INTEIN ENGINEERING FOR PROTEIN HYDROGEL SYNTHESIS AND PROTEIN PURIFICATION

A Dissertation

by

MIGUEL ANGEL RAMIREZ

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Zhilei Chen
Committee Members,	Arul Jayaraman
	Katy Kao
	Steve Lockless
Head of Department,	M. Nazmul Karim

December 2013

Major Subject: Chemical Engineering

Copyright 2013 Miguel Ramirez

ABSTRACT

Inteins are proteins encoded within a precursor gene that excise themselves after translation and ligate the surrounding proteins with a peptide bond. Since their discovery two decades ago, many inteins have been engineered for various biotechnology applications. This dissertation focuses on the use and development of intein-based technologies for applications in protein purification and immobilization. The highly efficient naturally split DnaE intein from *Nostoc punctiforme* (*Npu* DnaE) was incorporated into synthetic protein building blocks for the synthesis of protein hydrogels, and engineered to catalyze rapid C-terminal cleavage reaction and used in the rapid purification of tag-less protein.

In the first application, we developed protein hydrogels as general scaffolds for protein immobilization. Immobilization has been shown to increase protein stability and facilitate enzyme recovery-and-recycle tasks. These hydrogels are composed of artificial protein building-blocks expressed in bacterial hosts. Hydrogel gelation is catalyzed by intein-mediated protein *trans*-splicing reactions or disulfide bond formation between different protein building blocks. The resulting artificial protein hydrogels possess high solution stability at a wide range of pHs and temperatures, undergo shear-thinning, and are compatible with organic solvents. These self-assembled protein hydrogels can protect immobilized enzymes from organic solvent denaturation during biosynthesis, be used in enzymatic biofuel cells, and are suitable for the immobilization of multiple enzymes.

In the second application, we engineered the *Npu* DnaE intein to catalyze rapid thio-induced C-terminal cleavage reaction and subsequently developed a split intein mediated technology for recombinant protein purification (SIRP). SIRP enables efficient purification of tag-less recombinant protein from *E. coli* lysate in less than 1 hour – the hitherto fastest reported intein technology for protein purification.

DEDICATION

To my family

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor, Dr. Zhilei Chen, for her training, knowledge, mentoring, patience, guidance, enthusiasm and continuous encouragement. Through the years in the lab, Dr. Chen has helped me become a better researcher, a more professional individual, and a better person overall. I would also like to thank the members of my committee, Dr. Arul Jayaraman, Dr. Katy Kao, and Dr. Steve Lockless, for their time and support. I am grateful to Dr. Karuppiah Chockalingam who patiently taught me many experimental techniques at the beginning of my Ph.D. studies. My gratitude also goes to Dr. Dongli Guan who contributed greatly to many of my research projects, and for her insight during many experimental discussions. I would like to thank my lab companions for their contributions to my work theoretically and experimentally. I want to specially thank Ana Maria Chamoun for her support in my research, continuous discussions of research and science in general, and for the friendship that we have shared through the years in the lab. I would also like to thank members of the Kao, M. Hahn and Jayaraman labs for the friendship and support they have given me throughout the years. I extend my gratitude to my cousins and friends for all the years of laughter and fun. Thanks also go to my friends in College Station for making my time at Texas A&M University an enjoyable experience. I appreciate the support provided by the Texas A&M Chemical Engineering Department, and the National Science Foundation.

I would like to express my sincere gratitude to my parents, my brother and my two sisters for their love, encouragement, understanding, and unconditional support. Everything I have accomplished is due to my family.

TABLE OF CONTENTS

Page

Page

Page

LIST OF FIGURES

Page

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

1.1 Background

Inteins are proteins encoded within a precursor gene that excise themselves after translation and ligate the surrounding proteins with a peptide bond in a process called protein splicing (Fig. 1.1) (*1*). The term "intein" was initially conceived as an abbreviation for internal protein by Perler and co-workers who referred the surrounding protein sequences as "exteins" (*2, 3*). Inteins have been discovered in all forms of lives including Archaea, Eubacteria, Eukarya and viruses (*4, 5*). Inteins were first discovered when a discrepancy between the gene size encoding the subunit VMA1 of the *Saccharomyces cerevisiae* (*Sce* VMA1) ATPase and the size of the protein obtained after expression was noticed (*6*). The expected size of the translated gene (118 kDa) was found to be much bigger than the size of the expressed protein (67 kDa). The missing protein sequence is homologous to the yeast endonuclease protein encoded by the HO gene (*6*).

After resolving the crystal structure of the *Sce* VMA1 intein, Duan and co-workers identified separate domains responsible for the splicing and endonuclease activities. The endonuclease domain was sandwiched between the two halves of the splicing domains (Fig. 1.2A) (*7*). Based on the structural information, they hypothesized that the endonuclease and splicing domain may function independently. This hypothesis was later proven correct by Derbyshire and co-workers who showed that several inteins

retained splicing activity even after removal of the endonuclease domain (*8*). These smaller inteins consisting only of the splicing domain are termed mini-inteins, which also includes inteins naturally lacking the endonuclease domain (Fig. 1.2B) (*9-11*).

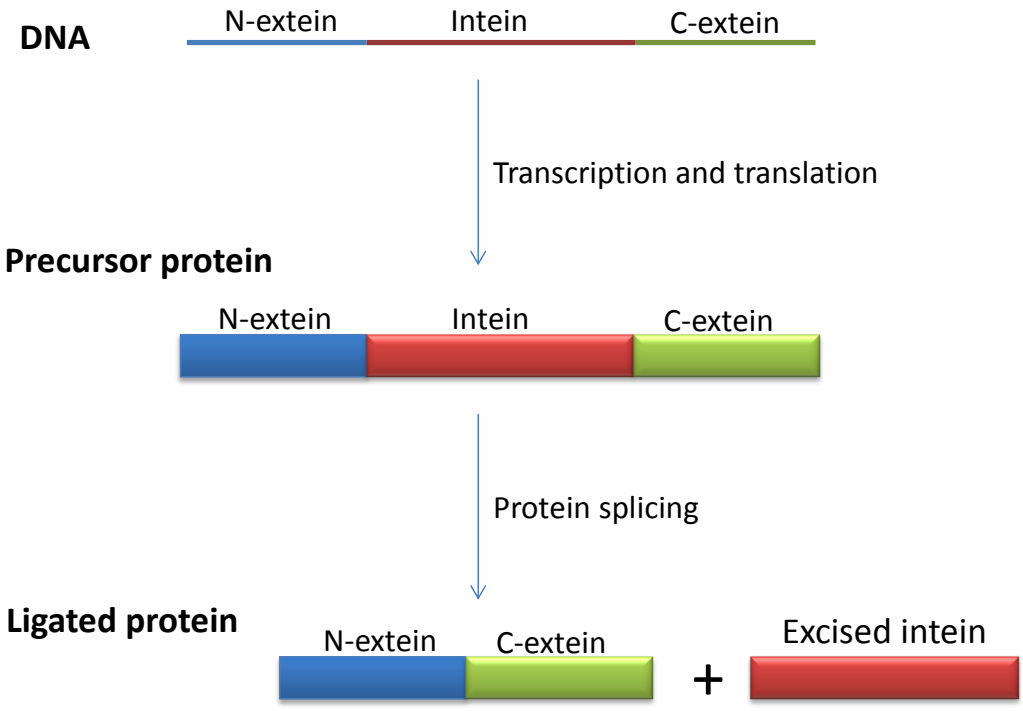


Figure 1.1 Intein-mediated protein splicing. Inteins are genetically encoded between the N- and C- extein fragments. After transcription and translation, the intein excises itself out and ligates the N- and C-exteins.

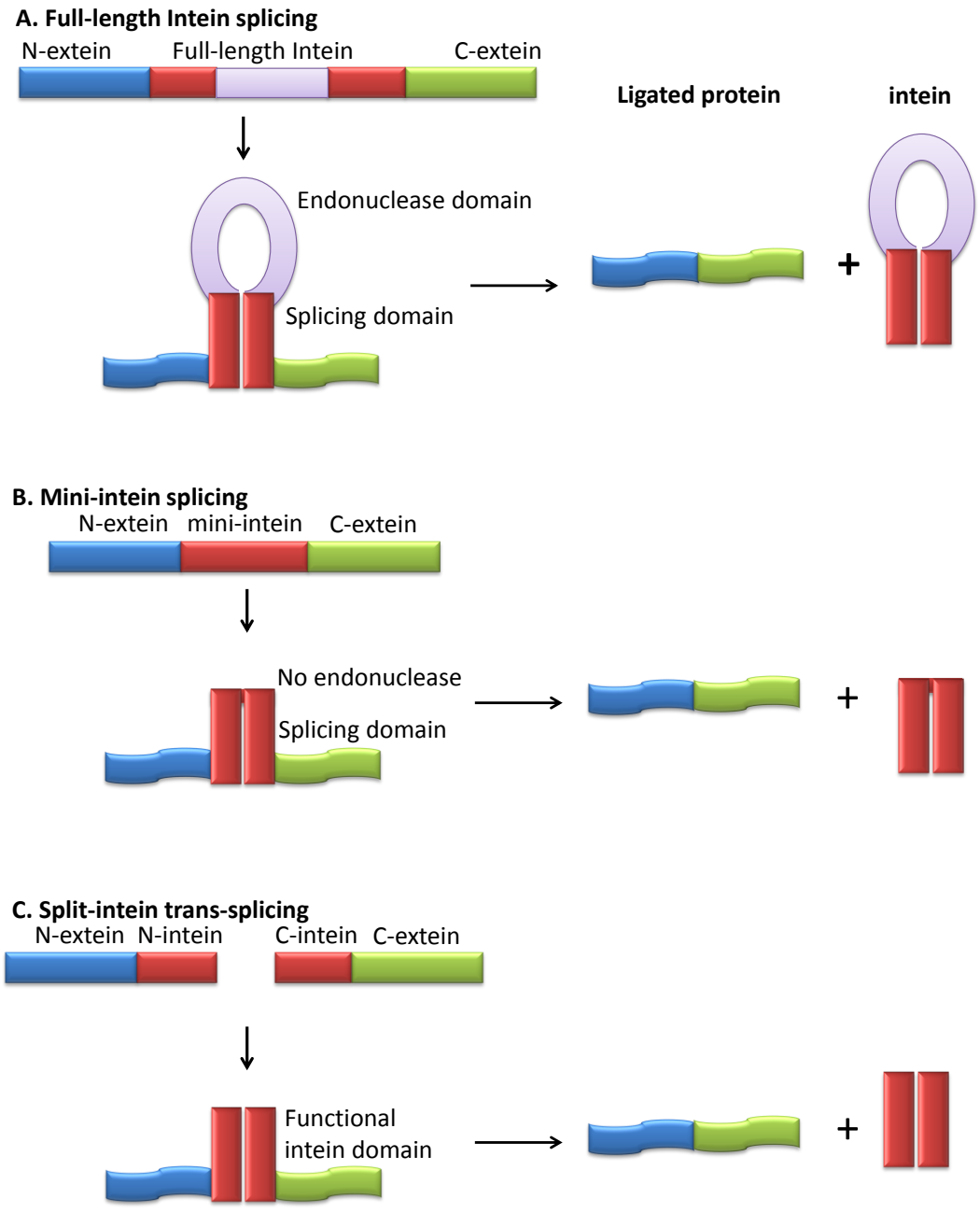


Figure 1.2 Different types of intein domains. (A) A full-length intein contains two independent domains: splicing domain (red) and endonuclease domain (purple) (B) Mini-inteins contain only the splicing domain and lack the endonuclease domain. (C) A split intein is a mini-intein whose N- and C-terminal fragments are expressed separately. Regardless of the type, intein splicing reaction results in the ligation of the exteins and the excision of the intein.

The two halves of a split intein can be expressed separately, and reconstituted via non-covalent interactions to restore the intein activity. The first artificially split mini-intein was constructed by the Paulus group using the RecA intein from *Mycobacterium tuberculosis* (*12, 13*). Many other artificially split inteins have been engineered by the same approach including the *Ssp* DnaB (*14*), PI-*Pfu*I (*15*), *Sce* VMA (*16*), and *Psp* Pol-1 (*17*).

In addition to artificially split mini-inteins, there are a number of naturally split inteins. The first discovered naturally split intein is the DnaE intein from the cyanobacterium *Synechocystis* sp. strain PCC6803 (*Ssp* DnaE) (*18*). Since then, several alleles of this intein have been identified in other cyanobateria with different levels of activity, efficiency, and flexibility (*19*). Recently, Dassa and co-workers identified a new group of small naturally split inteins using environmental metagenomic sequence analysis tools (*20*). These inteins have high trans-splicing rates and yields, as well as orthogonal activities (*21*).

A full-length intein (including the endonuclease domain) typically contains 350 to 550 amino acids (*17*), whereas a mini-intein, either continuous or split, usually consists of 135-200 amino acids (*22*). In a split intein, the N- and C-terminal subdomains are usually 100-150 and 35-50 amino acids, respectively. Inteins from different organisms usually have low sequence homology but share similar tertiary structure, and most of the catalytic residues are highly conserved (*5, 8, 23-26*). The tertiary structure of a typical intein resembles a compact horseshoe with 11-14 beta-strands (*8, 10, 27, 28*).

This structure also shares some similarities with the auto-processing signaling proteins encoded by the hedgehog gene family (*29, 30*).

In Class A inteins, the splicing reaction consists of four nucleophilic displacement steps (Fig. 1.3) (*31*). More than 90% of the inteins discovered thus far belong to this class (*23, 32*). The first step is an acyl rearrangement at the N-terminal of the intein where the side chain of residue Cys1 or Ser1 reacts with the last carbonyl carbon of the N-extein resulting in a thio-ester bond and a linear intermediate. In the second step, the first amino acid of the C-extein (Cys+1, Ser+1, or Thr+1) attacks the thio-ester bond at the N-terminal, cleaving the N-terminal intein junction and connecting the N-extein to the side chain of the +1 residue of the C-extein, resulting in the formation of a branched product with two amine-terminals. This branch product is resolved in the third step by the cyclization of the last intein residue Asn, resulting in the cleavage of the intein at its C-terminal. The final step involves an S (or O) to N re-arrangement and the formation of a native peptide bond connecting the exteins. The coordination of these four steps is essential for efficient splicing reaction. The adaptation of inteins in non-native exteins can lead to low splicing yields and incomplete splicing resulting in N- and/or C-terminal cleavage side products (*33*). Slight variation in the splicing mechanism has been reported for some Inteins that differ in some catalytic amino acids (*30, 34-37*).

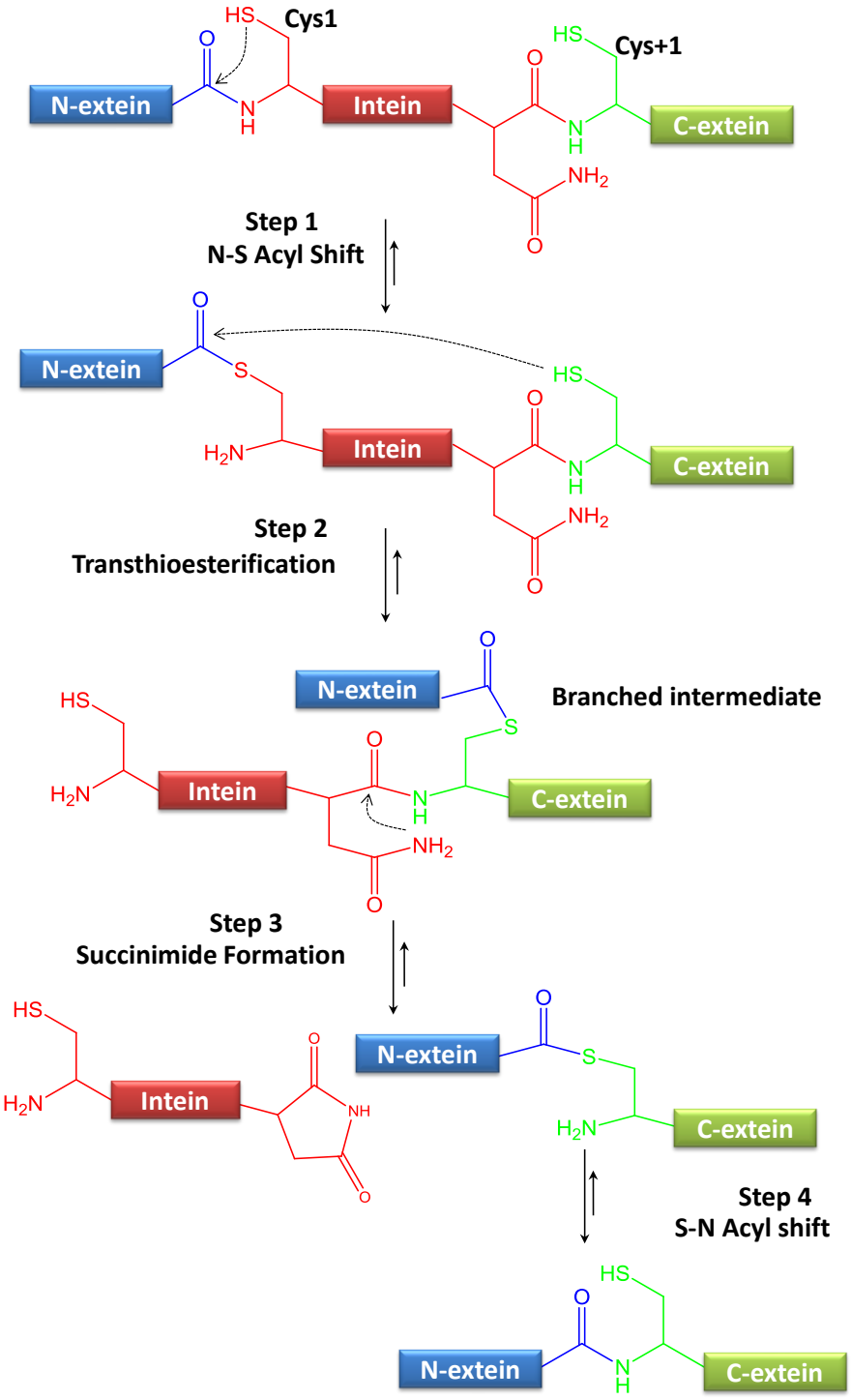


Figure 1.3 Intein splicing mechanism. Red residues correspond to the intein, blue to the N-extein and green to the C-extein.

The natural function of inteins is not well understood. It's generally believed that the splicing and endonuclease domains evolved independently. Mini-inteins are thought to have originated by combining two self-splicing domains (*1*). Endonucleases then invaded the mini-intein genes to provide mobility to these proteins (*8*). Split-inteins are thought to have originated from mini-inteins after genomic rearrangement (*38*). A potential explanation for the origin of full-length inteins was given by Liu in evolutionary terms (*1*). Evolutionary hypothesis was based on the ability of endonuclease-containing inteins to catalyze their homing: copying their own sequence into a gene containing a homologous endonuclease DNA (*4, 39*). Because of this property, inteins are believed to spread horizontally between different organisms in a parasitic nature. However, this hypothesis does not explain the natural relevance of intein existence, although recent findings suggest that inteins may post-translationally regulate the activity of many polymerases, helicases, and recombinases (*40*).

Over the last two decades inteins have become a very significant topic of research. Knowledge of intein's evolution and reaction mechanism are expected to elucidate intein's origin and native function. Moreover, the catalytic properties of inteins are being used as new biotechnology tools with significant potential applications in biomedicine, bioseperation and synthetic biology (*23*). By the summer of 2013, more than 450 inteins were listed in the intein data base (*32*), highlighting the rapid growth of the field and the importance of intein in various biotechnology applications.

1.2 Intein engineering and applications

1.2.1 Tag-less protein purification

Affinity tags has greatly simplified the purification of recombinant proteins from crude lysates, and is one of the most widely used techniques today (*41*). Aside from enabling affinity-based proteins purification, protein tags have also been used as linear epitopes for Western blots, as well as expression and solubility enhancers (*42*). However, in many cases the protein tags need to be removed after protein purification. This step is often done using site-specific endoproteases (*43*), a process that can be slow and costly (*44*). Chong and co-workers developed the first intein-based protein purification method using *Sce* VMA1 intein (*33, 45*). This intein contained a substitution of the first Cys with an Ala residue, abolishing the first two steps of intein splicing, and allowing the asparagine cyclization to proceed generating a C-terminal cleaving intein. This intein was further engineered by Wood and co-workers using directed evolution to attain some pH and temperature-controlled activity (*46*). Other engineered inteins with temperature, thio-, or pH-dependent N- or C-terminal cleavage activity have been engineered and used in affinity-based protein purification (*47-50*). Many of these purification technologies have been commercialized by New England Biolabs under the IMPACT acronym (intein mediated purification with affinity chitin-binding tag). In the IMPACT systems, an engineered intein is sandwiched between a protein of interest (POI) and a chitin-binding-domain (CBD). POI-fusion can be easily separated from cell lysate via interaction with a chitin column, and purified POI can be released from the column upon intein cleavage. Recently, a number of non-chromatographic purification

methods have been developed in which the CBD is replaced with the elastin-like polypeptides (ELP) capable of conditionally precipitating from solution (*51, 52*).

A significant limitation of the current intein-based purification methods is the premature *in vivo* cleavage of POI during expression, which can lead to > 90% loss of the target protein (*53*). This problem can be conveniently solved using split inteins. In a split intein-based protein purification system, the POI is fused to one half of a split intein and the purification tag is fused to the other half. These two proteins are expressed in separate hosts. During protein purification, lysates containing these two fusion proteins are mixed, rendering an active intein able to catalyze the tag-removal reaction. Chromatography-based purification systems using the artificially split *Ssp* DnaB intein have recently been reported. However, the current systems either require the use of large quantities of a synthetic peptide (*54*), or have a low cleavage efficiency and reaction kinetics (*55*).

A more recent split-intein based protein purification system was developed by Vila-Perello and co-workers (*56*). This work takes advantage of the formation of the first thio-ester bond at the N-terminal junction. The protein of interest is purified by the addition of a thio group, resulting in the purification of a protein containing a thio-ester at the C-terminal, which can be used for additional modification.

We have developed a similar split-intein based protein purification system using the highly efficient naturally split *Npu* DnaE intein. This work is presented in Chapter V of this dissertation (*57, 58*).

1.2.2 Protein cyclization

The synthesis of circular proteins has been an increasingly growing research task, since certain cyclic proteins show exceptional levels of stability against chemical, thermal and enzymatic degradation, as well as higher substrate specificity (*59*). The idea of using inteins for the formation of cyclic proteins was first conceived by Evans and co-workers with the use of two engineered inteins fused at each end of the target protein (*60*). However, this system resulted in low efficiency and unwanted protein polymerization. The same group later used the naturally split *Ssp* DnaE intein to successfully generate cyclic proteins (*61*). Similarly, Scott and co-workers used the split *Ssp* DnaE intein to generate libraries of stable circular peptides and small proteins *in vivo* (*62*). Thereafter, this cyclization methodology has been successfully carried out for diverse applications with the use of newly engineered split Inteins (*63, 64*).

1.2.3 Structural biology

Understanding the structure-function relationship of a protein has been one the main focuses of structural biologists. Protein structure can provide information about the protein's function, dynamic behavior, and inter-molecular interactions, and insights for drug design and discovery (*65-67*). Protein crystallization is one of the most important tools in structural biology. However, this technique usually takes a long time, is expensive and may not fully address the chemical and dynamic aspects of the protein (*68*). Optical methods have emerged as additional tools for the study of protein structure. Nuclear Magnetic Resonance (NMR) spectroscopy has been used to construct high

resolution images of proteins with the ability to capture real time interactions, dynamic behaviors and protein folding (*69*). A key component for the image resolution for NMR is the choice of probes: optical-chemical labels that can be traced within the target structure (*70*). The most commonly use probes are isotopes and fluorophores integrated into proteins at specific locations. NMR structures of proteins up to 25kDa have been resolved using heteronuclear NMR techniques (*71*). However, larger proteins generate signals with high levels of degeneracy, making it difficult to use NMR to resolve the structure of a protein larger than 25kDa (*71*).

Using the artificially split intein from the ribonucleotide reductase of *Pyrococcus furiosus* (*Pfu*I), Yamazaki and co-workers developed a methodology for segmental isotope labeling of proteins (*72*). To do this, a segment of the target protein was expressed in the presence of labeled amino acids. This fragment was then ligated to the unlabeled protein segment via intein trans-splicing reaction (*72*). Although this method enabled the successful construction of NMR images of larger proteins, the technique showed low yields of the labeled proteins, and required the purification of fusion proteins under denaturing conditions (*73*). In addition, intein ligation introduced non-native residues into the target protein, which inherently modified the target protein sequence (*68*).

The subsequent discovery and engineering of naturally split inteins significantly improved this technology. Orthogonal inteins have been used to ligate multiple fragments, enabling the selective isotope labeling of only the center blocks of full-length proteins, which significantly lowers the signal overlaps (*74*). Also, naturally split-inteins

have been used to produce labeled proteins *in vivo* (*74, 75*). Recently some inteins have been engineered to better tolerate extein residues near the splicing junction, thus minimizing the non-native residues in the ligated product (*76, 77*).

1.2.4 Expressed protein ligation

Protein semi-synthesis refers to the process of introducing synthetic chemical elements into proteins. The ability to modify proteins with chemical functional groups facilitates many applications in biotechnology as well as the study of structure-function relationship of proteins (*56, 78, 79*). Several methods for protein semi-synthesis have been developed. The most widely used technique to chemically modify proteins takes advantage of the side chains of Lys and Cys residues. However, this method lacks selectivity, thus indiscriminately modifying exposed and unprotected residues (*79*).

More sophisticated chemoselective methods have been used to selectively modify proteins at specific sites (*80, 81*). An effective method for chemoselective modification is expressed protein ligation (EPL) (*82*). EPL is a technology first introduced by Muir and co-workers (*83*) in which inteins are used for the generation of an N-terminal cysteine or a thio-ester group. These groups can undergo native chemical ligation (NCL) resulting in the formation of a native peptide bond (*84-86*). ELP has been used in many applications such as the conjugation of quantum dots to study a protease activity (*87*), attachment of PEG-like polymers at the C-terminal of proteins for pharmacological applications (*88*), fusion of proteins to cell-penetrating peptides for intracellular drug delivery (*89*), isotope labeling of protein termini for NMR

spectroscopy (*90*), synthesis and study of proteins containing different post-translational modifications such as phosphorylation, glycosidation, lipidation, ubiquitination, and acetylation (*91*). The most commonly used ELP procedure involves the ligation of a synthetic peptide containing the target modification the N-terminal of a protein (*92, 93*). EPL has also been used for the introduction of non-peptide bonds to the protein's backbone (*94, 95*).

1.2.5 Additional and specialized applications

Aside from the applications described previously, inteins have also been used for the synthesis of protein micro-arrays. In one example, the target protein is first biotinilated via EPL and then surface-immobilized (*96*). Similar micro-arrays have been constructed by directly immobilizing proteins onto solid surfaces using protein trans-splicing reactions or EPL (*97*). Intein trans-splicing reactions have enabled the conjugation of peptide ligands to adenoviruses to achieve targeted gene delivery (*98*). Other applications include the construction of high-quality *de novo* protein sequences libraries (*99, 100*). Split inteins were also used in the construction of a cell-based biosensor for the monitoring of caspase enzyme activity in living mammalian cells (*101*). In gene therapy, protein trans-splicing has the potential to overcome the size limitations of therapeutic genes in a viral genome. The adeno-associated virus vector is an efficient vehicle for therapeutic gene delivery. However, this vector can only carry genes up to 4 kbp in size. Li and co-workers divided the gene encoding the dystrophin prortein into two halves, each delivered by a separate adenovirus. Upon delivery into the

same cell, intein ligates the two halves of the protein to form the mature dystrophin therapeutic protein (*15*).

1.3 Historical perspective, limitations and current research

The elucidation of the major catalytic steps involved in the splicing reactions and the identification of multiple inteins with conserved residues (*102, 103*) allowed the development of intein-based protein purification technology (*45, 48*). A major challenge in intein-based purification technology was the high levels of premature protein cleavage in vivo, which led to the engineering of controllable inteins by directed evolution (*46*), and later to the use of artificially-split inteins (*54*).

Crystallographic studies identified independent endonuclease and splicing domains (*7*). Subsequent studies of the independent splicing domains led to the engineering of artificially-split inteins capable of catalyzing trans-splicing reactions (*14, 104*). However, artificially split inteins often have low solubility and require the use of denaturation buffer during purification (*14, 105*).

It wasn't long before the discovery of the first naturally split *Ssp* DnaE intein which allowed the purification of intein-fusion proteins under native conditions and expanded the use of intein technology (*34, 61, 106*). The major limitations of *Ssp* DnaE intein are the slow reaction kinetics and dependence of intein reaction on conserved extein residues adjacent to the splice junction (*34*). The earlier studies suggested that the *Ssp* DnaE intein requires approximately three native extein residues at each terminal, resulting in a protein containing up to six additional amino acids at the splice junction

after trans-splicing (*23*). New DnaE intein alleles were identified in different cyanobacteria species (*16*) through sequence alignment, including the *Npu* DnaE intein (*107*). *Npu* DnaE has a half-life close to one minute, a high *trans*-splicing yield of ~80% and can better tolerate non-native residues at the N-extein junction. However, the trans-splicing yield and kinetics of the *Npu* DnaE intein is heavily dependent on the C-terminal canonical CFN extein residues, resulting in a non-native footprint after ligation. Lockless and co-workers engineered *Npu* DnaE intein to have better trans-splicing efficiency with a non-native C-extein sequence (*77*). Recent studies by Cheriyan and co-workers used a similar directed evolution approach to engineer the same intein to catalyze traceless protein ligation and obtained intein variants with faster reaction kinetics against non-native extein sequence than the wild-type intein, suggesting that the native extein sequence is not always optimal (*108*). Collectively, these studies indicate that the intein trans-splicing activity can be tailored for specific applications using protein engineering approaches.

Protein trans-splicing reactions can also be used in protein chemical synthesis. However, one of the intein fragments has to be synthesized chemically along with the desired chemical modification. Due to limitation of solid state peptide synthesize, the synthesized fragment needs to be relatively small. Sun and co-workers showed that the split site of *Ssp* DnaB mini-intein is flexible and the N-terminal fragment can be as small as 11 amino acids (*17*). Using this intein, Ludwig and co-workers synthesized fluorescently labeled proteins (*109*). Although effective, the trans-splicing yield was low, mainly do the high rate of the C-terminal cleavage side reaction. Similarly, the C-

terminal fragment of the *Npu* DnaE intein can be as small as 6 amino acids (*110*). Most recently, Lin and co-workers identified a large number of 'atypical' inteins with very short N-terminal fragments (*111*). These short intein fragments are amenable to solid state synthesis, making them useful in many intein applications.

Natural intein-mediated splicing reactions are not regulated by any biological process (*56*). However, some groups have been able to control the onset of intein splicing or cleavage. Callahan and co-workers discovered a natural redox-trap that could be used to control the onset of intein activity by the addition of reducing agents (*40, 112*). Our group found that the activity *Npu* DnaE is thio-dependent (*57*). More sophisticated redox traps were subsequently developed for diverse applications (*113*). We also found that zinc ions can be used to reversibly regulate the activity of a mutant *Npu* DnaE intein (*58*). Additional intein-controlled systems include those responsive to small molecules, light, temperature or protease activity (*63, 82*).

The modification of protein internal (middle) blocks is still a difficult field of research, requiring complex sequential ligations. For this purpose, work is being developed to identify new inteins with high activity (*19*) with natural (*21*) and engineered orthogonally (*114*).

Inteins have been widely used in many biotechnology applications. The activity of intein can be tailored via protein engineering approaches. It is conceivable that new and more exciting applications of inteins will be developed in the near future.

1.4 Research objectives and organization

The focus of this research is to develop intein-based tools for applications that enable and enhance the use of enzymes in industrial settings. The highly efficient naturally split *Npu* DnaE intein was the starting point for the synthesis of protein-based hydrogels for protein immobilization. Similarly, this intein was engineered to mediate rapid and efficient tag-less protein purification.

The following chapters contain detailed descriptions of the specific projects. Chapter II contains a brief background and literature review of state-of-the art technologies involving the synthesis of polymer and protein based-hydrogels and enzyme immobilization. Current advances and challenges are addressed. A brief background on additional techniques for the purification of tag-less proteins is also given in chapter II.

Chapter III describes the design, synthesis, characterization and applications of an intein-triggered protein hydrogel. The resulting hydrogel is shown to have unprecedented stability, and is the first protein-based hydrogel used as a general enzyme immobilization scaffold. The excellent mechanical properties obtained were due to the intein mediated trans-splicing reactions that resulted in peptide bond formation. It was established that such peptide bonds allowed stable hydrogel formation independent of functionality.

In Chapter IV, we designed a novel, intein-free, system suitable for covalent bond formation. We envisioned that the high stability of the intein-mediated protein hydrogel was due to covalent bonds connecting trimeric protein cross-linkers. We

engineered a protein-ligand pair by site-directed mutagenesis to form disulfide bond with each other and used this pair in a protein hydrogel.

Chapter V describes the engineering, synthesis and successful application of the *Npu* DnaE intein for rapid and efficient tag-less protein purification. The Appendix contains additional characterization and optimization of the engineered *Npu* DnaE intein for tag-less protein purification. Findings shown in the Appendix were a direct continuation of the research presented in Chapter V and were carried out mainly by Dr. Dongli Guan, in which I collaborated in some experimental and conceptual steps.

1.5 Novelty

Protein-based hydrogels have been used as artificial scaffolds in tissue engineering, drug delivery, and biosensors (*115*). Many protein hydrogels are triggered by self-assembling interactions of different protein building blocks such as coiled-coils (*116*) and elastin-like polypeptides (*117*), and the gelation is driven by changes in pH, temperature, ionic strength or hydrophobicity (*118*). The coiled-coil artificial protein hydrogel have been used for proteins immobilization (*119, 120*), but the resulting gel showed low physical stability. To circumvent the use of reversible self-assembling peptides for hydrogel synthesis, we used inteins to generate covalent bonds between different cross-linkers. This approach enabled the synthesis of a highly stable hydrogel. Incorporation of docking station peptides in the hydrogel building blocks enabled the stable immobilization proteins in the hydrogel.

In the second study, we engineered the intein to catalyze an inducible C-terminal cleavage reaction and used it for tag-less protein purification. Compared to the previous technology that used the artificially split *Ssp* DnaB intein (*55*, *121*), this technology enables more rapid and efficient protein purification.

CHAPTER II

LITERATURE REVIEW

2.1 Hydrogels

Hydrogels are materials composed of polymeric networks dispersed in water with a soft and solid appearance (*122, 123*). In a typical hydrogel, over 90% of the mass corresponds to water, making these materials versatile platforms for diverse applications (*124-126*). Hydrogels are widely used in drug delivery (*127, 128*), tissue engineering (*129*), biosensors (*130*), enzyme immobilization (*119, 131*), self-healing materials (*132*), and bio-mimicking materials (*133*).

Most existing hydrogels are made of hydrophilic polymer chains capable of chemical or physical cross-linking. These polymer networks not only retain water but also provide the hydrogel with structural integrity and functionality (*123*). Polymer chemistry methods have been traditionally used for the synthesis of hydrogels with specific and uniform bulk properties (*123*). However, these methods cannot produce hydrogels with predetermined spatial architecture. Bulk polymer hydrogels have found extensive use in medical applications including implants (*134*) and contact lenses (*133*). Aside from the specialized applications of polymer hydrogels, hydrogels with superior mechanical properties have also been developed (*132, 135*).

Physical hydrogels self-assemble from block-copolymers driven by physical interactions (*136, 137*). These materials can undergo physical changes in response to external stimuli including changes in pH, temperature, light, solvent, electrical and

magnetic fields, and have found an extensive list of applications (*123*). Some physical hydrogels are self-assembling materials composed entirely of genetically engineered protein block-copolymers (*116, 138*). These protein block-copolymers are capable of assembling into three-dimensional structures via intramolecular folding (*137*), self-assembly of coiled-coil motifs (*116, 139, 140*), assembly of elastin-like proteins (*141-143*) or affinity of multimeric protein interactions (*144*). These physical systems possess interesting properties such as stimuli-responsiveness, reversible hydrogel formation (*145*) and injectability (*146*). The ability to precisely control the primary sequence of protein hydrogel building blocks enables the design of specific three-dimensional arrangements, resulting in protein hydrogel with pre-determined three-dimensional structure, properties and specific functionalities (*147*).

2.2 Enzyme immobilization

Despite the superb properties of enzymes, their widespread use as catalysts has been limited by poor long-term stability, high production costs, and difficulties in recovery. Immobilization of enzymes on solid support has been shown to improve their stability and allow recovery and recycle (*148*). Enzyme immobilization can be achieved through covalent bond formation between the protein surface cysteine or lysine residues and the supporting scaffolds (*149-159*), or by encapsulation of enzymes inside a polymer matrix or gel (*160-167*). Other methods for protein immobilization include inter-enzyme cross-linking (*168-174*), expressed protein ligation (*175*), Staudinger ligation (*176*) and "click" ligation (*177, 178*). However, most of these techniques require chemical

modifications of protein surface residues, which tend to reduce the enzyme's activity. In addition, none of these techniques is general, and individual optimization is required.

Protein immobilization in hydrogels has been explored by the Banta group using bioactive protein hydrogels (*119, 120*). The hydrogels are self-assembled from multimetic protein block copolyers and are relatively stable. Although efficient, this technique is not general and hydrogel formation is driven by changes in temperature or pH, limiting the range of conditions in which the hydrogels can be used.

2.3 Tag-less protein purification

Following protein purification, many applications require the removal of purification tags. This process is usually done via precision proteases such as the tobacco etch virus (TEV) protease (*43*). However, tag-removal by proteases often requires extended incubation periods and additional purification step(s) to remove the purification tag and the protease (*179*). Therefore, tag-removal remain costly for large scale protein purification (*180*), making the development of alternative protease-free tag removal technologies highly desirable.

In the last few years, different self-cleaving proteins have been adapted for the removal of purification tags in a number of applications. The self-excising FrpC protein from *Neisseria meningitides* (*181*) undergoes calcium responsive cleavage at an specific cite and has been fused with the CBD domain as self-cleaving tags. After cleavage, the target protein contains an additional aspartic acid at the C-terminal cite. This technique is effective, with almost 100% cleavage efficiency after 6 hours at room temperature

(*182*). The sortase A transpeptidase (SrtAc) (*183*) has also been used as a self-cleaving protein. The activity of this protein is controllable by calcium ions and cleavage can be achieved within 6 hours at room temperature. However, significant premature cleavage is often observed *in vivo*. The most widely used protease-free methodology for tag removal is based on engineered intein cleavage, as described in Section 1.2.1.

CHAPTER III

INTEIN-TRIGGERED ARTIFICIAL PROTEIN HYDROGELS THAT SUPPORT THE IMMOBILIZATION OF BIOACTIVE PROTEINS*

3.1 Overview

Protein hydrogels have important applications in tissue engineering, drug delivery and biofabrication. We present the development of a novel self-assembling protein hydrogel triggered by the mixing of two soluble protein block copolymers, each containing one half of a split intein. Mixing of these building blocks initiates an intein trans-splicing reaction that yields a hydrogel. This intein-triggered hydrogel is highly stable in a wide range of pHs (6-10) and temperatures (4-50 °C), instantaneously recovers its mechanical properties after shear-induced breakdown, and is compatible with both aqueous and organic solvents. Incorporation of a "docking station" peptide (DSP) into the hydrogel building blocks enables simple and stable immobilization of docking protein (DP)-fused bioactive proteins in the hydrogel. This intein-triggered protein hydrogel technology opens new avenues for both in vitro metabolic pathway construction and functional/bi compatible tissue engineering scaffolds and provides a convenient platform for enzymes immobilization in industrial biocatalysis.

---

* Reprinted with permission from "Intein-triggered artificial protein hydrogels that support the immobilization of bioactive proteins" by Miguel Ramirez, Dongli Guan, Victor Ugaz, and Zhilei Chen, 2013, Journal of the American Chemical Society, Copyright 2013, American Chemical Society, doi: 10.1021/Ja401075s.

3.2 Introduction

Enzymes are versatile catalysts due to their superior chemo-, regio- and stereo-specificity. However, lack of long-term stability under process conditions and difficulties in recovery and recycling have greatly hampered the usefulness of enzymes in industrial processes. Immobilization has been shown to alleviate some of these limitations (*148, 184*). In most cases, to achieve stable immobilization, target enzymes need to be covalently linked to a supporting matrix through a chemical conjugation reaction involving the side chain functionalities of the amino acids cysteine and lysine. The properties of the bio-conjugates thus generated are influenced by the frequency and location of lysine/serine residues, and are highly variable depending on the target enzymes (*148*). Recently, a number of protein derivatization techniques have been developed to achieve site-specific protein immobilization, including expressed protein ligation (*175*), Staudinger ligation (*176*) and "click" ligation (*177, 178*). However, given that it is often difficult to determine *a priori* which sites in a protein are essential for a function versus available for modification, there remains a major demand for the development of new and more general technologies for high-density, high-activity enzyme immobilization on solid supports.

In this study, specifically, we demonstrated the synthesis and application of a split-intein triggered protein hydrogel as a general scaffold for enzyme immobilization. The split-intein-containing protein building blocks we used for hydrogel formation are the copolymers CutA-*Npu*N (N) and *Npu*C-S-CutA (C). *Npu*N/C are the N-/C-fragments of the naturally split DnaE intein from *Nostoc punctiforme* (*Npu*). We chose this intein

due to its extraordinarily quick reaction kinetics ($t_{1/2}$ = 63 seconds) and very high *trans*-splicing yield (75-85%) (*21, 185*). CutA, a small trimeric protein (12 kDa) from *Pyrococcus horikoshii*, was used as the cross-linker protein (*186, 187*). CutA has an extremely high denaturation temperature of near 150 °C and retains its trimeric quaternary structure in solutions containing as much as 5 M GuHCl (*186*). The ultra-high stability of CutA is attributed to its very large number of intra- and inter-subunit ionic pairs forming extensive ion-pair networks (*188*). We reasoned that very strong inter-subunit interactions should discourage subunit exchange between different cross-linkers, thus minimizing hydrogel surface erosion via the formation of closed loops (*189*). A soluble (S) fragment ($[(AG)_3PEG]_{10}$) (*116, 190*) a flexible polyanionic linker, was incorporated as the midblock for water retention. Mixing of N and C initiates a *trans*-splicing reaction between *Npu*N and *Npu*C, generating a longer protein chain with the cross-linker CutA at both termini. Cross-linkers from multiple such molecular units interact with each other, forming a highly cross-linked gel-like network.

3.3 Materials and methods

3.3.1 Chemicals

All chemicals were reagent grade and purchased from either Sigma-Aldrich (St. Louis, MO), VWR International (Radnor, PA), or Thermo Fisher Scientific (Pittsburgh, PA) unless otherwise stated. *Escherichia coli* DH5α (Invitrogen, Grand Island, NY) was used for recombinant DNA cloning and manipulation. *Escherichia coli* BL21(DE3) (Novagen, Madison, WI) was used for the expression of recombinant proteins.

3.3.2 Plasmid construction

Table 3.1 contains a list of the construct used in this study. To generate NpuC-S-CutA (C): the NpuC gene was amplified from plasmid KanR-IntRBS-NpuNC-CFN (*77*) (gift from Prof. Tom Muir, Princeton University) using primers NpuC_F_NdeI and NpuC-R-HindIII, and inserted into pet26-IsceI plasmid (*191*) digested with NdeI and HindIII to generate pI-NpuC; the CutA gene was amplified from pet30-CutA-Tip1 (*192*) (gift from Dr. Takehisa Matsuda, Kanazawa Institute of Technology, Hakusan, Ishikawa, Japan) using primers SnaBI-CutA-F and XhoI-CutA-R and inserted in pI-NpuC digested with SnaBI and XhoI to generate pI-NpuC-CutA; the S fragment $[AG_3(PEG)]_{10}$ was amplified from plasmid pQE9 AC_{10}Atrp (*139*) (gift from Prof. David Tirrell, California Institute of Technology) using primers HindIII-[C10]-F and SnaBI-[C10]-R and inserted into pI-NpuC-CutA digested with HindIII and SnaBI to generate C.

Table 3.1 Protein constructs for intein hydrogel synthesis

Short name	Protein sequence	Molecular weight (kDa)
CutA-*Npu*N (N)	CutA-EAC-$(GGGGS)_2$-AS-NpuN-HHHHHH	26.3
*Npu*C-S-CutA (C)	NpuC-CFNKLYRDPMG- $[(AG)_3PEG]_{10}$-ARMPYV-CutA-HHHHHH	26.1
NpuC-S-$SH3_{lig}$-CutA (C-$SH3_{lig}$)	NpuC-CFNKLYRDPMG- $[(AG)_3PEG]_{10}$-ARMPYVGS-PPPALPPKRRR-$(GGGGS)_2$-AS-CutA-HHHHHH	28.3
SH3-GFP	SH3-KL-$(GGGGS)_2$-AS-GFP-HHHHHH	34.5

To generate NpuC-S-$SH3_{lig}$-CutA (C-$SH3_{lig}$): CutA gene was amplified from pet30-CutA-Tip1 (*192*) with primer L-NheI-CutA-F and XhoI-CutA-R to generate $(GGGGS)_2$-CutA; this PCR product was further amplified using primers SH3lig-L-F and XhoI-CutA-R to generate $SH3_{lig}$-$(GGGGS)_2$-CutA. Finally, this PCR product was amplified using primers SnaBI-BamHI-SH3-F and XhoI-CutA-R to introduce the appropriate restriction sites and inserted into C between SnaBI and XhoI restriction sites.

To generate CutA-NpuN (N): the CutA gene was amplified from pet30-CutA-Tip1 (*192*) using primers NdeI-CutA-F and HindIII-CutA-R and inserted into pet26-IsceI plasmid (*191*) digested with NdeI and HindIII to generate pI-CutA; NpuN was amplified from plasmid KanR-IntRBS-NpuNC-CFN (*77*) using primers HindIII-Link-Npu F and NpuN_R_XhoI to generate $(GGGGS)_2$-NpuN; and inserted into pI-CutA digested with HindIII and XhoI to generate N.

To generate SH3-GFP: the SH3 gene was amplified from plasmid pJD757 (*193*) (a gift from Dr. Jay D. Keasling, University of California, Berkeley) using primers NdeI-SH3-F and HindIII-SH3-R and inserted into pet26-IsceI plasmid (*191*) digested with NdeI and HindIII to generate pI-SH3. GFP was amplified with primers HindIII-L-GFP fwd and XhoI_GFP_R from pET26-GFP plasmid and inserted into pI-SH3 digested with HindIII and XhoI resulting in SH3-GFP. Table 3.2 contains the DNA sequence of the primers used for the previously mentioned cloning procedures.

Table 3.2 Primers used for intein hydrogel constructs

Primer	5’ to 3’ sequence
NpuC_F_NdeI	TTAGAAGGCATATGATCAAAATAGCCACACGTAAA TATTTAGG
NpuC_R_HindIII	ATTCGCAAGCTTATTGAAACAATTAGAAGCTATGAA GCC
HindIII_[C10]_F	TATTAAAAGCTTTATCGCGATCCGATGGGT
SnaBI_[C10]_R	TAATTATACGTACGGCATGCGGGC
SnaBI-CutA-F	ACACTGTTACGTAATAATAGTTTACACGACTTTTCC GG
XhoI-CutA-R	TAAAATCTCGAGTTTTTTCGTCTCTTCAATTAACCAT TT
L-NheI-CutA-F	GAGGCGGAGGGAGCGGAGGCGGAGGGAGCGCTAG C ATAATAGTTTACA CGACTTTTCCGG
SH3lig-L-F	CGCCGGCGCTGCCGCCGAAACGTCGTCGCGGAGGC G GAGGGAGCGG
SnaBI-BamHI-SH3-F	TGCCGTACGTAGGATCCCCGCCGCCGGCGCTGCC
NdeI-CutA-F	GGAATTCCATATGATAATAGTTTACACGACTTTTCC GG
HindIII-CutA-R	GGAATTCAAGCTTCTTTTTTCGTCTCTTCAATTAACC ATTT
HindIII-Link-Npu F	CCTGGAAGCTTGTGGAGGCGGAGGGAGCGGAGGCG GAGGGAGC GCTAGC TGTTTAAGCTATGAAACGGAAATATTGAC
NpuN_R_XhoI	ATATAGCTCGAGATTCGGCAAATTATCAACCCG
NdeI-SH3-F	TTGAAATTTACATATGGCAGAGTATGTGCGGG
HindIII-SH3-R	TAACTCCAAGCT TATACTTCTCCACGTAAGGGA
HindIII-L-GFP fwd	TTCAATAAGCTTGGAGGCGGAGGGAGCGGAGGCGG AGGGAGC GCTAGC GTGAGCAAGGGCGAGG
XhoI_GFP_R	TAA AATCTCGAGTAACTCGTCCATGCCGAGAG

3.3.3 Protein expression and purification

Escherichia coli BL21(DE3) was transformed with the appropriate expression plasmid and plated on an agar plate containing 50 μg/ml kanamycin. The next day, all colonies (50~100) were collected and transferred to 1L Luria-Bertani (LB) Broth containing kanamycin (50 μg/ml) and grown at 37 °C until OD_{600} ~0.8. For C and C-$SH3_{lig}$, protein expression was induced at 37 °C for 4 hours by the addition of Isopropyl β-D-1-thiogalactopyranoside (IPTG, 1 mM). All the other constructs were expressed at 18 °C for 14 hours with 1mM IPTG. After expression, cells were harvested by centrifugation at 6,000 *x g* at 4 °C for 20 minutes and stored at -80 °C until use.

Table 3.3 Buffer compositions for intein hydrogel process

Buffer	Composition
Buffer A	500mM NaCl, 10mM Tris-HCl, pH 8.0
Buffer DA	500mM NaCl, 8M Urea, 10mM Tris-HCl, pH 8.0
Buffer B	500mM NaCl, 50mM NaPOi, pH 6.0
Buffer DW	500mM NaCl, 8M Urea, 45mM Imidazole, 10mM Tris-HCl, pH 8.0
Buffer W	500mM NaCl, 45mM Imidazole, 50mM NaPOi, pH 6.0
DPBS	Dulbecco's PBS, 137.9mM NaCl, 2.7mM KCl, 1.5mM KH_2PO_4, 8.1mM Na_2HPO_4, pH 7.4
DMEM	Dulbecco's Modified Eagle Medium, Lonza, Allendale, NJ 07401, cat # 12-604F, with 0.5% NaN_3
Buffer E	DPBS supplemented with 5 mM DTT and 0.5% NaN_3
EB1	DPBS supplemented with 0.5 % NaN_3
EB2	137mM NaCl, 50mM NaPOi, 0.5% NaN_3, pH 6.0
EB3	137mM NaCl, 10mM Tris-HCl, 0.5% NaN_3, pH 10.0

Table 3.3 contains the composition of the different buffers used in this study. For the purification of N, cell pellets were resuspended in Buffer A at 10 mL per gram of wet pellet, and disrupted by sonication (QSonica Misonix 200, Amp 10, 16-20W, with 1

second pulse 6 seconds pause for 1 minute). The lysate was centrifuged at 16,000 x g for 20 minutes at 4 °C. The supernatant was discarded and the pellet was resuspended in Buffer DA and centrifuged at 16,000 *x g* for 20 minutes at 4 °C. The resulting supernatant was passed through a 5-mL Ni Sepharose™ High Performance HisTrap column (GE Healthcare Life Sciences, Piscataway, NJ) using a BioLogic LP System (Bio-Rad, Hercules, CA). The column was washed with Buffer DW and the purified protein was eluted in Buffer DA supplemented with 150 mM imidazole. Purified protein was buffer exchanged into DPBS buffer via a 30-kDa ultra-filtration spin column (Amicon Ultra, Millipore, Billerica, MA) and concentrated to ~100 mg/mL via the same column in DPBS supplemented with 2mM Dithiothreitol (DTT) and stored at -80 °C until use. Storage of purified protein without removing imidazole can significantly reduce the intein catalytic activity.

For the purification of C and C-$SH3_{lig}$, cell pellets were resuspended in Buffer B supplemented with 1x protease inhibitor cocktail (Roche Applied Science, Indianapolis, IN). Acidic buffer (pH 6) was used to minimize proteolytic degradation. The cells were disrupted by sonication. The soluble lysate was passed through a 5-mL Ni Sepharose™ High Performance HisTrap column using a BioLogic LP System, washed with Buffer W (0.5M NaCl, 50 mM NaPOi, 45 mM Imidazole, pH 6) and the target protein was eluted in Buffer B supplemented with 150 mM imidazole. Purified protein was buffer exchanged into phosphate buffer (50 mM NaPOi, pH 7.0) and loaded onto a 5-mL HiTrap SP Sepharose FF ion exchange column (GE Healthcare Life Sciences, Piscataway, NJ). Target protein eluted at ~200mM NaCl (50 mM NaPOi, pH 7.0), and

was buffer exchange into DPBS buffer via a 30-kDa ultra-filtration spin column and concentrated to ~100mg/ml via the same column in DPBS buffer supplemented with 2mM DTT and stored at -80 °C until use.

SH3-GFP was purified from the soluble fraction via a single pass in Ni Sepharose™ High Performance HisTrap column in Buffer A. Purified protein was buffer exchanged into DPBS via a 30-kDa ultra-filtration spin column and concentrated in DPBS buffer to ~150 mg/mL and stored at -80 °C.

For SDS-gels, sample protein was first diluted in distilled water to reduce the concentration of NaCl to ~1 mM. We found that at this NaCl concentration the CutA trimer proteins run as monomers in the SDS-gels. The diluted samples were mix with 2x SDS sample buffer (0.5 M Tris-HCl, pH 6.8, 20% Glycerol, 10% w/v SDS, 0.1 % w/v bromo-phenol blue, 2% β-mercaptoethanol), incubated at 95 ˚C for 5 minutes and analyzed using 12% SDS-PAGE gels. The gels were stained with Coomassie brilliant blue R250.

3.3.4 Rheological characterization

Rheological measurements were performed using a Paar-Physica MCR-300 parallel plate rheometer (Anton Paar, Ashland, VA) with a 25 mm plate fixture (PP25). Purified C was manually mixed with N on the rheometer plate by swirling motion. The final mixture contained a total of 200 µL protein at 1.6 mM N/C in Buffer E. The measuring gap was set at 0.2 mm. Excess protein was removed from the side of the fixture. The outer edge of the sample was coated in mineral oil in order to minimize

sample evaporation. The samples were incubated overnight to ensure hydrogel formation. The next day, strain sweep and frequency sweep were measured at 10 rad/sec frequency and 10% strain amplitude, respectively. Large-amplitude oscillatory shear (LAOS) was probed at 10 rad/s frequency.

3.3.5 Hydrogel erosion studies

To determine the stability of the intein-triggered hydrogel at different temperatures, purified C and N (total 50 µL in Buffer E) were manually mixed in a 1.7 mL micro-centrifuge tube by a swirling motion using a pipette tip. The hydrogel (1.6 mM J) was briefly centrifuged at 6000 x *g* to remove the air bubbles, and incubated at room temperature overnight to ensure gel formation. Erosion was initiated by the additional of 500 µL of Erosion Buffer 1 (EB1) to the hydrogel and the entire tube was incubated at the specified temperature. At specified times, EB1 was removed and replaced with 500 µL fresh EB1. The amount of protein in EB1 was determined using the Bradford method (Coomassie Plus Bradford Assay Reagent, Pierce Protein Biology Products, Rockford, IL). To determine the stability of intein hydrogel at acidic and basic conditions, hydrogel erosion was carried in EB2 and EB3, respectively.

3.3.6 Hydrogel stability in Dulbecco's Modified Eagle Medium (DMEM)

To determine the stability of hydrogel in DMEM, and the ability of DMEM to diffuse into the hydrogel, a 50 µL hydrogel in Buffer E (1.6 mM J) was immersed in 500 µL DMEM at room temperature. Images were taken immediately and 16 hours later.

3.3.7 Hydrogel synthesis in DMEM

One hundred microliters of concentrated protein (N and C) was diluted in 1 mL of DMEM (Table 3.3) and dialyzed overnight in 50 mL of DMEM at 4°C using a 12000 to 14000 kDa cut-off dialysis membrane (Spectrum Laboratories Inc, Rancho Dominguez, CA). The dialyzed proteins were concentrated via 10-kDa ultra-filtration spin columns to ~100mg/mL (estimated based on initial concentrations). Concentrated proteins were mixed at ~1:1 molar ratio in the presence of 5mM DTT. Hydrogel formation was confirmed via upside-down tube test. We were also unable to pipette the solid hydrogel using a pipette tip.

3.3.8 Hydrogel pore size estimation

Hydrogel building block C was first mixed with 20 kDa fluorescently labeled FITC-dextran molecules (FD20S, Sigma-Aldrich, St. Louis, MO) or pyranine (TCI America, Portland, OR) in Buffer E inside a 1.7 mL micro-centrifuge tube. Building block N was added to the mixture to induce hydrogel formation as described above. The total volume of hydrogel (1.6 mM J) was 50 µL and the final concentrations of FITC-dextran and pyranine in the hydrogel were both 0.1%. After hydrogel formation, 500 µL EB1 was added to the hydrogel to initiate the diffusion of entrapped molecules. EB1 was replaced periodically and the amount of FITC-dextran/pyranine present in EB1 was quantified by measuring the fluorescence intensity in a SpectraMax Gemini EM (Molecular Devices, Sunnyvale, CA) plate reader with excitation/emission wave lengths of 485/538 nm. For the control experiments using agarose gels, 0.5 g of agarose was

dispersed in 50 mL DPBS buffer and melted in a microwave oven. The solution was cooled to ~50 °C and 47.5 uL of the solution was added to a 1.7ml centrifuge tube containing 2.5 µL of 20X- FITC-Dextran or pyranine solutions (20 mg/mL). The mixture was briefly centrifuged at 6000 x g and incubated overnight. The next day, 500 µL EB1 was added to the gel to initiate diffusion of entrapped molecules.

3.3.9 GFP leaching studies

SH3-GFP was manually mixed with C-$SH3_{lig}$ or C at 1:1 molar ratio in Buffer E in a 1.7 mL micro-centrifuge tube. The mixture was incubated at 22 °C for 30 minutes. N was then added to the same tube and the solution was manually mixed with a pipette tip. The final mixture contains 1:1:1 molar ratio of SH3-GFP:C-SH3lig/C:N at 1.2 mM each component in 50 µL buffer E. This mixture was incubated at room temperature overnight. The next day, EB1 (500 µL) was added to the hydrogel. EB1 was replaced with fresh EB1 periodically at specified times. The amount GFP in EB1 was quantified by measuring the fluorescence intensity in a SpectraMax Gemini EM plate reader (Molecular Devices, Sunnyvale, CA) with excitation/emission wave lengths of 485/538 nm. The amount of total protein in EB1 was quantified via Bradford method.

3.3.10 Enzymatic reaction studies of HRP-containing hydrogel

Purified C was mixed with horseradish peroxidase (HRP) (Sigma-Aldrich, St. Louis, MO) in Buffer E in a 1.7 mL micro-centrifuge tube. Purified N was then added to the same tube and the solution was manually mixed using a pipette tip. The final mixture

contains 1.6 mM J and 0.042 mM HRP in 30 µL total volume. After hydrogel formation, all tubes were left at 4 °C. At different times, then hydrogel was submerged in 1 mL of n-heptane and manually disrupted into small piece to increase the contact surface area. For control samples, the same amount of HRP was dissolved in 30 µL Buffer E and the experiment was carried out as described above. All tubes were equilibrated to room temperature (~30 minutes) prior to the addition of substrate. Then, the n-heptane was replaced with 500 µL reaction cocktail containing *N*,*N*-dimethyl-*p*-phenylene diamine (5.8 mM), phenol (5.8 mM) and *tert*-butyl hydroperoxide (2.9 mM) in n-heptane (*194*). The formation of product, an indophenol-type dye, was monitored by measuring optical absorbance of the supernatant at 546 nm every 30 minutes for 2 hours in a transparent 96-well plate using a SpectraMax 340PC384 plate reader (Molecular Devices, Sunnyvale, CA).

3.4 Results

For hydrogel formation, purified N and C were manually mixed at 1:1 molar ratio via a swirling motion using a pipette tip in Dulbecco's PBS (DPBS) supplemented with 5 mM DTT (Table 3.3: Buffer E). DTT serves to activate the intein *trans*-splicing reaction between N and C (*57, 195*), triggering the formation of protein fragment J: CutA-S-CutA (Fig. 3.1A, C). However, due to the very high affinity between the intein N- and C-fragment, hydrogel can also form in the absence of DTT. The final concentration of cross-linked protein (J) is ~5% w/v (combined mass of N and C minus that of the spliced inteins) or 1.6 mM. Individually, the N and C proteins are in the form

of viscous liquid, but when the fragments are mixed, a gel-like material is formed (Fig. 3.1B). Densitometric analysis of SDS-PAGE gels revealed that ~80% of the input protein successfully underwent the *trans*-splicing reaction (Fig. 3.1C).

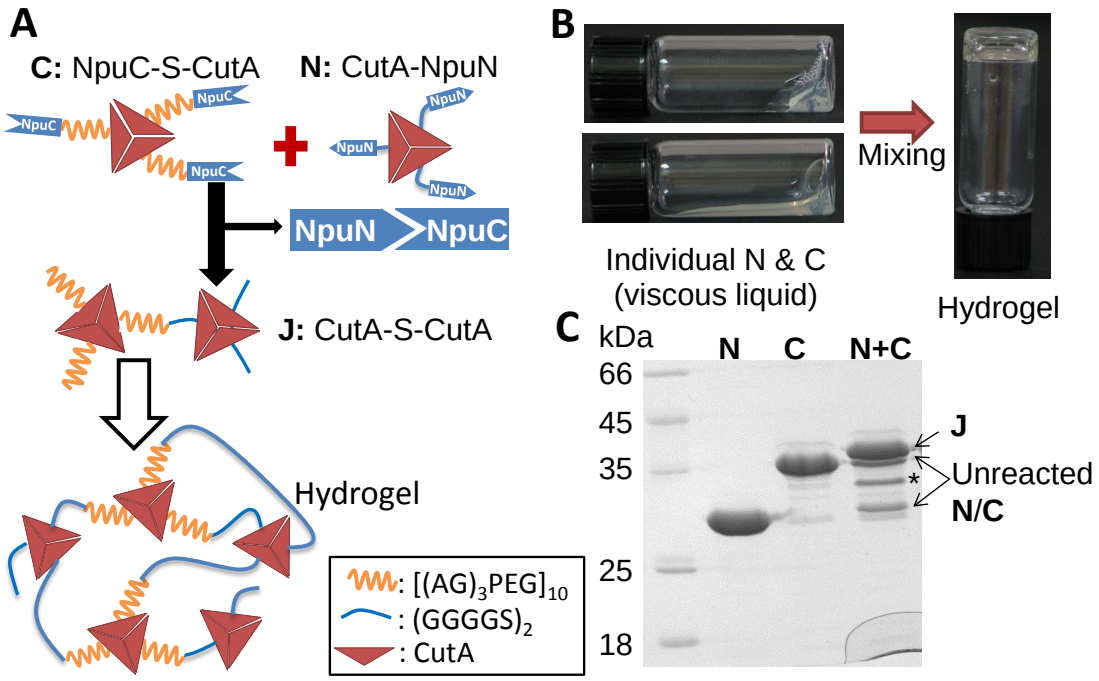


Figure 3.1 Intein-mediated protein hydrogel. (A) Schematics of intein trans-splicing reaction that triggers the formation of an extended protein chain (J) with cross-linker proteins at both termini. Cross-linker proteins from multiple J protein chains non-covalently interact with each other upon intein-mediated protein ligation, inducing the formation of a highly cross-linked protein network with hydrogel properties. NpuN/C: intein N-/C-fragment. (B) Mixing of purified N and C (8.3 % w/v) leads to the formation of a highly cross-linked hydrogel network (1.6 mM J). (C) SDS-PAGE analysis of purified N and C building blocks before and after mixing. "N+C" corresponds to a sample taken directly from a 1.6 mM hydrogel. "+" denotes an intein C-terminal cleavage side reaction product.

The gel-like material formed upon mixing N and C was confirmed to be a hydrogel by rheological analysis which showed a substantially greater plateaued storage modulus (G') compared to loss modulus (G") over a wide range of frequencies and strains (Fig. 3.2A, B).

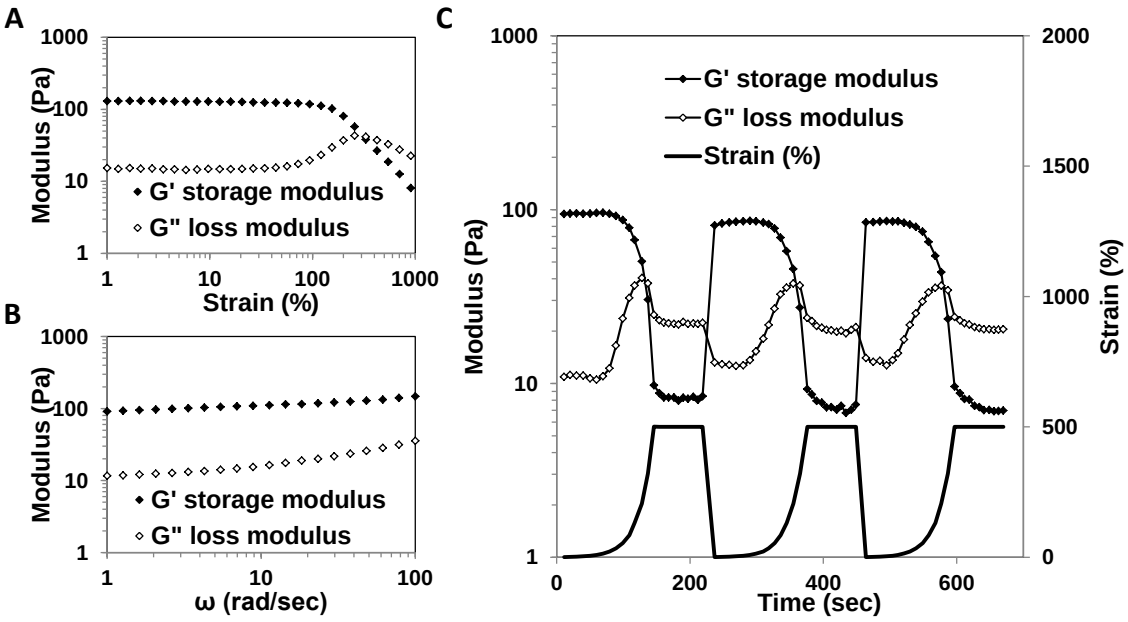


Figure 3.2 Rheological characterization of a hydrogel with 1.6 mM J. (A) Strain sweep at 10 rad/sec. (B) Angular frequency sweep at 10% strain. (C) Large-amplitude oscillatory shear cycles at 10 rad/sec.

An intein-mediated hydrogel with 1.6 mM J is able to retain elastic properties at ~100% strain amplitude (Fig. 3.2A). At strains >100%, G' decreases while G" increases until the two values cross each other and the hydrogel transitions from a predominantly elastic material to a viscous fluid. The increase in energy dissipation seen in G" at this transition point suggests the disruption of non-covalent associations within the network (*146*), likely the dissociation of CutA trimer into monomers. The plateau G'_{∞} of a

hydrogel with 1.6 mM J is 110 ± 27 Pa. Measurements were made on three separate samples, and errors are standard deviations. The hydrogel mechanical properties are influenced by the choice of both the cross-linker and the midblock (*146*). The cross-linker structure governs the number of intermolecular interactions in the hydrogel. Previously, it was shown that a hydrogel with tetrameric cross-linkers has lower G'_{∞} (~400 Pa) than one with pentameric cross-linkers (G'_{∞} >1000 Pa) (*189*). The trimeric cross-linker structure of this intein hydrogel may be partially responsible for the low G'_{∞} and a higher G'_{∞} may be obtained by using a different cross-linker protein that exhibits a higher order of multimerization.

Dynamic rheological characterization revealed that our intein-mediated hydrogel instantaneously recovers its mechanical properties after shear-induced breakdown (Fig. 3.2C). A small but permanent loss of elastic modulus (10%) is seen after the first cycle of large-amplitude oscillatory shear (LAOS). However, the G' does not decrease further after subsequent cycles of LAOS. The very rapid recovery kinetics is attributed to the use of a single protein, CutA, as the cross-linker allowing the disengaged monomer to very easily associate with other monomers to reform new trimer cross-linkers. Hydrogels with non-symmetrical cross-linkers typically exhibit slower recovery kinetics (*196*). The ability of the intein-triggered hydrogel to recover its mechanical properties after LAOS points to the injectability of this hydrogel which is highly desirable for controlled drug delivery and tissue engineering applications (*197, 198*).

The intein-triggered protein hydrogel (1.6 mM J) exhibits remarkable stability in aqueous solution. After 21 days at 22 °C, the total amount of protein released into the

solution only slightly exceeds the theoretical amount of the spliced out intein (calculated based on 100% *trans*-splicing efficiency), indicating little to no loss of cross-linked hydrogel scaffold (Fig. 3.3A). This result is confirmed by SDS-PAGE analysis which shows only trace amounts of the *trans*-spliced product in the surrounding buffer (Fig. 3.3B, band J). The vast majority of the protein in the surrounding buffer corresponds to the spliced intein that diffused out of the hydrogel. The hydrogel volume expanded significantly in the first 24 hours and slightly in the first week due to hydrogel swelling, but did not detectably change after that, likely due to negligible loss of cross-linked hydrogel scaffold. An undisturbed hydrogel is stable in aqueous solution at room temperature for over 3 months with essentially no erosion (Fig. 3.3A inlet). Small isolated air bubbles that occasionally became trapped at the surface of the hydrogel during hydrogel formation remained after 3 months, indicating that this hydrogel does not undergo surface erosion due to closed loop formation (*189, 199*).

We next determined the stability of this hydrogel under various conditions. The hydrogel exhibits a similar erosion profile in pH 6.0 and pH 10.0 buffer (Fig. 3.4A) and a hydrogel with as low as 0.8 mM J (2.7% w/v) was able to retain ~40% of its initial mass after 7 days in DPBS buffer at room temperature (Fig. 3.4B). The hydrogel is also stable at 37 °C but less so at 50 °C (Fig. 3.3C). The elevated hydrogel erosion rate at high temperatures is somewhat surprising since CutA retains its trimeric quaternary structure at temperatures near 150 °C (*186*).

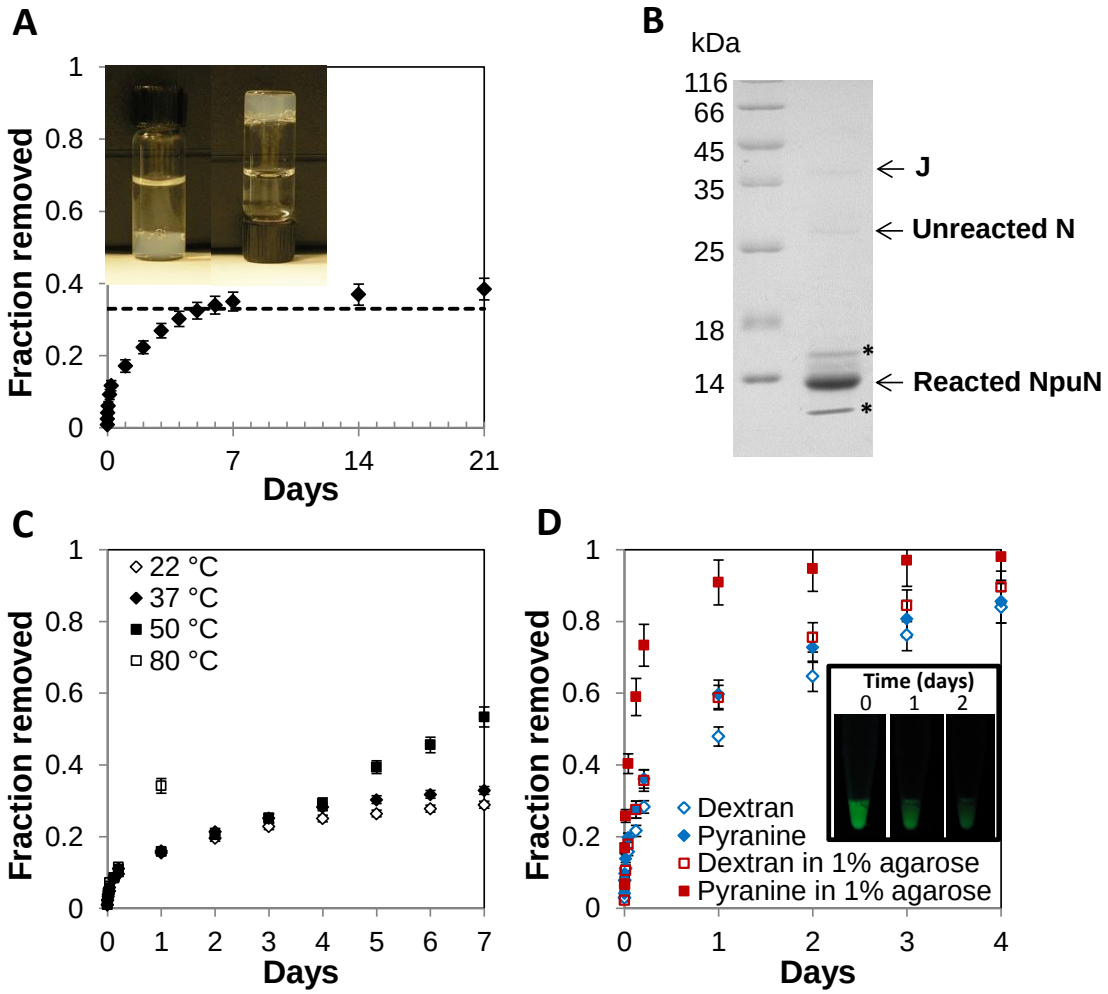


Figure 3.3 Stability of intein-mediated hydrogels in DPBS. A) Erosion of a 1.6 mM hydrogel at 22°C. Dotted line represents the theoretical mass corresponding to the cleaved inteins. The inlet shows an undisturbed hydrogel in DPBS after 3 months at room temperature. B) SDS-PAGE analysis of hydrogel's surrounding buffer. All the samples of the buffer in which the hydrogel was immersed in (A) were pooled (total 7.5 ml) and concentrated 75-fold via ultrafiltration through a 10 kDa membrane prior to gel loading. J: intein *trans*-spliced product. N: unreacted CutA-NpuN. NpuN: spliced out N intein fragment. Unreacted C and spliced out NpuC are not visible from the gel due to the small quantity and small size (4 kDa), respectively. The asterisk denotes unidentified bands. (C) Erosion profile of hydrogel incubated at different temperatures. (D) Diffusion kinetics of FITC-dextran (20 kDa) and pyranine (524 Da) from a 1.6 mM hydrogel and 1% agarose gels. The inlet shows the hydrogel containing pyranine under UV exposure at three time points. The error bars represent the standard deviation of 2 independent experiments.

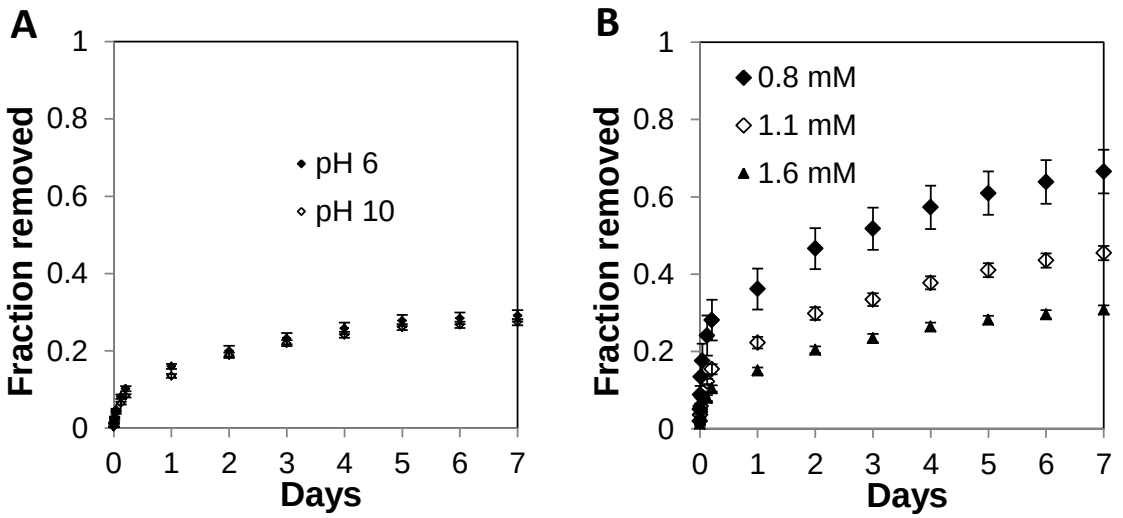


Figure 3.4 Erosion profiles of intein-triggered hydrogels under different conditions. A) Erosion profiles of a 1.6 mM hydrogel at pH 6 and pH 10. Hydrogels (50 µL) in micro-centrifuge tubes were immersed in 500 µL EB2 (pH 6.0) or EB3 (pH 10.0). The entire supernatant was removed and replaced with 500 µL of the same buffer at the specified times. B) Erosion profiles of intein-triggered hydrogels with different protein concentrations. These studies were carried out in Buffer EB1.

The normalized plateau storage modulus G'_∞/nkT (where G'_∞ is the plateau storage modulus, n is the chain number density, k is the Boltzmann constant and T is temperature) of our intein hydrogel was only 0.024, indicating that most of the cross-linkers are not productively connected and that there are extensive amount of independent loops in the hydrogel (*189*). Since the intein-triggered hydrogel exhibits little-to-no surface erosion, one possible reason for the low magnitude of the plateau storage modulus is hydrogel inhomogeneity due to manual mixing. Since the *Npu* intein has very rapid reaction kinetics, the interface between the two intein fragment solutions (N and C) may rapidly react to form hydrogel 'sheets' containing high level of crosslinking, while regions away from the interface are connected by a smaller number

of intermolecular interactions. The bulk rheology experiments measures the lowest storage modulus (i.e. 'weakest link') in a material, possibly explaining why our hydrogel exhibits a low overall measured G'_{∞} value. The formation of hydrogel sheets is supported by the observation that our intein hydrogel crumbled into small sheet-like structures when incubated in DPBS at elevated temperature (>50 °C) during the erosion experiment. This phenomenon also suggests that the mechanical properties of the hydrogel could be tuned over a wide range by controlling the speed of mixing.

Molecules of 20 kDa size can easily diffuse out of the hydrogel (Fig. 3.3D). However, both small (pyranine, 534 Da) and large (20 kDa dextran) molecules appear to diffuse out of the hydrogel at a similar rate. In contrast, agarose gels showed a much higher diffusion rate for pyranine than for dextran (Fig. 3.3D). The pore size of our intein-triggered hydrogel is primarily governed by the size and physicochemical properties of the midblock chain. The average hydrodynamic diameter of our midblock fragment – the S fragment chain – is 40 Å and the diagonal distance of CutA trimer is ~45 Å (measured from pdb code: 1V99). The average pore size of this hydrogel is expected to be similar to the cytosolic environment (*200, 201*). The similar diffusion rate of 20 kDa dextran and 534 Da pyranine may reflect the dynamic nature of the hydrogel in which molecular diffusion is limited by the vibrational motion of the S fragment. Use of proteins with elongated or rigid structures as a midblock should yield an expanded hydrogel pore size, thus facilitating molecular diffusion.

Next, we tested the ability of the intein-mediated hydrogel to function as a protein immobilization scaffold. For target protein immobilization in the hydrogel, we

chose the Src homology 3 domain from the adaptor protein CRK (SH3) and its ligand ($SH3_{lig}$) as the DP and DSP, respectively, because of their relatively small size (56 aa for SH3 and 11 aa for $SH3_{lig}$), high affinity (k_d = 0.1 μM) towards each other (*202*), and their previous application in intracellular protein docking (*193*). $SH3_{lig}$ was inserted between the *Npu*C and the CutA to form C-$SH3_{lig}$ (Table 3.1). The SH3 protein was fused to the N-terminus of green fluorescent protein (GFP) to form SH3-GFP. Purified C-$SH3_{lig}$ and SH3-GFP were mixed at 1:1 molar ratio to facilitate the docking of SH3 to $SH3_{lig}$. An equimolar amount of N was then added to the mixture to form a hydrogel containing embedded GFP (Fig. 3.5A). The final hydrogel contains 1.2 mM *trans*-spliced hydrogel backbone and 1.2 mM GFP (1:1:1 molar ratio of SH3-GFP: C-$SH3_{lig}$: N). Incorporation of GFP did not compromise the hydrogel's stability in solution (Fig. 3.5B, Fig. 3.6). The leaching of immobilized SH3-GFP from $SH3_{lig}$-containing hydrogel is ~30% after 3 weeks, significantly slower than that from hydrogel lacking $SH3_{lig}$, which lost >70% of the entrapped protein within the same period (Fig. 3.5C). The immobilized SH3-GFP in the hydrogel can be conveniently visualized under UV, with $SH3_{lig}$-containing hydrogel retaining most of the GFP fluorescence after 3 weeks while the hydrogel lacking $SH3_{lig}$ lost most of its fluorescence (Fig. 3.5D). It is possible that an even slower leaching rate of the immobilized protein can be achieved if a higher-affinity DP/DSP pair is used. Any DP-fused target protein should be able to be immobilized in a similar fashion. These results provide a first proof-of-principle that our intein-triggered protein hydrogel can be used as a general scaffold for protein immobilization through a method that does not significantly impact protein activity – the docking of a protein-

protein pair. This result is on contrast to chemical immobilization of proteins involving modification of the side chain functionalities of the amino acids cysteine and lysine – currently widely used for protein immobilization – which can negatively impact protein activity (*148*).

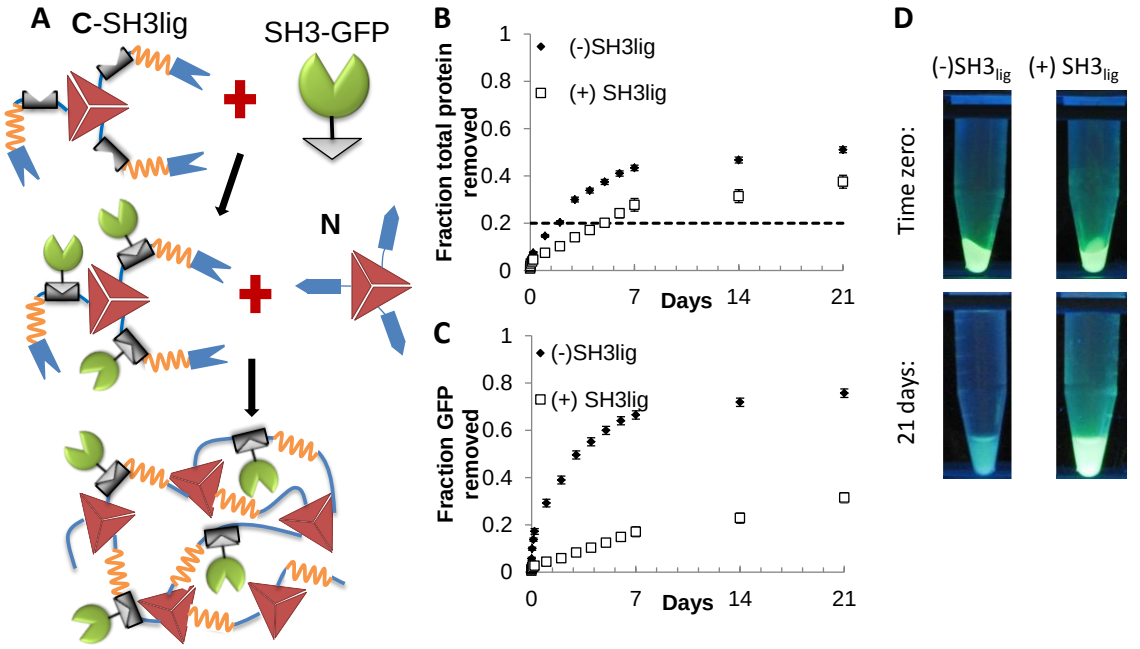


Figure 3.5 Intein-triggered hydrogel as a scaffold for protein immobilization. (A) Schematic of protein immobilization using GFP as a model globular protein. DSP-containing hydrogel building blocks are first mixed with DP-fused target protein. The complementary intein fragment-containing hydrogel building block is added to the mixture yielding a hydrogel with immobilized GFP. (B) Total protein erosion profile of hydrogel containing 1:1 molar ratio of SH3-GFP. Dotted line represents the theoretical mass corresponding to the spliced out inteins. The error bars represent the standard deviation of 2 independent experiments. (C) Leaching profile of SH3-GFP from hydrogel with and without the DSP. (D) Images of GFP containing hydrogels under UV exposure immediately after hydrogel formation and after 21 days.

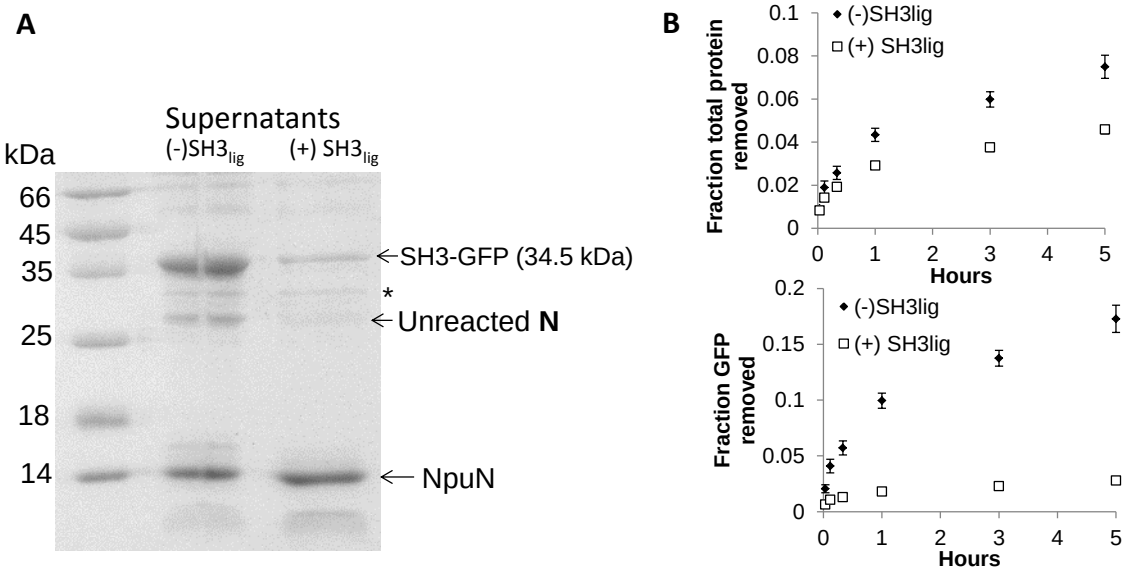


Figure 3.6 Leaching of encapsulated and immobilized GFP. (A) SDS-PAGE analysis of the surrounding buffer of SH3-GFP containing hydrogel whose building blocks contain or lack DSP (SH3lig). The hydrogel is immersed in EB1 for 21 days and the buffer is periodically replaced as indicated in Fig. 3.5B. All the buffers harvested at different times (total 7.5 mL) were pooled and analyzed as described in Fig. 3.3B. The asterisk denotes unidentified bands. (B) Erosion and leaching profiles of hydrogel harboring SH3-GFP in the first 5 hours of buffer contact.

The density of immobilized GFP in the current hydrogel is ~33 mol%. A higher immobilization density can be achieved by incorporating multiple DSP sites into the hydrogel building blocks. The docking of target proteins onto individual hydrogel building blocks prior to hydrogel formation ensures an even distribution of target protein throughout the hydrogel. Since the identity and location of the DSP are genetically encoded, specific ordering and ratios of different proteins can be conveniently achieved through incorporation of multiple different DSP's in the hydrogel building block. This approach also facilitates easy tuning of multi-enzyme reaction cascades to achieve high *in vitro* reaction rates and yields. Alternatively, the same hydrogel building block

containing a given DSP can be separately preloaded with different target proteins, and solutions of a hydrogel building block loaded with the different target proteins can be mixed at any desired ratio prior to the addition of the protein building block containing the complementary split intein fragment. This second approach enables convenient immobilization of different proteins at any desired ratio using a single DP-DSP pair, facilitating, for example, convenient multi-enzyme metabolic pathway optimization *in vitro* (*203*).

We next decided to determine the ability of our protein hydrogel to facilitate enzymatic reactions in an organic solvent. Enzymes hold enormous potential as catalysts for organic synthesis. However, enzymes are rarely used for the synthesis of organic compounds due to their low activity and stability in organic solvents (*204*). Synthetic polymer hydrogels have been used to prevent enzyme denaturation in organic solvents (*205, 206*). Using horseradish peroxidase (HRP) as the model enzyme and the oxidative coupling of *N*,*N*-dimethyl-*p*-phenylene diamine and phenol with *tert*-butyl hydroperoxide as the model reaction (*194*), we determined the ability of our hydrogel to protect the immobilized enzymes from the denaturing effect of the organic solvent heptane. Hydrogel entrapping 0.042 mM HRP was immersed in heptane for 24, 9 and zero hours (Set A, B, C respectively) prior to the addition of reaction substrates. To increase the area of the hydrophilic/hydrophobic interface, this hydrogel was manually disrupted into small pieces after immersion in the solvent. As anticipated, HRP entrapped in the hydrogel was able to catalyze rapid oxidation reactions, giving rise to a colorimetric product that is quantifiable via spectrophotometry. The product accumulates

linearly regardless of the amount of time in contact with the solvent indicating little-to-no enzyme inactivation by heptane (Fig. 3.7). On the other hand, HRP added directly to heptane exhibited very low catalytic activity due to enzyme inactivation. HRP dissolved in DPBS first and then added to heptane (Set C-E, incubated in heptane for 24, 9 and zero hours, respectively) was unable to catalyze significant conversion, likely due to the very limited surface area between the enzyme in the aqueous phase and the substrate in the organic phase. The ability of our protein hydrogel to withstand the denaturing effect of organic solvent is attributed to the incorporation of the hydrophilic S fragment in the hydrogel backbone, which effectively 'locks' water molecules inside the hydrogel, preventing organic solvent from accessing the hydrogel interior. These results demonstrate the potential of our intein-triggered protein hydrogel as a scaffold for enzymatic reactions in organic solvents.

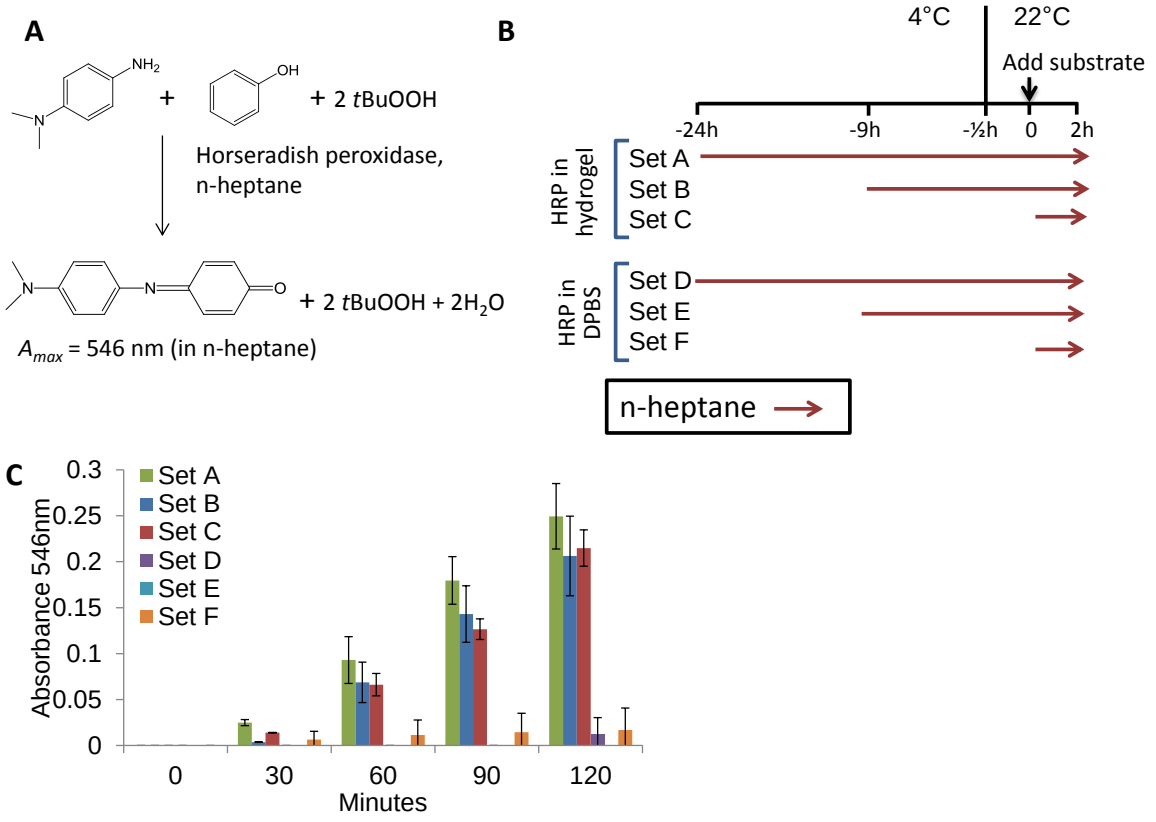


Figure 3.7 Intein-mediated protein hydrogel facilitates horseradish peroxidase (HRP)-catalyzed reaction in organic solvent n-heptane. (A) Model reaction catalyzed by HRP. (B) Schematics of the experimental procedure. HRP containing hydrogel or DPBS were submerged in equal volume of n-heptane for different period of times (indicated by the brown arrow) at 4 °C. Hydrogel was manually disrupted into small piece after immersion in the solvent. All samples were equilibrated to room temperature (22 °C) prior to the addition of reaction substrate. The formation of product, an indophenol-type dye, was monitored by measuring the optical absorbance of the organic phase supernatant for 2 hours. (C) Relative absorbance for each sample after subtracting the background.

In summary, we engineered a new protein hydrogel that conditionally assembles in response to a split-intein-triggered *trans*-splicing reaction. This hydrogel is formed under physiological conditions and shows unprecedented stability under a broad set of conditions. In addition, this intein-mediated hydrogel is compatible with cell culture

growth medium (Fig. 3.8), pointing to its potential for use as a scaffold for tissue engineering applications. Intein-triggered hydrogel technology provides a new platform for protein hydrogel design and synthesis, and should find use in many research and biofabrication applications including enzyme immobilization, bioelectrode fabrication, organic synthesis, injectable drug delivery, and functional tissue engineering scaffolds.

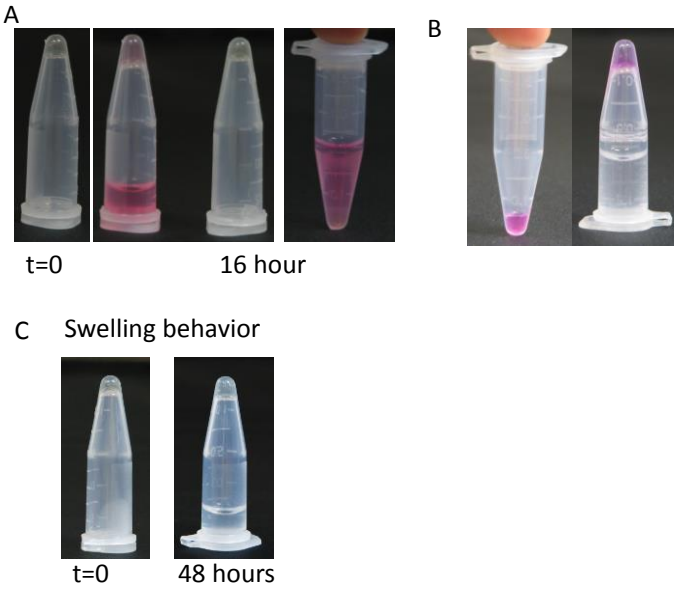


Figure 3.8 Hydrogels in DMEM media. (A) 1.6mM hydrogel synthesized using DPBS. Hydrogel was incubated with DMEM media to monitor diffusion of media into hydrogel. After overnight incubation, the coloration corresponding to phenol red is uniform throughout the hydrogel. Swelling behavior is evident by the increased volume of the hydrogel after media diffuses into hydrogel. (B) Synthesis of hydrogels with proteins dissolved in DMEM media: Concentrated protein (N and C) was dialyzed overnight against DMEM at 4°C. The proteins were concentrated using a 10-kDa ultra-filtration spin column and concentrated to ~100mg/ml. Proteins were mixed at 1:1 molar ratio supplemented with 5mM DTT and 0.2% NaN_3. Hydrogel formation was confirmed via upside-down tube test, and by the inability to pipette the solid hydrogel. C) Visual evidence of swelling of hydrogel. A 1.6mM J hydrogel was incubated for 48 hours in DPBS buffer.

CHAPTER IV

TWO-COMPONENT PROTEIN HYDROGELS ASSEMBLED USING AN ENGINEERED DISULFIDE-FORMING PROTEIN-LIGAND PAIR*

4.1 Overview

We present the development of a two-component self-assembling protein hydrogel. The building blocks of the hydrogel are two liquid-phase protein block copolymers each containing a subunit of the trimeric protein CutA as a cross-linker and one member of a PDZ-domain-containing protein-ligand pair whose interaction was reinforced by an engineered disulfide linkage. Mixing of the two building blocks reconstitutes a self-assembling polypeptide unit, triggering hydrogel formation. This hydrogel exhibits extremely high solution stability at neutral and acidic pHs and in a wide range of temperatures (4-50 °C). Incorporation of a "docking station peptide" binding motif into a hydrogel building block enables functionalization of the hydrogel with target proteins tagged with a "docking protein". We demonstrated the application of an enzyme-functionalized hydrogel in a direct electron transfer enzymatic biocathode. These disulfide-reinforced protein hydrogels provide a potential new material for diverse applications including industrial biocatalysis, biosynthesis, biofuels, tissue engineering, and controlled drug delivery.

---

* Reprinted with permission from "Two-component protein hydrogels assembled using an engineered disulfide-forming protein-ligand pair" by Dongli Guan, Miguel Ramirez, Shao Lin, Daniel Jacobsen, Ivan Barrera, Jodie Lutkenhaus, and Zhilei Chen, 2013, Biomacromolecules, Copyright 2013, American Chemical Society, doi: 10.1021/bm400814u. M. Ramirez designed and performed preliminary hydrogel-formation experiments. M. Ramirez designed, engineered and characterized disulfide-bond formation protein:ligand pair. Dongli Guan carried out hydrogel formation and characterization.

4.2 Introduction

Hydrogels are three-dimensional, hydrophilic polymer networks that hold large amounts of water. Based on the nature of the cross-linkers, hydrogels can be categorized as either chemical or physical hydrogels (*207*). Chemically cross-linked hydrogels are usually very stable in solution, typically undergo a large volume change during solution-to-gel transition, and contain covalently joined cross-linkers formed through chemical reactions such as disulfide formation, (photo) polymerization or the reaction between thiols and acrylates or sulfonates (*208*). Physical hydrogels contain non-covalently joined cross-linkers and often self-assemble in response to external stimuli such as changes in pH and temperature. These non-covalently joined cross-linkers render the physical hydrogel susceptible to shear-thinning under mechanical stress and self-healing upon cessation of the stress, making these hydrogels suitable for use as injectable materials. Additionally, gelation of self-assembled hydrogels does not rely on organic solvents or extraneous cross-linking reagents, making this material more favorable for biomedical applications including controlled drug delivery and tissue engineering (*209, 210*).

Many natural proteins can self-assemble into physical hydrogels, including elastin, collagen and gelatin. However, natural proteins often have a short shelf-life and purification from natural sources is non-trivial (*210*), limiting their broad applicability. In contrast, recombinant proteins are amenable to tuning of physical properties and can be purified on a large scale, making them a more desirable alternative to natural proteins

as building blocks for protein-based hydrogels (*119, 139, 192*). Diverse functional motifs can be easily incorporated into recombinant hydrogel-building-block proteins via genetic manipulation resulting in materials with specific and desired biological, chemical and mechanical properties. Recombinant protein hydrogels have been developed for applications in tissue engineering, biosynthesis and biofuel applications (*119, 146, 211-213*).

In this study, we report the synthesis of a novel self-assembling protein hydrogel based on the bioaffinity of a pair of engineered proteins. Gelation is triggered by the mixing of two liquid-phase protein building blocks, each containing a subunit of a trimeric protein (CutA) and a PDZ domain-containing protein (Tip1) or its peptide ligand ($Tip1_{lig}$). CutA serves as a crosslinking unit. A polypeptide containing a cross-linker on only one terminus is incapable of self-assembly into a network. On the other hand, a polypeptide containing cross-linkers on both termini is able to self-assemble into a network. Mixing of the two polypeptide building blocks, each containing a CutA cross-linker on one terminus and either Tip1 or $Tip1_{lig}$ on the other terminus, assembles into a polypeptide unit containing CutA cross-linker on both termini via the interaction between Tip1 and $Tip1_{lig}$, triggering a networking cascade and hydrogel formation. This hydrogel is stable at both acidic and neutral pHs (6-8) and in a wide range of temperatures (4-50 °C), exhibits high elasticity, and quickly recovers elasticity after shear-induced thinning. The high solution-stability of the hydrogel is attributed to the introduction of a disulfide bond between Tip1 and $Tip1_{lig}$ which reinforces the linkage between cross-linkers on the self-assembling polypeptide unit, and the use of the

extremely stable CutA trimer as the crosslinking agent. Incorporation of a docking station peptide (DSP) enables functionalization of the Tip1-mediated hydrogel with docking protein (DP)-tagged target proteins.

4.3 Materials and methods

4.3.1 Chemicals and bacterial strains

All chemicals were purchased from Sigma-Aldrich (St. Louis, MO), VWR International (Radnor, PA), or Thermo Fisher Scientific (Waltham, MA) unless otherwise stated. *Escherichia coli* strains DH5α (Invitrogen, Carlsbad, CA) and BL21(DE3) (New England Biolabs, Ipswich, MA) were used for recombinant DNA cloning and recombinant protein expression, respectively.

4.3.2 Plasmid construction

A schematic depiction of the amino acid sequences of all protein constructs and their numberings is shown in Table 4.1. Note that, in this chapter, the engineered disulfide-forming protein and its ligand, $Tip1_{T58C}$ and $Tip1lig_{D778C}$, are referred to as dsTip1 and $dsTip1_{lig}$, respectively.

Table 4.1 Protein constructs used in disulfide bond hydrogel

Construct	Short Name	Protein sequence	MW(kDa)
1	CutA-Tip1(*192*)	M-$(H)_7$-M-CutA-Tip1	24.8
2	CutA-$Tip1_{lig}$	$MG(S)_2$-$(H)_6$-$(S)_2$GLVPRGSH-CutA-EAYRDPMG-$[(AG)_3PEG]_{10}$-ARMPYVGS-$SH3_{lig}$-$[(G)_4S]_2$-AS-$Tip1_{lig}$	26.1
3	CutA-dsTip1	CutA-EAC-$[(G)_4S]_2$-AS-$Tip1_{T58C}$-LE-$(H)_6$	25.9
4	CutA-$dsTip1_{lig}$	$MG(S)_2$-$(H)_6$-$(S)_2$GLVPRGSH-CutA-EAYRDPMG-$[(AG)_3PEG]_{10}$-ARMPYVGS-$SH3_{lig}$-$[(G)_4S]_2$-AS-$Tip1_{ligD778C}$	26.1
5	SH3-Slac	SH3-KL-$[(G)_4S]_2$-AS-Slac-LE-$(H)_6$	45.9
6	SH3-GFP	SH3-KL-$[(G)_4S]_2$-AS-GFP-LE-$(H)_6$	35.5
7	dsTip1	$Tip1_{T58C}$-LE-$(H)_6$	12.6
8	GFP-dsTip1lig	$MG(S)_2$-$(H)_6$-$(S)_2$GLVPRGSH-GFP-$[(G)_4S]_2$-AS-$Tip1_{ligD778C}$	28.8

Sequences of primers used in this study are shown in Table 4.2. To generate CutA-dsTip1 (construct 3), the dsTip1 gene from pET30-CutA-Tip1 (*192*), a gift encoding CutA-Tip1(1) from Dr. Takehisa Matsuda (Kanazawa Institute of Technology, Japan), was modified by site-directed mutagenesis (overlap extension PCR) using the primers NheI-Tip1-F/T58C Tip R and T58C Tip F/XhoI-Tip1-R, and the mutant gene was inserted into CutA-NpuN (*213*) between the NheI and XhoI sites to replace NpuN with dsTip1.

Table 4.2 Primers for disulfide bond hydrogel constructs

Name	5' to 3' sequence
NheI-Tip1-F	ATTATCTGCTAGCGTAGTGCAAAGAGTTGAAA TAAGTT
T58C Tip R	CTCTGATACTCGGCAGACGTAAATGC
T58C Tip F	GCATTTACGTCTGCCGAGTATCAGAG
XhoI-Tip1-R	ATTAATACTCGAGAGACTGCCGAGTCACC
NdeI-CutA-F	GGAATTCCATATGATAATAGTTTACACGACT TTTCCGG
HindIII-CutA-R	GGAATTCAAGCTTCTTTTTTCGTCTCTTCAAT TAACCATTT
HindIII-[C10]-F	TATTAAAAGCTTATCGCGATCCGATGGGT
SnaBI-[C10]-R	TAATTATACGTACGGCATGCGGGC
L-NheI-CutA-F	GAGGCGGAGGGAGCGGAGGCGGAGGGAGCG CTAGCATAATAGTTTACACGACTTTTCCGG
XhoI-CutA-R	TAAAATCTCGAGTTTTTTCGTCTCTTCAATTA ACCATTT
SH3lig-L-F	CGCCGGCGCTGCCGCCGAAACGTCGTCGCGG AGGCGGAGGGAGCGG
SnaBI-SH3lig-F	TGCCGTACGTA GGATCCCCGCCGCCGGCGCTGCC
NheI-Tip1lig-XhoI	CTAGCCAGCTGGCGTGGTTTGATACCGATCT GTGATAAC
XhoI-Tip1lig-NheI	TCGAGTTATCACAGATCGGTATCAAACCACG CCAGCTGG
Tip_{lig} D778C top	CTAGCCAGCTGGCGTGGTTTTGCACCGATCT GTGATAAC
Tip_{lig} D778C bottom	TCGAGTTATCACAGATCGGTGCAAAACCACG CCAGCTGG
Slac-F	ATAAGCTAGCATGGACAGGCGAGGC
Slac-R	CATTCTCGAGGTGCTCGTGTTCGTG
NdeI-Tip1-F	GCATACATATGGTAGTGCAAAGAGTTGAAAT TCATAAGTT
XhoI-Tip1-R	ATTAATACTCGAGAGACTGCCGAGTCACC
NdeI-GFP-N	GTGGATATACATATGGTGAGCAAGGGCGAGG
HindIII-GFP-R	ATTTATAAGCTTCTAACTCGTCCATGCCGAGA GTG

To construct CutA-Tip1$_{lig}$, the CutA gene was amplified from pET30-CutA-Tip1 (*192*) using primers NdeI-CutA-F/HindIII-CutA-R, the S fragment $[(AG)_3PEG]_{10}$ was amplified from pQE9AC$_{10}$Atrp (*139*) (a gift from Prof. David Tirrell, California Institute of Technology) using primers HindIII-[C10]-F/SnaBI-[C10]-R, and inserted into pET26-ISceI (*214*) between NdeI and HindIII sites, and between HindIII and SnaBI sites, respectively to generate CutA-S; then, the CutA gene was amplified from pET30-CutA-Tip1 (*192*) with primers L-NheI-CutA-F/XhoI-CutA-R, SH3$_{lig}$ was then added to the N-terminus of CutA with another forward primer SH3lig-L-F, and a SnaBI site was further added at N-terminus with another forward primer SnaBI-SH3lig-F to introduce this SH3$_{lig}$-CutA into CutA-S between SnaBI and XhoI sites to generate CutA-S-SH3$_{lig}$-CutA; finally, this fragment was removed by NdeI and XhoI digestion and cloned into pET-15b (Novagen) at the same sites to introduce an N-terminal 6His-tag, and Tip1$_{lig}$ was annealed with primers NheI-Tip1lig-XhoI/XhoI-Tip1lig-NheI and inserted between NheI and XhoI sites to replace the C-terminal CutA with Tip1$_{lig}$. CutA-dsTip1$_{lig}$ was generated by site-directed mutagenesis using primers Tip$_{lig}$ D778C top/Tip$_{lig}$ D778C bottom annealed and inserted into CutA-Tip1$_{lig}$ between NheI and XhoI sites to replace Tip1$_{lig}$ with dsTip1$_{lig}$.

To construct dsTip1 and Tip1, the mutant dsTip1 and wild-type genes were amplified from CutA-Tip1 and CutA-dsTip1 respectively with primers NdeI-Tip1-F and XhoI-Tip1-R and inserted into pET26-ISceI (*214*) between NdeI and XhoI sites. To create GFP-Tip1$_{lig}$ we digested plasmid CBD-N with NdeI and HindIII and inserted GFP amplified with NdeI-GFP-N and HindIII-GFP-R into the plasmid backbone to generate

GFP-N (not used in these studies). Next, we digested GFP-N with NheI and XhoI and inserted the product of the hybridized primers NheI-Tip1lig-XhoI and XhoI-Tip1lig-R. The mutant GFP-ds$Tip1_{lig}$ was generated by inserting the product of the hybridized primers $Tip1_{lig}$ D778C top and $Tip1_{lig}$ D778C bottom into the backbone of the plasmid GFP-N digested with NheI and XhoI.

4.3.3 Protein expression and purification

E. coli BL21(DE3) was transformed with expression plasmid and plated on a Luria-Bertani (LB) agar plate containing either 50 μg/ml kanamycin (Construct 1, 3, 5-7) or 100 μg/ml ampicillin (Construct 2, 4, 8). The next day, all the colonies (50~100) from a plate were pooled, resuspended in 5 mL LB broth and transferred to 1 L LB broth containing the specific antibiotic. The cells were incubated at 37 ºC with shaking at 250 rpm to an optical absorbance (OD_{600}) of 0.6-0.9. For constructs 2 and 4, protein expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG, 1 mM) followed by incubation at 37 °C for 4 hours. For all the other protein expressions, the cell cultures were cooled to 18 °C prior to the addition of 1 mM IPTG, and followed by overnight incubation (~16 hours) at 18 °C. After expression cells were harvested by centrifugation at 8000 x *g* and 4 °C for 15 minutes and the cell pellets were stored at -80 °C until use.

For protein purification, cell pellets were resuspended in Buffer A (500 mM NaCl, 10 mM Tris-HCl, pH 8.0, 1 g wet pellet in 10 mL), and disrupted by sonication (QSonica Misonix 200, Amp 40, with 1 second pulse and 5 seconds pause for 2 minutes

total pulse). Whole cell lysate was centrifuged at 16,000 x *g* for 20 minutes at 4 °C. Target protein in the soluble lysate was purified by nickel affinity chromatography using a 5-ml Ni Sepharose™ High Performance HisTrap column (GE Healthcare Life Sciences, Piscataway, NJ). After equilibration with Buffer A and protein loading, the column was thoroughly washed with Buffer A containing 22.5 mM imidazole and the target protein was eluted in Buffer A containing 150 mM imidazole. For the purification of CutA-Tip1/dsTip1 (constructs 1 and 3), 2.5 mM EDTA was immediately added to the eluted protein to minimize heavy metal-induced protein precipitation (*186*). To remove imidazole from the protein samples, purified proteins were buffer-exchanged into Buffer A via 30-kDa ultra-filtration spin columns (Amicon Ultra, Millipore, Billerica, MA), as appropriate, concentrated to ~100 mg/ml using the same column and stored at -80 °C until use. The concentrations of purified proteins were determined by measuring the absorbance at 280 nm using a NanoDrop 1000 (Thermo Fisher Scientific). Proteins were stored at -80 °C until use. For sodium dodecyl sulfate(SDS)-polyacrylamide gel electrophoresis (PAGE) analysis, appropriately diluted protein samples were mixed with 2X SDS sample buffer (0.5 M Tris-HCl, pH 6.8, 20% glycerol, 10% w/v SDS, 0.1% w/v bromophenol blue) that included β-mercaptoethanol (2% final) in the case of reducing conditions, and the samples were incubated at 95 °C for 5 minutes prior to loading. The gels were stained with Coomassie brilliant blue R250 for visualization.

4.3.4 Hydrogel synthesis

The purified hydrogel building blocks CutA-dsTip1 (construct 3) and CutA-dsTip1$_{lig}$ (construct 4) were manually mixed with a pipette tip at 1:1 molar ratio at the specified concentration in Buffer B (500 mM NaCl, 10 mM Tris-HCl, pH 8.0, 0.1% w/v NaN_3) and the mixture was briefly centrifuged to remove trapped air bubbles prior to incubation at room temperature for 14-20 hours. Hydrogels form during incubation.

4.3.5 Characterization of interaction between dsTip1 and dsTip1lig

dsTip1 (construct 7) and GFP-dsTip1$_{lig}$ (construct 8) were mixed at 1:1 molar ratio (0.2 mM each) in Buffer A supplemented with 2 mM dithiothreitol (DTT) in a 1.5 mL centrifuge tube. The tube was incubated at 22 °C in a humidified chamber to facilitate DTT evaporation. Samples were taken at different times, mixed with 2X SDS loading buffer in the presence (reducing) or absence (non-reducing) of β-mercaptoethanol and incubated at 95 °C for 5 minutes. Samples were directly loaded onto a 12% SDS-PAGE gel (non-reducing) or incubated for an additional 3 minutes at 95 °C in the presence of 2% β-mercaptoethanol (reducing). The gels were stained with Coomassie brilliant blue R250 for visualization.

4.3.6 Hydrogel solution-stability, pore size, and rheological characterization

These experiments were performed essentially as described previously in Chapter III with the following modifications: the amount of protein present in the suspension

buffer used for the solution-stability study was determined by absorbance at 280 nm using a NanoDrop 1000 (Thermo Fisher Scientific).

Rheological characterization was carried out on a Paar-Physica MCR-300 (Anton Paar, Ashland, VA) parallel plate rheometer with a 25 mm plate fixture (PP25). Purified CutA-dsTip1 and CutA-dsTip1 were manually mixed on top of the rheometer plate by pipetting up and down. The final mixture contained 1.9 mM of each protein in 200 µL Buffer B. The measuring gap was set at 0.2 mm. The rheometer plate fixture was positioned in a humid chamber and mineral oil was added to the outer edge of the fixture to minimize evaporation. The sample on the plate fixture was incubated at room temperature overnight using settings of 1 rad/sec and 1% strain. The next day, strain sweeps and frequency sweeps were performed at 10 rad/sec frequency and 10% strain amplitude, respectively. The large-amplitude oscillatory shear (LAOS) experiment was performed at 10 rad/sec frequency.

4.4 Results and discussion

4.4.1 Self-assembling protein hydrogel design

The initial building blocks (before the incorporation of disulfide-forming functionality) of our self-assembling hydrogel are two protein block copolymers CutA-Tip1 and CutA-Tip1$_{lig}$ (Fig. 4.1, Table 4.1). CutA, a small trimeric protein (12 kDa) from *Pyrococcus horikoshii* (*186*), was chosen as the crosslinking protein due to its extremely high stability. CutA exhibits a denaturation temperature of nearly 150 °C and is able to retain stable trimeric structure on SDS-PAGE even after exposure to 0.1% SDS and

boiling for more than 1 hour (*187, 188*). The tax-interacting protein-1 (Tip1) and its peptide ligand ($Tip1_{lig}$, QLAWFDTDL) (*215*) were used to reconstitute a self-assembling polypeptide unit by connecting CutA cross-linkers located on the two different building blocks. A non-structural hydrophilic linker, S fragment ($[(AlaGly)_3ProGluGly]_{10}$), was inserted between CutA and Tip_{lig} to facilitate water retention (Table 4.1) (*116, 216*). To enable target protein immobilization, we inserted the small peptide $SH3_{lig}$ (PPPALPPKRRR) between the S fragment and $Tip1_{lig}$ as the DSP. $SH3_{lig}$ exhibits high affinity toward the Src homology 3 domain (SH3) protein (*217*) and thus mediates the immobilization of SH3-tagged target proteins.

The proteins CutA-Tip1 and CutA-$Tip1_{lig}$ were purified from *E. coli* lysate. Mixing of these proteins at concentrations >2 mM leads to the formation of a jelly-like material that weakly adheres to the containment vessel after inversion (*144*). However, this material is completely solubilized within 1 hour when suspended in Buffer B. Previously we successfully prepared a stable protein hydrogel using CutA as the cross-linker (*213*), in which different cross-linker units were covalently linked by protein trans-splicing reactions. Thus, the low solution-stability of the CutA-Tip1/CutA-$Tip1_{lig}$ material was attributed to the weak affinity between the wild-type Tip1 and $Tip1_{lig}$ (K_d ~0.19 μM) (*215*), making the material prone to intermolecular domain swapping (*189*).

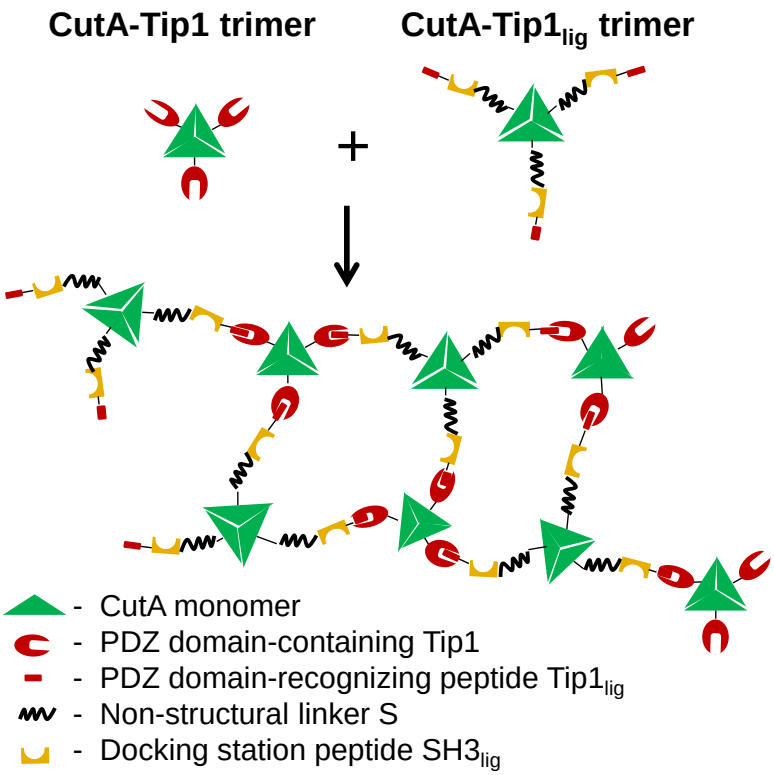


Figure 4.1 Schematic of dsTip1 hydrogel formation. Mediated by the interaction between two protein building blocks, each containing the trimeric cross-linker CutA and one member of the protein-ligand interaction pair Tip1/Tip1$_{lig}$.

4.4.2 Rational design of a disulfide bond between Tip1 and Tip1$_{lig}$

Disulfide bonds can conditionally form between cysteine residues on proteins under non-reducing conditions, providing a convenient way to seal a linkage between a cargo protein and carrier protein. Previously, Miyagi *et al.* reversibly conjugated a δPKC inhibitor peptide to a cell penetration peptide TAT via disulfide linkage to enable the delivery of δPKC into mice (*218*). Upon delivery to the reducing environment of the cytosol, the disulfide linkage breaks, releasing free δPKC.

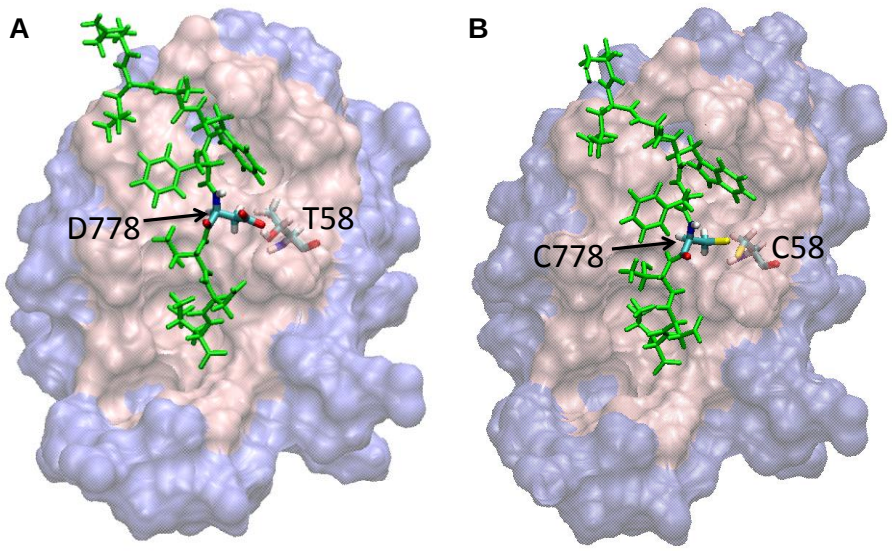


Figure 4.2 Crystal structure of Tip1-Tip1lig (pdb code: 3IDW). A) Representation of wild-type Tip1-Tip1$_{lig}$ interaction. Tip1 is shown in purple surface mode. Residues in Tip1 that are within 5 Å of Tip1$_{lig}$ (green) are shown in pink color. Tip$_{lig}$ is show in green licorice mode. Residues T58 and D778 are shown in element colors. B) Representation of dsTip1 and dsTip1$_{lig}$ interaction containing the substitutions T58C and D778C.

To increase the solution-stability of the self-assembled CutA-Tip1/CutA-Tip1$_{lig}$ material, we sought to convert the non-covalent Tip1-Tip1$_{lig}$ interaction to a covalent disulfide bond through the introduction of a pair of cysteine substitutions, one in the Tip1 binding pocket and one in Tip1$_{lig}$. The crystal structure of the Tip1 protein in complex with the Tip1$_{lig,}$ (PDB 3IDW) (*215*) reveals that the Tip1$_{lig}$ interacts with Tip1 through a combination of hydrophobic interactions and hydrogen bonds. Particularly, the oxygen (OD2) of D778 is within hydrogen bonding distance of the hydrogen (HG1) of T58 (Fig. 4.2A). We hypothesized that substituting both these residues with cysteines will facilitate disulfide bond formation (Fig. 4.2B) between the protein-peptide ligand

pair. Tip1 (T58C) and $Tip1_{lig}$ (D778C) are referred as dsTip1 and $dsTip1_{lig}$, respectively, in this paper.

To confirm the formation of disulfide bonds *in vitro*, we fused $dsTip1_{lig}$ to the C-terminus of GFP to form GFP-$dsTip1_{lig}$ (construct 8, Table 4.1). dsTip1 was expressed by itself with a 6x histidine tag at the C-terminus (construct 7). dsTip1 and GFP-$dsTip1_{lig}$ were mixed in the presence of dithiothreitol (DTT, 2 mM) in an open container. We anticipated that, as DTT evaporates over time, intermolecular disulfide bonds between dsTip1 and $dsTip1_{lig}$ should form gradually. Samples taken at different times were boiled in sample buffer either containing or lacking the reducing agent β-mercaptoethanol and analyzed via SDS-PAGE (Fig. 4.3). A band corresponding to the combined molecular weight of dsTip1 and GFP-$dsTip1_{lig}$ (41.4 kDa) appeared only on the gel loaded with samples prepared under non-reducing conditions. The emergence of this band coincided with the disappearance of the monomers GFP-$dsTip1_{lig}$ and dsTip1 on the same gel, suggesting that this band corresponds to the disulfide-linked complex of GFP-$dsTip1_{lig}$ and dsTip1. Additional bands of higher molecular weight also appeared on the non-reducing gel, likely indicating homodimers and/or multimers mediated by inter-cysteine disulfide linkage. No new bands or changes in band intensity were observed in the reducing gel. These results confirm that the engineered dsTip1 covalently attaches to its ligand $dsTip1_{lig}$ via a disulfide bond.

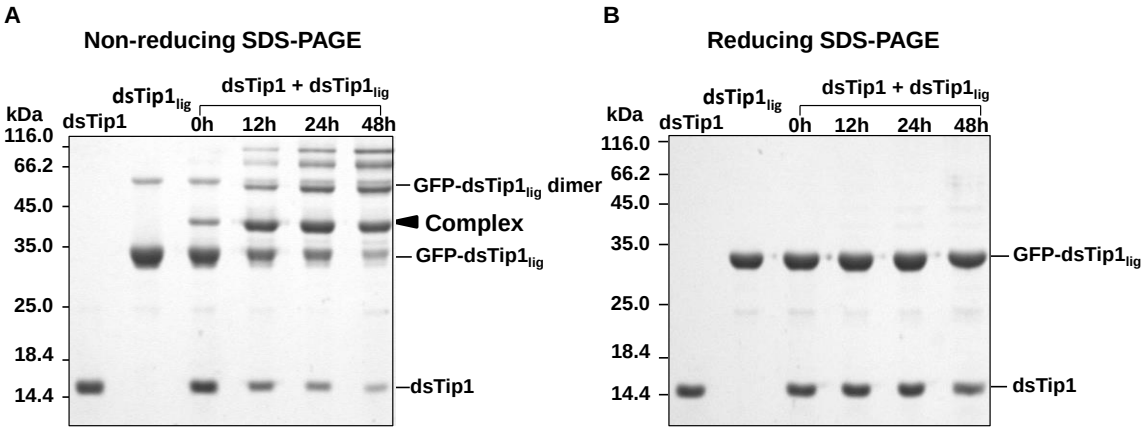


Figure 4.3 dsTip1 forms a disulfide bond with dsTip1lig. A mixture containing dsTip1 (12.6 kDa), GFP-dsTip1$_{lig}$ (28.8 kDa) and 2 mM DTT were analyzed at different times after initial mixing using SDS-PAGE conducted under non-reducing (A) and reducing (B) conditions. The DTT in the original mixture evaporates over time, simulating oxidizing conditions. The black triangle indicates the band corresponding to the disulfide-mediated complex of dsTip1:GFP-dsTip1$_{lig}$ (41.4 kDa). A band close to twice the molecular weight of GFP-dsTip1$_{lig}$ is also observed (57.6 kDa). Higher molecular weight bands observed on non-reducing gel were not identified.

4.4.3 Hydrogel stability characterization

The wild-type Tip1 and Tip1$_{lig}$ in the hydrogel building block were replaced with dsTip1 and dsTip1$_{lig}$ to form CutA-dsTip1 and CutA-dsTip1$_{lig}$ (constructs 3 and 4, Table 4.1). Both CutA-dsTip1 and CutA-dsTip1$_{lig}$ can be conveniently purified by one-step immobilized ion metal affinity chromatography (IMAC). Most of the purified CutA-dsTip1 and CutA-dsTip1$_{lig}$ exist as higher molecular weight multimers (Fig. 4.4A), with only a small fraction existing as monomer. This result is consistent with the observation that CutA retains trimeric structure during SDS-PAGE analysis even after boiling (*186*). Unlike CutA-dsTip1, which shows a single trimer band in the SDS-PAGE analysis, multiple high molecular weight bands coexist on the gel for CutA-dsTip1$_{lig}$. The

appearance of these multiple bands could be due to the presence of the non-structured S-fragment, which can cause associated proteins to migrate variably on SDS-PAGE gels.

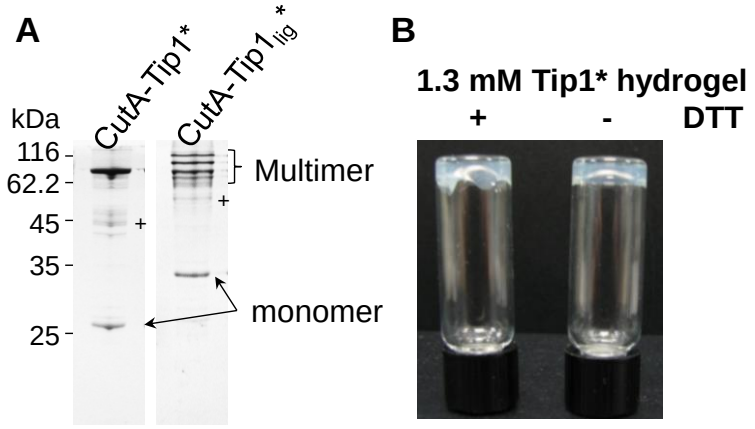


Figure 4.4 Characterization of 1.3 mM dsTip1 hydrogel. (A) SDS-PAGE analysis (12% acrylamide) of purified CutA-dsTip1 and CutA-dsTip1lig. "+" denotes unidentified proteins. (B) Formation of jelly-like material in the presence and absence of DTT. The material formed in the absence of DTT appeared to more strongly adhere to the bottom of glass vial after inversion.

For hydrogel formation, purified CutA-dsTip1 and CutA-dsTip1$_{lig}$ were manually mixed at 1:1 molar ratio (1.3 mM each) using a pipet tip in Buffer B (500 mM NaCl, 10 mM Tris-HCl, pH 8.0, 0.1% w/v NaN_3) either lacking or containing the reducing agent DTT. The vials were subsequently sealed. A highly viscous material immediately formed under both conditions. Both materials appeared to only weakly adhere to the bottom of the glass vial upon inversion, analogous to the observation with the hydrogel composed of wild-type Tip1 and Tip1$_{lig}$. These hydrogels were then incubated at room temperature for 20 hours and the vial test was repeated. At the later time point, the material formed in the absence of DTT appeared to more firmly adhere to the bottom of

the vial upon inversion, while the material formed in the presence of DTT still appeared relatively fluid (Fig. 4.4B). This result indicates that the hydrogel formed in the absence of DTT grew stronger over time likely due to the increased formation of disulfide bonds between dsTip1 and dsTip1$_{lig}$. For convenience, we refer to the material formed in the absence of DTT as dsTip1 hydrogel.

Unlike the material containing wild-type Tip1 and Tip1$_{lig}$, which is completely solubilized within 1 hour after immersion in buffer, the dsTip1 hydrogel is highly stable. After being immersed in buffer for one month at temperatures up to 50 °C, only 31% of the total protein was lost by erosion (Fig. 4.5A). The high solution-stability of the hydrogel is attributed to the use of the hyper-thermostable CutA trimer as the cross-linker and the use of a covalent (disulfide) linkage to bridge the two hydrogel building blocks. To our knowledge, this dsTip1 hydrogel is one of the most stable artificial protein hydrogels engineered to date (*119, 189*).

The stability of the dsTip1 hydrogel is not affected by acidic pH but is significantly weakened by basic pH (Fig. 4.5B). This phenomenon is likely caused by hydrolysis of disulfide bonds under alkaline conditions (*219, 220*), further underscoring the contribution of disulfide bonds to hydrogel stability. A dsTip1 hydrogel with as low as 1.3 mM of each building block (6.7% w/v total protein) displayed a slightly increased rate of erosion compared to the same hydrogel with 2.2 mM of each building block (11.3% w/v total protein) (Fig. 4.5C). Hydrogel with even lower building block concentrations could potentially be prepared but were not tested in this study.

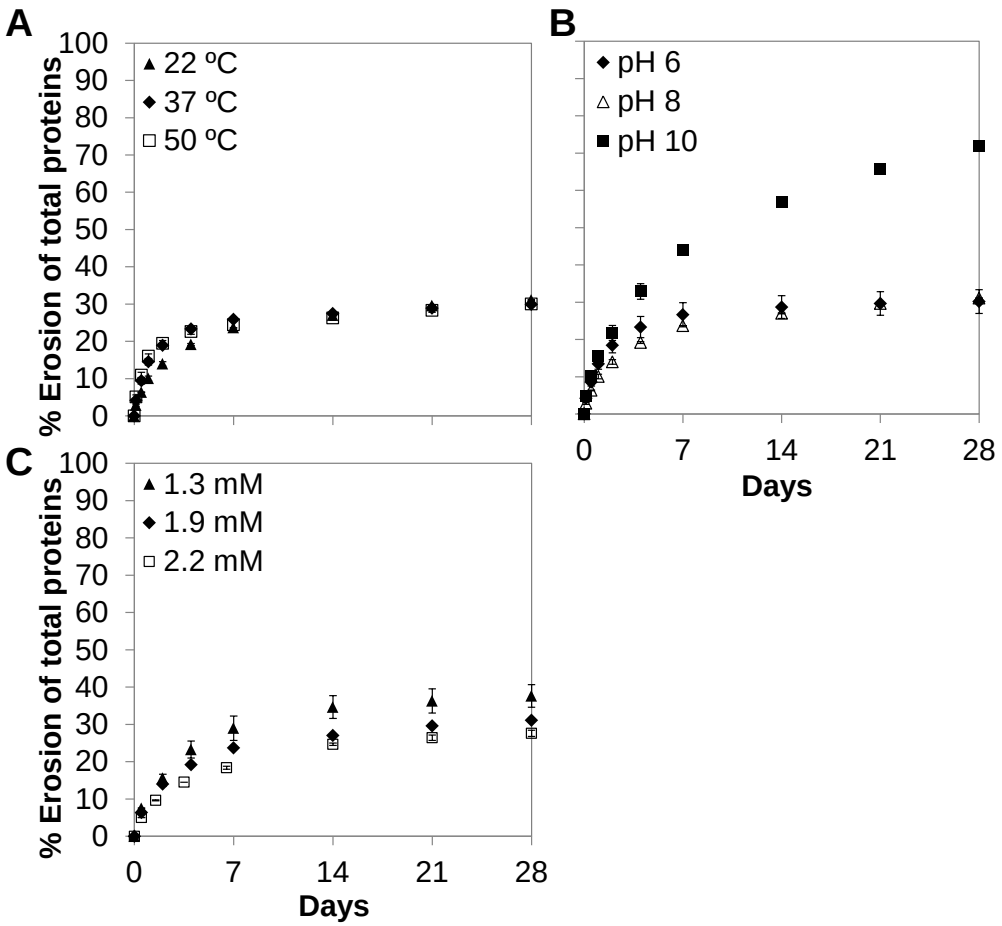


Figure 4.5 Solution stability of dsTip1 hydrogel. (A) Erosion profiles of hydrogels (1.9 mM) at different temperatures in pH 8 buffer. (B) Erosion profiles of hydrogels (1.9 mM) at different pHs at 22ºC. (C) Erosion profiles of hydrogels containing different protein concentrations in pH 8 buffer at 22ºC.

4.4.4 Rheological characterization of the dsTip1 hydrogel

The plateau storage modulus (G'_{∞}) for a dsTip1 hydrogel containing 1.9 mM of each building block is 262 ± 54 Pa, 15-fold greater than the plateau loss modulus (G''_{∞}) for the same gel (17 ± 1 Pa) (Fig. 4.6A), consistent with gel-like materials. The G'_{∞} value reflects the cross-linker density and strength. Hydrogels with higher G'_{∞} value can

potentially be obtained by increasing the building block concentration and/or using cross-linker proteins with a higher order of multimerization.

To evaluate the ability of dsTip1 hydrogel (1.9 mM) to re-form after shear-induced thinning, we conducted a large-amplitude oscillatory shear (LAOS) experiment. Although a permanent loss of storages modulus G' (~15%) occurred after the first cycle, no further decrease in G' was observed in subsequent LAOS cycles (Fig. 4.6B). Hydrogel thinning is attributed to the dissociation of CutA trimer into monomers under shear stress. Upon cessation of stress, disengaged monomers can rapidly re-associate to form new trimeric cross-linkers, restoring the hydrogel to full strength. The ability of the dsTip1 hydrogel to regain mechanical strength points to its suitability for applications that require injection, including controlled drug delivery and tissue engineering (*146, 197*).

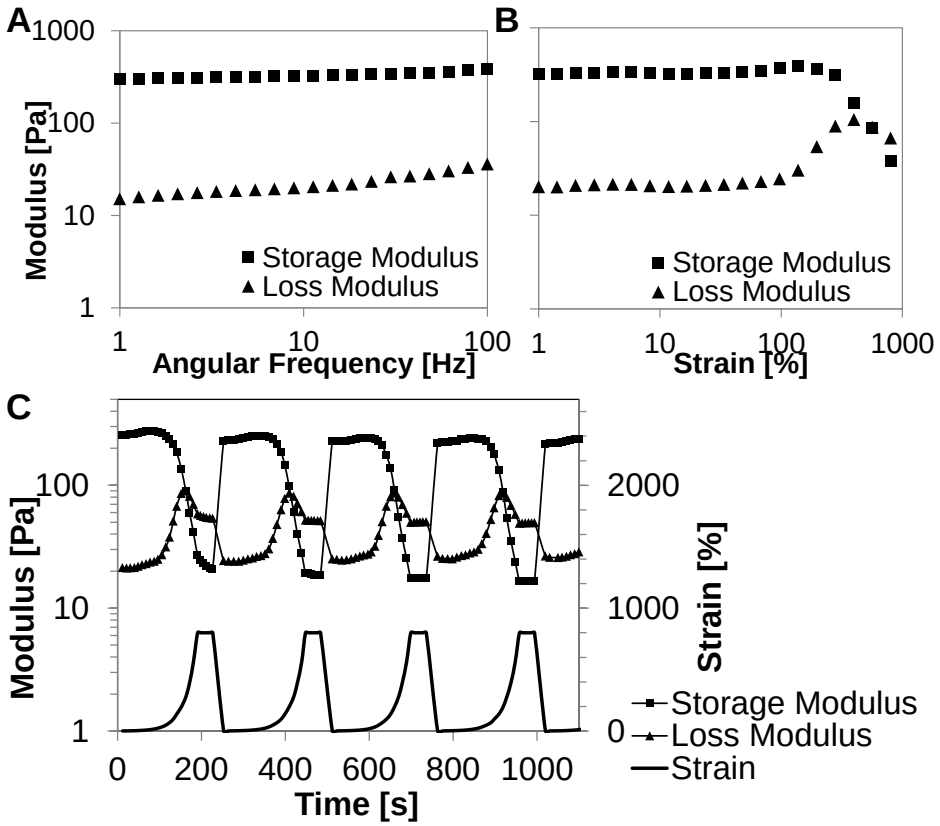


Figure 4.6 Rheological characterization of a 1.9 mM dsTip1 hydrogel. (A) Angular frequency sweep at 10% strain. (B) Strain sweep at 10 rad/sec. (C) Large-amplitude oscillatory shear cycles at 10 rad/sec.

4.5 Discussion

We demonstrated the synthesis of a disulfide-reinforced protein hydrogel that self-assembles upon mixing of two liquid-phase building blocks. Specifically, mixing of the two components reconstitutes a single self-assembling polypeptide unit via a Tip1 protein-ligand interaction that is reinforced by an engineered disulfide linkage. The use of a universal docking-station-peptide/docking-protein (DSP/DP) pair enables convenient incorporation of globular proteins in a "plug and play" fashion without changing the molecular architecture of the hydrogel backbone. This hydrogel retains its

integrity under a wide variety of conditions and does not require extraneous chemicals or alterations in pH/temperature/ionic strength to induce the gelation process, making this hydrogel especially suitable for tissue engineering as live cells can be conveniently encapsulated without any environmental triggers (*129*). The hydrogel can easily be functionalized with proteins by using the docking station peptide: docking protein method introduced in Chapter III. For example, growth hormones can potentially be incorporated to generate bioactive scaffolds for tissue engineering, and enzymes can be incorporated to create solid-phase biocatalysts for industrial biosynthesis. With the use of multiple orthogonal docking station peptide: docking protein pairs, it is conceivable that an entire metabolic pathway can be synthesized on the hydrogel for *in vitro* pathway construction (*221*, *222*). Finally, we anticipate that the physical properties of the hydrogel can be further tailored through the employment of linkers with varying structural rigidities and cross-linker proteins with different multimeric states (*189*), paving the way for the development of even more sophisticated hydrogels.

CHAPTER V

ENGINEERING SPLIT INTEIN DNAE FROM *NOSTOC PUNCTIFORME* FOR RAPID PROTEIN PURIFICATION[*]

5.1 Overview

We report the engineering of a DnaE intein able to catalyze rapid C-terminal cleavage in the absence of N-terminal cleavage. A single mutation in DnaE intein from *Nostoc punctiforme* PCC73102 (*Npu* DnaE), Asp118Gly, was introduced based on sequence alignment with a previously engineered C-terminal cleaving intein mini-*Mtu*RecA. This mutation was able to both suppress N-terminal cleavage and significantly elevate C-terminal cleavage efficiency. Molecular modeling suggests that in *Npu* DnaE Asp118 forms a hydrogen bond with the penultimate Asn, preventing its spontaneous cyclization prior to N-terminal cleavage. Mutation of Asp118 to Gly essentially abolishes this restriction leading to subsequent C-terminal cleavage in the absence of N-terminal cleavage. The Gly118 *Npu* DnaE mutant exhibits rapid thio-dependent C-terminal cleavage kinetics with 80% completion within 3 hours at room temperature. We used this newly engineered intein to develop both column-free and chromatography-based protein purification methods utilizing the elastin-like-peptide and chitin-binding protein as removable purification tags, respectively. We demonstrate

[*] Reprinted with permission from "Engineering split intein DnaE from Nostoc punctiforme for rapid protein purification" by Miguel Ramirez, Najla Valdes, Dongli Guan, and Zhilei Chen, 2013, Protein Engineering Design & Selection, Copyright 2013, Oxford University Press, doi:10.1093/protein/gzs097.

rapid target protein purification to electrophoretic purity at yields up to 84 mg per liter of *E. coli* culture.

5.2 Introduction

Affinity tags have greatly simplified the purification of recombinant proteins and are invaluable to modern biotechnology. However, the additional time and the high cost associated with proteases needed for tag removal have greatly hampered the usefulness of affinity tags in large-scale industrial processes. Many protease-free protein purification processes have been developed using inteins engineered to undergo N-/C-terminal cleavage reaction in acidic (pH 6) (*46, 47*) or reducing environments (*49, 105, 223, 224*). However, all pH-dependent inteins retain significant activity at neutral pH (*47, 105*) and the intracellular environment is naturally reducing. Thus, unless a slow-acting intein is used, target protein expression conditions often need to be optimized to minimize *in vivo* intein cleavage. In some cases, even after optimization of expression conditions, premature protein cleavage still significantly impacts target protein yields (*53, 225*).

One solution to prevent *in vivo* target protein cleavage is to use split inteins whose catalytic residues are split between two peptide chains: the N-terminal intein (I_N) and the C-terminal intein (I_C). Split inteins are only active when the two fragments are associated. Two protein purification systems using artificially split DnaB intein from *Synethocystis* sp (*Ssp*) have recently been developed. Volkmann *et al*. used the artificially split S1 DnaB intein consisting of an 11-aa N-intein (I_N) and a 144-aa C-

intein (I_C) (*54*). The target protein is fused to I_C and tag-removal is achieved by the addition of the I_N peptide. Since there is no N-extein present, wild-type catalytic residues are maintained. To achieve sufficient cleavage, a 40 to 1 molar ratio of I_N to peptide to I_C fused target protein is needed. Despite the small size of I_N, peptide synthesis is costly, prohibiting the application of this system in large-scale processes. Lu *et al*. developed a similar affinity-based purification system using the same *Ssp* DnaB intein with a different split junction (*55*). In this system, mutations of the appropriate catalytic residue at the N- and C-terminal are employed to achieve C- and N-terminal cleavage, respectively. However, despite the relatively rapid reaction rate of wild-type *Ssp* DnaB (Table 5.1), the mutant intein suffers reduced kinetics and requires extended incubation time (16 hours) at room temperature to achieve sufficient C-terminal cleavage. The long incubation times limit the usefulness of this system.

Moreover, artificially split inteins suffer from two major limitations: they are often less active than their continuous counterparts due in part to a lower affinity between the split fragments (*226*), and have a high tendency to form aggregates when expressed alone (*50*). Naturally split inteins such as DnaE from *Ssp* and *Npu*, on the other hand, are highly active, soluble and exhibit very high affinity between the two split fragments (*107, 114, 185, 227*). Unfortunately, despite these desirable traits, neither of the two naturally split DnaE inteins have been used for protein purification. The highly exposed hydrophobic surface on I_N of DnaE (102 and 123 aa for DnaE from *Npu* and *Ssp*, respectively) tends to interfere with the folding of N-extein, causing some fusion proteins to miss-fold and form insoluble aggregates and limiting the use of the N-

fragment as a general purification tag. On the other hand, despite its small size (36 aa for both *Npu* and *Ssp*) and no apparent interference with target protein solubility, the C-fragment of DnaE is also not suitable as a purification tag due to the tightly coupled C- and N-terminal cleavage reactions. In naturally split DnaE inteins, the C-terminal cleavage can only happen after the N-terminal cleavage, and mutation of the first Cys to Ala, which normally prevents N-terminal cleavage without interfering C-terminal cleavage activity, also abolishes the C-terminal cleavage (*34, 185*).

Table 5.1 Reported apparent half-lives of various continuous and split inteins

Trans-splicing				**N, C- Cleavage**			
Intein	**T (˚C)**	**$t_{1/2}$**	**Ref.**	**N/C**	**T(˚C)**	**$t_{1/2}$**	**Ref.**
Npu DnaE	37	63s	(*185*)	N	n. r.	--	--
				C	37	16 min	this work
Ssp DnaE	23	175 min	(*185*)	N	23	11.5 min[a]	(*34*)
	37	35 min	(*185*)	C	23	60.8 min[a,b]	(*34*)
*Sce*Vma[c]	25	6 min	(*185*)	N	23	6 min	(*228*)
				C	23	2-6hr	(*45*)
Split *Ssp* DnaB	25	12 min	(*185*)	N	RT	57.8 min[a]	(*54*)
				C	RT	39.4 min[a]	(*54*)
*Mtu*RecA[c]	30	60-120 min	(*105*)	N	n. r.	--	--
				C	37	<1hr	(*229*)
gp41-1	45	3.8 s	(*21*)	N	n. r.	5 min	(*21*)
				C	37		
gp41-8	37	15 s	(*21*)	N	n. r.	150 min	(*21*)
				C	37		
NrdJ-1	37	7 s	(*21*)	N	n. r.	48 min	(*21*)
				C	37		
IMPDH-1	37	8 s	(*21*)	N	n. r.	15 min	(*21*)
				C	37		

[a] Half-life for *Ssp* DnaE and split *Ssp* DnaB calculated from pseudo first order k_{obs}
[b] *Ssp* DnaE C-terminal cleavage rate after N-terminal cleavage
[c] Continuous intein

In this work, we engineered *Npu* DnaE (termed Npu*) to undergo C-terminal cleavage without N-terminal cleavage by introducing a single mutation, Asp118, based on the sequence alignment to mini-*Mtu*RecA intein (*46*). Npu* achieves ~80% C-terminal cleavage yield within 3 hours of reaction at 22 ˚C. In comparison, to achieve a similar extent of C-terminal cleavage, it takes ~16 hours at 23 ˚C for the IMPACT system (New England Biolabs) employing *Sce*Vma1 and *Ssp* DnaB intein (IMPACT Manual) (*45, 228*). Using Npu*, we further developed two protein purification methods and purified multiple target proteins to electrophoretic purity at high yields (up to 84 mg per liter of *E. coli* culture) within a short time (<4 hours), demonstrating the usefulness of these technologies and their potential for large-scale industrial protein purification.

5.3 Material and methods

5.3.1 Chemicals and strains

All chemicals were reagent grade and purchased from Sigma-Aldrich (St. Louis, MO) or VWR International (Radnor, PA), unless otherwise stated. *E. coli* DH5α (Invitrogen, Grand Island, NY) was used for recombinant DNA cloning and manipulation. *E. coli* BLR(DE3) (Novagen, Madison, WI) was used for the expression of recombinant protein. Ortho-Nitrophenyl-β-galactoside (ONPG) was purchased from Research Products International Corp. (Mount Prospect, IL). Chitin beads were purchased from New England Biolabs (Ipswich, MA).

5.3.2 Plasmid construction

To generate C-GFP (Construct 1), the NpuC gene was amplified from plasmid KanR-IntRBS-NpuNC-CFN (*77*) (gift from Prof. Tom Muir, Rockefeller University) using primers NpuC-F-NdeI and OXP-NC-G-Rev, joined to the N-terminal of GFP pet26-GFP with primers OXP-GFP-NC-FWD and XhoI-GFP-R by overlap extension PCR and cloned into pET-26b(+) (Novagen, Madison, WI) between NdeI and XhoI sites. Mutation D118G was introduced via site directed mutagenesis to generate C*-GFP (3) with the primers NpuCD17G-F and NpuCD17G-R.

To generate CBD-N, NpuN was also amplified from plasmid KanR-IntRBS-NpuNC-CFN (*77*) (gift from Prof. Tom Muir, Rockefeller University) using primers HindIII-Link-Npu F and NpuN-R-XhoI, joined to the chitin binding domain (CBD), amplified from pTWIN1 (New England Biolabs) with primers NdeI-CBD-F and HindIII-CBD-R via overlap extension PCR, and inserted into the pET-26b(+) (Novagen, Madison, WI) between NdeI and XhoI sites. CBD-N_{C1A} was generated by site directed mutagenesis using primers NheI-C1A-F and NpuN-R-XhoI.

ELP-N was constructed by inserting NpuN (aa 1-102) into plasmid pET-EI:MBP (*52*) (containing ELP, gift from Prof. David Wood, Ohio State University) between the EcoRI and HindIII sites. N was amplified first using primers HindIII-Link-Npu-F and HindIII-6H-NupN-R, then amplified again with primers EcorI-Linker-NpuN-F and HindIII-6H-NupN-R to include the restriction sites and flexible linker.

C*-DsRed was cloned into pET-26b(+) (Novagen, Madison, WI) between NdeI and XhoI sites. C* was amplified with primers NpuC-F-NdeI and NheI-NpuC CFN-R

from C*-GFP. DsRed was amplified from pTY24 plasmid (NCRR, YRS, Seattle, WA) with primers HindII-L-DsRed-fwd and XhoI-DsRed-R. The product was linked to C* by digestion with NheI enzyme resulting in a short linker peptide CFNAS. Aside from the canonical CFN sequence, the "AS" dipeptide corresponds to NheI restriction site and was included to facilitate subsequent cloning.

To clone C*-PTDH, The phosphate dehydrogenase "PTDH" was amplified from plasmid PTDH 12xA176R-pet15b (*230*) (gift from Prof. Huimin Zhao, University of Illinois) with primers NheI-PTDH-F and XhoI-PTDH12x-R and inserted into C*-DsRed digested with NheI and XhoI. Plasmid constructs C*-β-Gal, C*-CAT and C*-MBP were synthesized in the same way by insertion between NheI and XhoI sites with the appropriate primers. The β-galactosidase gene was amplified from plasmid pET-E-I: β-galactosidase (*52*) (gift from Prof. David Wood). Similarly, chloramphenicol acetyl transferase (CAT) and maltose binding protein (MBP) genes were amplified from plasmid pET-E-I:CAT (*52*) and pET-E-MBP (*52*), respectively. All constructs are listed in Table 5.2. Primer sequences used during the cloning of the constructs can be found in Table 5.3.

Table 5.2 Protein constructs for NpuC engineering

Construct	*Short Name*	*Protein sequence*	*Molecular Weight (kDa)*
1	C-GFP	NpuC-CFN-GFP-HHHHHH	32.1
2	CBD-N	ChBD-2x(GGGGS)-NpuN-HHHHHH	20.5
3	C*-GFP	NpuC D118G-CFN-GFP-HHHHHH	32.0
4	ChBD-N_{C1A}	ChBD-2x(GGGGS)-NpuN C1A-HHHHHH	20.5
5	ELP-N	ELP-2x(GGGGS)-NpuN-HHHHHH	59.8
6	C*-PTDH	NpuC D118G-CFNAS-PTDH-HHHHHH	42.2
7	C*-DsRed	NpuC D118G-CFNAS-DsRed-HHHHHH	31.3
8	C*-β_Gal	NpuC D118G-CFNAS-β_Gal-HHHHHH	122.1
9	C*-CAT	NpuC D118G-CFNAS-CAT-HHHHHH	31.3
10	C*-MBP	NpuC D118G-CFNAS-MBP-HHHHHH	48.2

Table 5.3 Primers used for cloning constructs for NpuC engineering

Primer	5’ to 3’ sequence
NpuC_F_NdeI	TTAGAAGGCATATGATCAAAATAGCCACACGTAAATATTTAGG
OXP-NC-G-Rev	CCTCGCCCTTGCTCACATTGAAACAATTAGAAGCTATGAAGCCAT
OXP-GFP-NC-FWD	ATAGCTTCTAATTGTTTCAATGTGAGCAAGGGCGAGG
XhoI_GFP_R	TAAAATCTCGAGTAACTCGTCCATGCCGAGAG
NpuCD17G-F	GGCAAACAAAATGTCTATGGCATTGGAGTT
NpuCD17G-R	GTCGCGCTCAACTCCAATGCCATAGACATT
HindIII-Link-Npu F	CCTGGAAGCTTGTGGAGGCGGAGGGAGCGGAGGCGGAGGGAG
NpuN_R_XhoI	CGCTAGCTGTTTAAGCTATGAAACGGAAATATTGACATATAGCTCGAGATTCGGCAAATTATCAACCCG
NdeI-CBD-F	TAATTTAACATATGAAAATCGAAGAAGGTAAACTGACAAATCCT
HindIII-CBD-R	AAGATTAAAGCTTCTTGAAGCTGCCACAAGGCA
NheI-C1A-F	AATTAAGCTAGCGCCTTAAGCTATGAAACGGAAATATTGACA
EcorI-Linker-NpuN F	AATATGGGAATTCGGAGGCGGAGGGAGCGG
HindIII-6H-NupN-R	GTACATTAAGCTTAGCAGCCGGATCTCAGT
NheI-NpuC CFN-R	ATTCGCGCTAGCATTGAAACAATTAGAAGCTATGAAGCC
XhoI_DsRed_R	TAAAATCTCGAGCAGGAACAGGTGGTGGC
HindII-L-DsRed-fwd	TTCAATAAGCTTGGAGGCGGAGGGAGCGGAGGCGGAGGGAGC
primers NheI-PTDH-F	GCTAGCGCCTCCTCCGAGGACG ATTTAACGCTAGCATGCTGCCGAAACTCGTTATAACTC
XhoI-PTDH12x-R	AGTTTAGCTCGAGGTCTGCGGCAGGATTGG
NheI-LacZ-F	ATTTCAATGCTAGCATGACCATGATTACGGATTCACT
XhoI-LacZ-R	TGATAATCTCGAGTTTTTGACACCAGACCAACTG
NheI-CAT-F	GTTTCAATGCTAGCATGGAGAAAAAAATCACTGGATATACCACCGTTGATATAT
XhoI-CAT-R	TAATAATTAACTCGAGCGCCCCGCCCTGCCAC
NheI-MBP-F	GTTTCAATGCTAGCATGAAAATCGAAGAAGGTAAACTGGTAATCT
XhoI-MBP-R	AAGTTATACTCGAGTCCCCTTCCCTCGATCC

5.3.3 Protein expression and purification

E. coli BLR(DE3) was transformed with the appropriate expression plasmid and plated on an agar plate containing 5 μg/mL tetracycline and 100 μg/mL ampicillin (Table 5.2, construct 5) or 5 μg/mL tetracycline and 50 μg/mL kanamycin (all other constructs). The next day, a single colony was picked and grown in 5 mL of Luria-Bertani (LB) broth to OD_{600} ~0.6. The culture was transferred to 1 L LB broth containing the same antibiotics and grown at 37 °C until OD_{600} ~0.6. Protein expression was induced at 18 °C for 14 hours by the addition of Isopropyl β-D-1-thiogalactopyranoside (IPTG, 0.2 mM). After expression, cells were harvested by centrifugation at 6000 *x g* at 4 °C for 15 minutes and stored at -80 °C until use.

For purification of CBD-N/N_{C1A} (construct 2, 4) cell pellets were resuspended in Buffer A (0.5 M NaCl, 10 mM Tris-HCl, pH 8.0) at 10 mL per gram of wet pellet, and disrupted by sonication (QSonica Misonix 200, Amp 10, 16-20W, with 1 sec pulse 6 sec pause for 1 minutes). Soluble lysates were collected after centrifugation at 16,000 *x g* for 20 minutes at 4 °C and passed through a 5-ml Ni-NTA column (GE Healthcare Life Sciences, Piscataway, NJ), washed with wash buffer (0.5 M NaCl, 10 mM Tris-HCl, 45 mM Imidazole, pH 8) and eluted in Buffer A containing 150 mM imidazole.

Proteins C/C*-GFP (Construct 1, 3) were purified in a similar way but with Buffer B (0.5M NaCl, 50 mM NaPOi, pH 6.0, 1x protease inhibitor cocktail (Roche Applied Science, Indianapolis, IN)) to minimize proteolytic degradation (*185*). Purified protein was buffer-exchanged into Buffer A and concentrated via 10 kDa ultra-filtration spin columns (Amicon Ultra, Millipore, Billerica, MA).

For sample purifications using Method 1, all cell pellets were resuspended in Buffer A. For sample purifications using Method 2, all cell pellets were resuspended in Column Buffer (1 M NaCl, 10 mM Tris-HCl, pH 8) to increase binding of target protein to chitin resin (*231*).

5.3.4 Intein reaction kinetics characterization

All intein characterization experiments were carried out using purified proteins diluted in Buffer A with the indicated amount of reducing agents at specified temperature. All reactions contained 20 μM of each intein fragment. Samples were taken at different time points after the initiation of the reaction, mixed with 2X SDS sample buffer (0.5 M Tris-HCl, pH 6.8, 20% Glycerol, 10% w/v SDS, 0.1 % w/v bromo-phenol blue, 2% β-mercaptoethanol), incubated at 95 °C for 5 minutes and analyzed using 12% SDS-PAGE gels, unless otherwise specified. The gels were stained with Coomassie brilliant blue R250. For samples corresponding to 0 min time points, purified C/C*-GFP (construct 1 and 3) protein was first mixed with 2X SDS sample in the absence of β-mercaptoethanol, and incubated at 95 °C for 3 minutes. CBD-N/N_{C1A} (construct 2 and 4) and β-mercaptoethanol were then added to the mixture. The entire mixture was incubated at 95 °C for additional 3 minutes to inactivate the protein. Band intensities corresponding to reactants and products were quantified using Trace Quantity module in Quantity One software (Bio-Rad, Hercules, CA).

5.3.4 Protein purification via reversible precipitation of elastin-like-peptide

In this study, we pre-purified ELP-N (construct 5) with one round of ammonium sulfate precipitation to facilitate the interpretation of SDS-PAGE gels, although this step was not found to improve the efficiency of protein purification. Clarified cell lysates of ELP-N and C*-POI (construct 3, 6-10) were thoroughly mixed and incubated at room temperature for 10 min to allow association of C* and N. Ammonium sulfate (0.4 M final concentration) (*232*) was added to the mixture to induce precipitation of the ELP complex. The pellet containing the target protein non-covalently linked to ELP was resuspended in Buffer A. The intein reaction was initiated by the addition of DTT (50 mM) and was carried out at room temperature for 3 or 20 hours. It is possible to use a much lower DTT concentration to induce C-terminal cleavage. At the end of the reaction, ammonium sulfate (0.4 M) was added to the mixture to precipitate out the ELP-N/C* complex. This precipitant was removed by centrifugation. The supernatant contained the highly purified target protein.

5.3.5 Protein purification via chitin resin

A slurry of chitin beads was first incubated with lysate of CBD-N (construct 2) at room temperature for 10 minutes, washed extensively with Column Buffer to remove all contaminating proteins, and then loaded with lysate containing C*-GFP (construct 3). After washing, DTT (5 mM) was added to the mixture to induce C-terminal cleavage. The reaction was essentially complete after 3 hours of reaction at room temperature and purified GFP was collected in the flow-through. CBD-N (construct 2), as well as trace

amount of unreacted C*-GFP (construct 3) remained bound to the chitin beads. Cleaved C*, although not visible on SDS-PAGE gels due to its very small size (4 kDa), was presumed to remain on the column due to interaction with N.

5.3.6 Molecular modeling

The structures of mini-*Mtu*RecA (pdb: 2IMZ), *Npu* DnaE (pdb: 2KEQ) and *Ssp* DnaE (pdb: 1ZD7) were visualized using Visual Molecular Dynamics (VMD) (*233*), and aligned using the MultiSeq module in VMD. Hydrogen bond interactions were identified by VMD. The NMR structure of *Npu* DnaE contains 20 different solution structures (*110*).

5.3.7 Purified protein content quantification

Target protein purification yield was quantified by measuring the concentration of purified sample using the Bradford method (Coomassie Plus Bradford Assay Reagent, Pierce Biotechnology, Rockford, IL). To estimate the percent recovery, soluble lysate and purified protein were loaded on the same SDS-PAGE stained with SimplyBlue SafeStain (Life Technology, Carlsbad, CA), and the band intensity corresponding to the target protein were measured using the Trace Quantity module in Quantity One software (Bio-Rad, Hercules, CA).

5.3.8 Pre-purification of ELP-NpuN

Ammonium sulfate (0.4 M) was added to the soluble lysate to induce ELP-N phase separation. The mixture was incubated at room temperature for ~3 minutes and centrifuged at 14,000 *x g* for 10 minutes. The resulting pellet was resuspended in one third of the original volume of buffer A. A low intensity water-bath sonicator (Ultrasonic Cleaner, GB 928) was used (5 minutes) to aid the re-suspension of ELP-N.

5.3.9 Sample protein activity assays

The activity of purified PTDH was confirmed by the NBT assay as described by Mayer and co-workers (*234*). Since DTT interferes with the NBT reaction at high concentration, the DTT concentration in purified protein was reduced to ~5 μM by buffer exchange using a 30-kDa cut-off spin column (Amicon Ultra-15 Centrifugal Filter Unit, Millipore, Billerica, MA) before the assay.

The activity of MBP was confirmed by binding to amylose resin (New England Biolabs, Ipswich, MA). Amylose beads (25 μL) were incubated with purified protein (500 μL) at room temperature for 10 minutes, washed twice with 500 μL buffer A and resuspended in 500 μL of Buffer A. Ten μL of this suspension was mixed with 10 μL of 2X SDS loading buffer, boiled at 95 °C for three minutes and analyzed via SDS-PAGE. The MBP protein was visible in the amylose beads suspension but not in the wash buffer.

The proteins GFP and DsRed were assayed by fluorescence measurements against the non-fluorescent protein CAT. Purified GFP or DsRed were diluted 2-fold and

transferred to a 96-well plate (150 μL/well). The fluorescence intensity was measured using a spectrofluorometer SpectraMax Gemini EM (Molecular Devices, Sunnyvale, CA) with excitation/emission wave lengths of 485/538 nm (GFP) or 544/590 nm (DsRed). The control protein CAT generated background values in both assays.

β-galactosidase activity was measured by the hydrolysis of ONPG to o-nitrophenol which absorbs at 420 nm. β-galactosidase was diluted 1000-fold in Z-buffer (0.06 M Na_2HPO_4, 0.04 M NaH_2PO_4, 0.01 M KCl, 0.001 M $MgSO_4$, and 0.27 % β-mercaptoethanol). Diluted protein (30 μL) was mixed with Z buffer (200 μL) and ONPG (70 μL, 4 mg/mL in 100 mM Potassium Phosphate buffer pH 7) and incubated at 22 °C for 15 or 30 minutes. At the end of the reaction, 500 μL stopping buffer (1 M Na_2CO_3) was added and the absorbance at 420 nm was measured in a Biomate 3 spectrophotometer (Thermo Electron Corporation).

To estimate the enzymatic units of β-galactosidase, the following formula was used:

$$nmoles\ ONPG\ hydrolized = \frac{(OD_{420}) * (8x10^5 nanoliters)}{\left(4500\ \frac{nl}{nmoles\ cm}\right)(1cm)}$$

$8x10^5$ nanoliters is the volume of the reaction

4500 $M^{-1}cm^{-1}$ is the extinction coefficient of o-nitrophenol

1-cm is the length of the light path.

One unit of β-galactosidase is defined as the amount of enzyme necessary to hydrolyze one micromole of ONPG at 22 °C per minute.

$$Unit\ number = \frac{\mu moles\ ONPG\ hydrolize}{time}$$

$$Specific\ Units = \frac{Unit\ number}{mg\ of\ enzyme\ used}$$

To estimate the sample recovery of β-galactosidase, a similar activity assay was carried out in a 96-well plate by diluting the purified β-galactosidase 1000-fold in Z-buffer. Diluted protein (50 µL) was mixed with Z buffer (50 µL) and 10 µl of ONPG solution. Absorbance at 420 nm was measured after 20-minute incubation using a SpectraMax 340PC384 Absorbance Microplate Reader (Molecular Devices, Sunnyvale, CA).

5.4 Results

5.4.1 Npu DnaE intein activity is thio-dependent

For the analysis of intein *in vitro* activity, various fusion proteins containing the I_N or I_C of *Npu* DnaE (N or C) were generated as illustrated in Table 5.2. For application as protein purification tag, it is important that the intein activity is regulated by an external stimulus. Although in theory, intein reaction does not require any thiol agents, there are a number of unpaired cysteine residues in *Npu* DnaE intein that could form intermolecular disulfide bonds and may prevent the intein reaction via redox trap

formation (*40, 227, 235*). Therefore, we first determined whether the *Npu* DnaE activity is thio-dependent. Purified CBD-N (construct 2) was mixed with C-GFP (construct 1) at equimolar concentrations in the absence or presence of 2 mM DTT. The trans-spliced product CBD-GFP was only present in reaction containing DTT, confirming that split *Npu* DnaE intein is thio-dependent (Fig. 5.1).

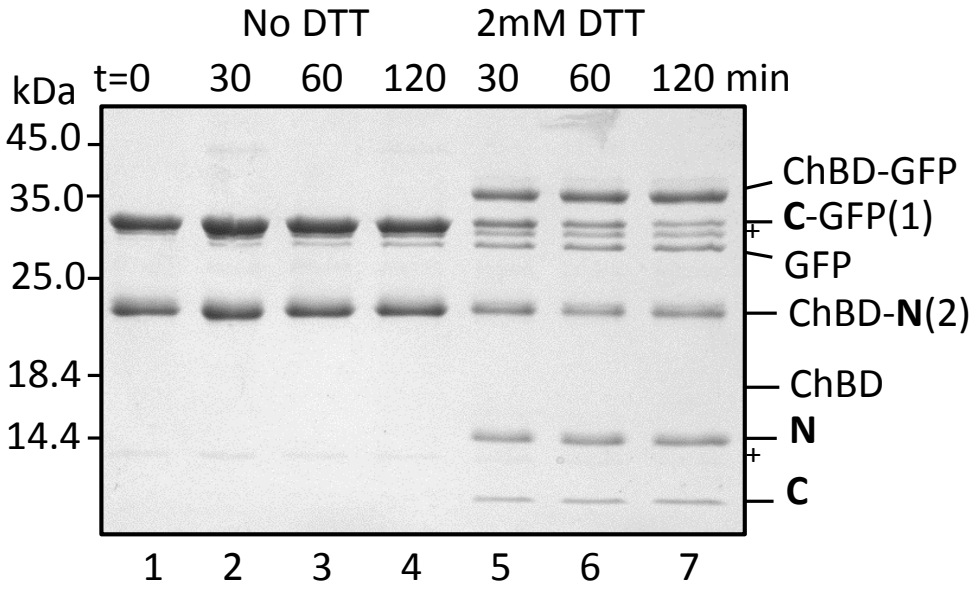


Figure 5.1 The trans-splicing activity of wild type Npu DnaE intein is thio-dependent. Reaction between CBD-N (2) and C-GFP (1) at 22 °C in the absence or presence of 2 mM DTT visualized on SDS-gel (12 % acrylamide). CBD-GFP is the trans-spliced product. GFP is the cleaved C- extein. C, N are the cleaved inteins. Trace amounts of the cleaved N-extein CBD are not visible from the SDS-gel due their low concentration, but are detectable by Western Blot (data not shown). "+" denotes unidentified bands.

The trans-splicing reaction is near completion after 30 minutes, consistent with previous findings (*185*). Trace amounts of cleaved N and C-extein are also visible, and

are slightly more pronounced in reactions carried out at 50 mM DTT concentration (Fig. 5.2A). DTT can launch nucleophilic attack at the thio-ester bond of the linear or branched intermediate (Fig. 5.3A, step 1 and 2, respectively), resulting in N-terminal cleavage. In a previous study on *Ssp* DnaE intein, the presence of 50 mM DTT was found to almost completely block protein trans-splicing and shunt the reaction to N-terminal cleavage and subsequent C-terminal cleavage (*34*). The limited amount of N-terminal cleavage seen in *Npu* DnaE intein, even at 50 mM DTT, may be due to the extremely rapid trans-splicing kinetics, given that the conversion of linear/branched intermediates to the trans-spliced product effectively competes with the nucleophilic attack by DTT.

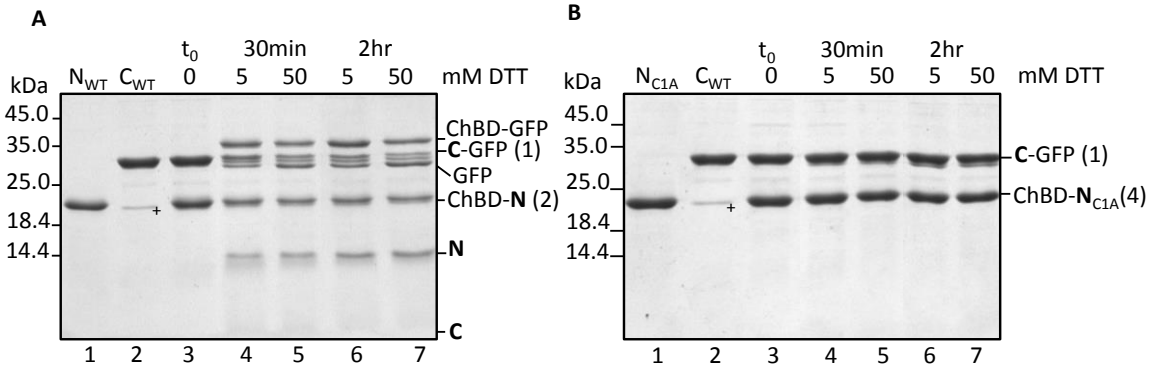


Figure 5.2 Reaction between C-GFP and CBD-N or ChBD-NC1A at 22 °C in the absence or presence of 5 or 50 mM DTT. CBD-GFP is the trans-spliced product and GFP is the cleaved C-extein. "+" denotes unidentified bands.

Mutant N with the first Cys replaced with Ala, CBD-N_{C1A} (construct 4), showed no trans-splicing activity and negligible C-terminal cleavage activity even in 50 mM DTT (Fig. 5.2B), consistent with previous findings that the C-terminal cleavage reaction

is tightly coupled to N-terminal cleavage in DnaE inteins (*34*, *185*). This unique property prevents the use of the wild-type DnaE intein in C-terminal protein purification applications.

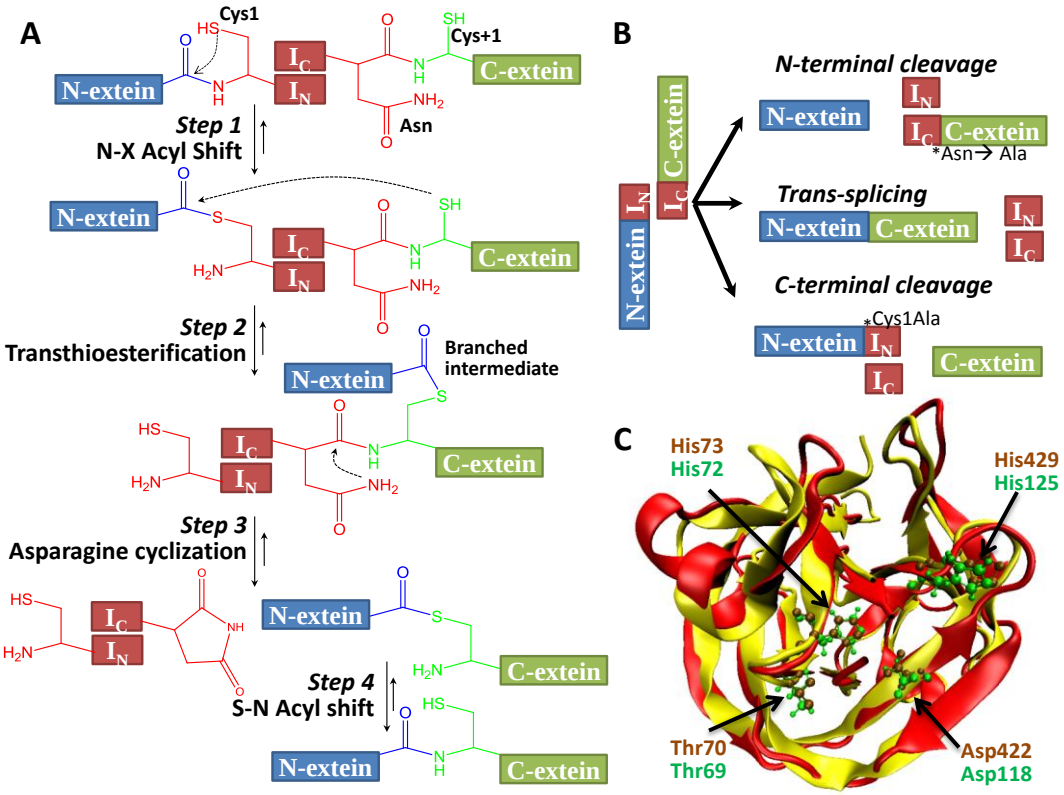


Figure 5.3 Intein trans-splicing mechanism. (B) Products resulting from different intein reactions. Mutation of the last asparagine and the first cystein to alanine render most inteins N- and C-terminal cleaving, respectively. (C) Structural alignment of *Npu* DnaE intein (red, pdb: 2keq) with mini-*Mtu*RecA intein (yellow, pdb: 2IMZ). Conserved catalytic residues for *Npu* DnaE and mini-*Mtu*RecA inteins are highlighted in color green and orche, respectively.

5.4.2 Rational design of C-terminal cleaving Npu DnaE

With a few exceptions (*35*), most intein splicing reactions comprise four highly coordinated nucleophilic replacements (Fig. 5.3A). The first step involves an N-X acyl shift (X: C or S) in which the first residue of the intein, a cysteine (Cys 1) or serine, attacks the proceeding peptide carbonyl, forming a linear (thio) ester. In the second step, the first residue of the C-extein, a cystein (Cys+1), serine, or threonine, attacks the (thio)ester carbonyl, cleaving the N-extein and forming a branched intermediate with two N-termini, one belonging to the intein and the other to the N-extein. In this state, the exteins are joined together but are still attached to the intein C-terminal. This branched intermediate is then cleaved from the intein by a transamidation reaction involving the last asparagine residue of the intein (step 3). In the final step, the free exteins undergo a spontaneous X-N acyl shift, which reverts the (thio)ester to a peptide bond and forms the spliced protein product.

In most inteins, the reactions at N- and C-termini are independent, such that the mutation of a catalytic residue that abolishes the reaction at one terminus results in a cleavage reaction at the other terminus (Table 5.1) (*33, 45-47, 229*). However, due to the tight coupling of N- and C-terminal cleavage reaction, the conventional method cannot be used to engineer a C-terminal cleaving DnaE intein (*34, 185*).

Previously, Wood, *et al.* engineered a C-terminal cleaving intein, mini-*Mtu*RecA, using directed evolution (*46*). A single mutation, D422G, was found to be responsible for the elevated C-terminal cleavage activity and suppressed N-terminal cleavage. Alignment of *Npu* DnaE and mini-*Mtu*RecA inteins revealed high homology on the

sequence level (Fig. 5.4) and even higher homology on the structural level (Fig. 5.3C). Most of the catalytic residues, including Asp422 (Asp118 in *Npu* DnaE) are conserved between *Npu* DnaE and mini-*Mtu*RecA intein (Fig. 5.4). Thus we hypothesized that mutation D118G may confer similar C-terminal cleavage activity to the *Npu* DnaE intein.

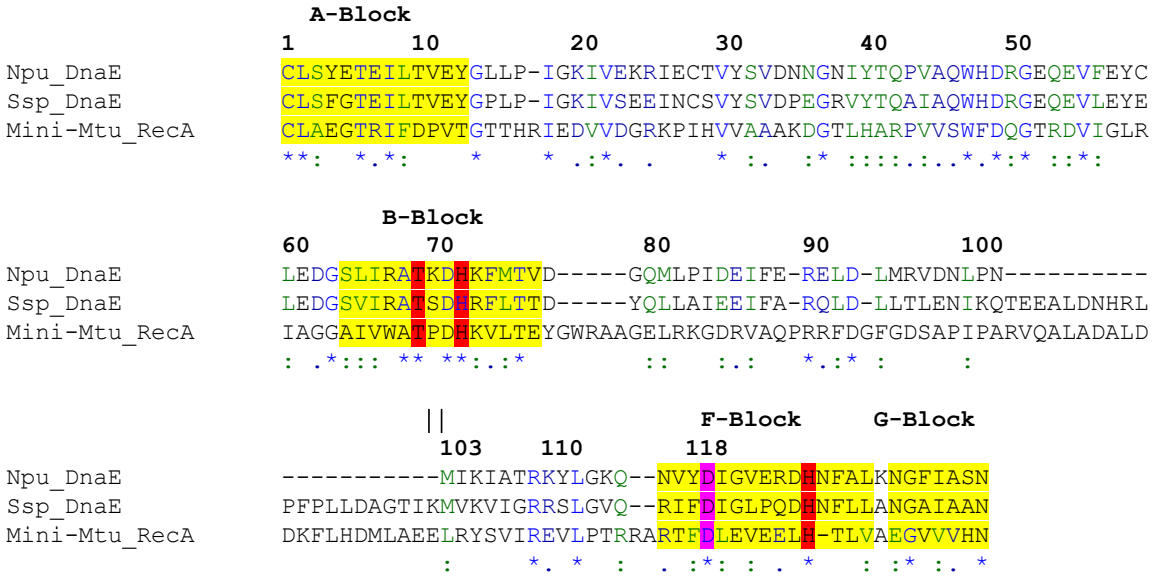


Figure 5.4 Sequence alignment of DnaE inteins from Ssp and Npu, and mini-MtuRecA intein. ||: break point for I_N and I_C for DnaE intein. Active site residues as shown in Figure 5.3C and Figure 5.11 are highlighted in red. The Asn118 residue is highlighted in magenta. Numbers correspond to the *Npu* DnaE residue.

5.4.3 Activity of Npu DnaE intein with Asp118Gly mutation

To test the effect of the D118G mutation, the amino acid substitution was introduced into C-GFP via site directed mutagenesis to form C*-GFP (construct 3).

Similar to wild type *Npu* DnaE, the activity of mutant Npu* is also thio-dependent. The D118G mutation completely abolished the trans-splicing reaction and induced rapid C-terminal cleavage under reducing conditions (Fig. 5.5). No spontaneous C-terminal cleavage was observed with C*-GFP incubated alone or with DTT at room temperature even after 20 hours, confirming the C-terminal cleavage activity is N dependent. Little free N-extein was observed when the reaction was carried out at low DTT concentration, and only very limited amount of free N-extein was observed in reaction at high DTT concentration (50 mM), suggesting that D118G mutation essentially abolished the first N-X acyl shift and induced C-terminal cleavage reaction independent of N-terminal cleavage. To further confirm that Npu* is able to undergo C-terminal cleavage in the absence of N-terminal cleavage, we carried out the reaction in the presence of trialkylphosphine (tris(2-carboxyethyl)phosphine, TCEP) that is capable of disrupting disulfide but not thio-ester bonds. TCEP also induced C-terminal cleavage reaction with no N-terminal cleavage at all (Fig. 5.5B), indicating that Npu* has uncoupled N- and C-terminal cleavage activity. We also determined the activity of C*-GFP when mixed with CBD-N_{C1A}. A similar rate of C-terminal cleavage was observed under reducing conditions. However, rapid C-terminal cleavage was also observed even in the absence of DTT, making it unsuitable for use as protein purification tag (Fig. 5.6). It is possible that, due to very close proximity, Cys1 and Cys+1 form disulfide bond immediately upon association of the two intein fragments, preventing further intein reaction and allowing control of the onset of C-terminal cleavage by reducing agents (*235*).

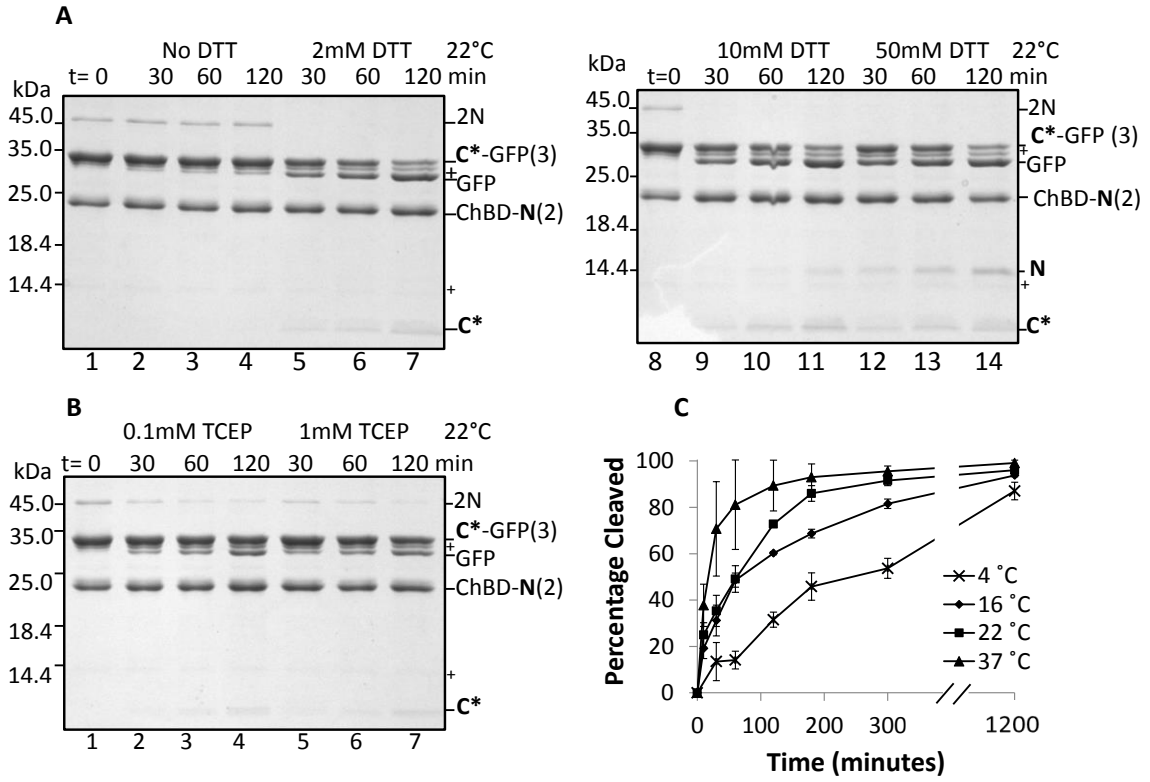


Figure 5.5 Catalytic activity of mutant C* Reaction between CBD-N and C*-GFP. At 22 °C in the absence or presence of reducing agent DTT (A) or TCEP (B), 2N: dimer complex of CBD-N (2). It disappears in samples treated with higher concentration of β-mercaptoethanol and boiled for longer period. GFP and C* are the cleaved C-extein and C-intein, respectively. N is the cleaved N-intein. Cleaved N-extein CBD is not visible on the SDS-gel but can be detected on Western Blot for samples incubated with DTT (data not shown). "+" denotes unidentified bands. (C) Time course of the disappearance of C*-GFP due to C-terminal cleavage at different temperatures. The error bar represents the standard deviation from 3 independent experiments.

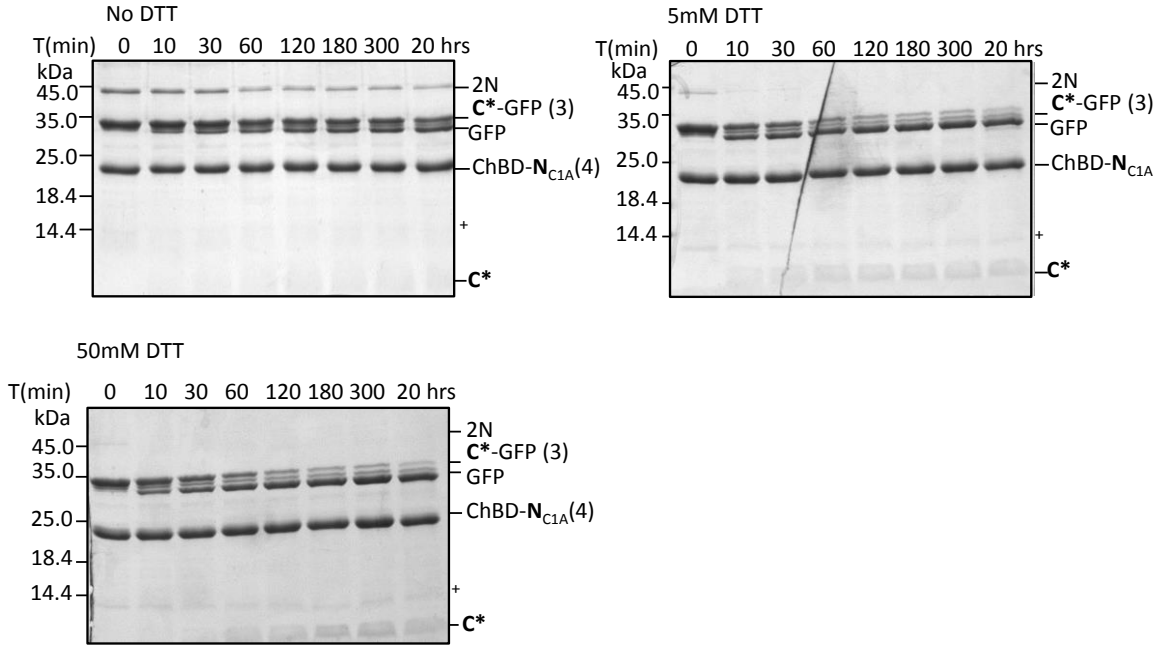


Figure 5.6 Catalytic activity of C* with NC1A. Reaction of C*-GFP (3) with CBD-N_{C1A} (4) at 22 °C in t/he absence or presence of DTT. GFP and NpuC* are the cleaved C-extein and C-intein, respectively. "+" denotes unidentified bands.

We next determined the C-terminal cleavage kinetics of Npu* at different temperatures (Fig. 5.5C and Fig. 5.7). The highest cleavage rate was obtained at 37 °C where ~80% cleavage was achieved in just 1 hour. To achieve the same 80% cleavage, 3 and 4.5 hours were needed for samples incubated at room temperature (22 °C) and 16 °C, respectively. Over 85% C-terminal cleavage was obtained at 4 °C after 20 hours. This cleavage rate is significantly higher than that of *Ssp* DnaB and *Sce*Vma1 inteins used currently in the IMPACT system (New England Biolab), which require about 16 hours incubation at 23 °C to achieve a similar cleavage efficiency (IMPACT Manual) (*45, 228*). It's possible that a higher cleavage rate may be achieved when an excess of N

is present (*54*). Taken together, these results demonstrate the potential of Npu* as a self-cleaving tag for protein purification.

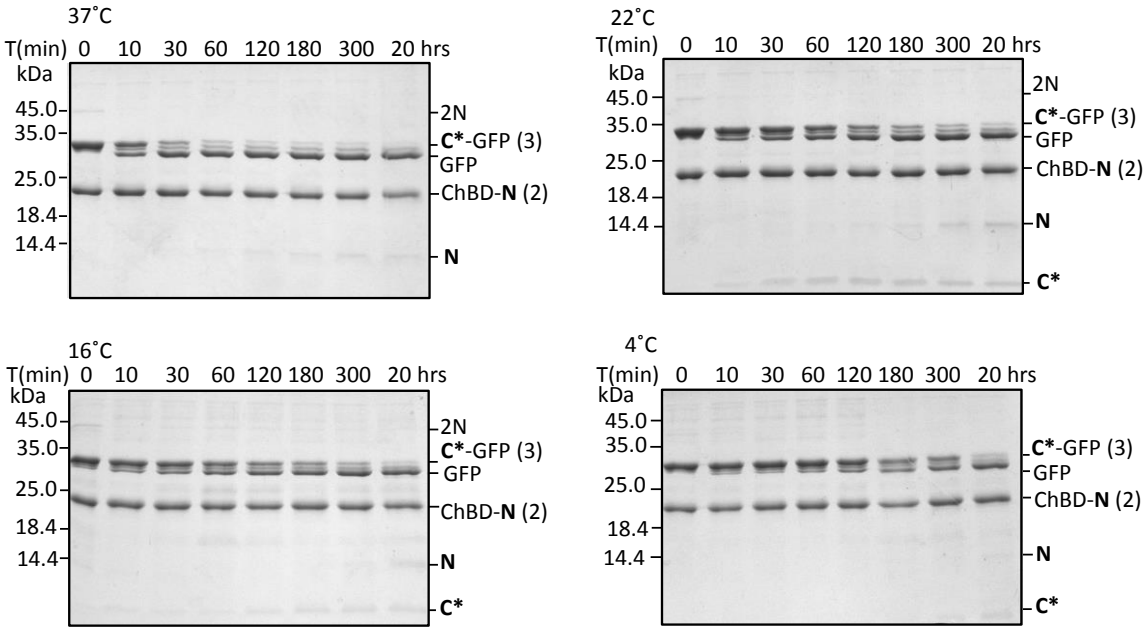


Figure 5.7 Reaction of C*-GFP with CBD-N was incubated in 5mM DTT at different temperatures. The reaction was stopped by mixing samples with SDS-sample buffer and boiled for 5 minutes.

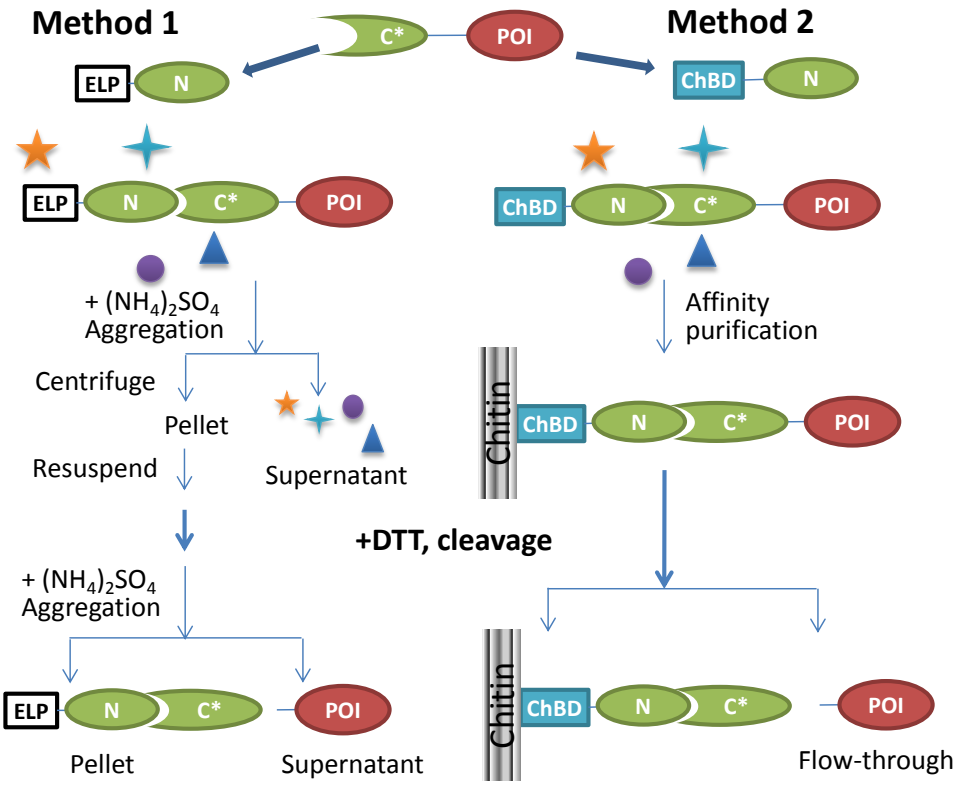
To determine the half-life for the C-terminal cleavage reaction of C*-GFP at different temperatures, trend line was generated using Lab Fit software package (Campina Grande, Paraíba, Brazil) that best fit all the data points from Fig. 5.7. Time corresponding to 50% cleavage was estimated based on the fit curve.

Table 5.4 Estimated half-lives for C*-GFP cleavage at different temperatures

Temperature	Half-life
4˚C	243 min
16˚C	70 min
22˚C	55 min
37˚C	16 min

5.4.4 Protein purification via reversible precipitation and chitin resin

To demonstrate the utility of Npu* in protein purification, we developed two protein purification methods (Fig. 5.8). The first method combines the reversible precipitation of the elastin-like-peptide (ELP) (*49, 52, 236-238*) with the controllable C-terminal cleavage of Npu* (Fig. 5.8, Method 1). N was joined to the C-terminus of the elastin-like polypeptide (ELP) via a flexible linker ELP-N (construct 5), and mutant C* was fused to the N-terminus of various sample target proteins C*-POI (constructs 3, 6-10). Under non-reducing conditions, N and C* non-covalently interact with each other without cleavage, physically associating the POI with the ELP. Addition of ammonium sulfate triggers phase separation of the ELP (*232*), resulting in the aggregation of the ELP and the associated POI. After removal of cellular proteins in the supernatant, this precipitant is then solubilized in low-salt reducing buffer, reversing the phase transition and inducing C-terminal cleavage of the intein. Upon cleavage, the POI is released from the ELP-intein complex which is selectively removed by a second round of ammonium sulfate precipitation and centrifugation, giving rise to POI of high purity in solution.



ELP: Elastin-like-peptide POI: Protein Of Interest ChBD: Chitin binding domain

Figure 5.8 Schematic of protein purification methods developed. Method 1: column-free approach. Method 2: chromatography-based approach. The symbols represent *E. coli* cellular proteins present in the lysate.

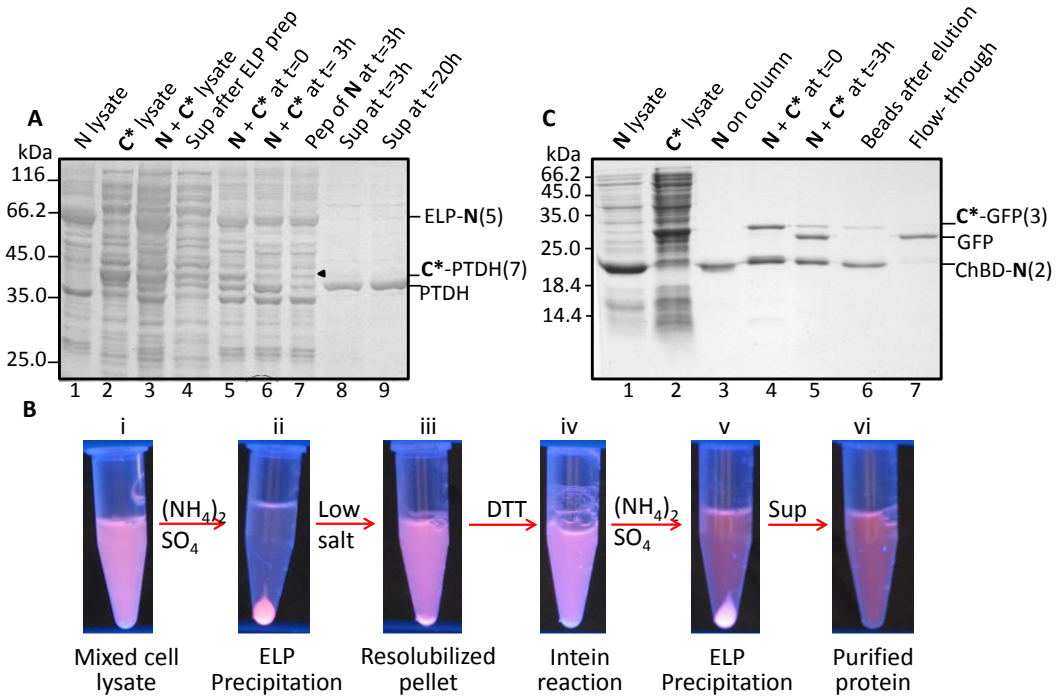


Figure 5.9 Purification of PTDH, DsRed and GFP using engineered Npu* intein. (A) SDS-PAGE (10% acrylamide) analysis of samples collected during the purification of PTDH via reversible precipitation of ELP. Lane 1, pre-purified ELP-N (5); lane 2, soluble lysate containing C*-PTDH (6); lane 3, mixture of samples from lanes 1 and 2; lane 4, supernatant after the precipitation of the ELP complex; lane 5 and 6, mixture of ELP-N and C*-PTDH at t=0 and 3 hours of intein reaction at 22 °C, respectively; lanes 7, ELP precipitant after 3 hours intein reaction; lanes 8 and 9, supernatant containing purified PTDH after ammonium sulfate precipitation after 3 and 20 hours intein reaction, respectively. An equivalent amount of protein was loaded into each lane. Black arrow indicates the un-cleaved C*-PTDH. (B) Images taken over the course of DsRed purification via reversible precipitation of ELP. (C) SDS-PAGE (12% acrylamide) analysis of samples collected during the purification of GFP via chitin column. Lane 1 and 2, soluble lysate containing CBD-N (2) and C*-GFP (3), respectively; lane 3, sample taken from chitin beads after binding of CBD-N (2); lane 4, sample taken from chitin beads immediately after binding of C*-GFP (3); lane 5, sample after 3 hours of intein reaction; lane 6, sample taken from chitin beads after elution of GFP; lane 7, elution containing purified GFP. An equivalent amount of protein was loaded into each lane, except for lanes 1 and 2.

Sample purification of a globular protein phosphite dehydrogenase (PTDH, construct 6) (*230*) is shown in Fig. 5.9A. The yields were 49.8 and 59.3 mg of purified PTDH per liter of *E. coli* culture with intein self-cleavage reaction times of 3 and 20 hours, respectively, at 22 °C. The purification of 5 additional proteins of various sizes and multimeric states (constructs 3, 7-10) are shown in Fig. 5.10. The purification process for DsRed (construct 7) can be conveniently monitored by visual inspection (Fig. 5.9B). The target protein purification yields and the percentage recovery from soluble lysate are summarized in Table 5.5. High purities were obtained for all the proteins tested. Both monomeric and multimeric proteins can be efficiently purified by this method. In most cases, the intein self-cleavage reaction is essentially complete in 4 hours, keeping the time to complete the entire procedure on par with conventional chromatographic protein purification processes.

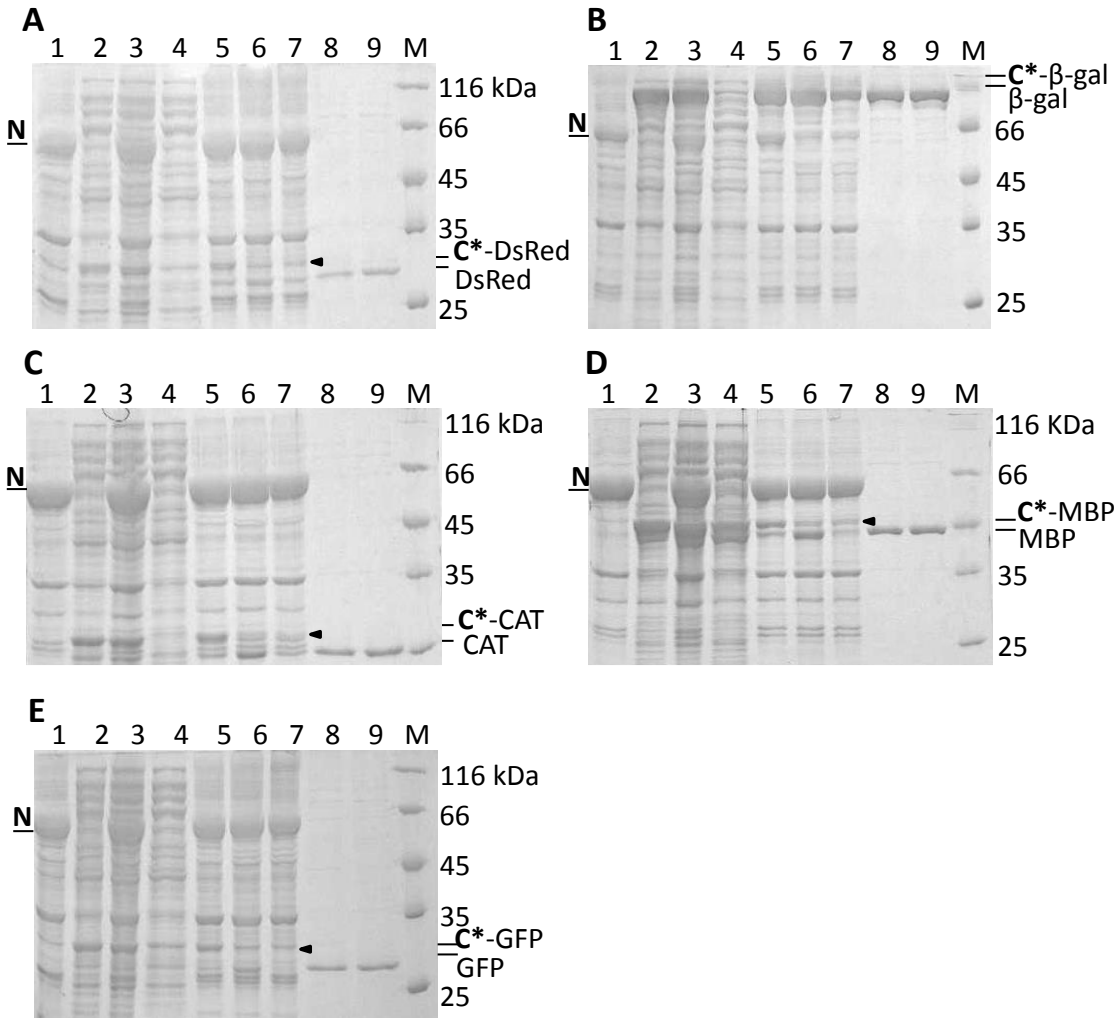


Figure 5.10 Additional sample purifications of recombinant proteins. Done using reversible precipitation of ELP and self-cleaving C*: Lane 1, pre-purified ELP-N; lane 2, soluble lysate containing C*-POI; lane 3, mixture of samples from lanes 1 and 2; lane 4, supernatant after the precipitation of the ELP complex; lane 5 and 6, ELP-N and C*-POI mixture at the beginning and after 3 hours of intein reaction, respectively; lanes 7, ELP precipitant after 3 hours intein cleavage reaction; lanes 8 and 9, supernatant containing purified POI after ammonium sulfate precipitation with 3 and 20 hours intein cleavage reaction times, respectively. An equivalent amount of protein was loaded into each lane. Black arrow indicates the un-cleaved NpuC*-POI (A) DsRed, (B) β-galactosidase (β-gal), (C) chloramphenicol acetyl transferase (CAT), (D) maltose binding protein (MBP), (E) green fluorescent protein (GFP).

Table 5.5 Protein purification and quantification via ELP-precipitation

Protein	Size (kDa)	Structure	Yield (mg/L) [b]		Percent recovery [c]		Activity [d]
			3 h[a]	20 h[a]	3 h[a]	20 h[a]	
PTDH	38.2	Dimer	49.8	59.3	34	42	NBT-methosulfate assay
β-Gal	117	Tetramer	68.1	84.5	68 (55)[f]	76 (57)[f]	261 Units/mg
CAT	27.2	Trimer	57.2	70.6	49	61	Not determined
MBP	44.1	Monomer	53.6	57.9	26	31	Affinity to Amylose
DsRed	27.2	Tetramer	48.9	54.4	44 (45)[f]	57 (55)[f]	Fluorescence 590nm
GFP	28.2	Monomer	56.2	62.2	43 (32)[f]	47 (35)[f]	Fluorescence 538nm

[a] : Intein reaction time at 22 °C.
[b] : Purification yields were determined from 25 mg of wet *E. coli* pellet.
[c] : Percent recovery was estimated using densitometry analysis of SDS-PAGE.
[d] : Protein purified with 3 hours intein reaction time is used. One β-Galactosidase unit is defined as the amount of protein needed to hydrolyze 1.0 μmole of ONPG per minute at 22°C. DTT interferes with the absorbance of the products of PTDH and CAT reaction, preventing accurate activity measurement of these proteins. Activity of MBP was analyzed qualitatively only.
[f] : Numbers in parenthesis represent percent recovery based on activity assay.

Table 5.6 ELP pull-down efficiency and purification yield calculated based on activity assay

Activity	**β-Gal**	**DsRed**	**GFP**
Pull-down efficiency [a]	93 %	93 %	41 %
Yield (3 h) [b]	56 %	45 %	32 %
Yield (20 h) [c]	59 %	48 %	37 %

[a]: Pull-down efficiency = (Lane 2 activity – Lane 4 activity) / Lane 2 activity x100 %
[b]: Yield (3 h) = Lane 8 activity / Lane 2 activity x100 %
[c]: Yield (20 h) = Lane 9 activity / Lane 2 activity x100 %

Not surprisingly, the ELP pull-down efficiency and intein cleavage kinetics are influenced by the target protein (Table 5.6). Since chromatography-based methods remain the mainstay for recombinant protein purification, we also developed a second affinity based purification method to further expand the utility of our engineered intein. In this method, the ELP is replaced with the chitin binding domain (CBD) (*239*) and purification is achieved through binding to the chitin beads (Fig. 5.8, method 2). Sample purification of GFP is shown in Fig. 5.9C. The yield of purified GFP was ~2mg from a 100μl chitin column. The binding capacity of chitin beads in our application appears to be much higher than that reported for maltose binding protein (0.2 mg/100 μL chitin beads, New England Biolabs website). The exact reason is unknown but may due in part to the much smaller size of N compared to the maltose binding protein.

5.5 Discussion

In this study, we engineered a split *Npu* DnaE intein that is able to undergo rapid C-terminal cleavage reaction without N-terminal cleavage. Although very recently four new split inteins were discovered with unprecedented rates of reaction (*20, 21*), naturally split *Npu* DnaE intein remains one of the most active inteins identified to date (Table 5.1). However, despite the rapid reaction kinetics and high solubility of *Npu* DnaE, the dependence of C-terminal asparagine cyclization on the acyl shift at the N-terminus prevents the use of DnaE intein in many applications such as protein purification. The multi-step catalytic pathway leading to intein trans-splicing is highly coordinated, but the precise mechanism involved in this series of reactions remains unclear. Intein N-/C-

terminal cleavage can result from either an increase in the rate of cleavage at that splice junction or a decrease in the reaction rate of another step. Previously, Wood, *et al.* identified a single mutation in mini-*Mtu*RecA intein, D422G, that is able to abolish the trans-splicing activity and significantly elevate the C-terminal cleavage activity (*46*). Asp422 lies in the conserved Block F region (also termed the C2 motif) within the intein active site and is 75% conserved among all inteins from different species, including split *Npu* DnaE and *Ssp* DnaE inteins (Fig. 5.4) (*103*). This Block F aspartate has previously been shown to be essential for both the first and second steps of intein reaction (Fig. 5.3). The crystal structures of multiple inteins, including the *Ssp* DnaE, show that this aspartate forms hydrogen bond with the oxyanion of the N-terminal Cys1, likely positioning this residue for the first step N-X acyl shift (*10, 24, 25*). In addition, quantum mechanics simulations and structural studies from other inteins suggest that this aspartate also deprotonates the thiol group of Cys+1, enabling it to launch nucleophilic attack to form branched intermediate (*30, 240*). Mutation of this Asp422 in mini-*Mtu*RecA significantly retards the N-terminal cleavage reaction and the formation of branched intermediate (*240*). The NMR structure of *Npu* DnaE intein showed that Asp118 (equivalent of Asp422 in mini-*Mtu*RecA) is within hydrogen bond distance with the oxyanion of the first residue (*110*), although the NMR structure contains an Ala at this position. Npu* with the corresponding aspartate changed to Gly exhibited very limited N-terminal cleavage even at 50 mM DTT (Fig. 5.5A), suggesting that this mutation likely also blocks the first step N-X acyl shift and formation of branched intermediate.

In mini-*Mtu*RecA intein, the C- and N-terminal cleavage reactions are not coupled, thus retardation of the first and second steps of intein reaction can account for the elevated C-terminal cleavage product. However, in DnaE intein, the C- and N-terminal cleavage reactions are highly coupled. Inhibition of the first two steps do not lead to elevated C-terminal cleavage product directly, as mutant DnaE inteins with C1A exhibited little to no C-terminal cleavage (*34, 185*). To understand how D118G induces C-terminal cleavage in *Npu* DnaE intein, we compared the solution structure of *Npu* DnaE with the crystal structure of its closest homolog *Ssp* DnaE (*25*). In *Ssp* DnaE intein, C-terminal asparagine cyclization is mediated by a charge relay process involving His147, Asn159, Arg73 and a water molecule near the C-terminal splicing junction (*25*) (Fig. 5.11A). The water molecule is within hydrogen bonding distance to the N^{δ} atoms of Asn159 and His147 and the backbone nitrogen of Leu143, and transfers a proton from Asn159 to His147 (Fig. 5.11A). The deprotonated Asn159 N^{δ} initiates nucleophilic attack to its carbonyl carbon atom, resulting in breakage of the C-terminal intein-extein bond.

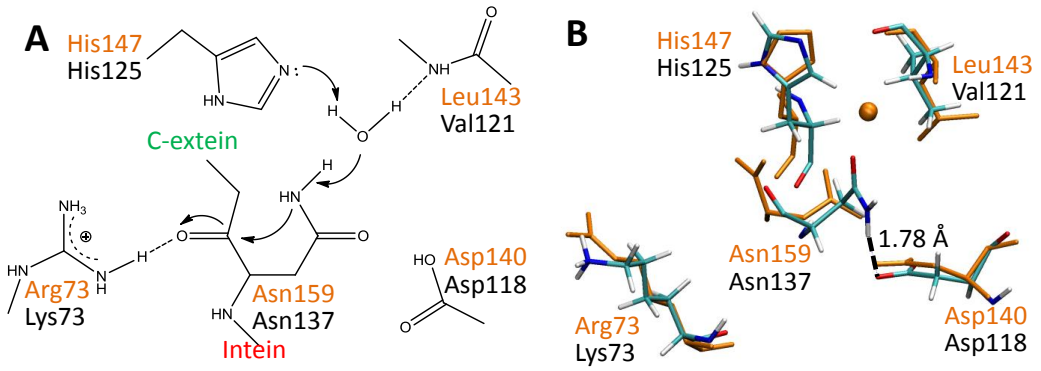


Figure 5.11 Residues that participate in a charge relay for C-terminal cleavage. (A) Schematic of the first step of charge relay responsible for C-terminal asparagine cyclization (*25*). Corresponding residues in *Ssp* DnaE (orange) and *Npu* DnaE (black) are indicated. (B) Structural alignment of charge relay residues in *Ssp* DnaE (orange, pdb: 1zd7) and *Npu* DnaE (elemental colors, pdb: 2keq). Asn137 forms H-bond with Asp117 in *Npu* DnaE, rendering it less suited to participation in the charge relay.

The same mechanism is involved in the asparagine cyclization in *Ssp* DnaB (*22*) and mini-*Mtu*RecA (*240*) which contain a water molecule in the similar position. The NMR structure of *Npu* DnaE shows that Asp118 can form a hydrogen bond with the N^{δ} of Asn137, rendering it unfavorable for charge relay and thus inhibiting C-terminal cleavage (Fig. 5.11B). Formation of a branched intermediate, which was shown to require protonation of Asp118 by the hydrogen from the thiol group in Cys+1 in some studies (*30, 240*), may break the H-bond interaction of Asn137 with Asp118, allowing Asn137 to participate in charge relay leading to C-terminal cleavage. Muir and co-workers showed that formation of a branched intermediate results in a subtle change of the intein structure that accelerates the C-terminal asparagine cyclization reaction (*241*). Mutation of Asp118 to the much smaller Gly eliminates the H-bond interaction and

enables Asn137 to freely participate in charge relay without the need of structural changes, leading to decoupled C-terminal cleavage. Thus, in addition to inhibiting the first two steps of intein reaction, D118G may also accelerate the asparagine cyclization in Npu*, leading to rapid C-terminal cleavage.

The cleavage kinetics of Npu* is slightly slower than that of the wild-type intein trans-splicing reaction (Table 5.1). This may due to the imperfect positioning of Asn137 in the absence of Asp118 side chain. Nevertheless, approximately 80% C-terminal cleavage can be achieved within 3 hours at room temperature, making this mutant intein valuable for tag-removal in protein purification. Using Npu* we developed one column-free and one chromatography-based protein purification methods by replacing the N-extein with ELP and CBD, respectively, and demonstrated rapid purification (< 4 hours) of various target proteins of various sizes and multimeric states to high purity and with high yield. However, since reducing agent is employed to trigger intein cleavage, our methods were not tested in the purification of proteins containing naturally occurring disulfide bonds. In the first method, the use of ELP eliminates the need for a costly column and should facilitate its use in large-scale industrial protein purification. In the second method, it is conceivable that other purification tags, such as his-tag, can be used in place of CBD to mediate affinity purification.

For all current studies, target protein contains the tripeptide CFN at the N-terminus after purification (Table 5.2). The AS present in constructs 6-10 corresponds to *Nhe*I site and was introduced to facilitate the cloning. The Cys+1 is important to render the intein inactive under non-reducing conditions, likely through disulfide-bond

formation with Cys1(*235*). The function of the Phe+2 and Asn+3 are unknown, but these residues likely do not play a significant role in Npu* C-terminal cleaving activity, as previously reported for the trans-splicing activity in wild-type *Npu* DnaE (*107*). Not surprisingly, the C-terminal cleavage efficiency depends not only on the immediate extein residues (*242*), but also on the target protein (Table 5.6). This variability may due to steric hindrance by different target proteins on the association of C* with N, affecting both the affinity between these two fragments as well as the intein catalytic efficiency.

In summary, we engineered an *Npu* DnaE intein via rational design. This intein exhibits rapid C-terminal cleavage kinetics independent of N-terminal cleavage. We demonstrated the application of this engineered intein for protein purification. Comparing our mutant *Npu* DnaE intein-based purification methods to the previously reported purification methods mediated by artificially split DnaB intein, our methods eliminate the dependence of small peptide (*54*) and achieve a much more rapid cleavage rate (*55*). Thus, our methods should find use in many large scale protein purification applications.

CHAPTER VI

CONCLUSIONS

The naturally split *Npu* DnaE intein was successfully engineered for the development of two different technologies. These advances are centered on applications that enable and enhance the use of enzymes in industrial settings.

First, the design, synthesis, characterization and applications of an intein-triggered protein hydrogel were shown. The resulting hydrogel had unprecedented solution stability, and was the first protein-based hydrogel used as a general enzyme immobilization scaffold. The excellent mechanical properties obtained were due to the intein-mediated trans-splicing reactions that result in peptide bond formation. It was established that such peptide bonds allowed stable hydrogel formation independent of functionality.

Since protein hydrogels can be used as artificial scaffolds in tissue engineering, drug delivery, and biosensors, the advances herein presented have the potential to greatly contribute in those applications. Most of the currently used protein hydrogels are formed via self-assembling interactions of different protein building blocks, but the resulting materials often show low physical stability. The use of inteins for hydrogel formation circumvents the use of reversible self-assembling peptides and results in covalent bonds between different cross-linkers. This approach enables the synthesis of a highly stable hydrogels. Incorporation of docking station peptides in the hydrogel building blocks enables the stable immobilization proteins in the hydrogel.

The results obtained during intein-mediated hydrogel formation were expanded for the development of a hydrogel that does not rely on intein trans-splicing reactions. Instead, this system is based on a protein-ligand pair engineered using site-directed mutagenesis. Protein interactions between the receptor and the ligand result in the formation of a disulfide bond. This interaction was adapted for protein hydrogel formation.

The second technology is based on the engineering and successful application of the *Npu* DnaE intein for rapid and efficient tag-less protein purification. This intein was modified to catalyze an inducible C-terminal cleavage reaction, and its catalytic properties were adapted for the creation of a novel tag-less protein purification methodology. The intein was engineered via rational design. This intein exhibits rapid C-terminal cleavage kinetics independent of N-terminal cleavage. Comparing the mutant *Npu* DnaE intein-based purification methods to the previously reported purification methods mediated by artificially split DnaB intein, this technology eliminates the dependence of small peptide and achieve a much more rapid cleavage rate.

REFERENCES

1. Liu, X. Q. (2000) Protein-splicing intein: genetic mobility, origin, and evolution, *Annu Rev Genet 34*, 61-76.
2. Perler, F. B., Davis, E. O., Dean, G. E., Gimble, F. S., Jack, W. E., Neff, N., Noren, C. J., Thorner, J., and Belfort, M. (1994) Protein splicing elements - inteins and exteins - a definition of terms and recommended nomenclature, *Nucleic Acids Res 22*, 1125-1127.
3. Belfort, M., Reaban, M. E., Coetzee, T., and Dalgaard, J. Z. (1995) Prokaryotic introns and inteins - a panoply of form and function, *J Bacteriol 177*, 3897-3903.
4. Belfort, M., and Roberts, R. J. (1997) Homing endonucleases: keeping the house in order, *Nucleic Acids Res 25*, 3379-3388.
5. Gogarten, J. P., Senejani, A. G., Zhaxybayeva, O., Olendzenski, L., and Hilario, E. (2002) Inteins: structure, function, and evolution, *Annu Rev Microbiol 56*, 263-287.
6. Hirata, R., Nakano, A., Kawasaki, H., Suzuki, K., and Anraku, Y. (1990) Molecular-structure of a gene, Vma1, encoding the catalytic subunit of H+-translocating adenosine-triphosphatase from vacuolar membranes of Saccharomyces cerevisiae, *J Biol Chem 265*, 6726-6733.
7. Duan, X., Gimble, F. S., and Quiocho, F. A. (1997) Crystal structure of PI-Scel, a homing endonuclease with protein splicing activity, *Cell 89*, 555-564.

8. Derbyshire, V., Wood, D. W., Wu, W., Dansereau, J. T., Dalgaard, J. Z., and Belfort, M. (1997) Genetic definition of a protein-splicing domain: functional mini-inteins support structure predictions and a model for intein evolution, *P Natl Acad Sci USA 94*, 11466-11471.

9. Perler, F. B., Olsen, G. J., and Adam, E. (1997) Compilation and analysis of intein sequences, *Nucleic Acids Res 25*, 1087-1093.

10. Klabunde, T., Sharma, S., Telenti, A., Jacobs, W. R., and Sacchettini, J. C. (1998) Crystal structure of GyrA intein from Mycobacterium xenopi reveals structural basis of protein splicing, *Nat Struct Biol 5*, 31-36.

11. Reith, M., and Munholland, J. (1995) Complete nucleotide sequence of the Porphyra purpurea chloroplast genome, *Plant Mol Biol Rep 13*, 333-335.

12. Shingledecker, K., Jiang, S. Q., and Paulus, H. (1998) Molecular dissection of the Mycobacterium tuberculosis RecA intein: design of a minimal intein and of a trans-splicing system involving two intein fragments, *Gene 207*, 187-195.

13. Mills, K. V., Lew, B. M., Jiang, S. Q., and Paulus, H. (1998) Protein splicing in trans by purified N- and C-terminal fragments of the Mycobacterium tuberculosis RecA intein, *P Natl Acad Sci USA 95*, 3543-3548.

14. Wu, H., Xu, M. Q., and Liu, X. Q. (1998) Protein trans-splicing and functional mini-inteins of a cyanobacterial DnaB intein, *Bba-Protein Struct M 1387*, 422-432.

15. Iwai, H., Lingel, A., and Plückthun, A. (2001) Cyclic green fluorescent protein produced in vivo using an artificially split PI-PfuI intein from Pyrococcus furiosus, *J Biol Chem 276*, 16548-16554.

16. Noren, C. J., Wang, J., and Perler, F. B. (2000) Dissecting the chemistry of protein splicing and its applications, *Angew Chem Int Edit 39*, 450-466.

17. Southworth, M. W., Adam, E., Panne, D., Byer, R., Kautz, R., and Perler, F. B. (1998) Control of protein splicing by intein fragment reassembly, *Embo J 17*, 918-926.

18. Wu, H., Hu, Z., and Liu, X.-Q. (1998) Protein trans-splicing by a split intein encoded in a split DnaE gene of Synechocystis sp. PCC6803, *P Natl Acad Sci USA 95*, 9226-9231.

19. Shah, N. H., Dann, G. P., Vila-Perello, M., Liu, Z. H., and Muir, T. W. (2012) Ultrafast protein splicing is common among cyanobacterial split inteins: implications for protein engineering, *J Am Chem Soc 134*, 11338-11341.

20. Dassa, B., London, N., Stoddard, B. L., Schueler-Furman, O., and Pietrokovski, S. (2009) Fractured genes: a novel genomic arrangement involving new split inteins and a new homing endonuclease family, *Nucleic Acids Res 37*, 2560-2573.

21. Carvajal-Vallejos, P., Pallisse, R., Mootz, H. D., and Schmidt, S. R. (2012) Unprecedented rates and efficiencies revealed for new natural split inteins from metagenomic sources, *J Biol Chem 287*, 28686-28696.

22. Ding, Y., Xu, M. Q., Ghosh, I., Chen, X., Ferrandon, S., Lesage, G., and Rao, Z. (2003) Crystal structure of a mini-intein reveals a conserved catalytic module involved in side chain cyclization of asparagine during protein splicing, *J Biol Chem 278*, 39133-39142.

23. Volkmann, G., and Mootz, H. D. (2013) Recent progress in intein research: from mechanism to directed evolution and applications, *Cell Mol Life Sci 70*, 1185-1206.

24. Van Roey, P., Pereira, B., Li, Z., Hiraga, K., Belfort, M., and Derbyshire, V. (2007) Crystallographic and mutational studies of Mycobacterium tuberculosis RecA mini-inteins suggest a pivotal role for a highly conserved aspartate residue, *J Mol Biol 367*, 162-173.

25. Sun, P., Ye, S., Ferrandon, S., Evans, T. C., Xu, M. Q., and Rao, Z. (2005) Crystal structures of an intein from the split DnaE gene of Synechocystis sp PCC6803 reveal the catalytic model without the penultimate histidine and the mechanism of zinc ion inhibition of protein splicing, *J Mol Biol 353*, 1093-1105.

26. Du, Z. M., Shemella, P. T., Liu, Y. Z., McCallum, S. A., Pereira, B., Nayak, S. K., Belfort, G., Belfort, M., and Wang, C. Y. (2009) Highly conserved histidine plays a dual catalytic role in protein splicing: a pK(a) shift mechanism, *J Am Chem Soc 131*, 11581-11589.

27. Hashimoto, H., Takahashi, H., Nishioka, M., Fujiwara, S., Takagi, M., Imanaka, T., Inoue, T., and Kai, Y. (2000) Crystallographic study of intein homing

endonuclease II encoded in the archaeal DNA polymerase gene, *Acta Crystallogr D 56*, 1185-1186.

28. Ichiyanagi, K., Ishino, Y., Ariyoshi, M., Komori, K., and Morikawa, K. (2000) Crystal structure of an archaeal intein-encoded homing endonuclease PI-PfuI, *J Mol Biol 300*, 889-901.

29. Perler, F. B. (1998) Protein splicing of inteins and hedgehog autoproteolysis: structure, function, and evolution, *Cell 92*, 1-4.

30. Johnson, M. A., Southworth, M. W., Herrmann, T., Brace, L., Perler, F. B., and Wuthrich, K. (2007) NMR structure of a KlbA intein precursor from Methanococcus jannaschii, *Protein Sci 16*, 1316-1328.

31. Mills, K. V., and Perler, F. B. (2005) The mechanism of intein-mediated protein splicing: variations on a theme, *Protein Peptide Lett 12*, 751-755.

32. Perler, F. B. (2000) InBase, the intein database, *Nucleic Acids Res 28*, 344-345.

33. Chong, S., Shao, Y., Paulus, H., Benner, J., Perler, F. B., and Xu, M. Q. (1996) Protein splicing involving the Saccharomyces cerevisiae VMA intein. The steps in the splicing pathway, side reactions leading to protein cleavage, and establishment of an in vitro splicing system, *J Biol Chem 271*, 22159-22168.

34. Martin, D. D., Xu, M. Q., and Evans, T. C. (2001) Characterization of a naturally occurring trans-splicing intein from Synechocystis sp PCC6803, *Biochemistry-Us 40*, 1393-1402.

35. Tori, K., Dassa, B., Johnson, M. A., Southworth, M. W., Brace, L. E., Ishino, Y., Pietrokovski, S., and Perler, F. B. (2010) Splicing of the mycobacteriophage

Bethlehem DnaB intein: identification of a new mechanistic class of inteins that contain an obligate block F nucleophile, *J Biol Chem 285*, 2515-2526.

36. Aranko, A. S., Oeemig, J. S., and Iwaï, H. (2013) Structural basis for protein trans-splicing by a bacterial intein-like domain – protein ligation without nucleophilic side chains, *FEBS Journal 280*, 3256-3269.

37. Chen, L. X., Benner, J., and Perler, F. B. (2000) Protein splicing in the absence of an intein penultimate histidine, *J Biol Chem 275*, 20431-20435.

38. Sun, W. C., Yang, J., and Liu, X. Q. (2004) Synthetic two-piece and three-piece split inteins for protein trans-splicing, *J Biol Chem 279*, 35281-35286.

39. Gimble, F. S., and Thorner, J. (1992) Homing of a DNA endonuclease gene by meiotic gene conversion in Saccharomyces cerevisiae, *Nature 357*, 301-306.

40. Callahan, B. P., Topilina, N. I., Stanger, M. J., Van Roey, P., and Belfort, M. (2011) Structure of catalytically competent intein caught in a redox trap with functional and evolutionary implications, *Nat Struct Mol Biol 18*, 630-633.

41. Young, C. L., Britton, Z. T., and Robinson, A. S. (2012) Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications, *Biotechnology Journal 7*, 620-634.

42. Terpe, K. (2003) Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems, *Appl Microbiol Biotechnol 60*, 523-533.

43. Kapust, R. B., Tozser, J., Fox, J. D., Anderson, D. E., Cherry, S., Copeland, T. D., and Waugh, D. S. (2001) Tobacco etch virus protease: mechanism of

autolysis and rational design of stable mutants with wild-type catalytic proficiency, *Protein Eng 14*, 993-1000.

44. Wood, D. W., Banki, M. R., and Feng, L. A. (2005) Simple bioseparations using self-cleaving elastin-like polypeptide tags, *Nat Methods 2*, 659-661.

45. Chong, S., Montello, G., Zhang, A., Cantor, E., Liao, W., Xu, M., and Benner, J. (1998) Utilizing the C-terminal cleavage activity of a protein splicing element to purify recombinant proteins in a single chromatographic step, *Nucleic Acids Res 26*, 5109-5115.

46. Wood, D. W., Wu, W., Belfort, G., Derbyshire, V., and Belfort, M. (1999) A genetic system yields self-cleaving inteins for bioseparations, *Nat Biotechnol 17*, 889-892.

47. Mathys, S., Evans, T. C., Chute, I. C., Wu, H., Chong, S., Benner, J., Liu, X. Q., and Xu, M. Q. (1999) Characterization of a self-splicing mini-intein and its conversion into autocatalytic N- and C-terminal cleavage elements: facile production of protein building blocks for protein ligation, *Gene 231*, 1-13.

48. Chong, S., Mersha, F., Comb, D., Scott, M., Landry, D., Vence, L., Perler, F., Benner, J., Kucera, R., Hirvonen, C., Pelletier, J., Paulus, H., and Xu, M. (1997) Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element, *Gene 192*, 271-281.

49. Ge, X., Yang, D. S. C., Trabbic-Carlson, K., Kim, B., Chilkoti, A., and Filipe, C. D. M. (2005) Self-cleavable stimulus responsive tags for protein purification without chromatography, *J Am Chem Soc 127*, 11228-11229.

50. Brenzel, S., Kurpiers, T., and Mootz, H. D. (2006) Engineering artificially split inteins for applications in protein chemistry: biochemical characterization of the split Ssp DnaB intein and comparison to the split Sce VMA intein, *Biochemistry-Us 45*, 1571-1578.

51. Wu, W. Y., and Wood, D. W. (2008) Application of the ELP-intein-based nonchromatographic protein purification technique to the CHO expression system, *Abstr Pap Am Chem S 236*.

52. Banki, M. R., Feng, L. A., and Wood, D. W. (2005) Simple bioseparations using self-cleaving elastin-like polypeptide tags, *Nat Methods 2*, 659-661.

53. Cui, C., Zhao, W., Chen, J., Wang, J., and Li, Q. (2006) Elimination of in vivo cleavage between target protein and intein in the intein-mediated protein purification systems, *Protein expres purif 50*, 74-81.

54. Volkmann, G., Sun, W., and Liu, X. (2009) Controllable protein cleavages through intein fragment complementation, *Protein Sci 18*, 2393-2402.

55. Lu, W., Sun, Z., Tang, Y., Chen, J., Tang, F., Zhang, J., and Liu, J. N. (2011) Split intein facilitated tag affinity purification for recombinant proteins with controllable tag removal by inducible auto-cleavage, *J Chromatogr A 1218*, 2553-2560.

56. Vila-Perello, M., Liu, Z. H., Shah, N. H., Willis, J. A., Idoyaga, J., and Muir, T. W. (2013) Streamlined expressed protein ligation using split inteins, *J Am Chem Soc 135*, 286-292.

57. Ramirez, M., Valdes, N., Guan, D., and Chen, Z. (2013) Engineering split intein DnaE from Nostoc punctiforme for rapid protein purification, *Protein Eng Des Sel 26*, 215-223.

58. Guan, D., Ramirez, M., and Chen, Z. (2013) Split intein mediated ultra-rapid purification of tagless protein (SIRP), *Biotechnol Bioeng 110*, 2471-2481.

59. Poggeler, S., and Elleuche, S. (2010) Inteins, valuable genetic elements in molecular biology and biotechnology, *Appl Microbiol Biotechnol 87*, 479-489.

60. Evans, T. C., Benner, J., and Xu, M. Q. (1999) The cyclization and polymerization of bacterially expressed proteins using modified self-splicing inteins, *J Biol Chem 274*, 18359-18363.

61. Evans, T. C., Martin, D., Kolly, R., Panne, D., Sun, L., Ghosh, I., Chen, L. X., Benner, J., Liu, X. Q., and Xu, M. Q. (2000) Protein trans-splicing and cyclization by a naturally split intein from the DnaE gene of Synechocystis species PCC6803, *J Biol Chem 275*, 9091-9094.

62. Scott, C. P., Abel-Santos, E., Wall, M., Wahnon, D. C., and Benkovic, S. J. (1999) Production of cyclic peptides and proteins in vivo, *P Natl Acad Sci USA 96*, 13638-13643.

63. Cheriyan, M., and Perler, F. B. (2009) Protein splicing: a versatile tool for drug discovery, *Adv Drug Deliver Rev 61*, 899-907.

64. Tavassoli, A., and Benkovic, S. J. (2007) Split-intein mediated circular ligation used in the synthesis of cyclic peptide libraries in E coli, *Nat Protoc 2*, 1126-1133.

65. Jubb, H., Higueruelo, A. P., Winter, A., and Blundell, T. L. (2012) Structural biology and drug discovery for protein-protein interactions, *Trends Pharmacol Sci 33*, 241-248.

66. Scapin, G. (2006) Structural biology and drug discovery, *Curr Pharm Design 12*, 2087-2097.

67. Congreve, M., Murray, C. W., and Blundell, T. L. (2005) Keynote review: structural biology and drug discovery, *Drug Discov Today 10*, 895-907.

68. Volkmann, G., and Iwai, H. (2010) Protein trans-splicing and its use in structural biology: opportunities and limitations, *Molecular BioSystems 6*, 2110-2121.

69. Mittermaier, A., and Kay, L. E. (2006) New tools provide new insights in NMR studies of protein dynamics, *Science 312*, 224-228.

70. Percy, A. J., Rey, M., Burns, K. M., and Schriemer, D. C. (2012) Probing protein interactions with hydrogen/deuterium exchange and mass spectrometry - a review, *Anal Chim Acta 721*, 7-21.

71. Clore, G. M., and Gronenborn, A. M. (1994) Structures of larger proteins and protein-ligand and protein-DNA complexes by heteronuclear multidimensional NMR, *J Protein Chem 13*, 441-442.

72. Yamazaki, T., Otomo, T., Oda, N., Kyogoku, Y., Uegaki, K., Ito, N., Ishino, Y., and Nakamura, H. (1998) Segmental isotope labeling for protein NMR using peptide splicing, *J Am Chem Soc 120*, 5591-5592.

73. Otomo, T., Ito, N., Kyogoku, Y., and Yamazaki, T. (1999) NMR observation of selected segments in a larger protein: central-segment isotope labeling through intein-mediated ligation, *Biochemistry-Us 38*, 16040-16044.

74. Busche, A. E., Aranko, A. S., Talebzadeh-Farooji, M., Bernhard, F., Dotsch, V., and Iwai, H. (2009) Segmental isotopic labeling of a central domain in a multidomain protein by protein trans-splicing using only one robust DnaE intein, *Angew Chem Int Edit 48*, 6128-6131.

75. Zuger, S., and Iwai, H. (2005) Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies, *Nat Biotechnol 23*, 736-740.

76. Oeemig, J. S., Zhou, D. W., Kajander, T., Wlodawer, A., and Iwai, H. (2012) NMR and crystal structures of the Pyrococcus horikoshii RadA intein guide a strategy for engineering a highly efficient and promiscuous intein, *J Mol Biol 421*, 85-99.

77. Lockless, S. W., and Muir, T. W. (2009) Traceless protein splicing utilizing evolved split inteins, *P Natl Acad Sci USA 106*, 10999-11004.

78. Vila-Perello, M., and Muir, T. W. (2010) Biological applications of protein splicing, *Cell 143*, 191-200.

79. Durek, T., and Becker, C. F. W. (2005) Protein semi-synthesis: new proteins for functional and structural studies, *Biomol Eng 22*, 153-172.

80. Liu, C. C., and Schultz, P. G. (2010) Adding new chemistries to the genetic code, *Annu Rev Biochem 79*, 413-444.

81. Hackenberger, C. P. R., and Schwarzer, D. (2008) Chemoselective ligation and modification strategies for peptides and proteins, *Angew Chem Int Edit 47*, 10030-10074.

82. Mootz, H. D. (2009) Split inteins as versatile tools for protein semisynthesis, *Chembiochem 10*, 2579-2589.

83. Muir, T. W., Sondhi, D., and Cole, P. A. (1998) Expressed protein ligation: a general method for protein engineering, *PNatl Acad Sci USA 95*, 6705-6710.

84. Dawson, P. E., Muir, T. W., Clarklewis, I., and Kent, S. B. H. (1994) Synthesis of proteins by native chemical ligation, *Science 266*, 776-779.

85. Kent, S. (2010) Origin of the chemical ligation concept for the total synthesis of enzymes (proteins), *Biopolymers 94*, iv-ix.

86. Kent, S. B. H. (2009) Total chemical synthesis of proteins, *Chem Soc Rev 38*, 338-351.

87. Xia, Z. Y., Xing, Y., So, M. K., Koh, A. L., Sinclair, R., and Rao, J. H. (2008) Multiplex detection of protease activity with quantum dot nanosensors prepared by intein-mediated specific bioconjugation, *Anal Chem 80*, 8649-8655.

88. Gao, W. P., Liu, W. G., Christensen, T., Zalutsky, M. R., and Chilkoti, A. (2010) In situ growth of a PEG-like polymer from the C terminus of an intein fusion protein improves pharmacokinetics and tumor accumulation, *P Natl Acad Sci USA 107*, 16432-16437.

89. Yu, H. H., Nakase, I., Pujals, S., Hirose, H., Tanaka, G., Katayama, S., Imanishi, M., and Futaki, S. (2010) Expressed protein ligation for the preparation of fusion

proteins with cell penetrating peptides for endotoxin removal and intracellular delivery, *Bba-Biomembranes 1798*, 2249-2257.

90. Anderson, L. L., Marshall, G. R., Crocker, E., Smith, S. O., and Baranski, T. J. (2005) Motion of carboxyl terminus of G alpha is restricted upon G protein activation, *J Biol Chem 280*, 31019-31026.

91. Chatterjee, C., and Muir, T. W. (2010) Chemical approaches for studying histone modifications, *J Biol Chem 285*, 11045-11050.

92. Schwarzer, D., and Cole, P. A. (2005) Protein semisynthesis and expressed protein ligation: chasing a protein's tail, *Curr Opin Chem Biol 9*, 561-569.

93. Lu, X., Olsen, S. K., Capili, A. D., Cisar, J. S., Lima, C. D., and Tan, D. S. (2010) Designed semisynthetic protein inhibitors of Ub/Ubl E1 activating enzymes, *J Am Chem Soc 132*, 1748-1749.

94. Valiyaveetil, F. I., Sekedat, M., MacKinnon, R., and Muir, T. W. (2006) Structural and functional consequences of an amide-to-ester substitution in the selectivity filter of a potassium channel, *J Am Chem Soc 128*, 11591-11599.

95. Valiyaveetil, F. I., Leonetti, M., Muir, T. W., and MacKinnon, R. (2006) Ion selectivity in a semisynthetic K+ channel locked in the conductive conformation, *Science 314*, 1004-1007.

96. Lesaicherre, M. L., Lue, R. Y. P., Chen, G. Y. J., Zhu, Q., and Yao, S. Q. (2002) Intein-mediated biotinylation of proteins and its application in a protein microarray, *J Am Chem Soc 124*, 8768-8769.

97. Chattopadhaya, S., Abu Bakar, F. B., and Yao, S. Q. (2009) Use of intein-mediated protein ligation strategies for the fabrication of functional protein arrays, *Method Enzymol 462*, 195-223.

98. Nyanguile, O., Dancik, C., Blakemore, J., Mulgrew, K., Kaleko, M., and Stevenson, S. C. (2003) Synthesis of adenoviral targeting molecules by intein-mediated protein ligation, *Gene Ther 10*, 1362-1369.

99. Bradley, L. H., Kleiner, R. E., Wang, A. F., Hecht, M. H., and Wood, D. W. (2005) An intein-based genetic selection allows the construction of a high-quality library of binary patterned de novo protein sequences, *Protein Eng Des Sel 18*, 201-207.

100. Bradley, L. H., Kleiner, R. E., Wang, A. F., Hecht, M. H., and Wood, D. W. (2004) Preselection of a library of in-frame sequences, independent of fold, using an intein-thymidylate synthase reporter system, *Protein Sci 13*, 213-214.

101. Kanno, A., Ozawa, T., and Umezawa, Y. (2006) Intein-mediated reporter gene assay for detecting protein-protein interactions in living mammalian cells, *Anal Chem 78*, 556-560.

102. Colston, M. J., and Davis, E. O. (1994) The ins and outs of protein splicing elements, *Mol Microbiol 12*, 359-363.

103. Pietrokovski, S. (1994) Conserved sequence features of inteins (protein introns) and their use in identifying new inteins and related proteins, *Protein Sci 3*, 2340-2350.

104. Iwai, H., Lingel, A., and Pluckthun, A. (2001) Cyclic green fluorescent protein produced in vivo using an artificially split PI-PfuI intein from Pyrococcus furiosus, *J Biol Chem 276*, 16548-16554.

105. Lew, B. M., Mills, K. V., and Paulus, H. (1999) Characteristics of protein splicing in trans mediated by a semisynthetic split intein, *Peptide Science 51*, 355-362.

106. Chen, L., Pradhan, S., and Evans Jr, T. C. (2001) Herbicide resistance from a divided EPSPS protein: the split Synechocystis DnaE intein as an in vivo affinity domain, *Gene 263*, 39-48.

107. Iwai, H., Zuger, S., Jin, J., and Tam, P. H. (2006) Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostoc punctiforme, *FEBS Lett 580*, 1853-1858.

108. Cheriyan, M., Pedamallu, C. S., Tori, K., and Perler, F. (2013) Faster protein splicing with the Nostoc punctiforme DnaE intein using non-native extein residues, *J Biol Chem 288*, 6202-6211.

109. Ludwig, C., Pfeiff, M., Linne, U., and Mootz, H. D. (2006) Ligation of a synthetic peptide to the N terminus of a recombinant protein using semisynthetic protein trans-splicing, *Angew Chem Int Edit 45*, 5218-5221.

110. Oeemig, J. S., Aranko, A. S., Djupsjobacka, J., Heinamaki, K., and Iwai, H. (2009) Solution structure of DnaE intein from Nostoc punctiforme: structural basis for the design of a new split intein suitable for site-specific chemical modification, *FEBS Lett 583*, 1451-1456.

111. Lin, Y., Li, M., Song, H., Xu, L., Meng, Q., and Liu, X. Q. (2013) Protein trans-splicing of multiple atypical split inteins engineered from natural inteins, *PLoS One 8*, e59516.

112. Callahan, B., Topilina, N., Stanger, M., Van Roey, P., and Belfort, M. (2011) Disulfide engineering catches a self-splicing intein in a redox trap, *FASEB J 25*, 715.

113. Callahan, B. P., Stanger, M., and Belfort, M. (2013) A redox trap to augment the intein toolbox, *Biotechnol Bioeng 110*, 1565-1573.

114. Shah, N. H., Vila-Perello, M., and Muir, T. W. (2011) Kinetic control of one-pot trans-splicing reactions by using a wild-type and designed split intein, *Angew Chem Int Edit 50*, 6511-6515.

115. Banta, S., Wheeldon, I. R., and Blenner, M. (2010) Protein engineering in the development of functional hydrogels, *Annu Rev Biomed Eng 12*, 167-186.

116. Petka, W. A., Harden, J. L., McGrath, K. P., Wirtz, D., and Tirrell, D. A. (1998) Reversible hydrogels from self-assembling artificial proteins, *Science 281*, 389-392.

117. Langer, R., and Tirrell, D. A. (2004) Designing materials for biology and medicine, *Nature 428*, 487-492.

118. Ulijn, R. V., and Smith, A. M. (2008) Designing peptide based nanomaterials, *Chem Soc Rev 37*, 664-675.

119. Wheeldon, I. R., Gallaway, J. W., Barton, S. C., and Banta, S. (2008) Bioelectrocatalytic hydrogels from electron-conducting metallopolypeptides

coassembled with bifunctional enzymatic building blocks, *P Natl Acad Sci USA 105*, 15275-15280.

120. Wheeldon, I. R., Barton, S. C., and Banta, S. (2007) Bioactive proteinaceous hydrogels from designed bifunctional building blocks, *Biomacromolecules 8*, 2990-2994.

121. Volkmann, G., and Liu, X. Q. (2011) Intein lacking conserved C-terminal motif G retains controllable N-cleavage activity, *FEBS Journal 278*, 3431-3446.

122. Almdal, K., Dyre, J., Hvidt, S., and Kramer, O. (1993) Towards a phenomenological definition of the term gel, *Polymer Gels and Networks 1*, 5-17.

123. Peppas, N. A., Huang, Y., Torres-Lugo, M., Ward, J. H., and Zhang, J. (2000) Pysicochemical foundations and structural design of hydrogels in medicine and biology, *Annu Rev Biomed Eng 2*, 9-29.

124. Peppas, N. A., Moynihan, H. J., and Lucht, L. M. (1985) The structure of highly crosslinked poly(2-hydroxyethyl methacrylate) hydrogels, *J Biomed Mater Res 19*, 397-411.

125. Langer, R., Fisher, O. Z., Khademhosseini, A., and Peppas, N. A. (2010) Bioinspired materials for controlling stem cell fate, *Accounts Chem Res 43*, 419-428.

126. Kopecek, J. (2007) Hydrogel biomaterials: a smart future?, *Biomaterials 28*, 5185-5192.

127. Gupta, P., Vermani, K., and Garg, S. (2002) Hydrogels: from controlled release to pH-responsive drug delivery, *Drug Discov Today 7*, 569-579.

128. Qiu, Y., and Park, K. (2001) Environment-sensitive hydrogels for drug delivery, *Adv Drug Deliver Rev 53*, 321-339.

129. Wong Po Foo, C. T., Lee, J. S., Mulyasasmita, W., Parisi-Amon, A., and Heilshorn, S. C. (2009) Two-component protein-engineered physical hydrogels for cell encapsulation, *P Natl Acad Sci USA 106*, 22067-22072.

130. Richter, A., Paschew, G., Klatt, S., Lienig, J., Arndt, K.-F., and Adler, H.-J. (2008) Review on hydrogel-based pH sensors and microsensors, *Sensors 8*, 561-581.

131. Kim, B., Lee, Y., Lee, K., and Koh, W.-G. (2009) Immobilization of enzymes within hydrogel microparticles to create optical biosensors for the detection of organophosphorus compounds, *Curr Applied Physics 9*, 225-228.

132. Wang, Q., Mynar, J. L., Yoshida, M., Lee, E., Lee, M., Okuro, K., Kinbara, K., and Aida, T. (2010) High-water-content mouldable hydrogels by mixing clay and a dendritic molecular binder, *Nature 463*, 339-343.

133. Ribeiro, A., Veiga, F., Santos, D., Torres-Labandeira, J. J., Concheiro, A., and Alvarez-Lorenzo, C. (2011) Bioinspired imprinted PHEMA-hydrogels for ocular delivery of carbonic anhydrase inhibitor drugs, *Biomacromolecules 12*, 701-709.

134. Voldřrich, Z., Tománek, Z., Vacík, J., and Kopečcek, J. (1975) Long-term experience with poly(glycol monomethacrylate) gel in plastic operations of the nose, *J Biomed Mater Res 9*, 675-685.

135. Okumura, Y., and Ito, K. (2001) The polyrotaxane gel: a topological gel by figure-of-eight cross-links, *Adv Mater 13*, 485-487.

136. Wright, E. R., McMillan, R. A., Cooper, A., Apkarian, R. P., and Conticello, V. P. (2002) Thermoplastic elastomer hydrogels via self-assembly of an elastin-mimetic triblock polypeptide, *Adv Funct Mater 12*, 149-154.

137. Pochan, D. J., Schneider, J. P., Kretsinger, J., Ozbas, B., Rajagopal, K., and Haines, L. (2003) Thermally reversible hydrogels via intramolecular folding and consequent self-assembly of a de novo designed peptide, *J Am Chem Soc 125*, 11802-11803.

138. Nagai, Y., Unsworth, L. D., Koutsopoulos, S., and Zhang, S. G. (2006) Slow release of molecules in self-assembling peptide nanofiber scaffold, *J Control Release 115*, 18-25.

139. Shen, W., Lammertink, R. G. H., Sakata, J. K., Kornfield, J. A., and Tirrell, D. A. (2005) Assembly of an artificial protein hydrogel through leucine zipper aggregation and disulfide bond formation, *Macromolecules 38*, 3909-3916.

140. Shen, W., Kornfield, J. A., and Tirrell, D. A. (2007) Structure and mechanical properties of artificial protein hydrogels assembled through aggregation of leucine zipper peptide domains, *Soft Matter 3*, 99-107.

141. Kopecek, J., and Yang, J. (2009) Peptide-directed self-assembly of hydrogels, *Acta Biomater 5*, 805-816.

142. Kopecek, J., Xu, C. Y., and Breedveld, V. (2005) Reversible hydrogels from self-assembling genetically engineered protein block copolymers, *Biomacromolecules 6*, 1739-1749.

143. Lao, U. L., Sun, M., Matsumoto, M., Mulchandani, A., and Chen, W. (2007) Genetic engineering of self-assembled protein hydrogel based on elastin-like sequences with metal binding functionality, *Biomacromolecules 8*, 3736-3739.

144. Lu, H. D., Wheeldon, I. R., and Banta, S. (2010) Catalytic biomaterials: engineering organophosphate hydrolase to form self-assembling enzymatic hydrogels, *Protein Eng Des Sel 23*, 559-566.

145. Diaz, D. D., Kuhbeck, D., and Koopmans, R. J. (2011) Stimuli-responsive gels as reaction vessels and reusable catalysts, *Chem Soc Rev 40*, 427-448.

146. Olsen, B. D., Kornfield, J. A., and Tirrell, D. A. (2010) Yielding behavior in injectable hydrogels from telechelic proteins, *Macromolecules 43*, 9094-9099.

147. Bahney, C. S., Lujan, T. J., Hsu, C. W., Bottlang, M., West, J. L., and Johnstone, B. (2011) Visible light photoinitiation of mesenchymal stem cell-laden bioresponsive hydrogels, *Eur Cells Mater 22*, 43-55.

148. Sheldon, R. A. (2007) Enzyme immobilization: the quest for optimum performance, *Adv Synth Catal 349*, 1289-1307.

149. Kallenberg, A. I., van Rantwijk, F., and Sheldon, R. A. (2005) Immobilization of penicillin G acylase: the key to optimum performance, *Adv Synth Catal 347*, 905-926.

150. Wegman, M. A., Janssen, M. H. A., van Rantwijk, F., and Sheldon, R. A. (2001) Towards biocatalytic synthesis of beta-lactam antibiotics, *Adv Synth Catal 343*, 559-576.

151. Katchalski-Katzir, E., and Kraemer, D. M. (2000) Eupergit (R) C, a carrier for immobilization of enzymes of industrial potential, *J Mol Catal B-Enzym 10*, 157-176.

152. Krajewska, B. (2004) Application of chitin- and chitosan-based materials for enzyme immobilizations: a review, *Enzyme Microb Tech 35*, 126-139.

153. de Alteriis, E., Parascandola, P., Pecorella, M., Scardi, V.,. (1988) Effect of gelatin-immobilization on the catalytic activity of enzymes and microbial cells, *Biotechnol Tech 2*, 205-210.

154. Belgoudi, J., and Fortier, G. (1999) Poly(ethylene glycol)-bovine serum albumin hydrogel as a matrix for enzyme immobilization. In vitro biochemical characterization, *J Bioact Compat Pol 14*, 31-53.

155. Petri, A., Marconcini, P., and Salvadori, P. (2005) Efficient immobilization of epoxide hydrolase onto silica gel and use in the enantio selective hydrolysis of racemic para-nitrostyrene oxide, *J Mol Catal B-Enzym 32*, 219-224.

156. Diaz, J. F., and Balkus, K. J. (1996) Enzyme immobilization in mesoporous MCM-41 type molecular sieves., *J Mol Catal B-Enzym 2*, 115-126.

157. Yan, A. X., Li, X. W., and Ye, Y. H. (2002) Recent progress on immobilization of enzymes on molecular sieves for reactions in organic solvents, *Appl Biochem Biotech 101*, 113-129.

158. Moelans, D., Cool, P., Baeyens, J., and Vansant, E. F. (2005) Immobilisation behaviour of biomolecules in mesoporous silica materials, *Catal Commun 6*, 591-595.

159. Rusmini, F., Zhong, Z., and Feijen, J. (2007) Protein immobilization strategies for protein biochips, *Biomacromolecules 8*, 1775-1789.

160. Frenkel-Mullerad, H., and Avnir, D. (2005) Sol-gel materials as efficient enzyme protectors: preserving the activity of phosphatases under extreme pH conditions, *J Am Chem Soc 127*, 8077-8081.

161. Itoh, T., Ishii, R., Ebina, T., Hanaoka, T., Fukushima, Y., and Mizukami, F. (2006) Encapsulation of myoglobin with a mesoporous silicate results in new capabilities, *Bioconjug Chem 17*, 236-240.

162. Patil, A. J., Muthusamy, E., and Mann, S. (2004) Synthesis and self-assembly of organoclay-wrapped biomolecules, *Angew Chem Int Edit 43*, 4928-4933.

163. Kim, J., and Grate, J. W. (2003) Single-enzyme nanoparticles armored by a nanometer-scale organic/inorganic network, *Nano Lett 3*, 1219-1222.

164. Yan, M., Ge, J., Liu, Z., and Ouyang, P. (2006) Encapsulation of single enzyme in nanogel with enhanced biocatalytic activity and stability, *J Am Chem Soc 128*, 11008-11009.

165. Ge, J., Lu, D., Wang, J., and Liu, Z. (2009) Lipase nanogel catalyzed transesterification in anhydrous dimethyl sulfoxide, *Biomacromolecules 10*, 1612-1618.

166. Ge, J., Lu, D., Wang, J., Yan, M., Lu, Y., and Liu, Z. (2008) Molecular fundamentals of enzyme nanogels, *J Phys Chem B 112*, 14319-14324.

167. Kim, J., Grate, J. W., and Wang, P. (2008) Nanobiocatalysis and its potential applications, *Trends Biotechnol 26*, 639-646.

168. Cao, L. Q., van Rantwijk, F., and Sheldon, R. A. (2000) Cross-linked enzyme aggregates: a simple and effective method for the immobilization of penicillin acylase, *Org Lett 2*, 1361-1364.

169. Cao, L., van Langen, L. M., van Rantwijk, F., and Sheldon, R. A. (2001) Cross-linked aggregates of penicillin acylase: robust catalysts for the synthesis of beta-lactam antibiotics, *J Mol Catal B-Enzym 11*, 665-670.

170. Sheldon, R. A., Schoevaart, R., and Van Langen, L. M. (2005) Cross-linked enzyme aggregates (CLEAs): A novel and versatile method for enzyme immobilization, *Biocatal Biotransfor 23*, 141-147.

171. Sheldon, R. A. (2007) Cross-linked enzyme aggregates (CLEA (R) s): stable and recyclable biocatalysts, *Biochem Soc T 35*, 1583-1587.

172. Vafiadi, C., Topakas, E., and Christakopoulos, P. (2008) Preparation of multipurpose cross-linked enzyme aggregates and their application to production of alkyl ferulates, *J Mol Catal B-Enzym 54*, 35-41.

173. Dalal, S., Kapoor, M., and Gupta, M. N. (2007) Preparation and characterization of combi-CLEAs catalyzing multiple non-cascade reactions, *J Mol Catal B-Enzym 44*, 128-132.

174. Mateo, C., Chmura, A., Rustler, S., van Rantwijk, F., Stolz, A., and Sheldon, R. A. (2006) Synthesis of enantiomerically pure (S)-mandelic acid using an oxynitrilase-nitrilase bienzymatic cascade: a nitrilase surprisingly shows nitrile hydratase activity, *Tetrahedron-Asymmetr 17*, 320-323.

175. Muralidharan, V., and Muir, T. W. (2006) Protein ligation: an enabling technology for the biophysical analysis of proteins, *Nat Methods 3*, 429-438.

176. Kalia, J., Abbott, N. L., and Raines, R. T. (2007) General method for site-specific protein immobilization by Staudinger ligation, *Bioconjug Chem 18*, 1064-1069.

177. Weikart, N. D., and Mootz, H. D. (2010) Generation of site-specific and enzymatically stable conjugates of recombinant proteins with ubiquitin-like modifiers by the Cu(I)-catalyzed azide-alkyne cycloaddition, *Chembiochem 11*, 774-777.

178. Wieczorek, B., Lemcke, B., Dijkstra, H. P., Egmond, M. R., Gebbink, R. J. M. K., and van Koten, G. (2010) Site-selective ser-hydrolase labelling with a luminescent organometallic NCN-platinum complex, *Eur J Inorg Chem*, 1929-1938.

179. Lim, D. W., Trabbic-Carlson, K., MacKay, J. A., and Chilkoti, A. (2007) Improved non-chromatographic purification of a recombinant protein by cationic elastin-like polypeptides, *Biomacromolecules 8*, 1417-1424.

180. Fong, B. A., Wu, W.-Y., and Wood, D. W. (2010) The potential role of self-cleaving purification tags in commercial-scale processes, *TrendsBiotechnol 28*, 272-279.

181. Osicka, R., Prochazkova, K., Sulc, M., Linhartova, I., Havlicek, V., and Sebo, P. (2004) A novel "clip-and-link" activity of repeat in toxin (RTX) proteins from gram-negative pathogens: covalent protein cross-linking by an Asp-Lys isopeptide bond upon calcium-dependent processing at an Asp-Pro bond, *J Biol Chem 279*, 24944-24956.

182. Sadilkova, L., Osicka, R., Sulc, M., Linhartova, I., Novak, P., and Sebo, P. (2008) Single-step affinity purification of recombinant proteins using a self-excising module from Neisseria meningitidis FrpC, *Protein Sci 17*, 1834-1843.

183. Mazmanian, S. K., Liu, G., Ton-That, H., and Schneewind, O. (1999) Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall, *Science 285*, 760-763.

184. Mateo, C., Palomo, J. M., Fernandez-Lorente, G., Guisan, J. M., and Fernandez-Lafuente, R. (2007) Improvement of enzyme activity, stability and selectivity via immobilization techniques, *Enzyme Microb Tech 40*, 1451-1463.

185. Zettler, J., Schutz, V., and Mootz, H. D. (2009) The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction, *FEBS Lett 583*, 909-914.

186. Tanaka, Y., Tsumoto, K., Nakanishi, T., Yasutake, Y., Sakai, N., Yao, M., Tanaka, I., and Kumagai, I. (2004) Structural implications for heavy metal-induced reversible assembly and aggregation of a protein: the case of Pyrococcus horikoshii CutA, *FEBS Lett 556*, 167-174.

187. Sawano, M., Yamamoto, H., Ogasahara, K., Kidokoro, S., Katoh, S., Ohnuma, T., Katoh, E., Yokoyama, S., and Yutani, K. (2008) Thermodynamic basis for the stabilities of three CutA1s from Pyrococcus horikoshii,Thermus thermophilus, and Oryza sativa, with unusually high denaturation temperatures, *Biochemistry-Us 47*, 721-730.

188. Tanaka, T., Sawano, M., Ogasahara, K., Sakaguchi, Y., Bagautdinov, B., Katoh, E., Kuroishi, C., Shinkai, A., Yokoyama, S., and Yutani, K. (2006) Hyper-thermostability of CutA1 protein, with a denaturation temperature of nearly 150 degrees C, *FEBS Lett 580*, 4224-4230.

189. Shen, W., Zhang, K., Kornfield, J. A., and Tirrell, D. A. (2006) Tuning the erosion rate of artificial protein hydrogels through control of network topology, *Nat Mater 5*, 153-158.

190. McGrath, K. P., Fournier, M. J., Mason, T. L., and Tirrell, D. A. (1992) Genetically directed syntheses of new polymeric materials. Expression of artificial genes encoding proteins with repeating -(AlaGly)3ProGluGly- elements, *J Am Chem Soc 114*, 727-733.

191. Chen, Z., Katzenellenbogen, B. S., Katzenellenbogen, J. A., and Zhao, H. (2004) Directed evolution of human estrogen receptor variants with significantly enhanced androgen specificity and affinity, *J Biol Chem 279*, 33855-33864.

192. Ito, F., Usui, K., Kawahara, D., Suenaga, A., Maki, T., Kidoaki, S., Suzuki, H., Taiji, M., Itoh, M., Hayashizaki, Y., and Matsuda, T. (2010) Reversible hydrogel

formation driven by protein-peptide-specific interaction and chondrocyte entrapment, *Biomaterials 31*, 58-66.

193. Dueber, J. E., Wu, G. C., Malmirchegini, G. R., Moon, T. S., Petzold, C. J., Ullal, A. V., Prather, K. L., and Keasling, J. D. (2009) Synthetic protein scaffolds provide modular control over metabolic flux, *Nat Biotechnol 27*, 753-U107.

194. Bruns, N., and Tiller, J. C. (2005) Amphiphilic network as nanoreactor for enzymes in organic solvents, *Nano Lett 5*, 45-48.

195. Mohlmann, S., Bringmann, P., Greven, S., and Harrenga, A. (2011) Site-specific modification of ED-B-targeting antibody using intein-fusion technology, *BMC Biotech 11*, 76.

196. Skrzeszewska, P. J., Sprakel, J., de Wolf, F. A., Fokkink, R., Stuart, M. A. C., and van der Gucht, J. (2010) Fracture and self-healing in a well-defined self-assembled polymer network, *Macromolecules 43*, 3542-3548.

197. Yu, L., and Ding, J. (2008) Injectable hydrogels as unique biomedical materials, *Chem Soc Rev 37*, 1473-1481.

198. Lu, H. D., Charati, M. B., Kim, I. L., and Burdick, J. A. (2012) Injectable shear-thinning hydrogels engineered with a self-assembling dock-and-lock mechanism, *Biomaterials 33*, 2145-2153.

199. Shen, W., Kornfield, J. A., and Tirrell, D. A. (2007) Dynamic properties of artificial protein hydrogels assembled through aggregation of leucine zipper peptide domains, *Macromolecules 40*, 689-692.

200. Novak, I. L., Kraikivski, P., and Slepchenko, B. M. (2009) Diffusion in cytoplasm: effects of excluded volume due to internal membranes and cytoskeletal structures, *Biophys J 97*, 758-767.

201. Svitkina, T. M., Verkhovsky, A. B., McQuade, K. M., and Borisy, G. G. (1997) Analysis of the actin-myosin II system in fish epidermal keratocytes: mechanism of cell body translocation, *J Cell Biol 139*, 397-415.

202. Nguyen, J. T., Turck, C. W., Cohen, F. E., Zuckermann, R. N., and Lim, W. A. (1998) Exploiting the basis of proline recognition by SH3 and WW domains: design of n-substituted inhibitors, *Science 282*, 2088-2092.

203. Jung, G. Y., and Stephanopoulos, G. (2004) A functional protein chip for pathway optimization and in vitro metabolic engineering, *Science 304*, 428-431.

204. Schmid, A., Dordick, J. S., Hauer, B., Kiener, A., Wubbolts, M., and Witholt, B. (2001) Industrial biocatalysis today and tomorrow, *Nature 409*, 258-268.

205. Liszka, M. J., Clark, M. E., Schneider, E., and Clark, D. S. (2012) Nature versus nurture: developing enzymes that function under extreme conditions, *Annu Rev Chem Biomol 3*, 77-102.

206. Iyer, P. V., and Ananthanarayan, L. (2008) Enzyme stability and stabilization - aqueous and non-aqueous environment, *Process Biochem 43*, 1019-1032.

207. Maity, G. C. (2008) Supramolecular hydrogels, *J Phys Sci 12*, 173-186.

208. Veronese, F. M., and Pasut, G. (2005) PEGylation, successful approach to drug delivery, *Drug Discov Today 10*, 1451-1458.

209. Nguyen, M. K., and Lee, D. S. (2010) Injectable biodegradable hydrogels, *Macromol Biosci 10*, 563-579.

210. Jonker, A. M., Lowik, D. W. P. M., and van Hest, J. C. M. (2012) Peptide- and protein-based hydrogels, *Chem Mater 24*, 759-773.

211. Byun, J. Y., Lee, K. H., Lee, K. Y., Kim, M. G., and Kim, D. M. (2013) In-gel expression and in situ immobilization of proteins for generation of three dimensional protein arrays in a hydrogel matrix, *Lab on a Chip 13*, 886-891.

212. Vazquez-Duhalt, R., Tinoco, R., D'Antonio, P., Topoleski, L. D., and Payne, G. F. (2001) Enzyme conjugation to the polysaccharide chitosan: smart biocatalysts and biocatalytic hydrogels, *Bioconjug Chem 12*, 301-306.

213. Ramirez, M., Guan, D., Ugaz, V., and Chen, Z. (2013) Intein-triggered artificial protein hydrogels that support the immobilization of bioactive proteins, *J Am Chem Soc 135*, 5290-5293.

214. Chen, Z., and Zhao, H. (2005) A highly sensitive selection method for directed evolution of homing endonucleases, *Nucleic Acids Res 33*, e154.

215. Zhang, J., Yan, X., Shi, C., Yang, X., Guo, Y., Tian, C., Long, J., and Shen, Y. (2008) Structural basis of beta-catenin recognition by Tax-interacting protein-1, *J Mol Biol 384*, 255-263.

216. Fournier, M. J., Creel, H. S., Krejchi, M. T., Mason, T. L., Tirrell, D. A., Mcgrath, K. P., and Atkins, E. D. T. (1991) Genetic synthesis of periodic protein materials, *J Bioact Compat Pol 6*, 326-338.

217. Wu, X., Knudsen, B., Feller, S. M., Zheng, J., Sali, A., Cowburn, D., Hanafusa, H., and Kuriyan, J. (1995) Structural basis for the specific interaction of lysine-containing proline-rich peptides with the N-terminal SH3 domain of c-Crk, *Structure 3*, 215-226.

218. Miyaji, Y., Walter, S., Chen, L., Kurihara, A., Ishizuka, T., Saito, M., Kawai, K., and Okazaki, O. (2011) Distribution of KAI-9803, a novel δ-protein kinase C inhibitor, after intravenous administration to rats, *Drug Metab Dispos 39*, 1946-1953.

219. Andersson, L. O., and Berg, G. (1969) Hydrolysis of disulfide bonds in weakly alkaline media. I. Oxidized glutathione, *Biochim Biophys Acta 192*, 534-536.

220. Du, C., Martin, P. A., and Nickerson, K. W. (1994) Comparison of disulfide contents and solubility at alkaline pH of insecticidal and noninsecticidal Bacillus thuringiensis protein crystals, *Appl Environ Microbiol 60*, 3847-3853.

221. Zhang, Y. H., Evans, B. R., Mielenz, J. R., Hopkins, R. C., and Adams, M. W. (2007) High-yield hydrogen production from starch and water by a synthetic enzymatic pathway, *PLoS ONE 2*, e456.

222. You, C., Myung, S., and Zhang, Y. H. (2012) Facilitated substrate channeling in a self-assembled trifunctional enzyme complex, *Angew Chem 124*, 8917-8920.

223. Southworth, M. W., Amaya, K., Evans, T. C., Xu, M. Q., and Perler, F. B. (1999) Purification of proteins fused to either the amino or carboxy terminus of the Mycobacterium xenopi gyrase A intein, *Biotechniques 27*, 110-114.

224. Evans, T. C., Benner, J., and Xu, M. Q. (1999) The in vitro ligation of bacterially expressed proteins using an intein from methanobacterium thermoautotrophicum, *J Biol Chem 274*, 3923-3926.

225. Li, Y. F. (2011) Self-cleaving fusion tags for recombinant protein production, *Biotechnol Lett 33*, 869-881.

226. Mootz, H. D., Blum, E. S., Tyszkiewicz, A. B., and Muir, T. W. (2003) Conditional protein splicing: a new tool to control protein structure and function in vitro and in vivo, *J Am Chem Soc 125*, 10561-10569.

227. Shi, J., and Muir, T. W. (2005) Development of a tandem protein trans-splicing system based on native and engineered split inteins, *J Am Chem Soc 127*, 6198-6206.

228. Chong, S., Williams, K. S., Wotkowicz, C., and Xu, M. Q. (1998) Modulation of protein splicing of the Saccharomyces cerevisiae vacuolar membrane ATPase intein, *J Biol Chem 273*, 10567-10577.

229. Wood, D. W., Derbyshire, V., Wu, W., Chartrain, M., Belfort, M., and Belfort, G. (2000) Optimized single-step affinity purification with a self-cleaving intein applied to human acidic fibroblast growth factor, *Biotechnol Progr 16*, 1055-1063.

230. Johannes, T. W., Woodyer, R. D., and Zhao, H. (2007) Efficient regeneration of NADPH using an engineered phosphite dehydrogenase, *Biotechnol Bioeng 96*, 18-26.

231. Ming-Qun Xu, S. M. F., Christopher H. Taron, Paul A. Colussi. (2006) Modified chitin binding domain and uses thereof (US Patent 6,984,505), New England Biolabs, Inc., USA.

232. Fong, B. A., Wu, W. Y., and Wood, D. W. (2009) Optimization of ELP-intein mediated protein purification by salt substitution, *Protein Expres Purif 66*, 198-202.

233. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics, *J Mol Graph 14*, 27-38.

234. Mayer, K. M., and Arnold, F. H. (2002) A colorimetric assay to quantify dehydrogenase activity in crude cell lysates, *J Biomol Screen 7*, 135-140.

235. Chen, W., Li, L., Du, Z., Liu, J., Reitter, J. N., Mills, K. V., Linhardt, R. J., and Wang, C. (2012) Intramolecular disulfide bond between catalytic cysteines in an intein precursor, *J Am Chem Soc 134*, 2500-2503.

236. Urry, D. W. (1988) Entropic elastic processes in protein mechanisms. 1. Elastic structure due to an inverse temperature transition and elasticity due to internal chain dynamics, *J Protein Chem 7*, 1-34.

237. Mcpherson, D. T., Morrow, C., Minehan, D. S., Wu, J. G., Hunter, E., and Urry, D. W. (1992) Production and purification of a recombinant elastomeric polypeptide, G-(Vpgvg)19-Vpgv, from Escherichia coli, *Biotechnol Progr 8*, 347-352.

238. McPherson, D. T., Xu, J., and Urry, D. W. (1996) Product purification by reversible phase transition following Escherichia coli expression of genes

encoding up to 251 repeats of the elastomeric pentapeptide GVGVP, *Protein Expres Purif 7*, 51-57.

239. Hashimoto, M., Ikegami, T., Seino, S., Ohuchi, N., Fukada, H., Sugiyama, J., Shirakawa, M., and Watanabe, T. (2000) Expression and characterization of the chitin-binding domain of chitinase A1 from Bacillus circulans WL-12, *J Bacteriol 182*, 3045-3054.

240. Pereira, B., Shemella, P. T., Amitai, G., Belfort, G., Nayak, S. K., and Belfort, M. (2011) Spontaneous proton transfer to a conserved intein residue determines on-pathway protein splicing, *J Mol Biol 406*, 430-442.

241. Frutos, S., Goger, M., Giovani, B., Cowburn, D., and Muir, T. W. (2010) Branched intermediate formation stimulates peptide bond cleavage in protein splicing, *Nat Chem Biol 6*, 527-533.

242. Amitai, G., Callahan, B. P., Stanger, M. J., Belfort, G., and Belfort, M. (2009) Modulation of intein activity by its neighboring extein substrates, *Proc Natl Acad Sci USA 106*, 11005-11010.

APPENDIX A

SPLIT INTEIN MEDIATED ULTRA-RAPID PURIFICATION OF TAGLESS PROTEIN (SIRP)*

Overview

Rapid and efficient tag removal remains a significant problem in recombinant protein purification. Using an engineered DnaE intein from *Nostoc punctiforme*, we developed a split intein mediated ultra-rapid purification (SIRP) method for the purification of tag-less recombinant protein from *E. coli* lysate in less than 1 hour. This system exhibits extraordinarily rapid thio-induced C-terminal cleavage with about 50% completion within 30 seconds at both 22 °C and 6 °C. This is the fastest C-terminal cleavage activity reported to date for inteins. Although the reaction kinetics slow down after the first minute, >90% cleavage completion is achieved within 30 minutes at 22 °C, or within 3 hours at 6 °C. The ultra-rapid cleavage kinetics are made possible by the positioning of the purification tag at the split junction to the C-terminus of the intein N-fragment, thus avoiding potential steric hindrance of the critical interaction between the N- and C-extein. Target proteins are cleaved to >72% completion after 1 hour of intein reaction regardless of the identity of the N-terminal amino acid except in the cases of threonine (50% cleavage) and proline. The C-terminal cleavage reaction can be effectively inhibited by divalent Zn^{2+} under non-reducing conditions. Importantly, the

* Reprinted with permission from "Split intein mediated ultra-rapid purification of tagless protein (SIRP)" by Dongli Guan, Miguel Ramirez, and Zhilei Chen, 2013, Biotechnology and bioengineering, Copyright 2010, Wiley Periodicals, Inc., doi: 10.1002/bit.24913. D. Guan conducted all experimental characterization. M. Ramirez contributed with material preparation and experimental design.

association of the intein N- and C-fragments is reversible, enabling the column-bound intein N-fragment bait protein to be regenerated for multiple usages and further reducing the cost of protein purification. SIRP technology should provide a useful tool for the purification of tag-less proteins and peptides.

Introduction

Affinity tags have greatly facilitated the detection and purification of recombinant proteins (Waugh 2005), but no simple and low-cost method exists to date for affinity tag removal. For many years, serine proteases such as factor Xa, enterokinase, and thrombin have been used in laboratories to remove affinity tags. These low-specificity proteases have mostly been replaced with viral proteases with more stringent sequence specificity such as TEV protease (Sigma) and Rhinovirus 3C protease (PreScission™ protease, GE Healthcare Biosciences). However, these viral proteases exhibit relatively low cleavage efficiency due to their low catalytic rate constants (k_{cat}) (Gasparian et al. 2003; Isetti and Maurer 2007; Kapust et al. 2001). In addition, except for enterokinase, all these proteases also leave certain amino acids at the N-terminus of the target protein after cleavage (Waugh 2011). The recently discovered SUMO-protease exhibits both high specificity and efficiency (>90% completion within 20 minutes at 22 °C) (Malakhov et al. 2004). However, the cost to cleave 1 g of recombinant protein using SUMO-protease (Life Technologies) is $673,000, prohibiting its use in most applications.

To efficiently and cost effectively remove purification tags, many inteins have been engineered to undergo a N- or C-terminal cleavage reaction in acidic (pH 6) or reducing environments (Evans et al. 1999; Lew et al. 1999; Mathys et al. 1999; Southworth et al. 1999). The N- and C-exteins are replaced with the protein of interest (POI) and a purification tag. However, most of these engineered inteins exhibit relatively low catalytic efficiency and require at least overnight incubation to achieve significant cleavage. We recently engineered a naturally split intein DnaE from *Nostoc punctiforme* (*Npu* DnaE) that exhibits rapid thio-induced C-terminal cleavage kinetics with 80% completion within 3 hours at room temperature (Ramirez et al. 2013). The use of split inteins in recombinant protein purification avoids premature cleavage *in vivo* during expression which, in some cases, can significantly impact target protein yields (Cui et al. 2006; Li 2011).

In this work, we redesigned our purification platform and greatly improved the C-terminal cleavage efficiency by strategically relocating the purification tag to within the split intein with a view to alleviating potential steric effects. We demonstrate that this configuration yields a tag removal step that is dramatically faster and at least as effective as the originally configured split intein in which the purification tag was 'traditionally' positioned at the N-terminus of the N intein. This new intein pair exhibited ultra-rapid thio-activated reaction kinetics with ~50% cleavage completion within 30 seconds at 22 °C or 6 °C, and >90% C-terminal cleavage within 30 minutes at 22 °C and >90% within 3 hours at 6 °C. The cleavage reaction can be effectively inhibited by 0.5 mM $ZnCl_2$ under non-reducing conditions, thus providing control over the on-sent of the cleavage

reaction. In addition, the C-terminal cleavage efficiency does not appear to be highly dependent on the identity of the first residue of C-extein, making it possible to completely remove the purification tag.

Using this reconfigured intein pair, we developed a rapid affinity chromatography-based purification method, and purified 3 model proteins, GFP, phosphite dehydrogenase (PTDH) (Johannes et al. 2005) and β-galactosidase (β-Gal)(Ramirez et al. 2013), to 80-90% purity in a single step. After 30 minutes of thio-induced cleavage at 22 °C, ~10 mg of purified protein was obtained per mL of chitin resin. Chromatography resin-bound intein complex can be regenerated for multiple uses, further reducing the cost for large-scale protein purification. This newly designed split intein-mediated protein purification system combines affinity purification and tag removal and enables ultra-rapid purification of tag-less target protein at low cost.

Material and methods

Chemicals and bacterial strains

All chemicals were purchased from either VWR International (Radnor, PA) or Thermo Fisher Scientific (Waltham, MA), unless otherwise stated. *E. coli* strains DH5α and BL21 (DE3) (Novagen, Madison, WI) were used for plasmid cloning and protein expression, respectively.

Plasmid construction

A schematic depiction of the amino acid sequences of all protein constructs and their numberings are shown in Table A.1. The construction of C*-β-Gal, C*-PTDH and C*-GFP were described previously (Ramirez et al. 2013). To generate N-CBD, NpuN and the chitin-binding domain (CBD) were amplified from plasmid KanR-IntRBS-NpuNC (Lockless and Muir 2009) (a generous gift from Prof. Tom Muir, Rockefeller University) and pTWIN1 (New England Biolabs) using primers NdeI-N-F/L-HindIII-N-R and L-NheI-CBD-F/XhoI-CBD-R, respectively, joined via overlap extension PCR, and inserted into the pET-26b(+) (Novagen) between *Nde*I and *Xho*I sites. N_{C1A}-CBD was generated by site-directed mutagenesis using primers NdeI-N C1A-F and L-HindIII-N-R. To generate the seventeen C*-X-GFP fusions, C*-X and GFP genes were amplified individually from the plasmid C*-GFP with the primers NpuC-F-NdeI/NpuC XFN-GFP-R and NpuC XFN-GFP-F/XhoI-GFP-R, respectively, joined by overlap extension PCR and cloned into pET-26b(+) between the *Nde*I and *Xho*I sites. To generate the last two C*-X-GFP (22 and 23), the GFP gene was amplified from plasmid C*-GFP with the primers NheI-GFP-F/XhoI-GFP-R, and cloned into C*-PTDH between the *Nhe*I and *Xho*I sites; then C*-X was introduced by site-directed mutagenesis using primers NpuC-F-NdeI and NheI-XFN-NpuC-R.

Table A.1 Protein constructs used in this study

Construct	Short Name	Protein sequence	Molecular Weight (kDa)
0	N-CBD	NpuN-2x(GGGGS)-CBD-6xH	20.6
1	N_{C1A}-CBD	$NpuN_{C1A}$-2x(GGGGS)-CBD-6xH	20.6
2	C*-β-Gal	$NpuC_{D118G}$-CFNAS-β-Gal-6xH	122.1
3	C*-PTDH	$NpuC_{D118G}$-CFNAS-PTDH-6xH	42.2
4	C*-GFP	$NpuC_{D118G}$-CFN-GFP-6xH	32.0
5	C*-A-GFP	$NpuC_{D118G}$-AFN-GFP-6xH	32.0
6	C*-D-GFP	$NpuC_{D118G}$-DFN-GFP-6xH	32.0
7	C*-L-GFP	$NpuC_{D118G}$-LFN-GFP-6xH	32.0
8	C*-P-GFP	$NpuC_{D118G}$-PFN-GFP-6xH	32.0
9	C*-R-GFP	$NpuC_{D118G}$-RFN-GFP-6xH	32.0
10	C*-E-GFP	$NpuC_{D118G}$-EFN-GFP-6xH	32.0
11	C*-H-GFP	$NpuC_{D118G}$-HFN-GFP-6xH	32.0
12	C*-I-GFP	$NpuC_{D118G}$-IFN-GFP-6xH	32.0
13	C*-K-GFP	$NpuC_{D118G}$-KFN-GFP-6xH	32.0
14	C*-M-GFP	$NpuC_{D118G}$-MFN-GFP-6xH	32.0
15	C*-N-GFP	$NpuC_{D118G}$-NFN-GFP-6xH	32.0
16	C*-Q-GFP	$NpuC_{D118G}$-QFN-GFP-6xH	32.0
17	C*-S-GFP	$NpuC_{D118G}$-SFN-GFP-6xH	32.0
18	C*-T-GFP	$NpuC_{D118G}$-TFN-GFP-6xH	32.0
19	C*-V-GFP	$NpuC_{D118G}$-VFN-GFP-6xH	32.0
20	C*-W-GFP	$NpuC_{D118G}$-WFN-GFP-6xH	32.0
21	C*-Y-GFP	$NpuC_{D118G}$-YFN-GFP-6xH	32.0
22	C*-F-GFP	$NpuC_{D118G}$-FFNAS-GFP-6xH	32.0
23	C*-G-GFP	$NpuC_{D118G}$-GFNAS-GFP-6xH	32.0

Sample purification via SIRP

A disposable column containing 150 μL chitin resin was loaded with soluble lysate containing **N**$_{C1A}$-CBD in Buffer A (0.5 M NaCl, 10 mM Tris-HCl, pH 8.0) and washed 4 times with 10 column volumes (CV) of Buffer B (0.5 M NaCl, 50 mM NaPOi, pH 6.0). All loading and washing steps were performed in batch phase. The last wash contains 0.5 mM $ZnCl_2$. The same concentration of $ZnCl_2$ was added to the soluble lysate of C*-PTDH/GFP/β-Gal in Buffer B immediately before the lysate was loaded

onto the same chitin resin. The column was subsequently washed with 10 CV of Buffer B with $ZnCl_2$ (0.5 mM) and finally incubated in 4 CV of Buffer a containing 50 mM DTT at room temperature for 30 minutes or at 6 °C for 3 hours. Purified PTDH, GFP and β-Gal were collected in the flow-through. Trace amount N_{C1A}-CBD in the flow-through can be removed by passing through a fresh chitin column.

The concentrations of purified POIs were determined by absorbance at 280 nm using a NanoDrop 1000 (Thermo Fisher Scientific) for yield calculation. The purity was measured by analyzing SDS-PAGE gels using the Trace Quantity module in Quantity One software (BioRad, Hercules, CA).

To determine the ability of high pH buffer to reverse intein association, chitin column containing N_{C1A}-CBD and C*-β-Gal/PTDH/GFP was incubated in 4 CV of Buffer C (1.5 M NaCl, 50 mM Na_2HPO_4/NaOH, pH 11.4, 0.5 mM $ZnCl_2$) for 10 minutes at room temperature. The flow-through containing dissociated C*-β-Gal/PTDH/GFP was used directly for SDS-PAGE analysis.

For regeneration of the N_{C1A}-CBD-chitin affinity matrix, the used resin was thoroughly washed with Buffer C. The regenerated column can be stored in storage buffer (0.5 M NaCl, 10 mM Tris-HCl, 1 mM EDTA, 0.15% NaN_3, pH 8.0) at 4 °C for about a week without significant loss of activity.

A.3.4 Ni-affinity-mediated purification of protein fragments for intein reaction kinetics studies

E. coli BL21 (DE3) was transformed with the appropriate expression plasmid and plated on a Luria-Bertani (LB) agarose plate containing 50 μg/ml kanamycin. The next day, a single colony was picked and grown in 4 ml of LB broth containing the same antibiotic to OD_{600} ~0.6. The culture was transferred to 1 L LB broth and grown at 37 °C until OD_{600} ~0.6. Protein expression was induced at 18 °C for 14-15 hours by the addition of IPTG (1 mM). After expression, cells were harvested by centrifugation at 8000 x *g* and 4 °C for 15 minutes and stored at -80 °C until use.

For purification of construct containing N, cell pellets were resuspended in Buffer A at 10 ml per gram of wet pellet, and disrupted by sonication (QSonica Misonix 200, Amp 20, with 1 second pulse 4 seconds pause for 2 minutes total pulse). Soluble lysates were collected after centrifugation at 16,000 x *g* for 20 minutes at 4 °C, passed through a 0.5 mL Ni-NTA column by gravity (Qiagen, Valencia, CA), washed extensively with Buffer A containing 20 mM imidazole and finally eluted in Buffer A containing 150 mM imidazole. The same procedure was used to purify fusion proteins containing C except that Buffer A was replaced with buffer B. This low-pH buffer is intended to reduce the proteolytic degradation by cellular proteases.

Purified proteins were concentrated via 10 kDa ultra-filtration spin columns (Amicon Ultra, Millipore, Billerica, MA) and concentrations were determined by absorbance at 280 nm using NanoDrop 1000 (Thermo Fisher Scientific).

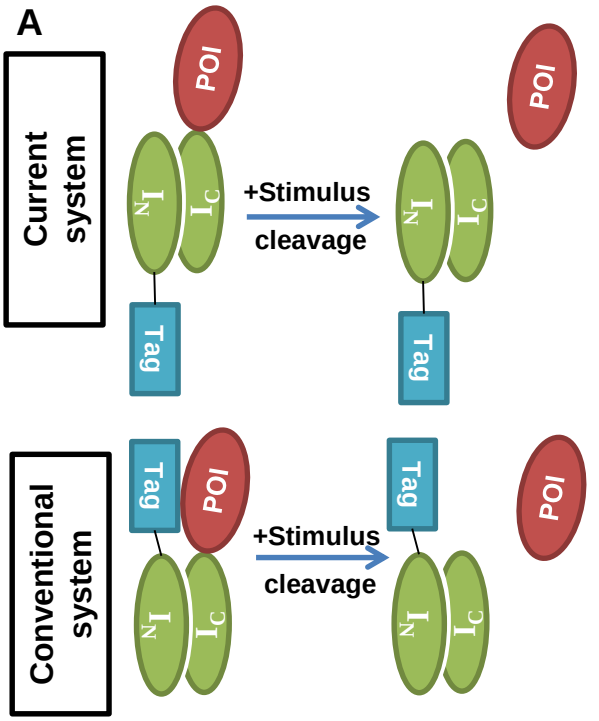
Intein reaction kinetics characterization

All characterization experiments for intein kinetics were carried out using Ni-NTA-purified proteins under specified conditions. All reactions contained 15 μM of each intein fragment. Samples were taken at different time points after the initiation of the reaction, immediately frozen in liquid N_2, boiled with 2X SDS sample buffer at 95 °C for 10 minutes and analyzed on 12% SDS-PAGE gels, unless otherwise specified. The gels were stained with Coomassie brilliant blue R250. Band intensities corresponding to reactants and products were quantified using the Trace Quantity module in Quantity One software (BioRad, Hercules, CA).

Results

Design new split-intein cleavage system

The natural function of inteins is to mediate a trans-splicing reaction that joins the associated N- and C-exteins via a peptide bond. Successful trans-splicing activity requires that the N- and C-exteins are in close proximity prior to trans-splicing. In conventional intein-based protein purification systems, the N- and C-exteins are replaced with a protein purification tag and the protein of interest (POI) (Fig. A.1) (Chong et al. 1997; Chong et al. 1998a; Hong et al. 2008; Lu et al. 2011; Ramirez et al. 2013; Wu et al. 2006). Depending on the POI and the choice of purification tag/protein, steric hindrance between the N- and C-exteins can significantly reduce the cleavage efficiency and, in the case of split inteins, the association of the N- and C-intein fragments as well (Ramirez et al. 2013).



I_N/I_C: intein N-/C-fragment
CBD: Chitin binding domain
POI: Protein of Interest

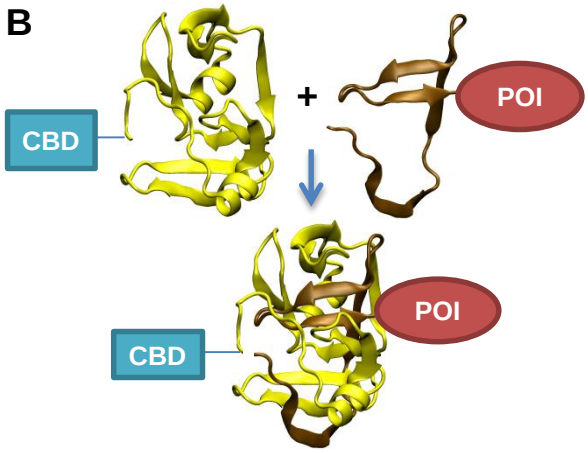


Figure A.1 Schematic of the engineered intein pair. (A) Comparison of the current intein design for tag removal vs. the conventional design. (B) Cartoon representation of fusion proteins before and after intein association. The intein N-fragment (yellow) and C-fragment (brown) are adapted from the NMR structure of *Npu*DnaE (PDB code: 2keq) (Oeemig et al. 2009).

Since some split inteins can tolerate insertions at the split junction (Busche et al. 2009), we hypothesized that moving the purification tag from the N-extein to the split intein junction may bypass any steric hindrance of critical interactions required for the C-terminal cleavage reaction without affecting the cleavage efficiency. Using our previously engineered thio-induced C-terminal cleaving NpuDnaE, we inserted a purification tag, a chitin binding domain (CBD) (Hashimoto et al. 2000) to the C-terminus of the intein N-fragment (N) (Fig. A.1), leaving a single Met as the N-extein. The first residue of N was further mutated to Ala (NC1A) to completely abolish any N-terminal cleavage activity (Ramirez et al. 2013). The resulting construct is named NC1A-CBD. The POI was attached to the C-terminus of the intein C-fragment (C*) to form C*-POI. C* contains a single mutation, D118G, which facilitates rapid thio-induced C-terminal cleavage in the absence of the N-terminal cleavage (Ramirez et al. 2013; Wood et al. 1999).

N_{C1A}-CBD enables controllable ultra-rapid C-terminal cleavage

We generated various fusion proteins containing the engineered intein pairs as listed in Table A.1. To determine the C-terminal cleavage kinetics, C*-PTDH, containing the C-fragment of intein C* fused to the globular protein phosphite dehydrogenase (PTDH), was mixed with N_{C1A}-CBD at 1:1 molar ratio in the presence of 50 mM DTT. ~50% of PTDH was cleaved from C*-PTDH within 30 seconds at both 22 °C and 6 °C (Fig. A.2). Since 30 seconds is the earliest time point we could measure accurately, it is possible that even shorter time is needed to achieve this much cleavage.

Although the reaction kinetics slowed down after the first minute, over 90% C-terminal cleavage is achieved within 30 minutes at 22 °C and within 3 hours at 6 °C (Fig. A.2).

Unfortunately, N_{C1A}-CBD also induced C-terminal cleavage of C*-PTDH under non-reducing conditions, albeit at a much slower rate than under reducing conditions (Fig. A3, 0 mM Zn^{2+}). Over 50% C*-PTDH is cleaved within 30 minutes at both neutral and acidic pH even in the absence of DTT. This is likely due to the absence of Cys1 which could potentially form a disulfide bond with Cys+1 and inhibit the intein reaction (Chen et al. 2012). Since many enzymatic reactions can be inhibited by divalent ions such as Zn^{2+} (Costello et al. 1997; Larsen and Auld 1989; Perry et al. 1997), and Zn^{2+} was found to inhibit the first N-S acyl shift reaction of *Ssp* DnaE, a close homolog of *Npu* DnaE(Nichols et al. 2003), we tested the ability of Zn^{2+} to inhibit the C-terminal cleavage reaction of our engineered *Npu* DnaE construct. $ZnCl_2$ (0.5 mM) can effectively inhibit the C-terminal cleavage reaction under non-reducing conditions but has little inhibitory effect in the presence of DTT (Fig. A.3). After 3 hours incubation at 22 °C, there was only ~14% intein cleavage in both pH 6 and pH 8 buffer. These results indicate that Zn^{2+} and DTT can be used as effective switches to turn off and on, respectively, the C-terminal cleavage.

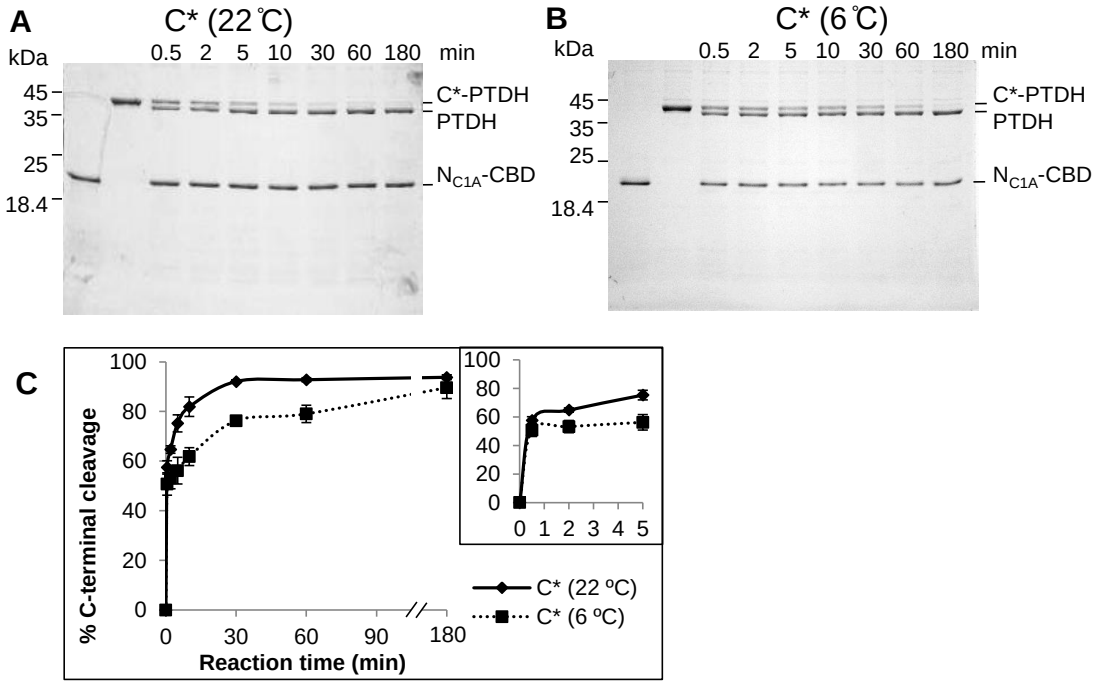


Figure A.2 Intein C-terminal cleavage kinetics characterization. (A and B) SDS-gels (12% acrylamide) of the reaction between N_{C1A}-CBD (construct 1) and C*-PTDH (construct 3) performed at 22 or 6 °C in pH 8 buffer containing 50 mM DTT. The cleaved PTDH is 38.2 kDa. The cleaved C* is not visible from the gel (4.1 kDa). (C) Time course of the disappearance of C*-PTDH due to C-terminal cleavage at different temperatures. Time course for the first 5 minutes is shown in the inset. The values error bars represent the mean and standard deviation, respectively, from 3 independent experiments.

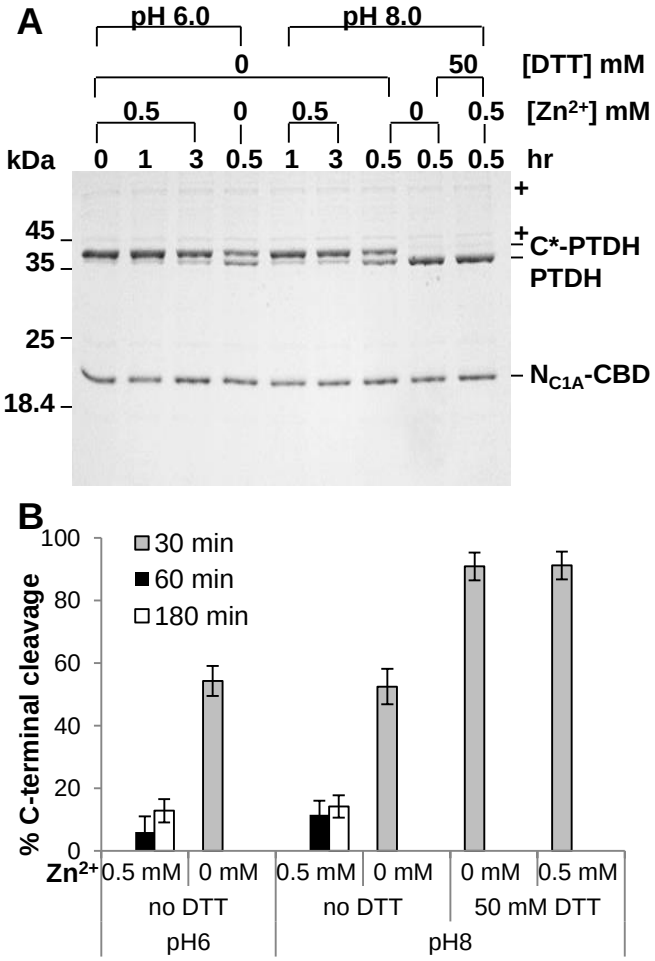


Figure A.3 C-terminal cleavage kinetics under different conditions. (A) An SDS-gel of the reaction between N_{C1A}-CBD (construct 1) and C*-PTDH (construct 3) performed under different buffer conditions at 22 °C. '+' denotes impurities. (B) Calculated percentage of C-terminal cleavage under different conditions. The values and error bars represent the mean and standard deviation, respectively, from 3 independent experiments.

Intein cleavage is mostly unaffected by the amino termini of the target protein

In many applications, it is desirable to completely remove any non-native amino acids from the target protein. For intein trans-splicing reactions, a cysteine is required at the +1 position to complete the transesterification and S/O-N acyl shift reactions (Nichols and Evans 2004; Xu and Perler 1996; Zettler et al. 2009). However, Cys+1 is not needed for the asparagine cyclization reaction responsible for C-terminal cleavage. We constructed various C*-X-GFP fusion proteins (Table A.1, construct 5-23) in which the first residue (X) of the C-extein was replaced with the other 19 amino acids. Among all 20 amino acid substitutions, 18 of them (all except for Pro and Thr) yielded mostly complete C-terminal cleavage (>72% completion) after 1 hour at room temperature. Fifteen amino acids were able to achieve >80% cleavage completion within 30 min at room temperature (except for Lys, Ser, Gly, Thr, Pro) (Fig. A.4). Thus, most N-terminal amino acids exhibited a cleavage profile comparable to that observed with the original C* construct which contains a Cys at the +1 position (Fig. A.2). The cleavage of proteins with N-terminal Thr and Pro are 50 % and 7 %, respectively, after one hour at room temperature.

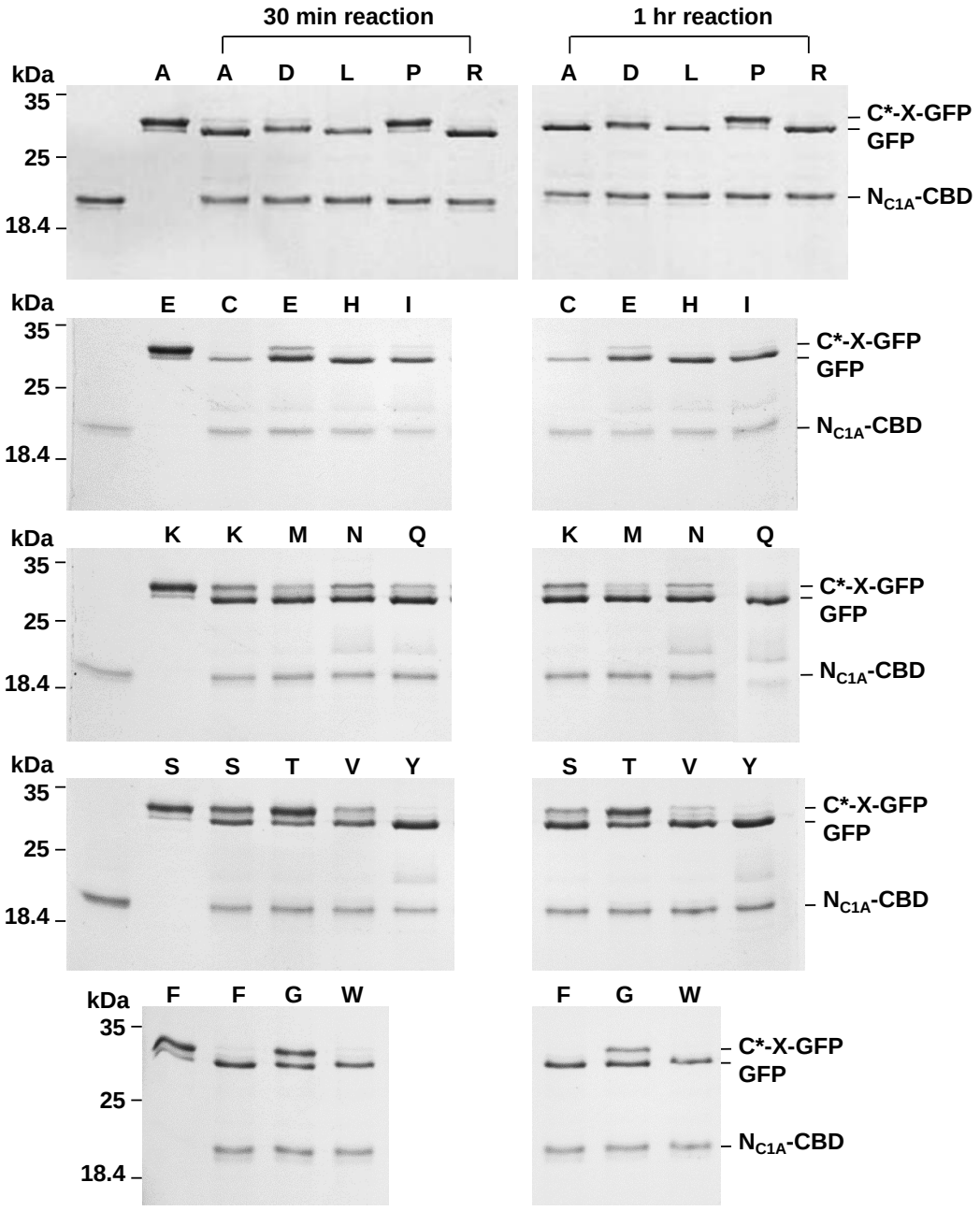


Figure A.4 Effects of +1 residue on C-terminal cleavage. SDS-gels of reactions of N_{C1A}-CBD and C*-X-GFP at pH8 and 22 °C with 50 mM DTT. The cleaved GFP is 28.0 kDa.

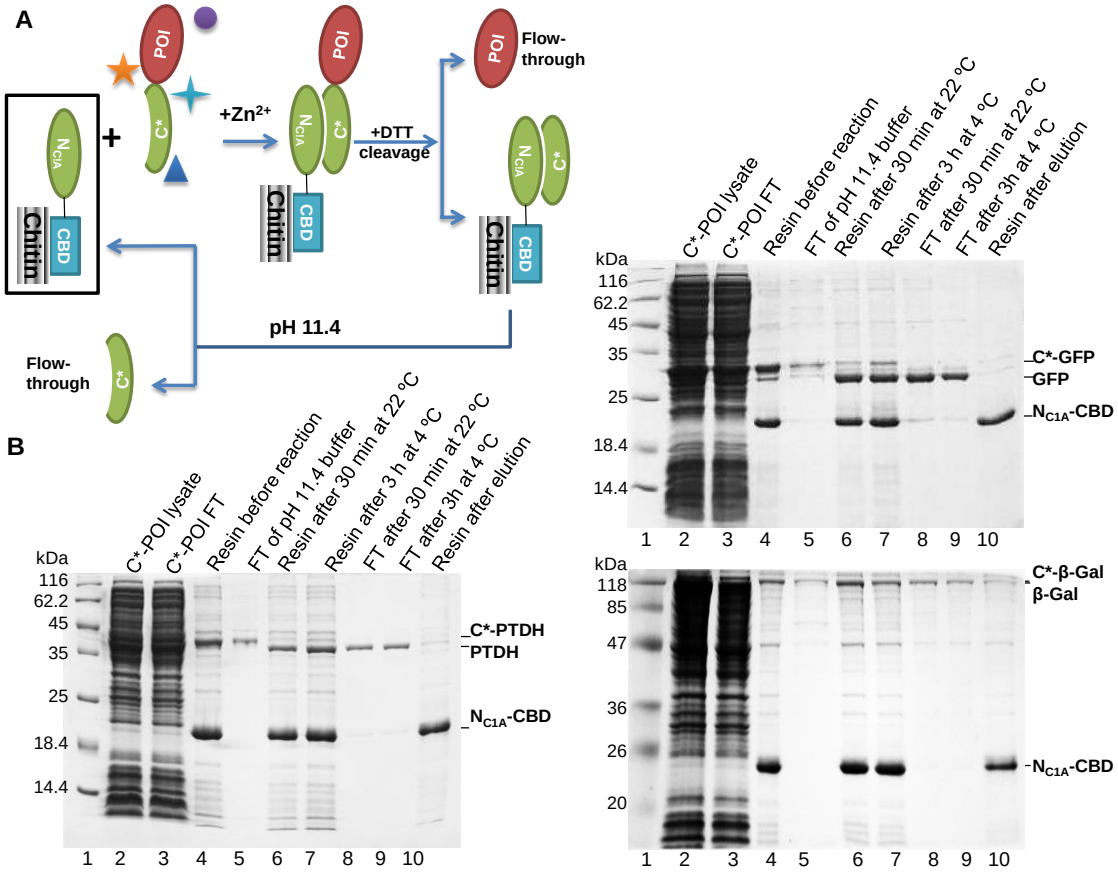


Figure A.5 Protein purification using SRIP (A) Schematic of chitin-mediated chromatography purification method. Lysate of C*-POI is passed through column prebound with N_{C1A}-CBD in the presence of 0.5 mM $ZnCl_2$. After washing, intein C-terminal cleavage reaction is induced by the addition of DTT and purified POI can be collected in the flow-through. The column is then regenerated by washing in pH 11.4 buffer to dissociate the intein complex. (B) Purification of PTDH, GFP and β-Gal using N_{C1A}-CBD-chitin resin. SDS-PAGE analysis of the purification of PTDH, GFP and β-Gal. Lane 1, EZ-Run Protein Ladder; lane 2, soluble fractions of lysate containing C*-POI; lane 3, flow through of soluble lysates; lane 4, chitin resin after loading lysate and washing with buffer containing 0.5 mM $ZnCl_2$; lane 5, elution of C*-POI in pH 11.4 buffer; lane 6, chitin resin in cleavage buffer incubated at 22 °C for 30 min; lane 7, chitin resin in cleavage buffer incubated at 6 °C for 3 h; lane 8, flow-through after incubation in cleavage buffer at 22 °C for 30 min; lane 9, flow-through after incubation at 6 °C for 3 h; lane 10, chitin resin after elution of target protein.

Protein purification via chitin resin

To demonstrate the utility of our engineered intein pairs, we designed a protein purification approach based on the chitin binding domain (CBD) (Fig. A.5A), and purified three proteins via chitin affinity chromatography (Fig. A.5B). In most cases, the yield of purified protein was ~10 mg per mL of chitin resin (Table A.2). The molar ratios of bound C*-GFP, C*-PTDH and C*-β-Gal to N_{C1A}-CBD, as determined from the SDS-PAGE analysis, were 0.57, 0.28 and 0.12, respectively. The column binding capacity is inversely proportional to the size of the POI. The binding capacity of tetrameric proteins (β-Gal) is significantly smaller than that of dimeric proteins (PTDH), which is smaller than monomeric proteins (GFP). The intein C-terminal cleavage efficiencies for all proteins when immobilized on the affinity resin are comparable to that observed in solution with >80% cleavage in 30 minutes at 22 °C and in 3 hours at 6 °C (Fig. A.5B, lanes 6, 7). Under the conditions tested, the final purified protein contains 80-90% tag-less POI and 10-15% N_{C1A}-CBD that detach from the chitin beads (Table A.2). In the case of GFP, we were unable to detect any un-cleaved C*-GFP via densitometric analysis of SDS-PAGE. The un-cleaved C*-PTDH comprised ~3% of the final protein. Due to the very large size of β-Gal, the un-cleaved and cleaved proteins were indistinguishable using SDS-PAGE analysis and thus we were unable to determine the amount of un-cleaved β-Gal. Since N_{C1A}-CBD is the major impurity in the purified protein, the use of a higher affinity tag, such as 6xHistidine, may further increase the sample purity and reduce the amount of un-cleaved protein in the sample. The

association of C* and N_{C1A} is at least partially reversible as buffer at pH 11.4 can selectively elute un-cleaved C*-POI from chitin-bound N_{C1A}-CBD (Fig. A.5B, lane 5).

Table A.2 Protein purification and quantification via chitin resin (SIRP)[a]

Protein	Quaternary structure	Size kDa	Yield (mg/ml chitin)[b]	Purity (mass %)[c]	un-cleaved C*-POI (mass %)	Activity [f]
GFP	Monomer	28.0	12.9 ± 4.1	81.0 ± 3.0	ND[d]	Fluorescence at 538nm
PTDH	Dimer	38.2	10.4 ± 3.4	88.3 ± 1.3	3.4 ± 5.9	NBT-methosulfate assay
β-Gal	Tetramer	118.1	3.1 ± 0.5	87.8 ±3.2	UD[e]	573.3 ± 70.4Units/mg

[a]: All values and errors represent the mean and standard deviations, respectively, from 3 independent purification experiments.
[b]: Purification yields were determined from 150 µl chitin resin
[c]: Mass % were estimated using densitometry analysis of SDS-PAGE.
[d]: ND, undetected. C*-GFP was not detected.
[e]: UD, undistinguishable. C*-β-Gal was not distinguishable from cleaved β-Gal.
[f]: Details of each activity assay were described in (Ramirez et al. 2013). One β-Galactosidase unit is defined as the amount of protein needed to hydrolyze 1.0 µmole of ONPG per minute at 22 °C. DTT was removed from purified PTDH by dialysis for accurate activity measurement.

Regeneration of N_{C1A}-CBD-chitin affinity matrix

Although an intein is an enzyme, it can only catalyze a single round of reaction. For large-scale protein purification, it would be desirable to reuse chitin-bound N_{C1A}-CBD for multiple rounds of purification to reduce costs. We showed above that bound N_{C1A} and C* complex can be partially dissociated with pH 11.4 buffer. To characterize the recyclability of chitin-bound N_{C1A}-CBD, we repeated the purification of PTDH using the same chitin column for 4 times (Fig. A.6). After elution of cleaved PTDH and before

the addition of fresh lysate containing C*-PTDH, the chitin resin was thoroughly washed with pH 11.4 buffer to remove cleaved C* from N_{C1A}-CBD on the column. The yields of purified PTDH are comparable for all 4 cycles, confirming the ability of our N_{C1A}-CBD-chitin affinity matrix to be regenerated for multiple cycles. It appears that cleaved C* may dissociate more readily from N_{C1A}-CBD in pH 11.4 buffer than the full-length C*-PTDH.

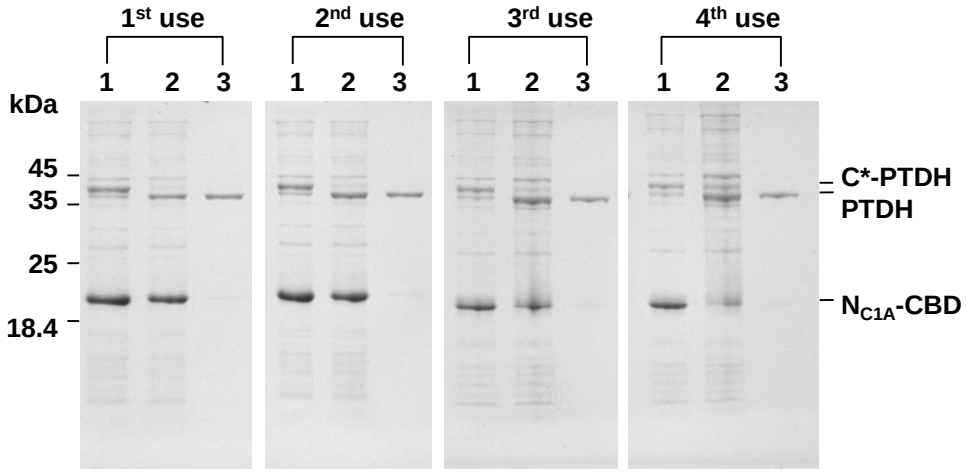


Figure A.6 Chitin-bound NC1A-CBD can be regenerated after purification. SDS-gels of samples collected during the purification of PTDH using regenerated column, Lane 1, chitin resin before cleavage; lane 2, chitin resin in cleavage buffer incubated at 22 °C for 30 min; lane 3, flow through containing purified PTDH. The cleaved C* is 4 kDa and is not visible from the gel.

Discussion

In this study, we developed a new protein purification method – split intein mediated ultra-rapid purification of tag-less protein (SIRP) – using an engineered C-terminal-cleaving naturally split *Npu* DnaE intein. Unlike most current intein-mediated

protein purification methods that have N- and C-exteins replaced with a protein purification tag and a protein of interest (POI), we inserted the purification tag in the split intein junction (Fig. A.1). Many large inteins are bipartite elements consisting of a protein splicing domain interrupted by an endonuclease domain (Derbyshire et al. 1997). Since the endonuclease activity is not required for protein splicing, mini-inteins lacking the endonuclease domain have been generated. Previously, the endonuclease domain of *Rec*A intein was replaced with a CBD, and the resulting chimeric protein was inserted into a target protein upstream of a cysteine so that the target protein can be purified via chitin resin followed by an intein-mediated *trans*-splicing reaction (Wu et al. 2002). However, due to the uncontrollable intein side reaction, an additional size exclusion chromatography step was needed to separate the cleave product from the *trans*-spliced product. In this work, we demonstrate that our strategic positioning of the purification tag in the intein split junction likely minimizes steric hindrance between the purification tag and POI and greatly improves the intein reaction kinetics. Remarkably, C-terminal cleavage is approximately 50% complete within 30 seconds at both 22 °C and 6 °C (Fig. A.2). In comparison, the fastest reported C-terminal-cleaving intein, gp41-1_{C1A}, has a $t_{1/2}$ of 5 minutes at 37 °C (Carvajal-Vallejos et al. 2012). Much slower cleavage kinetics were observed when the N-extein of the same intein was replaced with a purification tag (Ramirez et al. 2013). It is presently unclear whether the two-pronged strategy of introducing a mutation into the C-intein equivalent to the D118G in C* and moving the N-extein from the N-terminus to the C-terminus of the N-intein can be used to engineer a faster C-terminal cleaving gp41-1.

Although the initial ultra-rapid C-terminal cleavage kinetics of our engineered intein slowed down after the first minute, over 90% C-terminal cleavage was achieved within 30 minutes at 22 °C and within 3 h at 6 °C (Fig. A.2C). To achieve a similar extent of C-terminal cleavage, it takes about 16 h at 23 °C and 16-40 h at 4 °C for the IMPACT system (New England Biolabs) employing the *Sce*VMA1 or *Ssp*DnaB intein (Table A.1) (Chong et al. 1998a; Chong et al. 1998b).

For many applications it is desirable to completely remove the purification tag from the target protein after purification. Our engineered intein exhibits fast C-terminal cleavage to >72% completion after 1 hour at 22 °C when the first residue of the target protein is any of the natural amino acids except Pro and Thr. Among these, POI with 15 different N-terminal amino acids (all except for Lys, Ser, Gly, Thr and Pro) were cleaved to >80% completion after 30 minutes at 22 °C (Fig. A.4). The cleavage efficiency of protein with N-terminal Thr is ~50% completion after 1 hour at 22 °C. This result indicates that the cleavage reaction is tolerant to the presence of most amino acids at the +1 position. In comparison, the commercially available IMPACT system employing *Sce*VMA1 intein requires 16 hour at 23 °C to achieve a similar level of C-terminal cleavage (IMPACT-CN system manual; (Xu et al. 2000)). Little to no cleavage was observed when the first amino acid is Pro. This result is likely due to proline-induced structural alteration of the protein backbone.

In some cases the intein cleavage rate is affected not only by the first residue, but the entire C-extein structure, likely due to steric hindrance between the C- and N-exteins (Ramirez et al. 2013). We did not observe any noticeable difference in C-terminal

cleavage between the two target proteins in this study, GFP and PTDH, representing monomeric and dimeric proteins, respectively. More target proteins with diverse structures (i.e. globular vs. extended, dimer, tetramer, etc.) need to be investigated to determine the effect of protein structure on C-terminal cleavage efficiency.

Although all target proteins exhibit similar C-terminal cleavage efficiency, their ability to bind to N_{C1A}-CBD-functionalized chitin resin decreases with the increase of protein size. Interestingly, the binding capacity of monomeric C*-GFP is double that of dimeric C*-PTDH, which in turn is double that of tetrameric C*-β-Gal. The difference in binding capacity is likely due to steric hindrance as larger molecules occupy more surface area on the chitin resin. This steric effect may account for the reduced yield of purified β-Gal (3.1 mg/ml resin) compared to PTDH (10.4 mg/mL resin) and GFP (12.9 mg/mL resin).

Reaction between N_{C1A}-CBD and C*-POI is thio-activated but not thio-dependent. Significant C-terminal cleavage (>50% after 30 minutes at 22℃) is observed under non-reducing conditions (Fig. A.3, 0 mM $ZnCl_2$). However, this reaction can be effectively inhibited by $ZnCl_2$ (0.5 mM) with only ~14% C-terminal cleavage after 3 hours at 22 °C. The extent of inhibition varies slightly depending on the first residue of the C-extein but in all cases less than 20% and 25% C-terminal cleavage was observed after 1 and 3 hours, respectively, at 22 °C. Higher Zn^{2+} concentration can more efficiently inhibit the C-terminal cleavage reaction and may help with preventing loss of product during extended washing steps. However, Zn^{2+} ions at ≥ 1 mM can cause precipitation of some cellular proteins and thus should not be used directly in the cell

lysate. According to the crystal structure of *Ssp*DnaE, Zn^{2+} is coordinated by Asp140, His48 (equivalent to Asp118, His48 in *Npu*DnaE) and Cys+1(Sun et al. 2005). However, Asp118 is mutated to Gly in C* to confer C-terminal cleavage activity in the absence of N-terminal cleavage (Ramirez et al. 2013), and this protein as well as proteins lacking Cys+1 (Construct 5-23) also potently inhibited by Zn^{2+}. In addition, we tested the ability of cisplatin, an inhibitor of the *Rec*A intein presumably through targeting the Cys1 (Zhang et al. 2011), can also inhibit C-terminal of our intein at low micromolar concentration. It is worth noting that our engineered intein has Cys1 mutated to Ala to abolish the first thio-ester transfer step. Thus, we believe that there may be another site(s) on *Npu*DnaE for Zn^{2+} binding. Other divalent metals such as Cu^{2+}, Ca^{2+}, Mg^{2+}, Mn^{2+} and Fe^{2+} also inhibit the intein C-terminal cleavage under non-reducing conditions with varying efficiencies.

The use of DTT as a cleavage inducer for protein elution may not be desirable for proteins that rely on surface-exposed disulfide bonds for their tertiary and quaternary structure. In these cases, EDTA can be used as an inducer of C-terminal cleavage since it strongly chelates the Zn^{2+} ions that suppress basal cleavage, and release of the POI in SIRP (Fig. A.3). The cleavage kinetics induced by EDTA are slightly slower than that induced by DTT, but in most cases 5 mM EDTA induces nearly 100% cleavage after 1 h incubation at room temperature.

The high affinity between C* and N_{C1A} appears to be largely dictated by electrostatic interactions and can be significantly reduced by exposure to pH 11.4 buffer. This property enables regeneration of chitin-bound N_{C1A}-CBD for multiple rounds of

target protein purification. However, due to the relatively weak binding affinity of CBD, some N_{C1A}-CBD dissociates from the column during washing. In this study, we used CBD as the purification tag due to the low cost of chitin resin and CBD-binding amorphous cellulosic matrices (Hong et al. 2008). Higher affinity tags such as a poly histidine tag can potentially be used in place of a CBD to achieve more specific and higher affinity binding between the intein and the resin. For example, the interaction between the His-tag and nickel resin may yield a higher purification yield and regeneration efficiency.

In conclusion, we developed an ultra-rapid method for the purification of tag-less proteins – SIRP – that allows the use of an engineered split intein to purify tag-less proteins in less than 1 hour. This technology provides a powerful new tool for the purification of tag-less proteins in bench-scale applications. All proteins used for human therapy need to be rendered tag-less due to the potential for immunogenic responses. Thus currently most therapeutic proteins are purified via multiple time-consuming and expensive chromatography steps. SIRP provides a rapid, convenient and low-cost affinity-based method for the purification of tag-less proteins. Thus, application of SIRP can potentially significantly lower the purification cost for therapeutic proteins and other tag-less proteins. Furthermore, the low cost of chitin beads and the possibility of using even lower cost CBD-binding amorphous cellulosic matrices (Hong et al. 2008) as affinity supports for CBD makes the application of this intein-mediated approach as an affinity-based step for large-scale packed bed continuous protein purification an attractive prospect.

APPENDIX B

ADDITIONAL RESOURCES

Busche AE, Aranko AS, Talebzadeh-Farooji M, Bernhard F, Dotsch V, Iwai H. 2009. Segmental Isotopic Labeling of a Central Domain in a Multidomain Protein by Protein Trans-Splicing Using Only One Robust DnaE Intein. Angewandte Chemie-International Edition 48(33):6128-6131.

Carvajal-Vallejos P, Pallisse R, Mootz HD, Schmidt SR. 2012. Unprecedented rates and efficiencies revealed for new natural split inteins from metagenomic sources. J Biol Chem 287(34):28686-96.

Chen W, Li L, Du Z, Liu J, Reitter JN, Mills KV, Linhardt RJ, Wang C. 2012. Intramolecular disulfide bond between catalytic cysteines in an intein precursor. Journal of the American Chemical Society 134(5):2500-3.

Chong S, Mersha F, Comb D, Scott M, Landry D, Vence L, Perler F, Benner J, Kucera R, Hirvonen C and others. 1997. Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. Gene 192(2):271-81.

Chong S, Montello G, Zhang A, Cantor E, Liao W, Xu M, Benner J. 1998a. Utilizing the C-terminal cleavage activity of a protein splicing element to purify recombinant proteins in a single chromatographic step. Nucleic Acids Research 26(22):5109-5115.

Chong S, Williams KS, Wotkowicz C, Xu MQ. 1998b. Modulation of protein splicing of the Saccharomyces cerevisiae vacuolar membrane ATPase intein. Journal of Biological Chemistry 273(17):10567-10577.

Costello LC, Liu Y, Franklin RB, Kennedy MC. 1997. Zinc inhibition of mitochondrial aconitase and its importance in citrate metabolism of prostate epithelial cells. J Biol Chem 272(46):28875-81.

Cui C, Zhao W, Chen J, Wang J, Li Q. 2006. Elimination of in vivo cleavage between target protein and intein in the intein-mediated protein purification systems. Protein Expr Purif 50(1):74-81.

Derbyshire V, Wood DW, Wu W, Dansereau JT, Dalgaard JZ, Belfort M. 1997. Genetic definition of a protein-splicing domain: Functional mini-inteins support structure predictions and a model for intein evolution. Proceedings of the National Academy of Sciences of the United States of America 94(21):11466-11471.

Evans TC, Benner J, Xu MQ. 1999. The in vitro ligation of bacterially expressed proteins using an intein from Methanobacterium thermoautotrophicum. Journal of Biological Chemistry 274(7):3923-3926.

Gasparian ME, Ostapchenko VG, Schulga AA, Dolgikh DA, Kirpichnikov MP. 2003. Expression, purification, and characterization of human enteropeptidase catalytic subunit in Escherichia coli. Protein Expr. Purif. 31(1):133-9.

Hashimoto M, Ikegami T, Seino S, Ohuchi N, Fukada H, Sugiyama J, Shirakawa M, Watanabe T. 2000. Expression and characterization of the chitin-binding domain

of chitinase A1 from Bacillus circulans WL-12. Journal of bacteriology 182(11):3045-54.

Hong J, Wang Y, Ye X, Zhang YH. 2008. Simple protein purification through affinity adsorption on regenerated amorphous cellulose followed by intein self-cleavage. J Chromatogr A 1194(2):150-4.

Isetti G, Maurer MC. 2007. Employing mutants to study thrombin residues responsible for factor XIII activation peptide recognition: a kinetic study. Biochemistry 46(9):2444-52.

Johannes TW, Woodyer RD, Zhao H. 2005. Directed evolution of a thermostable phosphite dehydrogenase for NAD(P)H regeneration. Appl Environ Microbiol 71(10):5728-34.

Kapust RB, Tozser J, Fox JD, Anderson DE, Cherry S, Copeland TD, Waugh DS. 2001. Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. Protein Eng 14(12):993-1000.

Larsen KS, Auld DS. 1989. Carboxypeptidase A: mechanism of zinc inhibition. Biochemistry 28(25):9620-5.

Lew BM, Mills KV, Paulus H. 1999. Characteristics of protein splicing in trans mediated by a semisynthetic split intein. Peptide Science 51(5):355-362.

Li YF. 2011. Self-cleaving fusion tags for recombinant protein production. Biotechnology Letters 33(5):869-881.

Lockless SW, Muir TW. 2009. Traceless protein splicing utilizing evolved split inteins. Proceedings of the National Academy of Sciences of the United States of America 106(27):10999-11004.

Lu W, Sun Z, Tang Y, Chen J, Tang F, Zhang J, Liu JN. 2011. Split intein facilitated tag affinity purification for recombinant proteins with controllable tag removal by inducible auto-cleavage. Journal of Chromatography A 1218(18):2553-2560.

Malakhov MP, Mattern MR, Malakhova OA, Drinker M, Weeks SD, Butt TR. 2004. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. J Struct. Funct. Genomics 5(1-2):75-86.

Mathys S, Evans TC, Chute IC, Wu H, Chong S, Benner J, Liu XQ, Xu MQ. 1999. Characterization of a self-splicing mini-intein and its conversion into autocatalytic N- and C-terminal cleavage elements: facile production of protein building blocks for protein ligation. Gene 231(1–2):1-13.

Nichols NM, Benner JS, Martin DD, Evans TC, Jr. 2003. Zinc ion effects on individual Ssp DnaE intein splicing steps: regulating pathway progression. Biochemistry 42(18):5301-11.

Nichols NM, Evans TC, Jr. 2004. Mutational analysis of protein splicing, cleavage, and self-association reactions mediated by the naturally split Ssp DnaE intein. Biochemistry 43(31):10265-76.

Oeemig JS, Aranko AS, Djupsjobacka J, Heinamaki K, Iwai H. 2009. Solution structure of DnaE intein from Nostoc punctiforme: Structural basis for the design of a new

split intein suitable for site-specific chemical modification. Febs Letters 583(9):1451-1456.

Perry DK, Smyth MJ, Stennicke HR, Salvesen GS, Duriez P, Poirier GG, Hannun YA. 1997. Zinc is a potent inhibitor of the apoptotic protease, caspase-3. A novel target for zinc in the inhibition of apoptosis. J Biol Chem 272(30):18530-3.

Ramirez M, Valdes N, Guan D, Chen Z. 2013. Engineering split intein DnaE from Nostoc punctiforme for rapid protein purification. Protein Eng Des Sel 26(3):215-23.

Southworth MW, Amaya K, Evans TC, Xu MQ, Perler FB. 1999. Purification of proteins fused to either the amino or carboxy terminus of the Mycobacterium xenopi gyrase A intein. Biotechniques 27(1):110-4.

Sun P, Ye S, Ferrandon S, Evans TC, Xu MQ, Rao Z. 2005. Crystal structures of an intein from the split dnaE gene of Synechocystis sp PCC6803 reveal the catalytic model without the penultimate histidine and the mechanism of zinc ion inhibition of protein splicing. Journal of Molecular Biology 353(5):1093-1105.

Waugh DS. 2005. Making the most of affinity tags. Trends Biotechnol 23(6):316-20.

Waugh DS. 2011. An overview of enzymatic reagents for the removal of affinity tags. Protein Expr. Purif. 80:283-293.

Wood DW, Wu W, Belfort G, Derbyshire V, Belfort M. 1999. A genetic system yields self-cleaving inteins for bioseparations. Nature Biotechnology 17(9):889-892.

Wu W, Wood DW, Belfort G, Derbyshire V, Belfort M. 2002. Intein-mediated purification of cytotoxic endonuclease I-TevI by insertional inactivation and pH-controllable splicing. Nucleic Acids Res 30(22):4864-71.

Wu WY, Mee C, Califano F, Banki R, Wood DW. 2006. Recombinant protein purification by self-cleaving aggregation tag. Nature Protocols 1(5):2257-62.

Xu MQ, Paulus H, Chong S. 2000. Fusions to self-splicing inteins for protein purification. Methods Enzymol 326:376-418.

Xu MQ, Perler FB. 1996. The mechanism of protein splicing and its modulation by mutation. EMBO J. 15(19):5146-53.

Zettler J, Schutz V, Mootz HD. 2009. The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. FEBS Lett 583(5):909-14.

Zhang L, Zheng Y, Callahan B, Belfort M, Liu Y. 2011. Cisplatin inhibits protein splicing, suggesting inteins as therapeutic targets in mycobacteria. J Biol Chem 286(2):1277-82.