at specific positions on the N-intein and basic residues dominated the C-intein and these charged residues participate in intermolecular ion-pair interactions.[43b] This was also found to be more prevalent in naturally split inteins. The experimental realization of this hypothesis has been recently carried out by Shah *et al.* where the authors performed charge swapping experiments between the inteins to probe the role of intermolecular ion-pair interactions and in the process also designed a new mutant intein which displayed low cross reactivity with the wild type intein.[170] This further allowed the authors to perform one-pot three piece fragment ligation.

The high efficiency of the PTS reaction on the cell surface at a concentration of only 2 µM was most likely a result of this high affinity and fast association of the split intein fragments. For the same reason, ruling out the possibility of only an associated complex and no splicing due to this affinity was extremely important in this study. This interaction was strong enough to survive simple washing steps and specific enough as to abolish splicing with splice mutants. Both wild type and the splice mutants showed interactions between the inteins in the microscopy study post washing steps. The difference was revealed only on western blot analysis where the complex is believed to break apart due to the loss of the non-covalent interactions by boiling. This was in accordance with earlier reports with the *Ssp* DnaE intein.[52a]

The extra band observed in the immunoblot data of **Figure 28** was assigned as the starting material band by molecular weight analysis. Decrease of this residual protein band when lower concentrations of protein were used and/or harder wash conditions were applied hinted at unspecific binding to the cells or to the 6-well plate. This could be due to the reaction conditions used in this study and leaves room for further optimizations of the washing steps. Unreacted EGFP-$Int^N$ and incomplete washing would also contribute to the signal. However, confocal microscopy data, as seen earlier had not shown any significant fluorescence with the control expression plasmid (**Figure 27C**). Hence, it was concluded that this band was more pronounced in the western blot simply because of the extremely high sensitivity of this detection procedure.

Previous *in vitro* work by Zettler *et al.* reported an almost complete abolition of PTS with the Cys+1Ser mutation while the same substitution in the *Ssp* DnaE intein did not affect the yield of the reaction although it reduced the rate of the reaction by a factor of 10.[49b, 52a] The authors suggested an altered geometry and/or polarization of the key residues in the active site of the inteins as an explanation for a difference in reactivity. A second plausible explanation could be the superior folding properties of this intein into its active conformation after fragment association. In another study by Yang and coworkers demonstrated that the Cys+1Ser mutant was sufficiently splicing active, although it showed slower kinetics in comparison to the native cysteine residue.[158] They could successfully perform FRET with a dual- colored labeled proteins using this mutant intein. Results obtained with the splice mutant for performing PTS on the cell surface (**Figure 31**) seemed to corroborate these earlier observations. The lower yield of splice product could therefore be attributed to a block in the transesterification step in the splicing pathway. However, this effect would be more pronounced with purified proteins *in vitro* in comparison to a highly sensitive western blot procedure. Thioester cleavage may also occur due to the mercaptoethanol being present in the SDS-sample buffer, thereby resulting in N-terminal hydrolysis. Such a product obtained will however occur in the supernatant and will be lost during the washing steps prior to cell lysis.

The Cys1Ala mutant of this intein was reported to completely abolish PTS and was also in accordance with the effect observed for the *Ssp* homologue.[49b, 52a] This could be explained by a block in the N-S acyl shift reaction, thereby completely inhibiting the PTS reaction. The C-terminal cleavage reaction was also shown to be completely blocked and was accounted for by the dependence of the asparagine cyclization at the C-terminal splice junction on the acyl shift at the N-terminus. The results obtained for the Cys1Ser mutant in **Figure 31** could also

be similarly explained. *In vitro* data with this splice mutant showed no splice product formation (unpublished data, Schütz *et al.*) and consistently no splice product was observed with the PTS reaction on the cell surface. One might argue that Ser could partially be able to substitute as a nucleophile and might result in low amounts of splice product formation. However, such an effect was not observed even with the immunoblotting procedure. This would highlight the critical nature of the Cys1 residue and would represent an intricately evolved balance of chemical reactivities. These data were also consistent with earlier work using the *Sce* VMA1 intein where the Cys1Ser mutant had resulted in a complete loss of the protein splicing activity.[32f]

**III. Effect of reducing agent**

In theory, PTS does not require any thiol reagents for the reaction.[27c] Nevertheless, both the splice junctions of the *Npu* DnaE intein contain unpaired cysteines which in theory could form intermolecular disulfide bonds thereby preventing correct association. Therefore, it is desirable to keep the reaction under reducing conditions. However, surfaces of cells contain numerous proteins and addition of a reducing agent would lead to reduction of disulfide bridges of these proteins leading to structural and functional defects. In such a scenario, the reduced thiol group of the cysteine of 'CFN' may interact with erroneous proteins and participate in disulfide bond formation. This may however be prevented due to the highly specific and very strong affinity between the intein partners. An additional cause may also be the relatively low abundance of this amino acid in mammalian proteins.

Omission of DTT from the reaction assay did not show any significant difference in the PTS reaction *in vitro* and *in vivo*, as seen from **Figure 34**, if the exogenous protein was reduced prior to the assay and the excess DTT was removed. The catalytic cysteine of the C-extein is expected to be in its non-reduced state since it is exposed to the extracellular oxidizing environment. However, since free cysteines are unlikely to be available on the extracellular environment (only present as disulfides), the thiol group of the 'CFN' cysteine is unable to form disulfide bridges with other cell surface proteins and thus remains in its reduced form. It could thus be speculated from previous literature and data obtained from this study that one catalytic cysteine in its reduced form is sufficient to drive the PTS reaction. An additional consideration would be the concentration of the DTT used in the reaction, as it was previously

reported with the *Ssp* DnaE intein that the presence of 50 mM DTT would almost totally block PTS and instead shunt the reaction towards cleavage products.[45]

The type of reducing agent used has been earlier reported to greatly influence the yield of the reaction. This is because the thiol group of DTT is a nucleophile competing with the thiol of the first cysteine of the C-intein and induces dominant cleavage reactions. Side reactions in the presence of TCEP are significantly slower because it has no thiol group. Although one might exchange DTT with TCEP in an *in vitro* reaction; Tris of TCEP would most likely be toxic for living cells.[137]

## IV. Effect of the extein sequences

PTS was found to be dependent on the protein sequences of the exteins even if the sequences around the splicing junctions were identical but how the exteins influence the PTS remains unclear. Variation in reaction yields is not only dependent on the intein and the directly flanking sequences but also on the N-extein sequence. This was demonstrated with purified proteins by Zettler and coworkers.[49b] The authors had reported that this intein tolerated varying N-extein sequences with no loss in the reaction rate which was corroborated in the present study. A similar trend was observed with the various N-exteins used in this study as is seen in **Figure 44**. However, it is still unclear how the N-extein sequence affects the splicing reaction. The size, stability, solubility and folding properties of the N-extein sequence may play a critical role in the formation of the active conformation of the $Int^N$ sequence required for the PTS reaction. The results obtained in the present study (**Figure 44**) hint at such a possibility. The gpD sequence due to its small size, high solubility and an independently folding domain is extremely well tolerated and may assist in the correct folding of the entire recombinant fusion protein. This was evident from the high expression levels of this protein and the near completion of the PTS reaction both *in vitro* and on the cell surface. The MBP sequence although contributes to protein solubility is a rather big protein and may introduce structural changes or instability in the fusion protein. The PTS reaction although highly efficient on the cell surface also appears to suffer from degradation of the splice product which could result from the action of proteases or due to instability caused by the large size of the MBP. Reduction of the overall splice product yield to about 50-55 % was reported by Zettler *et al.* on swapping the N-extein to EGFP.[49b] The result obtained in the present study was in accordance with the *in vitro* results. The PCP fusion protein (*in vitro*) was also

effectively able to undergo splicing reaction inspite of a lower concentration of the purified protein obtained. The splice product formation at 5 min was a confirmation of the rapidity at which PTS occurs. The overall charge of all the N-extein sequences besides PCP used was negative (see **Table 5**), which further ruled out non-specific interactions with the negatively charged membrane. The PCP sequence has a net positive charge which might lead to unspecific binding of the labeled protein to the plasma membrane resulting in a false positive result. This might be a hindrance in performing PTS on the surface of mammalian cells.

**Table 5 Charge of the N-extein sequence at pH 7.0**

| N-extein | Charge |
|---|---|
| **Strep-EGFP** | -5.465 |
| **MBP** | -10.552 |
| **Strep-gpD** | -1.584 |
| **Strep-PCP** | +2.415 |

## V. Protein folding

Trans membrane domain proteins, regardless of function, are synthesized and assembled in the rough endoplasmic reticulum. Proteins destined for translocation into the ER usually contain amino terminal signal sequences recognized by a cytosolic ribonucleoprotein complex, the signal recognition particle (SRP). A range of ER resident molecular chaperones and folding factors that function to assist the folding and oligomeric assembly of the newly synthesized protein is encountered by the polypeptide as it enters the lumen of the ER.[171] Proteins which do not assemble correctly with partner subunits or are unable to reach their properly folded state are retained and degraded by an ER ‘quality control’ system that functions to eliminate such non-functional polypeptides.

These points are highlighted when one looks at the expression and targeting pattern of the various C-exteins used in this study. A plausible explanation for the results obtained in **Figure 37A** could be the correct folding of the C-extein fusion proteins. Trx is a highly soluble tag which enhances the folding and solubility of the fusion protein.[172] Understandably, the loss of this domain resulted in misfolding of the intein fusion protein.

Expression levels of the protein appear to be unaffected; however plasma membrane targeting is severely impaired. The mCherry fluorescent protein is an autocatalytic domain that folds extremely rapidly accounting for the observed red fluorescence. The partial rescue by the gpD protein was most likely due to its assistance in folding.

Fusion of the gpD domain at the N-terminus of the fusion protein, construct **7**, was unable to completely salvage the protein and redirect it to the plasma membrane in the absence of the Trx domain. The reasons for this could only be speculated upon. It could either arise due to the gpD not being well tolerated at this position or the inappropriate distance of the $Int^C$ split intein half from the transmembrane domain. An *in vitro* reaction with purified proteins might be important in further answering this question. If PTS is observed efficiently with the gpD-$Int^C$ fusion protein, this might rule out the possibility of non-tolerance of this sequence at the C-terminus of the $Int^C$ fragment.

A confounding observation made in the time lapse experiments was the loss of the HA-tagged fusion protein over time. Although the extremely fast PTS reaction could account for the gradual loss of the HA epitope from the cell surface protein over time, it did not however justify for the fusion protein present in the secretory pathway. This observation can be interpreted by assuming the majority of the protein to be present on the cell surface. Another possible explanation could be the inaccessibility of the HA epitope due to incomplete folding. The fluorescence in the red channel could be due to the cleaved mCherry in the secretory pathway and the remaining protein being degraded. The immunofluorescence data (**Figure 26**) may provide a justification for these explanations. The intensity of the green fluorescence in permeabilized cells was much higher at the plasma membrane than in the subcellular compartments for fusion protein **1**. Likewise, misfolded proteins resulted in a higher green fluorescence in the secretory pathway (**Figure 37A**). In contrast to the immunofluorescence method, the immunoblot technique should however allow access to even denatured proteins and therefore the above explanation may not be completely plausible.

What is evident here is, that the folding state of the protein plays a crucial role in determining the targeting, translocation and degradation of the protein. A temporal relationship exists between protein translocation, folding and polypeptide chain modification.

**f. Broad applicability**

The extent of modification can well be influenced by the type of host cell line, the level of expression and the physiological status of the cell. This was clearly demonstrated by the results obtained with the three different cell lines (N2a, CHO, COS-7) used in this study. The N2a cell line was most efficient. Although each cell line grew as a monolayer, the transfection efficiency of the N2a cells was highest. This could be due the small circular size of the cell which might offer least resistance to transfection. Also, the cell density obtained with this cell line was much higher than with the CHO or COS-7 cells where the cells are more flat and larger in size. This maybe due to the rapid proliferation rate of the neuro-2a cells (~10 h doubling time) in comparison to the COS-7 cells (~18h) and CHO cells (~17h). Thus, this might account for the higher number of transfected cells obtained after a 24 h time span in neuro-2a cells.

In theory, the PTS reaction described in this work could be used for labeling both the N-and C-terminal cell surface exposed proteins in contrast to the methods adopted by Volkmann *et al.*[90a] and Ando *et al.*[90c] However, labeling transmembrane proteins with their C-terminal end exposed to the extracellular environment would require fusion of the protein of interest to the N-terminus of the $Int^N$ half and expression of this fusion construct on the extracellular surface. A look at the results obtained with this experimental setup, points towards the fact that fusion constructs to the $Int^N$ terminal half are not as well tolerated as fusions to the $Int^C$ terminal (**Figure 45**). Presence of the majority of this protein in the secretory pathway suggests misfolding. One reason for this could be the large size of the $Int^N$ fragment which might fold inefficiently with non-natural extein sequences within cellular systems. Protein expression of this fusion construct was also suboptimal in a bacterial expression system where large amounts of the protein was found in the inclusion bodies. Fusion of small or soluble tag proteins (e.g., Trx, gpD, MBP) might offer a solution to this problem.

Interestingly, the splicing reaction was highly efficient on the cell surface both with a trans membrane domain fused protein and a GPI anchored protein. In fact, higher yields were obtained for both the expressed protein and the splice product formed with the $Int^C$-GPI fusion protein (**Figure 46**). However, a direct comparison of protein expression is not possible between construct **2** and **13** since the indirect immunostaining method is much more sensitive in contrast to the direct mCherry fluorescence. GPI-anchored proteins are anchored to the outer leaflet of the plasma membrane. This may allow better access of the purified $Int^N$ fusion protein to the $Int^C$ terminal half exposed on the extracellular surface. The GPI anchored

proteins are also processed via the secretory pathway and would explain the fluorescence observed in the subcellular compartments. (**Figure 46A** and **B**) These proteins have also been implicated to exit the ER via a mechanism different for transmembrane proteins.[173] It has been postulated that the GPI-anchored proteins partition into sphingolipid-containing 'raft' subdomains in the Golgi and plasma membrane.[174]

In conclusion we have shown a new method for the fast and efficient modification of cell surface proteins on living mammalian cells based on protein *trans*-splicing at physiological temperatures. Due to the autocatalytic character and self removal of the split intein halves, it is possible to ligate even entire protein domains with this approach. This approach allows direct backbone modification of POI in a post translational fashion. PTS is also a promising approach for site-specific labeling of fluorophores in transmembrane regions of membrane proteins in living cells, a currently challenging technology. Likewise, the strategy shown here will enable a new route to the preparation of proteins modified with a GPI anchor.

# 6 ABBREVIATIONS

| | |
|---|---|
| AA | Amino acid |
| Bla | Betalactamase |
| BRET | Bioluminiscence Resonance Energy Transfer |
| CHO | Chinese Hamster Ovary |
| CMV | Cytomegalovirus |
| CoA | Coenzyme A |
| DNA | Deoxyribonucleicacid |
| DTT | 1,4-Dithiothreitol |
| *E.coli* | *Escherichia coli* |
| EDTA | Ethylenediaminetetraaceticacid |
| EGFP | Enhanced Green Fluorescent Protein |
| EPL | Expressed Protein Ligation |
| ER | Endoplasmic Reticulum |
| FCS | Fluorescence Correlation Spectroscopy |
| FCCS | Fluorescnce Cross-correlation Spectroscopy |
| FKBP | FK506-Binding Protein |
| FRET | Fluorescence Resonance Energy Transfer |
| gpD | Headprotein D of Bacteriophage λ |
| GPI | Glycosylphosphatidylinositol |
| HA | Hemagglutinin |
| $His_6$ | Hexahistidine |
| $Int^N$ | Intein N-terminal half |
| $Int^C$ | Intein C-terminal half |
| IPTG | β-Isopropylthiogalactoside |
| LB | Luria Bertani |
| MBP | Maltose Binding Protein |
| MW | Molecular Weight |
| *Mxe* | *Mycobacterium xenopi* |

| | |
|---|---|
| *Npu* | *Nostoc punctiforme* |
| NRPS | Non-ribosomal Peptidesynthesis |
| NTA | Nitrilotriacetate |
| OD | Optical Density |
| ORF | Open Reading Frame |
| PAGE | Polyacrylamide Gel Electrophoresis |
| PCP | Peptidyl Carrier Protein |
| PCR | Polymerase Chain Reaction |
| PDB | Protein Data Bank |
| POI | Protein of Interest |
| Ppant | 4'-phosphopantetheinyl |
| PTS | Protein *trans*-splicing |
| QD | Quantum Dot |
| rpm | Rotations per minute |
| RNA | Ribonucleicacid |
| RT | Room Temperature |
| *Sce* | *Saccharomyces cerevisiae* |
| SDS | Sodium Dodecyl Sulphate |
| SPPS | Solid Phase Peptidesynthesis |
| SRP | Signal Recognition Peptide |
| $t_{1/2}$ | Half-life |
| TCEP | Tris-(2-carboxyethyl)-phosphine |
| TM | Transmembrane |
| TMR | Tetramethylrhodamine |
| Tris | Tris-(hydroxymethyl)-aminomethane |
| t-RNA | Transfer Ribonucleicacid |
| Trx | Thioredoxin |
| UV | Ultra Violet |
| v/v | Volume by Volume |
| w/v | Weight by Volume |

# 7 CURRICULUM VITAE

**Name: Tulika Dhar**

**Address:** Oespeler Kirchweg 6, 44379, Dortmund, Germany

**Telephone:** +49-176-831-81024

**Email:** tulika.dhar@tu-dortmund.de

**Date of Birth:** 07 May 1983

**Nationality:** Indian

**University Matriculation Number:** 125744(TU Dortmund)

**Education:**

*September 2007-Present:* **Graduate Fellow (Chemical Biology)**

**Institution:** TU Dortmund, Germany

**Supervisor:** Prof.Henning D.Mootz

*July 2005-June 2007:* **Masters of Science (Biomedical Science)**

**Institution:** Ambedkar Center for Biomedical Research, Delhi University

**Grade:** First class with distinction (70 %)

*July 2002-June 2005:* **Bachelors of Science (Biomedical Science)**

**Institution:** Acharya Narendra Dev College, Delhi University

**Grade:** First class with distinction (75 %)

# 8 REFERENCES

[1] H. Lodish, A. Berk, S. L. Zipursky, *Molecular Cell Biology*, W.H.Freeman, New York, **2000**.
[2] a) M. Z. Lin and L. Wang, *Physiology* **2008**, *23*, 131-141; b) Y. Yano and K. Matsuzaki, *Biochimica et Biophysica Acta* **2009**, *1788*, 2124-2131.
[3] H. Iwai, S. Zuger, J. Jin and P. H. Tam, *FEBS Letters* **2006**, *580*, 1853-1858.
[4] T. Dhar and H. D. Mootz, *Chemical Communications* **2011**, *47*, 3063-3065.
[5] a) K. Hanada, J. W. Yewdell and J. C. Yang, *Nature* **2004**, *427*, 252-256; b) N. Vigneron, V. Stroobant, J. Chapiro, A. Ooms, G. Degiovanni, S. Morel, P. van der Bruggen, T. Boon and B. J. Van den Eynde, *Science* **2004**, *304*, 587-590.
[6] a) R. Hirata, Y. Ohsumk, A. Nakano, H. Kawasaki, K. Suzuki and Y. Anraku, *Journal of Biological Chemistry* **1990**, *265*, 6726-6733; b) P. M. Kane, C. T. Yamashiro, D. F. Wolczyk, N. Neff, M. Goebl and T. H. Stevens, *Science* **1990**, *250*, 651-657.
[7] B. Dujon, *Gene* **1989**, *82*, 91-114.
[8] F. B. Perler, E. O. Davis, G. E. Dean, F. S. Gimble, W. E. Jack, N. Neff, C. J. Noren, J. Thorner and M. Belfort, *Nucleic Acids Research* **1994**, *22*, 1125-1127.
[9] A. A. Cooper and T. H. Stevens, *Trends in Biochemical Sciences* **1995**, *20*, 351-356.
[10] a) M. W. Southworth and F. B. Perler, *Journal of Bacteriology* **2002**, *184*, 6387-6388; b) S. Elleuche, C. Pelikan, N. Nolting and S. Poggeler, *Journal of Basic Microbiology* **2009**, *49*, 52-57; c) R. T. Poulter, T. J. Goodwin and M. I. Butler, *Fungal Genetics and Biology* **2007**, *44*, 153-179.
[11] F. B. Perler, *Nucleic Acids Research* **2000**, *28*, 344-345.
[12] F. B. Perler, G. J. Olsen and E. Adam, *Nucleic Acids Research* **1997**, *25*, 1087-1093.
[13] X. Q. Liu, J. Yang and Q. Meng, *Journal of Biological Chemistry* **2003**, *278*, 46826-46831.
[14] J. P. Gogarten, A. G. Senejani, O. Zhaxybayeva, L. Olendzenski and E. Hilario, *Annual Review of Microbiology* **2002**, *56*, 263-287.
[15] V. Derbyshire and M. Belfort, *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95*, 1356-1357.
[16] T. J. T. Evans and M. Q. Xu, *Chemical Reviews* **2002**, *102*, 4869-4884.
[17] a) M. Belfort, V. Derbyshire, B. L. Stoddard, D. W. Wood, *Homing Endonucleases and Inteins*, Springer, **2005** b) M. Belfort and R. J. Roberts, *Nucleic Acids Research* **1997**, *25*, 3379-3388; c) F. S. Gimble and J. Thorner, *Nature* **1992**, *357*, 301-306.
[18] a) K. Shingledecker, S. Q. Jiang and H. Paulus, *Gene* **1998**, *207*, 187-195; b) V. Derbyshire, D. W. Wood, W. Wu, J. T. Dansereau, J. Z. Dalgaard and M. Belfort, *Proceedings of the National Academy of Sciences of the United States of America* **1997**, *94*, 11466-11471.
[19] A. Telenti, M. Southworth, F. Alcaide, S. Daugelat, W. R. Jacobs, Jr. and F. B. Perler, *Journal of Bacteriology* **1997**, *179*, 6378-6382.
[20] D. R. Smith, L. A. Doucette-Stamm, C. Deloughery, H. Lee, J. Dubois, T. Aldredge, R. Bashirzadeh, D. Blakely, R. Cook, K. Gilbert, D. Harrison, L. Hoang, P. Keagle, W. Lumm, B. Pothier, D. Qiu, R. Spadafora, R. Vicaire, Y. Wang, J. Wierzbowski, R. Gibson, N. Jiwani, A. Caruso, D. Bush and J. N. Reeve, *Journal of Bacteriology* **1997**, *179*, 7135-7155.
[21] a) K. V. Mills, B. M. Lew, S. Jiang and H. Paulus, *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95*, 3543-3548; b) M. W. Southworth, E. Adam, D. Panne, R. Byer, R. Kautz and F. B. Perler, *EMBO Journal* **1998**, *17*, 918-926; c) H. Wu, M. Q. Xu and X. Q. Liu, *Biochimica et Biophysica Acta* **1998**, *1387*, 422-432.
[22] a) F. B. Perler, *Nucleic Acids Research* **2002**, *30*, 383-384; b) S. Pietrokovski, *Protein science* **1998**, *7*, 64-71.
[23] S. Pietrokovski, *Protein science* **1994**, *3*, 2340-2350.
[24] K. V. Mills and F. B. Perler, *Protein and Peptide Letters* **2005**, *12*, 751-755.
[25] G. Amitai, B. P. Callahan, M. J. Stanger, G. Belfort and M. Belfort, *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 11005-11010.
[26] a) F. B. Perler, *Cell* **1998**, *92*, 1-4; b) T. M. Hall, J. A. Porter, K. E. Young, E. V. Koonin, P. A. Beachy and D. J. Leahy, *Cell* **1997**, *91*, 85-97; c) J. J. Lee, S. C. Ekker, D. P. von Kessler, J. A. Porter, B. I. Sun and P. A. Beachy, *Science* **1994**, *266*, 1528-1537; d) E. V. Koonin, *Trends in Biochemical*

*Sciences* **1995**, *20*, 141-142; e) J. A. Porter, S. C. Ekker, W. J. Park, D. P. von Kessler, K. E. Young, C. H. Chen, Y. Ma, A. S. Woods, R. J. Cotter, E. V. Koonin and P. A. Beachy, *Cell* **1996**, *86*, 21-34.
[27] a) X. Q. Liu, *Annual Review of Genetics* **2000**, *34*, 61-76; b) C. J. Noren, J. Wang and F. B. Perler, *Angewandte Chemie International Edition English* **2000**, *39*, 450-466; c) H. Paulus, *Annual Review of Biochemistry* **2000**, *69*, 447-496; d) L. Saleh and F. B. Perler, *Chemical Record* **2006**, *6*, 183-193.
[28] a) H. Paulus, *Bioorganic Chemistry* **2001**, *29*, 119-129; b) R. David, M. P. Richter and A. G. Beck-Sickinger, *European Journal of Biochemistry* **2004**, *271*, 663-677; c) B. W. Poland, M. Q. Xu and F. A. Quiocho, *The Journal of Biological Chemistry* **2000**, *275*, 16408-16413; d) T. Klabunde, S. Sharma, A. Telenti, W. R. Jacobs, Jr. and J. C. Sacchettini, *Nature Structural Biology* **1998**, *5*, 31-36; e) A. Romanelli, A. Shekhtman, D. Cowburn and T. W. Muir, *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 6397-6402.
[29] a) M. A. Johnson, M. W. Southworth, T. Herrmann, L. Brace, F. B. Perler and K. Wuthrich, *Protein Science* **2007**, *16*, 1316-1328; b) M. W. Southworth, J. Benner and F. B. Perler, *EMBO Journal* **2000**, *19*, 5019-5026.
[30] K. Tori, B. Dassa, M. A. Johnson, M. W. Southworth, L. E. Brace, Y. Ishino, S. Pietrokovski and F. B. Perler, *Journal of Biological Chemistry* **2010**, *285*, 2515-2526.
[31] a) S. Pietrokovski, *Current Biology* **1998**, *8*, 634-635; b) G. Amitai, B. Dassa and S. Pietrokovski, *Journal of Biological Chemistry* **2004**, *279*, 3121-3131; c) K. V. Mills, J. S. Manning, A. M. Garcia and L. A. Wuerdeman, *Journal of Biological Chemistry* **2004**, *279*, 20685-20691.
[32] a) S. Chong, K. S. Williams, C. Wotkowicz and M. Q. Xu, *Journal of Biological Chemistry* **1998**, *273*, 10567-10577; b) A. A. Cooper, Y. J. Chen, M. A. Lindorfer and T. H. Stevens, *EMBO Journal* **1993**, *12*, 2575-2583; c) M. W. Southworth, K. Amaya, T. C. Evans, M. Q. Xu and F. B. Perler, *BioTechniques* **1999**, *27*, 110-114, 116, 118-120; d) M. Q. Xu and F. B. Perler, *EMBO Journal* **1996**, *15*, 5146-5153; e) S. Chong, Y. Shao, H. Paulus, J. Benner, F. B. Perler and M. Q. Xu, *Journal of Biological Chemistry* **1996**, *271*, 22159-22168; f) R. Hirata and Y. Anraku, *Biochemical and Biophysical Research Communications* **1992**, *188*, 40-47.
[33] M. Q. Xu and T. C. Evans, Jr., *Methods in Molecular Biology* **2003**, *205*, 43-68.
[34] M. Q. Xu, M. W. Southworth, F. B. Mersha, L. J. Hornstra and F. B. Perler, *Cell* **1993**, *75*, 1371-1377.
[35] a) S. Chong, G. E. Montello, A. Zhang, E. J. Cantor, W. Liao, M. Q. Xu and J. Benner, *Nucleic Acids Research* **1998**, *26*, 5109-5115; b) E. J. Pearl, A. A. Bokor, M. I. Butler, R. T. Poulter and S. M. Wilbanks, *Biochimica et Biophysica Acta* **2007**, *1774*, 995-1001.
[36] a) K. V. Mills and H. Paulus, *Journal of Biological Chemistry* **2001**, *276*, 10832-10838; b) I. Ghosh, L. Sun and M. Q. Xu, *Journal of Biological Chemistry* **2001**, *276*, 24051-24058.
[37] L. Zhang, N. Xiao, Y. Pan, Y. Zheng, Z. Pan, Z. Luo, X. Xu and Y. Liu, *Chemistry* **2010**, *16*, 4297-4306.
[38] L. Zhang, Y. Zheng, Z. Xi, Z. Luo, X. Xu, C. Wang and Y. Liu, *Molecular BioSystems* **2009**, *5*, 644-650.
[39] W. Sun, J. Yang and X. Q. Liu, *Journal of Biological Chemistry* **2004**, *279*, 35281-35286.
[40] T. Yamazaki, T. Otomo, N. Oda, Y. Kyogoku, K. Uegaki, N. Ito, Y. Ishino, H. Nakamura, *Journal of the American Chemical Society* **1998**, *120*, 5591-5592.
[41] S. Brenzel, T. Kurpiers and H. D. Mootz, *Biochemistry* **2006**, *45*, 1571-1578.
[42] T. Kurpiers and H. D. Mootz, *Chembiochem* **2008**, *9*, 2317-2325.
[43] a) H. Wu, Z. Hu and X. Q. Liu, *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95*, 9226-9231; b) J. S. Oeemig, A. S. Aranko, J. Djupsjobacka, K. Heinamaki and H. Iwai, *FEBS Letters* **2009**, *583*, 1451-1456; c) Y. Ding, M. Q. Xu, I. Ghosh, X. Chen, S. Ferrandon, G. Lesage and Z. Rao, *Journal of Biological Chemistry* **2003**, *278*, 39133-39142; d) P. Sun, S. Ye, S. Ferrandon, T. C. Evans, M. Q. Xu and Z. Rao, *Journal of Molecular Biology* **2005**, *353*, 1093-1105.
[44] J. Shi and T. W. Muir, *Journal of the American Chemical Society* **2005**, *127*, 6198-6206.
[45] A. S. Aranko, S. Zuger, E. Buchinger and H. Iwai, *PloS One* **2009**, *4*, e5185.
[46] C. Ludwig, M. Pfeiff, U. Linne and H. D. Mootz, *Angewandte Chemie International Edition English* **2006**, *45*, 5218-5221.
[47] J. H. Appleby, K. Zhou, G. Volkmann and X. Q. Liu, *Journal of Biological Chemistry* **2009**, *284*, 6194-6199.

[48] J. Caspi, G. Amitai, O. Belenkiy and S. Pietrokovski, *Molecular Microbiology* **2003**, *50*, 1569-1577.
[49] a) B. Dassa, G. Amitai, J. Caspi, O. Schueler-Furman and S. Pietrokovski, *Biochemistry* **2007**, *46*, 322-330; b) J. Zettler, V. Schütz and H. D. Mootz, *FEBS Letters* **2009**, *583*, 909-914.
[50] a) C. P. Scott, E. Abel-Santos, M. Wall, D. C. Wahnon and S. J. Benkovic, *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 13638-13643; b) T. C. Evans, Jr., D. Martin, R. Kolly, D. Panne, L. Sun, I. Ghosh, L. Chen, J. Benner, X. Q. Liu and M. Q. Xu, *Journal of Biological Chemistry* **2000**, *275*, 9091-9094.
[51] N. M. Nichols, J. S. Benner, D. D. Martin and T. C. Evans, Jr., *Biochemistry* **2003**, *42*, 5301-5311.
[52] a) N. M. Nichols and T. C. Evans, Jr., *Biochemistry* **2004**, *43*, 10265-10276; b) D. D. Martin, M. Q. Xu and T. C. Evans, Jr., *Biochemistry* **2001**, *40*, 1393-1402.
[53] a) M. A. Marahiel, T. Stachelhaus and H. D. Mootz, *Chemical Reviews* **1997**, *97*, 2651-2674; b) H. von Dohren, U. Keller, J. Vater and R. Zocher, *Chemical Reviews* **1997**, *97*, 2675-2706; c) T. Stein, J. Vater, V. Kruft, A. Otto, B. Wittmann-Liebold, P. Franke, M. Panico, R. McDowell and H. R. Morris, *Journal of Biological Chemistry* **1996**, *271*, 15428-15435.
[54] a) T. Stachelhaus and M. A. Marahiel, *Journal of Biological Chemistry* **1995**, *270*, 6163-6169; b) R. Dieckmann, M. Pavela-Vrancic, H. von Dohren and H. Kleinkauf, *Journal of Molecular Biology* **1999**, *288*, 129-140.
[55] K. Turgay, M. Krause and M. A. Marahiel, *Molecular Microbiology* **1992**, *6*, 2743-2744.
[56] T. Stachelhaus, A. Huser and M. A. Marahiel, *Chemistry & Biology* **1996**, *3*, 913-921.
[57] a) C. T. Walsh, A. M. Gehring, P. H. Weinreb, L. E. Quadri and R. S. Flugel, *Current Opinion in Chemical Biology* **1997**, *1*, 309-315; b) V. De Crecy-Lagard, P. Marliere and W. Saurin, *Comptes Rendus de l'Academie des Sciences. Serie III, Sciences de la vie* **1995**, *318*, 927-936.
[58] a) R. H. Lambalot, A. M. Gehring, R. S. Flugel, P. Zuber, M. LaCelle, M. A. Marahiel, R. Reid, C. Khosla and C. T. Walsh, *Chemistry & Biology* **1996**, *3*, 923-936; b) T. Stein, J. Vater, V. Kruft, B. Wittmann-Liebold, P. Franke, M. Panico, R. Mc Dowell and H. R. Morris, *FEBS Letters* **1994**, *340*, 39-44.
[59] T. Stachelhaus, H. D. Mootz, V. Bergendahl and M. A. Marahiel, *Journal of Biological Chemistry* **1998**, *273*, 22773-22781.
[60] D. Konz and M. A. Marahiel, *Chemistry & Biology* **1999**, *6*, R39-48.
[61] a) J. J. Meng, M. Rojas, W. Bacon, J. T. Stickney and W. Ip, *Methods in Molecular Biology* **2005**, *289*, 341-358; b) E. M. Phizicky and S. Fields, *Microbiological Reviews* **1995**, *59*, 94-123.
[62] K. H. Young, *Biology of Reproduction* **1998**, *58*, 302-311.
[63] J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*, **2006**, p. 954.
[64] P. Schwille, F. J. Meyer-Almes and R. Rigler, *Biophysical Journal* **1997**, *72*, 1878-1886.
[65] R. Rigler and E. S Elson, *Fluorescence Correlation Spectroscopy: Theory and Applications*, Springer, **2001**, p. 487.
[66] M. Eigen and R. Rigler, *Proceedings of the National Academy of Sciences of the United States of America* **1994**, *91*, 5740-5747.
[67] K. M. Berland, *Methods in Molecular Biology* **2004**, *261*, 383-398.
[68] U. Meseth, T. Wohland, R. Rigler and H. Vogel, *Biophysical Journal* **1999**, *76*, 1619-1631.
[69] a) P. Schwille, *Fluorescence Correlation Spectroscopy.Theory and applications*, Springer, **2001**, pp. 360-378; b) T. Wiedemann, M. Wachsmuth, M. Tewes, K. Rippe and J. Langowski, *Single Molecules* **2002**, *3*, 49-61.
[70] a) M. Rarbach, U. Kettling, A. Koltermann and M. Eigen, *Methods* **2001**, *24*, 104-116; b) S. A. Kim, K. G. Heinze, M. N. Waxham and P. Schwille, *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 105-110.
[71] a) M. Burkhardt, K. G. Heinze and P. Schwille, *Optics Letters* **2005**, *30*, 2266-2268; b) J. Korlach, T. Baumgart, W. W. Webb and G. W. Feigenson, *Biochimica et Biophysica Acta* **2005**, *1668*, 158-163; c) S. M. Stavis, J. B. Edel, K. T. Samiee and H. G. Craighead, *Lab on a Chip* **2005**, *5*, 337-343; d) L. C. Hwang and T. Wohland, *Chemphyschem* **2004**, *5*, 549-551; e) J. L. Swift, R. Heuff and D. T. Cramb, *Biophysical Journal* **2006**, *90*, 1396-1410.
[72] a) A. Watson, X. Wu and M. Bruchez, *BioTechniques* **2003**, *34*, 296-300, 302-293; b) J. Riegler and T. Nann, *Analytical and Bioanalytical Chemistry* **2004**, *379*, 913-919; c) R. L. Orndorff and S. J.

Rosenthal, *Nano letters* **2009**, *9*, 2589-2599; d) M. Bruchez, Jr., M. Moronne, P. Gin, S. Weiss and A. P. Alivisatos, *Science* **1998**, *281*, 2013-2016.
[73] a) L. E. Brus, *Journal of Chemical Physics* **1984**, *80*, 4403-4409; b) A. P. Alivisatos, *Science* **1996**, *271*, 933-937.
[74] a) M. Nirmal, B. O. Dabbousi, M. G. Bawendi, J. J. Macklin, J. K. Trautman, T. D. Haris and L. E. Brus, *Nature* **1996**, *383*, 802-804; b) M. Kuno, D. P. Fromm, H. F. Hamann, A. Gallagher and D. J. Nesbitt, *Journal of Chemial Physics* **2001**, *115*, 1028-1040; c) A. P. Alivisatos, W. Gu and C. Larabell, *Annual Review of Biomedical Engineering* **2005**, *7*, 55-76.
[75] a) T. C. Evans, Jr. and M. Q. Xu, *Biopolymers* **1999**, *51*, 333-342; b) T. W. Muir, D. Sondhi and P. A. Cole, *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95*, 6705-6710; c) K. Severinov and T. W. Muir, *Journal of Biological Chemistry* **1998**, *273*, 16205-16209.
[76] S. Zuger and H. Iwai, *Nature Biotechnology* **2005**, *23*, 736-740.
[77] Y. Q. Chen, S. Q. Zhang, B. C. Li, W. Qiu, B. Jiao, J. Zhang and Z. Y. Diao, *Protein Expression and Purification* **2008**, *57*, 303-311.
[78] A. Charalambous, M. Andreou and P. A. Skourides, *Journal of Nanobiotechnology* **2009**, *7*, 9.
[79] U. Arnold, *Biotechnology Letters* **2009**, *31*, 1129-1139.
[80] T. Ozawa and Y. Umezawa, *Current Opinion in Chemical Biology* **2001**, *5*, 578-583.
[81] T. Ozawa and Y. Umezawa, *Methods in Molecular Biology* **2007**, *390*, 269-280.
[82] S. B. Kim, T. Ozawa, S. Watanabe and Y. Umezawa, *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 11542-11547.
[83] a) H. G. Chin, G. D. Kim, I. Marin, F. Mersha, T. C. Evans, Jr., L. Chen, M. Q. Xu and S. Pradhan, *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 4510-4515; b) T. C. Evans, Jr., M. Q. Xu and S. Pradhan, *Annual Review of Plant Biology* **2005**, *56*, 375-392.
[84] a) T. Ozawa, Y. Sako, M. Sato, T. Kitamura and Y. Umezawa, *Nature Biotechnology* **2003**, *21*, 287-293; b) T. Ozawa, K. Nishitani, Y. Sako and Y. Umezawa, *Nucleic Acids Research* **2005**, *33*, e34.
[85] J. Li, W. Sun, B. Wang, X. Xiao and X. Q. Liu, *Human Gene Therapy* **2008**, *19*, 958-964.
[86] S. Brenzel, M. Cebi, P. Reiss, U. Koert and H. D. Mootz, *Chembiochem* **2009**, *10*, 983-986.
[87] T. W. Muir, *Annual Review of Biochemistry* **2003**, *72*, 249-289.
[88] M. Lovrinovic, R. Seidel, R. Wacker, H. Schroeder, O. Seitz, M. Engelhard, R. S. Goody and C. M. Niemeyer, *Chemical Communications* **2003**, 822-823.
[89] I. Burbulis, K. Yamaguchi, A. Gordon, R. Carlson and R. Brent, *Nature Methods* **2005**, *2*, 31-37.
[90] a) G. Volkmann and X. Q. Liu, *PloS One* **2009**, *4*, e8381; b) C. Ludwig, D. Schwarzer and H. D. Mootz, *Journal of Biological Chemistry* **2008**, *283*, 25264-25272; c) T. Ando, S. Tsukiji, T. Tanaka and T. Nagamune, *Chemical Communications* **2007**, 4995-4997.
[91] I. Giriat, T. W. Muir and F. B. Perler, *Genetic Engineering* **2001**, *23*, 171-199.
[92] a) T. Kurpiers and H. D. Mootz, *Angewandte Chemie International Edition English* **2007**, *46*, 5234-5237; b) T. Dhar, T. Kurpiers and H. D. Mootz, *Methods in Molecular Biology* **2011**, *751*, 131-142.
[93] T. Otomo, N. Ito, Y. Kyogoku and T. Yamazaki, *Biochemistry* **1999**, *38*, 16040-16044.
[94] a) T. Otomo, K. Teruya, K. Uegaki, T. Yamazaki and Y. Kyogoku, *Journal of Biomolecular NMR* **1999**, *14*, 105-114; b) H. Yagi, T. Tsujimoto, T. Yamazaki, M. Yoshida and H. Akutsu, *Journal of the American Chemical Society* **2004**, *126*, 16632-16638; c) M. Muona, A. S. Aranko and H. Iwai, *Chembiochem* **2008**, *9*, 2958-2961; d) M. Muona, A. S. Aranko, V. Raulinaitis and H. Iwai, *Nature Protocols* **2010**, *5*, 574-587.
[95] A. E. Busche, A. S. Aranko, M. Talebzadeh-Farooji, F. Bernhard, V. Dotsch and H. Iwai, *Angewandte Chemie International Edition English* **2009**, *48*, 6128-6131.
[96] a) M. M. Bastings, I. van Baal, E. W. Meijer and M. Merkx, *BMC Biotechnology* **2008**, *8*, 76; b) S. Chong and M. Q. Xu, *Journal of Biological Chemistry* **1997**, *272*, 15587-15590; c) A. R. Gillies, J. F. Hsii, S. Oak and D. W. Wood, *Biotechnology and Bioengineering* **2008**, *101*, 229-240; d) J. R. Liu, C. H. Duan, X. Zhao, J. T. Tzen, K. J. Cheng and C. K. Pai, *Applied Microbiology and Biotechnology* **2008**, *79*, 225-233; e) S. S. Sharma, S. Chong and S. W. Harcum, *Journal of Biotechnology* **2006**, *125*, 48-56; f) S. F. Singleton, R. A. Simonette, N. C. Sharma and A. I. Roca, *Protein Expression and Purification* **2002**, *26*, 476-488; g) K. Srinivasa Babu, T. Muthukumaran, A. Antony, S. D. Prem Singh Samuel, M. Balamurali, V. Murugan and S. Meenakshisundaram, *Biotechnology Letters* **2009**,

*31*, 659-664; h) Z. Zhao, W. Lu, B. Dun, D. Jin, S. Ping, W. Zhang, M. Chen, M. Q. Xu and M. Lin, *Applied Microbiology and Biotechnology* **2008**, *77*, 1175-1180.
[97] H. D. Mootz, E. S. Blum, A. B. Tyszkiewicz and T. W. Muir, *Journal of the American Chemical Society* **2003**, *125*, 10561-10569.
[98] a) A. R. Buskirk, Y. C. Ong, Z. J. Gartner and D. R. Liu, *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 10505-10510; b) G. Skretas and D. W. Wood, *Protein Science* **2005**, *14*, 523-532.
[99] M. P. Zeidler, C. Tan, Y. Bellaiche, S. Cherry, S. Hader, U. Gayko and N. Perrimon, *Nature Biotechnology* **2004**, *22*, 871-876.
[100] a) H. Iwai, A. Lingel and A. Pluckthun, *Journal of Biological Chemistry* **2001**, *276*, 16548-16554; b) H. Iwai and A. Pluckthun, *FEBS Letters* **1999**, *459*, 166-172; c) N. K. Williams, P. Prosselkov, E. Liepinsh, I. Line, A. Sharipo, D. R. Littler, P. M. Curmi, G. Otting and N. E. Dixon, *Journal of Biological Chemistry* **2002**, *277*, 7790-7798; d) G. Volkmann, P. W. Murphy, E. E. Rowland, J. E. Cronan, Jr., X. Q. Liu, C. Blouin and D. M. Byers, *Journal of Biological Chemistry* **2010**, *285*, 8605-8614.
[101] J. Lippincott-Schwartz and G. H. Patterson, *Science* **2003**, *300*, 87-91.
[102] a) R. Y. Tsien, *Annual Review of Biochemistry* **1998**, *67*, 509-544; b) R. Y. Tsien, *Methods in Cell Biology* **1989**, *30*, 127-156.
[103] R. E. Campbell, O. Tour, A. E. Palmer, P. A. Steinbach, G. S. Baird, D. A. Zacharias and R. Y. Tsien, *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 7877-7882.
[104] a) I. Chen and A. Y. Ting, *Current Opinion in Biotechnology* **2005**, *16*, 35-40; b) Z. Hao, S. Hong, X. Chen and P. R. Chen, *Accounts of Chemical Research* **2011**, DOI: 10.1021/ar200067r; c) S. Chattopadhaya, F. B. Abu Bakar and S. Q. Yao, *Current Medicinal Chemistry* **2009**, *16*, 4527-4543; d) M. Sunbul and J. Yin, *Organic & Biomolecular Chemistry* **2009**, *7*, 3361-3371; e) K. M. Marks and G. P. Nolan, *Nature Methods* **2006**, *3*, 591-596; f) T. L. Foley and M. D. Burkart, *Current Opinion in Chemical Biology* **2007**, *11*, 12-19.
[105] B. A. Griffin, S. R. Adams and R. Y. Tsien, *Science* **1998**, *281*, 269-272.
[106] S. R. Adams, R. E. Campbell, L. A. Gross, B. R. Martin, G. K. Walkup, Y. Yao, J. Llopis and R. Y. Tsien, *Journal of the American Chemical Society* **2002**, *124*, 6063-6076.
[107] J. P. Vilardaga, V. O. Nikolaev, K. Lorenz, S. Ferrandon, Z. Zhuang and M. J. Lohse, *Nature Chemical Biology* **2008**, *4*, 126-131.
[108] C. Hoffmann, G. Gaietta, M. Bunemann, S. R. Adams, S. Oberdorff-Maass, B. Behr, J. P. Vilardaga, R. Y. Tsien, M. H. Ellisman and M. J. Lohse, *Nature Methods* **2005**, *2*, 171-176.
[109] a) L. W. Miller, J. Sable, P. Goelet, M. P. Sheetz and V. W. Cornish, *Angewandte Chemie International Edition English* **2004**, *43*, 1672-1675; b) L. W. Miller, Y. Cai, M. P. Sheetz and V. W. Cornish, *Nature Methods* **2005**, *2*, 255-257.
[110] E. G. Guignet, R. Hovius and H. Vogel, *Nature Biotechnology* **2004**, *22*, 440-444.
[111] K. M. Marks, P. D. Braun and G. P. Nolan, *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 9982-9987.
[112] M. Robers, P. Pinson, L. Leong, R. H. Batchelor, K. R. Gee and T. Machleidt, *Cytometry* **2009**, *75*, 207-224.
[113] A. Ojida, K. Honda, D. Shinmi, S. Kiyonaka, Y. Mori and I. Hamachi, *Journal of the American Chemical Society* **2006**, *128*, 10452-10459.
[114] A. Juillerat, C. Heinis, I. Sielaff, J. Barnikow, H. Jaccard, B. Kunz, A. Terskikh and K. Johnsson, *Chembiochem* **2005**, *6*, 1263-1269.
[115] A. Keppler, H. Pick, C. Arrivoli, H. Vogel and K. Johnsson, *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 9955-9959.
[116] D. Maurel, L. Comps-Agrar, C. Brock, M. L. Rives, E. Bourrier, M. A. Ayoub, H. Bazin, N. Tinel, T. Durroux, L. Prezeau, E. Trinquet and J. P. Pin, *Nature Methods* **2008**, *5*, 561-567.
[117] G. V. Los, L. P. Encell, M. G. McDougall, D. D. Hartzell, N. Karassina, C. Zimprich, M. G. Wood, R. Learish, R. F. Ohana, M. Urh, D. Simpson, J. Mendez, K. Zimmerman, P. Otto, G. Vidugiris, J. Zhu, A. Darzins, D. H. Klaubert, R. F. Bulleit and K. V. Wood, *ACS Chemical Biology* **2008**, *3*, 373-382.
[118] M. K. So, H. Yao and J. Rao, *Biochemical and Biophysical Research Communications* **2008**, *374*, 419-423.

[119] M. L. Mannesse, R. C. Cox, B. C. Koops, H. M. Verheij, G. H. de Haas, M. R. Egmond, H. T. van der Hijden and J. de Vlieg, *Biochemistry* **1995**, *34*, 6400-6407.
[120] M. L. Mannesse, J. W. Boots, R. Dijkman, A. J. Slotboom, H. T. van der Hijden, M. R. Egmond, H. M. Verheij and G. H. de Haas, *Biochimica et Biophysica Acta* **1995**, *1259*, 56-64.
[121] R. Bonasio, C. V. Carman, E. Kim, P. T. Sage, K. R. Love, T. R. Mempel, T. A. Springer and U. H. von Andrian, *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104*, 14753-14758.
[122] M. Fernandez-Suarez, H. Baruah, L. Martinez-Hernandez, K. T. Xie, J. M. Baskin, C. R. Bertozzi and A. Y. Ting, *Nature Biotechnology* **2007**, *25*, 1483-1487.
[123] L. Lorand and R. M. Graham, *Nature Reviews. Molecular Cell Biology* **2003**, *4*, 140-156.
[124] H. Sato, M. Ikeda, K. Suzuki and K. Hirayama, *Biochemistry* **1996**, *35*, 13072-13080.
[125] C. W. Lin and A. Y. Ting, *Journal of the American Chemical Society* **2006**, *128*, 4542-4543.
[126] a) N. George, H. Pick, H. Vogel, N. Johnsson and K. Johnsson, *Journal of the American Chemical Society* **2004**, *126*, 8896-8897; b) J. Yin, F. Liu, X. Li and C. T. Walsh, *Journal of the American Chemical Society* **2004**, *126*, 7754-7755.
[127] J. Yin, A. J. Lin, P. D. Buckett, M. Wessling-Resnick, D. E. Golan and C. T. Walsh, *Chemistry & Biology* **2005**, *12*, 999-1006.
[128] B. H. Meyer, J. M. Segura, K. L. Martinez, R. Hovius, N. George, K. Johnsson and H. Vogel, *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 2138-2143.
[129] a) A. M. Perry, H. Ton-That, S. K. Mazmanian and O. Schneewind, *Journal of Biological Chemistry* **2002**, *277*, 16241-16248; b) S. K. Mazmanian, G. Liu, H. Ton-That and O. Schneewind, *Science* **1999**, *285*, 760-763.
[130] a) H. Mao, S. A. Hart, A. Schink and B. A. Pollok, *Journal of the American Chemical Society* **2004**, *126*, 2670-2671; b) R. Parthasarathy, S. Subramanian, E. T. Boder, *Bioconjugate Chemistry* **2007**, *18*, 469-476; c) T. Tanaka, T. Yamamoto, S. Tsukiji and T. Nagamune, *Chembiochem* **2008**, *9*, 802-807.
[131] M. W. Popp, J. M. Antos, G. M. Grotenbreg, E. Spooner and H. L. Ploegh, *Nature Chemical Biology* **2007**, *3*, 707-708.
[132] L. Wang and P. G. Schultz, *Angewandte Chemie International Edition English* **2005**, *44*, 34-66.
[133] Z. Zhang, B. A. Smith, L. Wang, A. Brock, C. Cho and P. G. Schultz, *Biochemistry* **2003**, *42*, 6735-6746.
[134] C. T. Hauser and R. Y. Tsien, *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104*, 3693-3697.
[135] a) C. M. McCann, F. M. Bareyre, J. W. Lichtman and J. R. Sanes, *BioTechniques* **2005**, *38*, 945-952; b) Y. Sekine-Aizawa and R. L. Huganir, *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 17114-17119.
[136] a) J. R. Litowski and R. S. Hodges, *Journal of Biological Chemistry* **2002**, *277*, 37272-37279; b) Y. Yano, A. Yano, S. Oishi, Y. Sugimoto, G. Tsujimoto, N. Fujii and K. Matsuzaki, *ACS Chemical Biology* **2008**, *3*, 341-345.
[137] J. Sambrook, E. F. Fritsch, T.Maniatis, *Molecular Cloning: A Lab Manual*, Cold Spring Harbor Laboratory Press, New York, **1989**.
[138] C. G. Kantzer in *Chemische Modifikation von Proteinen in vitro und in vivo mit Hilfe des Npu DnaE-Inteins*, *Vol. Bachelor* TU Dortmund, Dortmund, **2009**.
[139] L. Ellgaard, M. Molinari and A. Helenius, *Science* **1999**, *286*, 1882-1888.
[140] R. G. Gronwald, F. J. Grant, B. A. Haldeman, C. E. Hart, P. J. O'Hara, F. S. Hagen, R. Ross, D. F. Bowen-Pope and M. J. Murray, *Proceedings of the National Academy of Sciences of the United States of America* **1988**, *85*, 3435-3439.
[141] J. Y. Yang and W. Y. Yang, *Journal of the American Chemical Society* **2009**, *131*, 11644-11645.
[142] P. Forrer and R. Jaussi, *Gene* **1998**, *224*, 45-52.
[143] a) M. A. Ferguson, *Journal of Cell Science* **1999**, *112*, 2799-2809; b) M. Nozaki, K. Ohishi, N. Yamada, T. Kinoshita, A. Nagy and J. Takeda, *Laboratory Investigation* **1999**, *79*, 293-299; c) M. Tarutani, S. Itami, M. Okabe, M. Ikawa, T. Tezuka, K. Yoshikawa, T. Kinoshita and J. Takeda, *Proceedings of the National Academy of Sciences of the United States of America* **1997**, *94*, 7400-7405.

[144] M. G. Paulick and C. R. Bertozzi, *Biochemistry* **2008**, *47*, 6991-7000.
[145] a) B. S. Glick, *Current Biology* **2001**, *11*, 361-363; b) H. Ikezawa, *Biological & Pharmaceutical Bulletin* **2002**, *25*, 409-417; c) S. Chatterjee and S. Mayor, *Cellular and Molecular Life Sciences* **2001**, *58*, 1969-1987.
[146] P. Hillmen, J. M. Hows and L. Luzzatto, *British Journal of Haematology* **1992**, *80*, 399-405.
[147] S. D. Leidich, D. A. Drapp and P. Orlean, *Journal of Biological Chemistry* **1994**, *269*, 10193-10196.
[148] L. Popolo and M. Vai, *Biochimica et Biophysica Acta* **1999**, *1426*, 385-400.
[149] a) C. Nuoffer, A. Horvath and H. Riezman, *Journal of Biological Chemistry* **1993**, *268*, 10558-10563; b) C. Nuoffer, P. Jeno, A. Conzelmann and H. Riezman, *Molecular and Cellular Biology* **1991**, *11*, 27-37.
[150] M. C. Schumacher, U. Resenberger, R. P. Seidel, C. F. Becker, K. F. Winklhofer, D. Oesterhelt, J. Tatzelt and M. Engelhard, *Biopolymers* **2010**, *94*, 457-464.
[151] L. E. Quadri, P. H. Weinreb, M. Lei, M. M. Nakano, P. Zuber and C. T. Walsh, *Biochemistry* **1998**, *37*, 1585-1595.
[152] S. A. Sieber, C. T. Walsh and M. A. Marahiel, *Journal of the American Chemical Society* **2003**, *125*, 10862-10866.
[153] H. D. Mootz and M. A. Marahiel, *Journal of Bacteriology* **1997**, *179*, 6843-6850.
[154] E. Zamir, P. H. Lommerse, A. Kinkhabwala, H. E. Grecco and P. I. Bastiaens, *Nature Methods* **2010**, *7*, 295-298.
[155] J. Yin, A. J. Lin, D. E. Golan and C. T. Walsh, *Nature Protocols* **2006**, *1*, 280-285.
[156] G. T. Hermanson, *Bioconjugate Techniques*, Academic Press, **2008**.
[157] A. K. Miller and F. Tausig, *Biochemical and Biophysical Research Communications* **1964**, *14*, 210-214.
[158] Z. Xia, Y. Xing, M. K. So, A. L. Koh, R. Sinclair and J. Rao, *Analytical Chemistry* **2008**, *80*, 8649-8655.
[159] S. B. Kent, *Chemical Society Reviews* **2009**, *38*, 338-351.
[160] L. Wang, J. Xie and P. G. Schultz, *Annual Review of Biophysics and Biomolecular Structure* **2006**, *35*, 225-249.
[161] G. T. Hermanson, *Bioconjugate Techniques*, Academic Press, San Diego,CA, **2008**
[162] C. P. Hackenberger and D. Schwarzer, *Angewandte Chemie International Edition English* **2008**, *47*, 10030-10074.
[163] A. Keppler, S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel and K. Johnsson, *Nature Biotechnology* **2003**, *21*, 86-89.
[164] I. Chen, M. Howarth, W. Lin and A. Y. Ting, *Nature Methods* **2005**, *2*, 99-104.
[165] a) H. D. Mootz, *Chembiochem* **2009**, *10*, 2579-2589; b) V. Muralidharan and T. W. Muir, *Nature Methods* **2006**, *3*, 429-438.
[166] I. Giriat and T. W. Muir, *Journal of the American Chemical Society* **2003**, *125*, 7180-7181.
[167] I. Saves, V. Ozanne, J. Dietrich and J. M. Masson, *Journal of Biological Chemistry* **2000**, *275*, 2335-2341.
[168] K. M. Marks, M. Rosinov and G. P. Nolan, *Chemistry & Biology* **2004**, *11*, 347-356.
[169] N. K. Chu, D. Olschewski, R. Seidel, K. F. Winklhofer, J. Tatzelt, M. Engelhard and C. F. Becker, *Journal of Peptide Science* **2010**, *16*, 582-588.
[170] N. H. Shah, M. Vila-Perello and T. W. Muir, *Angewandte Chemie International Edition English* **2011**, *50*, 6511-6515.
[171] K. Leitzgen, I.G. Haas, *Chemtracts-Biochemistry and Molecular Biology* **1998**, 423-445.
[172] J. McCoy and E. Lavallie, *Current Protocols in Protein Science* **2001**
[173] M. Muniz, P. Morsomme and H. Riezman, *Cell* **2001**, *104*, 313-320.
[174] K. Simons and E. Ikonen, *Nature* **1997**, *387*, 569-572.