METHODS AND COMPOSITIONS RELATED TO AN INTEIN DUAL MUTANT FOR ACCELERATED CLEAVING

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to, and the benefit of, U.S. Provisional Application No. 63/564,246, filed March 12, 2024, which is hereby incorporated herein by reference in its entirety.

REFERENCE TO SEQUENCE LISTING

[0002] The Sequence Listing submitted on March 11, 2025, as an .xml file named "103361-635WO1_ST26.xml" created March 12, 2025, and having a file size of 105,639 bytes is hereby incorporated by reference pursuant to 37 CFR § 1.52(e)(5).

BACKGROUND

[0003] Inteins are naturally occurring, self-splicing protein subdomains that are capable of excising out their own protein subdomain from a larger protein structure while simultaneously joining the two formerly flanking peptide regions ("exteins") together to form a mature host protein.

[0004] The ability of inteins to rearrange flanking peptide bonds and retain activity when fused to proteins other than their native exteins, has led to a number of intein-based biotechnologies. These include various types of protein ligation and activation applications, as well as protein labeling and tracing applications. An important application of inteins is in the production of purified recombinant proteins. In particular, inteins have the ability to impart self-cleaving activity to a number of conventional affinity and purification tags and thus provide a major advance in the production of recombinant protein products for research, medical and other commercial applications.

[0005] Conventional purification tags provide a simple and robust means for purifying any tagged target protein and are commonly added to desired target proteins through simple genetic fusions. These tags are now ubiquitous in research and have formed a major platform for research and manufacturing of these important products. Once the tagged target protein is expressed in an appropriate host cell and purified via the tag, however, the presence of the tag on the purified target can lead to compromised activity, and potentially unwanted immunogenicity in the case of therapeutic proteins. For these reasons, the ability to remove

the affinity tag after purification is of critical importance in many applications, which is conventionally done through the addition of highly specific endopeptidase enzymes. Although these enzymes are generally effective, they are too expensive to scale up for manufacturing, and their use requires an additional step for their removal.

[0006] Thus, the ability of inteins to impart self-cleaving activity to conventional tags is a significant advance, and early implementations of intein-based self-cleaving affinity tag systems have been published in several patents and hundreds of journal papers in the biological sciences. Despite their strength, however, several substantial weaknesses remain that inhibit the full implementation of intein methods. In particular, the ability to tightly control the cleaving reaction in a variety of highly relevant contexts has been elusive. In order to be useful, the intein self-cleaving reaction must be tightly suppressed during protein expression and purification, but very rapid once the tagged target protein is pure. Of the two initially available classes of conventional inteins, one is highly controllable and is triggered to cleave by the addition of thiol compounds, while the other is more loosely controlled and is triggered by small changes in pH and temperature. To better control the activity of pH sensitive inteins, split intein systems have been developed where the intein is initially expressed as two separate segments that can only achieve cleaving capability after assembly. In these systems, one segment of the intein is covalently coupled to a convenient chromatographic backbone, while the other acts as an affinity tag in fusion to the target protein. In at least one of these systems, the assembled intein exhibits cleaving that is highly sensitive to pH, allowing assembly of the intein and purification of the target, followed by controlled release of the target protein from the affinity resin (Prabhala 2022).

[0007] Split intein systems have been effective in controlling premature cleavage and maximizing yields of their target proteins, but some challenges remain. A significant challenge is that some target proteins cleave more quickly or slowly than others, requiring a degree of optimization when first applying an intein purification system to a specific target. Although it is possible to roughly predict the cleavage rate of different proteins, usually based on the identities of their first two amino acids (Prabhala 2022) some proteins cleave at an inconveniently slow rate. In these cases, process times can become unacceptably long, and the split intein system cannot be used for that target.

[0008] Therefore, what is needed is a more reliable method for the selective purification of a broader range of proteins using a stable, transformative intein system, wherein the intein system comprises mutants which can increase cleavage rate (tag self-removal) for currently

difficult proteins, while still remaining strongly sensitive under permissive conditions, such as pH.

SUMMARY

[0009] In accordance with the purpose(s) of the invention, as embodied and broadly described herein, the invention, in one aspect, relates to an amino acid sequence comprising 90% or more identity to SEQ ID NO: 2, wherein SEQ ID NO: 2 comprises a tryptophan at position 74.

[0010] Also disclosed is an amino acid sequence comprising 90% or more identity to SEQ ID NO: 12, wherein SEQ ID NO: 12 comprises a threonine at position 34.

[0011] Described herein is a protein purification system, wherein the system comprises a split intein comprising two separate peptides: an N-terminal intein segment and a C-terminal intein segment, wherein the N-terminal intein segment comprises 90% or more identity to SEQ ID NO: 2, wherein SEQ ID NO: 2 comprises a tryptophan at position 74.

[0012] Also described herein is a protein purification system, wherein the system comprises a split intein comprising two separate peptides: an N-terminal intein segment and a C-terminal intein segment, wherein the C-terminal intein segment comprises 90% or more identity to SEQ ID NO: 12, wherein SEQ ID NO: 12 comprises a threonine at position 34.

[0013] While aspects of the present invention can be described and claimed in a particular statutory class, such as the system statutory class, this is for convenience only and one of skill in the art will understand that each aspect of the present invention can be described and claimed in any statutory class. Unless otherwise expressly stated, it is in no way intended that any method or aspect set forth herein be construed as requiring that its steps be performed in a specific order. Accordingly, where a method claim does not specifically state in the claims or descriptions that the steps are to be limited to a specific order, it is in no way intended that an order be inferred, in any respect. This holds for any possible non-express basis for interpretation, including matters of logic with respect to arrangement of steps or operational flow, plain meaning derived from grammatical organization or punctuation, or the number or type of aspects described in the specification.

BRIEF DESCRIPTION OF THE FIGURES

[0014] The accompanying figures, which are incorporated in and constitute a part of this specification, illustrate several aspects and together with the description serve to explain the principles of the invention.

[0015] Figure 1 shows the general splicing mechanisms for different classes of inteins. Examples of Class 1 inteins include the *N. punctiforme* DnaE split intein and the previously reported ΔICM contiguous mini intein. Examples of Class 3 inteins include the *M. smegmatis* DnaB1 contiguous intein and the *M. leprae* contiguous intein. In this figure, "X" represents either T or S. A key difference between Class 1 and Class 3 inteins is that the C (SH) nucleophile is buried in the middle of the intein for Class 3. By eliminating the first nucleophile, the first two steps of the splicing mechanisms can be circumvented, and the Asn cyclization step (Step III) can proceed without regulation of previous steps, thus generating a C-terminal cleaving mutant.

[0016] Figure 2 shows a purification method using a Npu split intein, where half of the intein (a) can be immobilized on an affinity chromatography stationary phase, while the other half intein (b) is expressed as recombinant fusion tag. This basic design has been commercialized as the iCapTag™ system by Protein Capture Science.

[0017] Figure 3 shows the Npu self-cleaving affinity purification system scheme.

[0018] Figure 4 shows that Npu intein associates in two reversible steps, typically referred to as capture and collapse. Also shown is SEQ ID NO: 2 (NupN1), SEQ ID NO: 24 (Npu_{N2}), and SEQ ID NO: 12 (Npu_C).

[0019] Figure 5 shows protein sequence alignments of different classes of inteins where the letter height indicates the degree of conservation of a given amino acid within each group (Tori, K., et al. *Journal of Biological Chemistry*, *285*(4), 2515-2526.). Note the highly conserved W-C-T of the Class 3 inteins, which is not observed in Class 1 and 2 inteins.

[0020] Figure 6A-B shows (A) structure alignments of three Class 1 inteins with a class 3 intein, *indicating* that the W67 and T137 residue locations are relatively preserved in the structure across different classes of inteins. Since the positions overlap, these mutations potentially have similar impacts in other Class 1 inteins. (B) Class 1 inteins structure alignment with class 3 intein shows that the W67,T137 residue locations are relatively preserved in the structure across different classes of inteins as the positions have overlapping-these mutations may potentially have a similar impact in other class 1 inteins/

[0021] Figure 7 shows the rational design of Class 3, the F65-T137 from the Class 3 WCT motif aligns with the Npu DnaE Class 1 split intein structure. Class 3 inteins have a WCT highly conserved triplet feature. In this case, Npu_N(F74W) combined with Npu_C(A33T), which precedes the catalytic Npuc(N35), may potentially interact with catalytic residue Npu_N(H65) in the active site. In this case, C is omitted to prevent any chances for

unwanted disulfide bond oxidation or N-terminal cleavage, also that is where D16G is located in NpuC.

[0022] Figure 8 shows that the mutations disclosed herein are located at the core of the Npu DNA split intein structure.

[0023] Figure 9 shows the protein engineering approach: rational design for the F74W,A33T dual mutation in the split Npu DnaE intein.

[0024] Figure 10 shows the approach for experimental characterization of C-terminal cleavage phenotypes.

[0025] Figure 11 shows that the NpuC (A33T) single mutant promotes faster C-terminal cleavage, while the NpuN (F74W) single mutant decreases cleaving activity.

[0026] Figure 12 shows that the F74W, A33T dual mutant promotes faster cleaving, while the F74W,A33S dual mutants exhibit decreased cleaving activity.

[0027] Figure 13 shows that the F74W, A33T dual mutation enhances C-terminal cleavage relative to the control intein and other intein variants.

[0028] Figure 14 shows that the F74W,A33T dual mutation shows a 2.08x-fold faster cleaving rate than the current Npu version.

[0029] Figure 15 shows that the F74W,A33T dual mutant maintains pH controllability.

[0030] Figure 16 shows how the first amino acids of the fusion target protein can influence the speed of the Npu C-terminal cleavage reaction. The second amino acid of the target has a similar effect, where the effects of the first two amino acids of the target protein are additive.

[0031] Figure 17 shows F74W,A33T dual mutant shows faster C-terminal cleavage across different +1,+2 extein sequences that are typically slower cleaving.

[0032] Figure 18 shows a comparison of cleaving activities and half-life times of the mutant with other split inteins.

[0033] Figure 19 shows F74W, A33T dual mutant consistently shows over 50% cleavage by 3 hours.

[0034] Figure 20 shows the predicted protein structure of F74W, A33T dual mutant, indicating conservation of intein native structure with very high confidence.

[0035] Figure 21 shows F74W,A33T dual mutations may promote better organization of the intein's catalytic core. Hydrophobic amino acids like alanine or valine show similar C-terminal cleavage rates. Hydrophilic amino acids cannot participate in protein folding (hydrophobic molten globule) and can decrease the strength of the hydrophobic collapse. A33T provides both a non-polar group and a polar group so that intein activity is not only

preserved but enhanced. F74W offers a bulky hydrophobic surface area, participates in protein folding, and can align A33T towards H72, promoting better coordination and/or chemistry of the latter. The A33T sidechain can create a hydrogen bond with the K73 peptide backbone and H72 sidechain.

[0036] Figure 22 shows the experimental design and setup for testing the intein self-cleavage phenotypes with eGFP as a model protein with the different mutations in F74 and A33 and their factorial combinations.

[0037] Figure 23 shows the intein self-cleavage phenotypes with eGFP as a model protein with the different mutations in F74 and A33 and their factorial combinations.

[0038] Figure 24 shows the intein self-cleavage phenotypes for the +1 amino acid effect study for control, single mutant A33T, and double mutant F74W, A33T.

[0039] Figure 25 shows the experimental setup for Npu ligand variant conjugation on a bromoacetic activated resin (Avitide) through a free reduced thiol on the C-terminal cysteine of Npu ligand.

[0040] Figure 26 shows the ONEWAY ANOVA and Tukey-Kramer all means comparison statistical analysis (JMP software) for the control and F74W ligand variant ligand immobilization efficiency.

[0041] Figure 27 shows intein self-cleavage phenotypes for single mutants F74W and A33T, and double mutant F74W, A33T against the control in pH 6.2 and pH 8.5.

[0042] Figure 28 shows intein self-cleavage phenotypes for single mutants F74W and A33T, and double mutant F74W,A33T against the control in cold temperature (4°C).

[0043] Additional advantages of the invention will be set forth in part in the description which follows, and in part will be obvious from the description, or can be learned by practice of the invention. The advantages of the invention will be realized and attained by means of the elements and combinations particularly pointed out in the appended claims. It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of the invention, as claimed.

DESCRIPTION

[0044] The present invention can be understood more readily by reference to the following detailed description of the invention and the Examples included therein.

[0045] Before the present compounds, compositions, articles, systems, devices, and/or methods are disclosed and described, it is to be understood that they are not limited to specific synthetic methods unless otherwise specified, or to particular reagents unless

otherwise specified, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular aspects only and is not intended to be limiting. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, example methods and materials are now described.

[0046] All publications mentioned herein are incorporated herein by reference to disclose and describe the methods and/or materials in connection with which the publications are cited. The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention. Further, the dates of publication provided herein can be different from the actual publication dates, which can require independent confirmation.

A. DEFINITIONS

[0047] As used in the specification and the appended claims, the singular forms "a," "an" and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a functional group," "an alkyl," or "a residue" includes mixtures of two or more such functional groups, alkyls, or residues, and the like.

[0048] Ranges can be expressed herein as from "about" one particular value, and/or to "about" another particular value. When such a range is expressed, a further aspect includes from the one particular value and/or to the other particular value. Similarly, when values are expressed as approximations, by use of the antecedent "about," it will be understood that the particular value forms a further aspect. It will be further understood that the endpoints of each of the ranges are significant both in relation to the other endpoint, and independently of the other endpoint. It is also understood that there are a number of values disclosed herein, and that each value is also herein disclosed as "about" that particular value in addition to the value itself. For example, if the value "10" is disclosed, then "about 10" is also disclosed. It is also understood that each unit between two particular units are also disclosed. For example, if 10 and 15 are disclosed, then 11, 12, 13, and 14 are also disclosed.

[0049] A weight percent (wt. %) of a component, unless specifically stated to the contrary, is based on the total weight of the formulation or composition in which the component is included.

[0050] As used herein, the terms "optional" or "optionally" means that the subsequently described event or circumstance can or cannot occur, and that the description includes instances where said event or circumstance occurs and instances where it does not.

[0051] The term "contacting" as used herein refers to bringing two biological entities together in such a manner that the compound can affect the activity of the target, either directly; i.e., by interacting with the target itself, or indirectly; i.e., by interacting with another molecule, co-factor, factor, or protein on which the activity of the target is dependent. "Contacting" can also mean facilitating the interaction of two biological entities, such as peptides, to bond covalently or otherwise.

[0052] As used herein, "kit" means a collection of at least two components constituting the kit. Together, the components constitute a functional unit for a given purpose. Individual member components may be physically packaged together or separately. For example, a kit comprising an instruction for using the kit may or may not physically include the instruction with other individual member components. Instead, the instruction can be supplied as a separate member component, either in a paper form or an electronic form which may be supplied on computer readable memory device or downloaded from an internet website, or as recorded presentation.

[0053] As used herein, "instruction(s)" means documents describing relevant materials or methodologies pertaining to a kit. These materials may include any combination of the following: background information, list of components and their availability information (purchase information, etc.), brief or detailed protocols for using the kit, trouble-shooting, references, technical support, and any other related documents. Instructions can be supplied with the kit or as a separate member component, either as a paper form or an electronic form which may be supplied on computer readable memory device or downloaded from an internet website, or as recorded presentation. Instructions can comprise one or multiple documents and are meant to include future updates.

[0054] As used herein, the terms "target protein, " "protein of interest," and "therapeutic agent" include any synthetic or naturally occurring protein or peptide. The term therefore encompasses those compounds traditionally regarded as drugs, vaccines, and biopharmaceuticals including molecules such as proteins, peptides, and the like. Examples of therapeutic agents are described in well-known literature references such as the Merck Index (14th edition), the Physicians' Desk Reference (64th edition), and The Pharmacological Basis of Therapeutics (1st edition), and they include, without limitation, medicaments; substances used for the treatment, prevention, diagnosis, cure or mitigation of a disease or illness; substances that affect the structure or function of the body, or pro-drugs, which become biologically active or more active after they have been placed in a physiological environment.

[0055] As used herein, "variant" refers to a molecule that retains a biological activity that is the same or substantially similar to that of the original sequence. The variant may be from the same or different species or be a synthetic sequence based on a natural or prior molecule. Moreover, as used herein, "variant" refers to a molecule having a structure attained from the structure of a parent molecule (e.g., a protein or peptide disclosed herein) and whose structure or sequence is sufficiently similar to those disclosed herein that based upon that similarity, would be expected by one skilled in the art to exhibit the same or similar activities and utilities compared to the parent molecule. For example, substituting specific amino acids in a given peptide can yield a variant peptide with similar activity to the parent.

[0056] As used herein, "mutant" refers to a molecule that has been modified in a way which confers a different structural and/or functional characteristic to the molecule. This can be a modification to the nucleic acid which affects the amino acid sequence, or, in the case of protein synthesis, can be a direct modification to an amino acid sequence. This modification can result in a difference in function between the mutant and the original, or naturally occurring, molecule. This modification can confer benefits to the mutant which are not seen in the original, or naturally occurring, molecule.

[0057] As used herein, the term "amino acid sequence" refers to a list of abbreviations, letters, characters, or words representing amino acid residues. The amino acid abbreviations used herein are conventional one letter codes for the amino acids and are expressed as follows: A, alanine; C, cysteine; D aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H histidine; I isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine.

[0058] "Peptide" as used herein refers to any peptide, oligopeptide, polypeptide, gene product, expression product, or protein. A peptide is comprised of consecutive amino acids. The term "peptide" encompasses naturally occurring or synthetic molecules.

[0059] In addition, as used herein, the term "peptide" refers to amino acids joined to each other by peptide bonds or modified peptide bonds, e.g., peptide isosteres, etc. and may contain modified amino acids other than the 20 gene-encoded amino acids. The peptides can be modified by either natural processes, such as post-translational processing, or by chemical modification techniques which are well known in the art. Modifications can occur anywhere in the peptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl terminus. The same type of modification can be present in the same or varying degrees at several sites in a given polypeptide. Also, a given peptide can have many types of

modifications. Modifications include, without limitation, linkage of distinct domains or motifs, acetylation, acylation, ADP-ribosylation, amidation, covalent cross-linking or cyclization, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of a phosphytidylinositol, disulfide bond formation, demethylation, formation of cysteine or pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristolyation, oxidation, pergylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, and transfer-RNA mediated addition of amino acids to protein such as arginylation. (See Proteins—Structure and Molecular Properties 2nd Ed., T. E. Creighton, W.H. Freeman and Company, New York (1993); Posttranslational Covalent Modification of Proteins, B. C. Johnson, Ed., Academic Press, New York, pp. 1-12 (1983)).

[0060] As used herein, "isolated peptide" or "purified peptide" is meant to mean a peptide (or a fragment thereof) that is substantially free from the materials with which the peptide is normally associated in nature, or from the materials with which the peptide is associated in an artificial expression or production system, including but not limited to an expression host cell lysate, growth medium components, buffer components, cell culture supernatant, or components of a synthetic *in vitro* translation system. The peptides disclosed herein, or fragments thereof, can be obtained, for example, by extraction from a natural source (for example, a mammalian cell), by expression of a recombinant nucleic acid encoding the peptide (for example, in a cell or in a cell-free translation system), or by chemically synthesizing the peptide. In addition, peptide fragments may be obtained by any of these methods, or by cleaving full length proteins and/or peptides.

[0061] The word "or" as used herein means any one member of a particular list and also includes any combination of members of that list.

[0062] The phrase "nucleic acid" as used herein refers to a naturally occurring or synthetic oligonucleotide or polynucleotide, whether DNA or RNA or DNA-RNA hybrid, single-stranded or double-stranded, sense or antisense, which is capable of hybridization to a complementary nucleic acid by Watson-Crick base-pairing. Nucleic acids of the invention can also include nucleotide analogs (e.g., BrdU), and non-phosphodiester internucleoside linkages (e.g., peptide nucleic acid (PNA) or thiodiester linkages). In particular, nucleic acids can include, without limitation, DNA, RNA, cDNA, gDNA, ssDNA, dsDNA, or any combination thereof.

[0063] As used herein, "isolated nucleic acid" or "purified nucleic acid" is meant to mean DNA that is free of the genes that, in the naturally-occurring genome of the organism from which the DNA of the invention is derived, flank the gene. The term therefore includes, for example, a recombinant DNA which is incorporated into a vector, such as an autonomously replicating plasmid or virus; or incorporated into the genomic DNA of a prokaryote or eukaryote (e.g., a transgene); or which exists as a separate molecule (for example, a cDNA or a genomic or cDNA fragment produced by PCR, restriction endonuclease digestion, or chemical or *in vitro* synthesis). It also includes a recombinant DNA which is part of a hybrid gene encoding additional polypeptide sequences. The term "isolated nucleic acid" also refers to RNA, e.g., an mRNA molecule that is encoded by an isolated DNA molecule, or that is chemically synthesized, or that is separated or substantially free from at least some cellular components, for example, other types of RNA molecules or peptide molecules.

[0064] As used herein, "extein" refers to the portion of an intein-modified protein that is not part of the intein and which can be spliced or cleaved upon excision of the intein.

[0065] "Intein" refers to an in-frame intervening sequence in a protein. An intein can catalyze its own excision from the protein through a post-translational protein splicing process to yield the free intein and a mature protein. An intein can also catalyze the cleavage of the intein-extein bond at either the intein N-terminus, or the intein C-terminus, or both of the intein-extein termini. As used herein, "intein" encompasses mini-inteins, modified or mutated inteins, and split inteins.

[0066] A "split intein" is an intein that is comprised of two or more separate components not fused to one another. Split inteins can occur naturally or can be engineered by splitting contiguous inteins, and are inactive when split but regain their intein activity after reassembly through a "Capture and Collapse" mechanism (see Figure 4).

[0067] As used herein, the term "splice" or "splices" means to excise a central portion of a polypeptide to form two or more smaller polypeptide molecules. In some cases, splicing also includes the step of fusing together two or more of the smaller polypeptides to form a new polypeptide. Splicing can also refer to the joining of two polypeptides encoded on two separate gene products through the action of a split intein.

[0068] As used herein, the term "cleave" or "cleaves" means to divide a single polypeptide to form two or more smaller polypeptide molecules. In some cases, cleavage is mediated by the addition of an extrinsic endopeptidase, which is often referred to as "proteolytic cleavage." In other cases, cleaving can be mediated by the intrinsic activity of one or both of the cleaved peptide sequences, which is often referred to as "self-cleavage."

Cleavage can also refer to the self-cleavage of two polypeptides that is induced by the addition of a non-proteolytic third peptide, as in the action of split intein system described herein.

[0069] By the term "fused" is meant covalently bonded to. For example, a first peptide is fused to a second peptide when the two peptides are covalently bonded to each other (e.g., via a peptide bond).

[0070] As used herein an "isolated" or "substantially pure" substance is one that has been separated from components which naturally accompany it. Typically, a polypeptide is substantially pure when it is at least 50% (e.g., 60%, 70%, 80%, 90%, 95%, and 99%) by weight free from the other proteins and naturally-occurring organic molecules with which it is naturally associated.

[0071] Herein, "bind" or "binds" means that one molecule recognizes and adheres to another molecule in a sample but does not substantially recognize or adhere to other molecules in the sample. One molecule "specifically binds" another molecule if it has a binding affinity greater than about 10^5 to 10^6 liters/mole for the other molecule.

[0072] Nucleic acids, nucleotide sequences, proteins or amino acid sequences referred to herein can be isolated, purified, synthesized chemically, or produced through recombinant DNA technology. All of these methods are well known in the art.

[0073] As used herein, the terms “modified” or “mutated,” as in “modified intein” or “mutated intein,” refer to one or more modifications in either the nucleic acid or amino acid sequence being referred to, such as an intein, when compared to the native, or naturally occurring structure. Such modification can be a substitution, addition, or deletion. The modification can occur in one or more amino acid residues or one or more nucleotides of the structure being referred to, such as an intein.

[0074] As used herein, the term “modified peptide,” “modified protein” or “modified protein of interest” or “modified target protein” refers to a protein which has been modified.

[0075] As used herein, “operably linked” refers to the association of two or more biomolecules in a configuration relative to one another such that the normal function of the biomolecules can be performed. In relation to nucleotide sequences, “operably linked” refers to the association of two or more nucleic acid sequences, by means of enzymatic ligation or otherwise, in a configuration relative to one another such that the normal function of the sequences can be performed. For example, the nucleotide sequence encoding a pre-sequence or secretory leader is operably linked to a nucleotide sequence for a polypeptide if it is expressed as a pre-protein that participates in the secretion of the polypeptide; a promoter or

enhancer is operably linked to a coding sequence if it affects the transcription of the coding sequence; and a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation of the sequence.

[0076] "Sequence homology" can refer to the situation where nucleic acid or protein sequences are similar because they have a common evolutionary origin. "Sequence homology" can indicate that sequences are very similar. Sequence similarity is observable; homology can be based on the observation. "Very similar" can mean at least 70% identity, homology or similarity; at least 75% identity, homology or similarity; at least 80% identity, homology or similarity; at least 85% identity, homology or similarity; at least 90% identity, homology or similarity, such as at least 93% or at least 95% or even at least 97% identity, homology or similarity. The nucleotide sequence similarity or homology or identity can be determined using the "Align" program of Myers et al. (1988) CABIOS 4:11-17 and available at NCBI. Additionally or alternatively, amino acid sequence similarity or identity or homology can be determined using the BlastP program (Altschul et al. Nucl. Acids Res. 25:3389-3402), and available at NCBI. Alternatively or additionally, the terms "similarity" or "identity" or "homology," for instance, with respect to a nucleotide sequence, are intended to indicate a quantitative measure of homology between two sequences.

[0077] Alternatively or additionally, "similarity" with respect to sequences refers to the number of positions with identical nucleotides divided by the number of nucleotides in the shorter of the two sequences wherein alignment of the two sequences can be determined in accordance with the Wilbur and Lipman algorithm. (1983) Proc. Natl. Acad. Sci. USA 80:726. For example, using a window size of 20 nucleotides, a word length of 4 nucleotides, and a gap penalty of 4, and computer-assisted analysis and interpretation of the sequence data including alignment can be conveniently performed using commercially available programs (e.g., Intelligenetics™ Suite, Intelligenetics Inc. CA). When RNA sequences are said to be similar or have a degree of sequence identity with DNA sequences, thymidine (T) in the DNA sequence is considered equal to uracil (U) in the RNA sequence. The following references also provide algorithms for comparing the relative identity or homology or similarity of amino acid residues of two proteins, and additionally or alternatively with respect to the foregoing, the references can be used for determining percent homology, identity, or similarity. Needleman et al. (1970) J. Mol. Biol. 48:444-453; Smith et al. (1983) Advances App. Math. 2:482-489; Smith et al. (1981) Nuc. Acids Res. 11:2205-2220; Feng et al. (1987) J. Molec. Evol. 25:351-360; Higgins et al. (1989) CABIOS 5:151-153; Thompson et al. (1994) Nuc. Acids Res. 22:4673-480; and Devereux et al. (1984) 12:387-395.

"Stringent hybridization conditions" is a term which is well known in the art; see, for example, Sambrook, "Molecular Cloning, A Laboratory Manual" second ed., CSH Press, Cold Spring Harbor, 1989; "Nucleic Acid Hybridization, A Practical Approach", Hames and Higgins eds., IRL Press, Oxford, 1985; see also FIG. 2 and description thereof herein wherein there is a sequence comparison.

[0078] The terms "plasmid" and "vector" and "cassette" refer to an extrachromosomal element often carrying genes which are not part of the central metabolism of the cell and usually in the form of circular double-stranded DNA molecules. Such elements may be autonomously replicating sequences, genome integrating sequences, phage or nucleotide sequences, linear or circular, of a single- or double-stranded DNA or RNA, derived from any source, in which a number of nucleotide sequences have been joined or recombined into a unique construction which is capable of introducing a promoter fragment and DNA sequence for a selected gene product along with appropriate 3′ untranslated sequence into a cell. Typically, a "vector" is a modified plasmid that contains additional multiple insertion sites for cloning and an "expression cassette" that contains a DNA sequence for a selected gene product (i.e., a transgene) for expression in the host cell. This "expression cassette" typically includes a 5′ promoter region, the transgene ORF, and a 3′ terminator region, with all necessary regulatory sequences required for transcription and translation of the ORF. Thus, integration of the expression cassette into the host permits expression of the transgene ORF in the cassette.

[0079] The term "buffer" or "buffered solution" refers to solutions which resist changes in pH by the action of its conjugate acid-base range.

[0080] The term "loading buffer" or "equilibrium buffer" refers to the buffer containing the salt or salts which is mixed with the protein preparation for loading the protein preparation onto a column. This buffer is also used to equilibrate the column before loading, and to wash to column after loading the protein.

[0081] The term "wash buffer" is used herein to refer to the buffer that is passed over a column (for example) following loading of a protein of interest (such as one coupled to a C-terminal intein fragment, for example) and prior to elution of the protein of interest. The wash buffer may serve to remove one or more contaminants without substantial elution of the desired protein.

[0082] The term "elution buffer" refers to the buffer used to elute the desired protein from the column. As used herein, the term "solution" refers to either a buffered or a non-buffered solution, including water.

[0083] The term "washing" means passing an appropriate buffer through or over a solid support, such as a chromatographic resin.

[0084] The term "eluting" a molecule (e.g., a desired protein or contaminant) from a solid support means removing the molecule from such material.

[0085] The term "contaminant" or "impurity" refers to any foreign or objectionable molecule, particularly a biological macromolecule such as a DNA, an RNA, or a protein, other than the protein being purified, that is present in a sample of a protein being purified. Contaminants include, for example, other proteins from cells that express and/or secrete the protein being purified.

[0086] The term "separate" or "isolate" as used in connection with protein purification refers to the separation of a desired protein from a second protein or other contaminant or mixture of impurities in a mixture comprising both the desired protein and a second protein or other contaminant or impurity mixture, such that at least the majority of the molecules of the desired protein are removed from that portion of the mixture that comprises at least the majority of the molecules of the second protein or other contaminant or mixture of impurities.

[0087] The term "purify" or "purifying" a desired protein from a composition or solution comprising the desired protein and one or more contaminants means increasing the degree of purity of the desired protein in the composition or solution by removing (completely or partially) at least one contaminant from the composition or solution.

[0088] Disclosed are the components to be used to prepare the compositions of the invention as well as the compositions themselves to be used within the methods disclosed herein. These and other materials are disclosed herein, and it is understood that when combinations, subsets, interactions, groups, etc. of these materials are disclosed that while specific reference of each various individual and collective combinations and permutation of these compounds cannot be explicitly disclosed, each is specifically contemplated and described herein. For example, if a particular compound is disclosed and discussed and a number of modifications that can be made to a number of molecules including the compounds are discussed, specifically contemplated is each and every combination and permutation of the compound and the modifications that are possible unless specifically indicated to the contrary. Thus, if a class of molecules A, B, and C are disclosed as well as a class of molecules D, E, and F and an example of a combination molecule, A-D is disclosed, then even if each is not individually recited each is individually and collectively contemplated meaning combinations, A-E, A-F, B-D, B-E, B-F, C-D, C-E, and C-F are considered disclosed. Likewise, any subset or combination of these is also disclosed. Thus,

for example, the sub-group of A-E, B-F, and C-E would be considered disclosed. This concept applies to all aspects of this application including, but not limited to, steps in methods of making and using the compositions of the invention. Thus, if there are a variety of additional steps that can be performed it is understood that each of these additional steps can be performed with any specific embodiment or combination of embodiments of the methods of the invention.

[0089] It is understood that the compositions disclosed herein have certain functions. Disclosed herein are certain structural requirements for performing the disclosed functions, and it is understood that there are a variety of structures that can perform the same function that are related to the disclosed structures, and that these structures will typically achieve the same result. For example, compounds used to control pH in the examples shown can be substituted with other buffering compounds to control pH, since pH is the critical variable to be controlled and the specific buffering compounds can vary.

B. Protein Purification Systems

[0090] Disclosed herein is a protein purification system, wherein the system comprises an engineered split intein comprising two separate peptides: an N-terminal intein segment and a C-terminal intein segment, wherein the N-terminal intein segment can be linked to a solid support, and wherein the C-terminal intein segment DNA is genetically fused to a desired target protein DNA such that the expressed target protein is tagged with the C-terminal intein segment, and wherein cleaving of the C-terminal intein segment is suppressed in the absence of the N-terminal intein segment, and wherein the N-terminal intein segment and C-terminal intein segment associate strongly to form an immobilized complex when contacted to each other on the solid support, and wherein the C-terminal cleaving activity of the C-terminal intein segment in the assembled N-terminal and C-terminal intein segment complex is highly sensitive to extrinsic conditions when compared to a native intein complex, and wherein the immobilized intein complex with the fused target protein can be purified through washing away of unimmobilized contaminants, and wherein the C-terminal intein segment can be induced to cleave and thereby release the untagged and substantially purified target protein from the immobilized intein complex through a controlled change in extrinsic conditions. Figure 3 shows an example of the split intein scheme.

[0091] Intein-based methods of protein modification and ligation have been developed. An intein is an internal protein sequence capable of catalyzing a protein splicing reaction that excises the intein sequence from a precursor protein and joins the flanking sequences (N- and

C-exteins) with a peptide bond (Perler et al. (1994)). Hundreds of intein and intein-like sequences have been found in a wide variety of organisms and proteins (Perler et al. (2002); Liu et al. (2003)), they are typically 350-550 amino acids in size and also contain a homing endonuclease domain, but natural and engineered mini-inteins having only the ~140-aa splicing domain are sufficient for protein splicing (Liu et al. (2003); Yang et al. (2004); Telenti et al. (1997); Wu et al. (1998); Derbyshire et al. (1997)). The conserved crystal structure of mini-inteins (or the protein splicing domain) consists of ~12 beta-strands that form a disk-like structure with the two splicing junctions located in a central cleft (Duan et al. (1997); Ichiyanagi et al. (2000); Klabunde et al. (1998); Ding et al. (2003); Xu et al. (1996)). Depending on the locations of specific residues involved in the splicing reaction, inteins can be classified as Class 1, Class 2 or Class 3 inteins (Figure 1).

[0092] As used herein, the term "split intein" refers to any intein in which one or more peptide bond breaks exists between the N-terminal intein segment and the C-terminal intein segment such that the N-terminal and C-terminal intein segments become separate molecules that can non-covalently reassociate, or reconstitute, into an intein that is functional for splicing or cleaving reactions. Any catalytically active intein, or fragment thereof, may be used to derive a split intein for use in the systems and methods disclosed herein. For example, in one aspect the split intein may be derived from a eukaryotic intein. In another aspect, the split intein may be derived from a bacterial intein. In another aspect, the split intein may be derived from an archaeal intein. Preferably, the split intein so-derived will possess only the amino acid sequences essential for catalyzing splicing reactions. Examples of inteins which can be used with the methods and systems disclosed herein can be found in Table 2.

[0093] Specifically disclosed herein are split inteins which have been engineered to possess unexpected and desirable properties. For example, SEQ ID NOS: 2 and 12 represent modified split inteins. These specific inteins are discussed in more detail below.

[0094] As used herein, the "N-terminal intein segment" refers to any intein sequence that comprises an N- terminal amino acid sequence that is functional for splicing and/or cleaving reactions when combined with a corresponding C-terminal intein segment. An N-terminal intein segment thus also comprises a sequence that is spliced out when splicing occurs. An N-terminal intein segment can comprise a sequence that is a modification of the N-terminal portion of a naturally occurring (native) intein sequence. For example, an N-terminal intein segment can comprise additional amino acid residues and/or mutated residues so long as the inclusion of such additional and/or mutated residues does not render the intein non-functional for splicing or cleaving. Such modifications are discussed in more detail below. Preferably,

the inclusion of the additional and/or mutated residues improves or enhances the splicing and/or cleaving activity and/or controllability of the intein. Non-intein residues can also be genetically fused to intein segments to provide additional functionality, such as the ability to be affinity purified or to be covalently immobilized. SEQ ID NOS: 1 and 2 represent N-terminal intein segments. SEQ ID NO: 1 is a native N terminal intein segment, and SEQ ID NO: 2 is a modified version of SEQ ID NO: 1, wherein the phenylalanine ("F") at position 74 is substituted for a tryptophan ("W"). This provides unexpectedly better results which are discussed in more detail below.

[0095] As used herein, the "C-terminal intein segment" refers to any intein sequence that comprises a C- terminal amino acid sequence that is functional for splicing or cleaving reactions when combined with a corresponding N-terminal intein segment. In one aspect, the C-terminal intein segment comprises a sequence that is spliced out when splicing occurs. In another aspect, the C-terminal intein segment is cleaved from a peptide sequence fused to its C-terminus. The sequence which is cleaved from the C-terminal intein's C-terminus is referred to herein as a "protein of interest" or "target protein" and is discussed in more detail below. A C-terminal intein segment can comprise a sequence that is a modification of the C-terminal portion of a naturally occurring (native) intein sequence. For example, a C terminal intein segment can comprise additional amino acid residues and/or mutated residues so long as the inclusion of such additional and/or mutated residues does not render the C-terminal intein segment non-functional for splicing or cleaving. Preferably, the inclusion of the additional and/or mutated residues improves or enhances the splicing and/or cleaving activity of the intein. SEQ ID NOS: 11 and 12 represent C-intein segments. SEQ ID NO: 11 is a native C intein segment, and SEQ ID NO: 12 is a modified version, wherein the alanine ("A") at position 34 is substituted for a threonine ("T"). It is noted that in some instances, this mutation is referred to as "A33T". It is noted that this occurs when methionine ("M") is not considered in the first position. So "A33T" is considered the same as "A34T" when described herein, with "A33T" referring to SEQ ID NO: 12 with no methionine present. These mutations (shown in Figure 8) provide unexpectedly better results which are discussed in more detail below.

[0096] The intein can be derived, for example, from an *Npu* DnaE intein, as shown in Figure 2. Npu_N refers to the N-terminal intein segment, while Npu_C refers to the C-terminal intein segment. "TP" refers to the target protein (also referred to herein as the protein of interest), which can be coupled to Npu_C. In Figure 3, a scheme is shown in which Npu_N is coupled to a solid support, such as a commercialized resin. Npu_C, while attached to a target

protein, is exposed to immobilized NpuN. The NpuN and Npuc then associate, allowing purification of the fused target protein. When the proper conditions are present, a cleaving reaction can occur, which allows for cleaving (and subsequent elution) of the target protein from the immobilized intein segments.

[0097] Specifically disclosed herein is a split intein, which can comprise at least one mutation in the N-terminal intein segment, at least one mutation in the C-terminal intein segment, or both mutations in the N and C-terminal intein segments. As described above, these mutations can be in the NpuN polypeptide represented by SEQ ID NO: 2, and/or in the Npuc polypeptide represented by SEQ ID NO: 12. These modifications can, for example, greatly speed up cleaving reactions with different proteins that consistently display slower cleaving rates with the unmutated intein segments. Because these intein variants can substantially accelerate the cleaving reaction, the number of native proteins that can be purified employing this method can be greatly increased due to its improved tolerance to different target protein amino acid sequences. These modifications and their benefits are demonstrated and discussed in detail in Examples 2 and 3.

[0098] The modified intein segments are found in SEQ ID NOS: 2 and 12 and can be combined with other mutations with respect to SEQ ID NOS: 1 and 11 as well. These are given as examples in Table 1.

[0099] Described herein is an N-terminal intein segment comprising SEQ ID NO: 2. Also described is SEQ ID NO: 2, with an additional 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more modifications that also provide desired properties which are preferable to the native SEQ ID NO: 1. Examples of these additional modifications can be found in SEQ ID NOS: 3-10. Put another way, described herein is SEQ ID NO: 2 which comprises 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%, or any amount in between, above, or below this amount, with respect to SEQ ID NO: 2, wherein the variation does not include changing tryptophan at position 74 of SEQ ID NO: 2.

[00100] Also described herein is a C-terminal intein segment comprising SEQ ID NO: 12. Also described is SEQ ID NO: 12, with an additional 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more modifications that also provide desired properties which are preferable to the native SEQ ID NO: 1. Examples of these additional modifications can be found in SEQ ID NOS: 13 and 14. Put another way, described herein is SEQ ID NO: 12 which comprises 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%, or any amount in between, above, or below this amount, with respect to SEQ ID NO:

12, wherein the variation does not include changing threonine at position 34 of SEQ ID NO: 12.

[00101] As mentioned above, the cleavage reaction can be sped up by using SEQ ID NO: 2 as the N-terminal intein segment, and/or by using SEQ ID NO: 12 as the C-terminal intein segment. By "sped up" is meant that the reaction rate is reduced by 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% compared to the native intein. More specifically, the reaction rate is sped up as compared to using SEQ ID NOS: 1 and 11 in an identical reaction.

[00102] Generally speaking, split inteins have many practical uses including the production of recombinant proteins from fragments, the circularization of recombinant proteins, and the fixation of proteins on protein chips (Scott et al., (1999); Xu et al. (2001); Evans et al. (2000); Kwon et al., (2006)). The advantages of intein-based protein cleavage methods, compared to others such as protease-based methods, have been noted previously (Xu et al. (2001)). The intein-based N- and C-cleavage methods can also be used together on a single target protein to produce precise and tag-free ends at both the N- and the C-termini, or to achieve cyclization of the target protein (ligation of the N- and C-termini) using the expressed protein ligation approach.

[00103] Some examples of other modifications to the N-terminal intein segment include those which alter the amino acid sequence so that it does not comprise any internal cysteine residues. This is desirous so as to eliminate side reactions associated with immobilization of the Npu$_N$ intein segment onto a solid support. For example, the residues can be mutated to serine. An example of Npu$_N$ in which the cysteine residues have been mutated can be found in SEQ ID NO: 3. It is noted that the first cysteine residue which is replaced (the first amino acid on the intein N-terminus) can be replaced with either alanine or glycine so as to eliminate intein splicing in the assembled intein complex. The Npu$_N$ intein can also be modified by the addition of an internal affinity tag to facilitate its purification, and the addition of specific peptides to its C-terminus to facilitate covalent immobilization onto a solid support. For example, an included His tag was used to purify the Npu$_N$-segment described herein, while a Cys residue was appended to the C-terminus of Npu$_N$ to facilitate

covalent chemical immobilization of the NpuN-segment. The modified NpuN segment that incorporates both the His tag and Cys residue is shown by way of example in SEQ ID NO: 6. Importantly, any tag can be used to purify the N-segment (not just His), and several different immobilization chemistries can be used to immobilize the N-segment. "Tags" are more generally referred to herein as purification tags. In one example, the N-segment could be purified without a tag, using conventional chromatography. Furthermore, it is also possible to add a His tag mutation in the C-terminal intein segment, and a "sensitivity enhancing domain" to the N-terminal intein segment.

[00104] The N-terminal intein segment can also comprise a purification, or affinity tag, attached to its C-terminus. This can include an affinity resin reagent. The purification tag can comprise, for example, one or more histidine residues. The purification tag can comprise, for example, a chitin binding domain protein with highly specific affinity for chitin. The purification tag can further comprise, for example, a reversibly precipitating elastin-like peptide tag, which can be induced to selectively precipitate under known conditions of buffer composition and temperature. Affinity tags are discussed in more detail below. The N-terminal intein segment can also comprise amino acids at its C-terminus which allow for covalent immobilization. For example, one or more amino acids at the C-terminus can be cysteine residues.

[00105] The N-terminal intein segment can be immobilized onto a solid support. A variety of supports can be used. For example, the solid support can a polymer or substance that allows for immobilization of the N-terminal intein fragment, which can occur covalently or via an affinity tag with or without an appropriate linker. When a linker is used, the linker can be additional amino acid residues engineered to the C-terminus of the N-terminal intein segment or can be other known linkers for attachment of a peptide to a support.

[00106] The N-terminal intein segment disclosed herein can be attached to an affinity tag through a linker sequence. The linker sequence can be designed to create distance between the intein and affinity tag, while providing minimal steric interference to the intein cleaving active site. It is generally accepted that linkers involve a relatively unstructured amino acid sequence, and the design and use of linkers are common in the art of designing fusion peptides. There is a variety of protein linker databases which one of skill in the art will recognize. This includes those found in Argos et al. J Mol Biol 1990 Feb 20; 211(4) 943-58; Crasto et al. Protein Eng 2000 May; 13(5) 309-12; George et al. Protein Eng 2002 Nov; 15(11) 871-9; Arai et al. Protein Eng 2001 Aug; 14(8) 529-32; and Robinson et al. PNAS

May 26, 1998, vol. 95 no. 11 5929-5934, hereby incorporated by reference in their entirety for their teaching of examples of linkers.

[00107] Table 1 shows exemplary sequences of the N-terminal intein segment and the C-terminal intein segment:

Table 1

SEQ ID	Construct Name	Construct Name (Description)	Amino Acid Sequence
SEQ ID NO: 1	Npu_N^{WT}	Wild-type *Npu* DNAE (INT_N segment) capable of splicing events	CLSYETEILTVEYGLLPIGKIVEKRI ECTVYSVDNNGNIYTQPVAQWHD RGEQEVFEYCLEDGSLIRATKDHK FMTVDGQMLPIDEIFERELDLMRV DNLPN
SEQ ID NO: 2	Npu_N^{F74W}	Variant of SEQ ID NO: 1 with mutation at F74W	CLSYETEILTVEYGLLPIGKIVEKRI ECTVYSVDNNGNIYTQPVAQWHD RGEQEVFEYCLEDGSLIRATKDHK WMTVDGQMLPIDEIFERELDLMR VDNLPN
SEQ ID NO: 3	$Npu_N^{C1X, F74W}$	Cleaving variant of SEQ ID NO: 2; cleaving phenotype resulting from a C1X mutation, where "X" = A or G	XLSYETEILTVEYGLLPIGKIVEKRI ECTVYSVDNNGNIYTQPVAQWHD RGEQEVFEYCLEDGSLIRATKDHK WMTVDGQMLPIDEIFERELDLMR VDNLPN
SEQ ID NO: 4	$Npu_N^{C1X,C-S. F74W}$	Thiol-knockout variant of SEQ ID NO: 3, derived by mutating remaining Cysteine residues to Serine residues.	XLSYETEILTVEYGLLPIGKIVEKRI ESTVYSVDNNGNIYTQPVAQWHD RGEQEVFEYSLEDGSLIRATKDHK WMTVDGQMLPIDEIFERELDLMR VDNLPN
SEQ ID NO: 5	S_{04b}-$Npu_N^{C1X,C-S. F74W}$	Variant of SEQ ID NO: 4 modified with Sensitivity-Enhancing Motif expressed as a fusion partner at the N-terminus	MGDGHGXLSYETEILTVEYGLLPI GKIVEKRIESTVYSVDNNGNIYTQP VAQWHDRGEQEVFEYSLEDGSLIR ATKDHKWMTVDGQMLPIDEIFER ELDLMRVDNLPN
SEQ ID NO: 6	S_{04b}-$Npu_N^{C1X,C-S_F74W}$-G_4S-His_6-Cys	Variant derived from SEQ ID NO: 5; constructed by adding linker, tag, and immobilization	MGDGHGXLSYETEILTVEYGLLPI GKIVEKRIESTVYSVDNNGNIYTQP VAQWHDRGEQEVFEYSLEDGSLIR ATKDHKWMTVDGQMLPIDEIFER ELDLMRVDNLPNGGGGSHHHHHH

		moiety fusion partners at the C-terminus of the N-Intein Ligand.	C
SEQ ID NO: 7	S_{04b}-$Npu_N^{C1X,C-S,F74W}G_4S$-Cys-His_6	Variant of SEQ ID NO: 6; derived from an alternate arrangement of the I-L-T fusion partners at the C-terminus.	MGDGHGALSYETEILTVEYGLLPIGKIVEKRIESTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYSLEDGSLIRATKDHKWMTVDGQMLPIDEIFERELDLMRVDNLPNGGGGSCHHHHHH
SEQ ID NO: 8	S_{04b}-$Npu_N^{C1X,C-S,F74W}$-G_4S-Cys-G_4S-His_6	Variant of SEQ ID NO: 7 created by adding an additional linker between I-L-T moieties at the C-terminus.	MGDGHGALSYETEILTVEYGLLPIGKIVEKRIESTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYSLEDGSLIRATKDHKWMTVDGQMLPIDEIFERELDLMRVDNLPNGGGGSCGGGGSHHHHHH
SEQ ID NO: 9	S_{04b}-$Npu_N^{C1X,C-S,F74W}$-$(G_4S)_2$-Cys-G_4S-His_6	Variant of SEQ ID NO: 8 created by adding an additional linker between I-L-T moieties at the C-terminus.	MGDGHGALSYETEILTVEYGLLPIGKIVEKRIESTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYSLEDGSLIRATKDHKWMTVDGQMLPIDEIFERELDLMRVDNLPNGGGGSGGGGSCGGGGSHHHHHH
SEQ ID NO: 10	S_{04b}-$Npu_N^{C1X,C-S,F74W}$-$(G_4S)_2$-Cys-G_4S	Variant of SEQ ID NO: 8 created by removing the His6 purification tag moiety from the C-terminus.	MGDGHGALSYETEILTVEYGLLPIGKIVEKRIESTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYSLEDGSLIRATKDHKWMTVDGQMLPIDEIFERELDLMRVDNLPNGGGGSGGGGSCGGGGS
SEQ ID NO: 11	Npu_C^{WT}	Wild-type *Npu* DNAE (INT_C segment) capable of splicing events	MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN
SEQ ID NO: 12	Npu_C^{A34T}	Variant of SEQ ID NO: 11 with A34T mutation	MIKIATRKYLGKQNVYDIGVERDHNFALKNGFITSN
SEQ ID NO: 13	$Npu_C^{D118G,\ A34T}$	Cleaving variant of SEQ ID NO: 12; Accelerated cleaving phenotype resulting from D118G mutation	MIKIATRKYLGKQNVYGIGVERDHNFALKNGFITSN
SEQ ID NO: 14	$Npu_C^{DG,HN,A34T}$	Variant derived from SEQ ID NO: 12	MIKIATRKYLGKQNVYGIGVERDHNFALKNGFITHN

		comprising D118G and S136H mutations, producing a cleaving phenotype with enhanced sensitivity to extrinsic conditions	
SEQ ID NO: 15	$Npu^{CDG,HN-FFN}$-sfGFP-His6	Variant derived from SEQ ID NO: 14; A Cognate Binding Partner comprising a rapid-cleaving INTC variant expressed with GFP and His6 as fusion partner tag moieties	MIKIATRKYLGKQNVYGIGVERDH NFALKNGFIAHNFFNGTVSKGEEL FTGVVPILVELDGDVNGHKFSVRG EGEGDATNGKLTLKFICTTGKLPV PWPTLVTTLTYGVQCFSRYPDHM KQHDFFKSAMPEGYVQERTISFKD DGTYKTRAEVKFEGDTLVNRIELK GIDFKEDGNILGHKLEYNFNSHNV YITADKQKNGIKANFKIRHNVEDG SVQLADHYQQNTPIGDGPVLLPDN HYLSTQSKLSKDPNEKRDHMVLL EFVTAAGITLGMDELYKLEHHHH HH
SEQ ID NO: 16	$Npuc^{DG,HA}$	Variant of SEQ ID NO: 14, A Cognate Binding Partner modified with an N137A mutation that produces a binding-only (non-cleaving) phenotype	MIKIATRKYLGKQNVYGIGVERDH NFALKNGFIAHA
SEQ ID NO: 17	His_6-$Npuc^{DG,HA}$	Variant of SEQ ID NO: 16, A non-cleaving Cognate Binding Partner expressed with a His_6 purification tag as an N-terminal fusion partner	MHHHHHHIKIATRKYLGKQNVYG IGVERDHNFALKNGFIAHA
SEQ ID NO: 18	$Npuc^{DG,HA}$-His_6	Variant of SEQ ID NO: 16, A non-cleaving Cognate Binding Partner expressed with a His_6 purification tag as an C-terminal fusion partner	MIKIATRKYLGKQNVYGIGVERDH NFALKNGFIAHAHHHHHH
SEQ ID NO: 19	$Npuc^{DG,HN}$(A33T)_(VSK)sfGFP-WSHPQFEK_HHHHHH	An INTc-POI fusion construct for testing split intein-mediated affinity capture with	MIKIATRKYLGKQNVYGIGVERDH NFALKNGFITHNVSKGEELFTGVV PILVELDGDVNGHKFSVRGEGEGD ATNGKLTLKFICTTGKLPVPWPTL

		a target protein of interest (sfGFP)	VTTLTYGVQCFSRYPDHMKQHDF FKSAMPEGYVQERTISFKDDGTYK TRAEVKFEGDTLVNRIELKGIDFKE DGNILGHKLEYNFNSHNVYITADK QKNGIKANFKIRHNVEDGSVQLAD HYQQNTPIGDGPVLLPDNHYLSTQ SKLSKDPNEKRDHMVLLEFVTAA GITLGMDELYKLEWSHPQFEKHHH HHH
SEQ ID NO: 20	S_{04b}-$Npu_N^{C1X,C-S}$-G_4S-Cys-His_6	Variant derived from SEQ ID NO: 19; constructed by adding linker, immobilization moiety, and purification tag fusion partners at the C-terminus of the N-Intein Ligand.	MGDGHGXLSYETEILTVEYGLLPI GKIVEKRIESTVYSVDNNGNIYTQP VAQWHDRGEQEVFEYSLEDGSLIR ATKDHKWMTVDGQMLPIDEIFER ELDLMRVDNLPNGGGGSCHHHHH H
SEQ ID NO: 21	M-Npuc(A33)_(MSD)TrxA_WSHPQFEK_HHHHHH (control)		MIKIATRKYLGKQNVYGIGVERDH NFALKNGFIAHNMSDKIIHLTDDSF DTDVLKADGAILVDFWAEWCGPC KMIAPILDEIADEYQGKLTVAKLNI DQNPGTAPKYGIRGIPTLLLFKNGE VAATKVGALSKGQLKEFLDANLA WSHPQFEKHHHHHH*
SEQ ID NO: 22	M-Npuc(A33T)_(MSD)TrxA_WSHPQFEK_HHHHHH		MIKIATRKYLGKQNVYGIGVERDH NFALKNGFITHNMSDKIIHLTDDSF DTDVLKADGAILVDFWAEWCGPC KMIAPILDEIADEYQGKLTVAKLNI DQNPGTAPKYGIRGIPTLLLFKNGE VAATKVGALSKGQLKEFLDANLA WSHPQFEKHHHHHH*
SEQ ID NO: 23	M-Npuc(A33S)_(MSD)TrxA_WSHPQFEK_HHHHHH		MIKIATRKYLGKQNVYGIGVERDH NFALKNGFISHNMSDKIIHLTDDSF DTDVLKADGAILVDFWAEWCGPC KMIAPILDEIADEYQGKLTVAKLNI DQNPGTAPKYGIRGIPTLLLFKNGE VAATKVGALSKGQLKEFLDANLA WSHPQFEKHHHHHH*

SEQ ID NO: 24	M-Npuc(A33V)_(MSD)TrxA_WSHPQFEK_HHHHHH		MIKIATRKYLGKQNVYGIGVERDHNFALKNGFI**V**HNMSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAWSHPQFEKHHHHHH*
SEQ ID NO: 25	M-Npuc(A33)_(EVQ)TrxA_WSHPQFEK_HHHHHH (control)		MIKIATRKYLGKQNVYGIGVERDHNFALKNGFI**A**HN**EVQ**KIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAWSHPQFEKHHHHHH*
SEQ ID NO: 26	M-Npuc(A33)_(DIG)TrxA_WSHPQFEK_HHHHHH (Control)		MIKIATRKYLGKQNVYGIGVERDHNFALKNGFI**A**HN**DIG**KIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAWSHPQFEKHHHHHH*
SEQ ID NO: 27	M-Npuc(A33T)_(EVQ)TrxA_WSHPQFEK_HHHHHH		MIKIATRKYLGKQNVYGIGVERDHNFALKNGFI**T**HN**EVQ**KIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAWSHPQFEKHHHHHH*

SEQ ID NO: 28	M-Npuc(A33T)_(DIG)TrxA_WSHPQFEK_HHHHHH		MIKIATRKYLGKQNVYGIGVERDHNFALKNGFITHN**DIG**KIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAWSHPQFEKHHHHHH*
SEQ ID NO: 29	pET_MGDGHG-NpuN(F74)_GGGS_CBD (control)		MGDGHGALSYETEILTVEYGLLPIGKIVEKRIESTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYSLEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPNGGGGSMTTNPGVSAWQVNTAYTAGQLVTYNGKTYKCLQPHTSLAGWEPSNVPALWQLQTSSSNNNNNNNNNLGIEGRISEF*
SEQ ID NO: 30	pET_MGDGHG-NpuN(F74W)_GGGS_CBD		MGDGHGALSYETEILTVEYGLLPIGKIVEKRIESTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYSLEDGSLIRATKDHKWMTVDGQMLPIDEIFERELDLMRVDNLPNGGGGSMTTNPGVSAWQVNTAYTAGQLVTYNGKTYKCLQPHTSLAGWEPSNVPALWQLQTSSSNNNNNNNNNLGIEGRISEF*
SEQ ID NO: 31	NpuC DG,HN(A33)_(VSK)sfGFP_WSHPQFEK_HHHHHH (Control)		MIKIATRKYLGKQNVYGIGVERDHNFALKNGFIAHNVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKLEWSHPQFEKHHHHHH

SEQ ID NO: 32	Npuc(A33)_(MSD)TrxA_strepII_His6x (control)		MIKIATRKYLGKQNVYGIGVERDHNFALKNGFIAHNMSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAWSHPQFEKHHHHHH*
SEQ ID NO: 33	Npuc(A33T)_(MSD)TrxA_strepII_His6x		MIKIATRKYLGKQNVYGIGVERDHNFALKNGFITHNMSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAWSHPQFEKHHHHHH*
SEQ ID NO: 34	Npuc(A33S)_(MSD)TrxA_strepII_His6x		MIKIATRKYLGKQNVYGIGVERDHNFALKNGFISHNMSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAWSHPQFEKHHHHHH*
SEQ ID NO: 35	Npuc(A33V)_(MSD)TrxA_strepII_His6x		MIKIATRKYLGKQNVYGIGVERDHNFALKNGFIVHNMSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAWSHPQFEKHHHHHH*
SEQ ID NO: 36	Npuc(A33)_(EVQ)TrxA_strepII_His6x (control)		MIKIATRKYLGKQNVYGIGVERDHNFALKNGFIAHNEVQKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAWSHPQFEKHHHHHH*

SEQ ID NO: 37	Npuc(A33)_(DIG)TrxA_strepII_His6x (Control)		MIKIATRKYLGKQNVYGIGVERDHNFALKNGFIAHNDIGKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAWSHPQFEKHHHHHH*
SEQ ID NO: 38	Npuc(A33T)_(EVQ)TrxA_strepII_His6x		MIKIATRKYLGKQNVYGIGVERDHNFALKNGFITHNEVQKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAWSHPQFEKHHHHHH*
SEQ ID NO: 39	Npuc(A33T)_(DIG)TrxA_strepII_His6x		MIKIATRKYLGKQNVYGIGVERDHNFALKNGFITHNDIGKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAWSHPQFEKHHHHHH*
SEQ ID NO: 40	SED-NpuN(F74)-GGGGS_CBD (control)		MGDGHGALSYETEILTVEYGLLPIGKIVEKRIESTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYSLEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPNGGGGSMTTNPGVSAWQVNTAYTAGQLVTYNGKTYKCLQPHTSLAGWEPSNVPALWQLQTSSSNNNNNNNNNLGIEGRISEF*
SEQ ID NO: 41	SED-NpuN(F74W)-GGGGS_CBD		MGDGHGALSYETEILTVEYGLLPIGKIVEKRIESTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYSLEDGSLIRATKDHKWMTVDGQMLPIDEIFERELDLMRVDNLPNGGGGSMTTNPGVSAWQVNTAYTAGQLVTYNGKTYKCLQPHTSLAGWEPSNVPALWQLQTSSSNNNNNNNNNLGIEGRISEF*

SEQ ID NO: 42	SED-Npun(F74Y)-GGGGS_CBD		MGDGHGALSYETEILTVEYGLLPI GKIVEKRIESTVYSVDNNGNIYTQP VAQWHDRGEQEVFEYSLEDGSLIR ATKDHKYMTVDGQMLPIDEIFERE LDLMRVDNLPNGGGGSMTTNPGV SAWQVNTAYTAGQLVTYNGKTY KCLQPHTSLAGWEPSNVPALWQL QTSSSNNNNNNNNNNLGIEGRISEF
SEQ ID NO: 43	SED-Npun(F74V)-GGGGS_CBD		MGDGHGALSYETEILTVEYGLLPI GKIVEKRIESTVYSVDNNGNIYTQP VAQWHDRGEQEVFEYSLEDGSLIR ATKDHKYMTVDGQMLPIDEIFERE LDLMRVDNLPNGGGGSMTTNPGV SAWQVNTAYTAGQLVTYNGKTY KCLQPHTSLAGWEPSNVPALWQL QTSSSNNNNNNNNNNLGIEGRISEF *
SEQ ID NO: 44	Npuc(A33)_(VSK)eGFP_His6x		MIKIATRKYLGKQNVYGIGVERDH NFALKNGFIAHNVSKGEELFTGVD HNFALKNGFINHNVSKGEELFTGV VPILVELDGDVNGHKFSVRGEGEG DATNGKLTLKFICTTGKLPVPWPT LVTTLTYGVQCFSRYPDHMKQHD FFKSAMPEGYVQERTISFKDDGTY KTRAEVKFEGDTLVNRIELKGIDF KEDGNILGHKLEYNFNSHNVYITA DKQENGIKANFKIRHNVEDGSVQL ADHYQQNTPIGDGPVLLPDNHYLS TQSKLSKDPNEKRDHMVLLEFVT AAGITLGMDELYKLEHHHHHH*
SEQ ID NO: 45	Npuc(A33T)_(VSK)eGFP_His6x		MIKIATRKYLGKQNVYGIGVERDH NFALKNGFITHNVSKGEELFTGVD HNFALKNGFINHNVSKGEELFTGV VPILVELDGDVNGHKFSVRGEGEG DATNGKLTLKFICTTGKLPVPWPT LVTTLTYGVQCFSRYPDHMKQHD FFKSAMPEGYVQERTISFKDDGTY KTRAEVKFEGDTLVNRIELKGIDF KEDGNILGHKLEYNFNSHNVYITA DKQENGIKANFKIRHNVEDGSVQL ADHYQQNTPIGDGPVLLPDNHYLS TQSKLSKDPNEKRDHMVLLEFVT AAGITLGMDELYKLEHHHHHH*

SEQ ID NO: 46	Npuc(A33V)_(VSK)eGFP_His6x		MIKIATRKYLGKQNVYGIGVERDH NFALKNGFI**V**HN**VSK**GEELFTGVD HNFALKNGFINHNVSKGEELFTGV VPILVELDGDVNGHKFSVRGEGEG DATNGKLTLKFICTTGKLPVPWPT LVTTLTYGVQCFSRYPDHMKQHD FFKSAMPEGYVQERTISFKDDGTY KTRAEVKFEGDTLVNRIELKGIDF KEDGNILGHKLEYNFNSHNVYITA DKQENGIKANFKIRHNVEDGSVQL ADHYQQNTPIGDGPVLLPDNHYLS TQSKLSKDPNEKRDHMVLLEFVT AAGITLGMDELYKLEHHHHHH*
SEQ ID NO: 47	Npuc(A33S)_(VSK)eGFP_His6x		MIKIATRKYLGKQNVYGIGVERDH NFALKNGFI**S**HN**VSK**GEELFTGVD HNFALKNGFINHNVSKGEELFTGV VPILVELDGDVNGHKFSVRGEGEG DATNGKLTLKFICTTGKLPVPWPT LVTTLTYGVQCFSRYPDHMKQHD FFKSAMPEGYVQERTISFKDDGTY KTRAEVKFEGDTLVNRIELKGIDF KEDGNILGHKLEYNFNSHNVYITA DKQENGIKANFKIRHNVEDGSVQL ADHYQQNTPIGDGPVLLPDNHYLS TQSKLSKDPNEKRDHMVLLEFVT AAGITLGMDELYKLEHHHHHH*
SEQ ID NO: 48	Npuc(A33C)_(VSK)eGFP_His6x		MIKIATRKYLGKQNVYGIGVERDH NFALKNGFI**C**HN**VSK**GEELFTGVD HNFALKNGFINHNVSKGEELFTGV VPILVELDGDVNGHKFSVRGEGEG DATNGKLTLKFICTTGKLPVPWPT LVTTLTYGVQCFSRYPDHMKQHD FFKSAMPEGYVQERTISFKDDGTY KTRAEVKFEGDTLVNRIELKGIDF KEDGNILGHKLEYNFNSHNVYITA DKQENGIKANFKIRHNVEDGSVQL ADHYQQNTPIGDGPVLLPDNHYLS TQSKLSKDPNEKRDHMVLLEFVT AAGITLGMDELYKLEHHHHHH*

SEQ ID NO: 49	Npuc(A33**N**)_(VSK)eGFP_His6x		MIKIATRKYLGKQNVYGIGVERDH NFALKNGFI**N**HN**VSK**GEELFTGVD HNFALKNGFINHNVSKGEELFTGV VPILVELDGDVNGHKFSVRGEGEG DATNGKLTLKFICTTGKLPVPWPT LVTTLTYGVQCFSRYPDHMKQHD FFKSAMPEGYVQERTISFKDDGTY KTRAEVKFEGDTLVNRIELKGIDF KEDGNILGHKLEYNFNSHNVYITA DKQENGIKANFKIRHNVEDGSVQL ADHYQQNTPIGDGPVLLPDNHYLS TQSKLSKDPNEKRDHMVLLEFVT AAGITLGMDELYKLEHHHHHH*
SEQ ID NO: 50	Npuc(A33**L**)_(VSK)eGFP_His6x		MIKIATRKYLGKQNVYGIGVERDH NFALKNGFI**L**HN**VSK**GEELFTGVD HNFALKNGFINHNVSKGEELFTGV VPILVELDGDVNGHKFSVRGEGEG DATNGKLTLKFICTTGKLPVPWPT LVTTLTYGVQCFSRYPDHMKQHD FFKSAMPEGYVQERTISFKDDGTY KTRAEVKFEGDTLVNRIELKGIDF KEDGNILGHKLEYNFNSHNVYITA DKQENGIKANFKIRHNVEDGSVQL ADHYQQNTPIGDGPVLLPDNHYLS TQSKLSKDPNEKRDHMVLLEFVT AAGITLGMDELYKLEHHHHHH*
SEQ ID NO: 51	Npuc(A33)_(FSK)eGFP_His6x		MIKIATRKYLGKQNVYGIGVERDH NFALKNGFI**A**HN**FSK**GEELFTGVD HNFALKNGFINHNVSKGEELFTGV VPILVELDGDVNGHKFSVRGEGEG DATNGKLTLKFICTTGKLPVPWPT LVTTLTYGVQCFSRYPDHMKQHD FFKSAMPEGYVQERTISFKDDGTY KTRAEVKFEGDTLVNRIELKGIDF KEDGNILGHKLEYNFNSHNVYITA DKQENGIKANFKIRHNVEDGSVQL ADHYQQNTPIGDGPVLLPDNHYLS TQSKLSKDPNEKRDHMVLLEFVT AAGITLGMDELYKLEHHHHHH*

SEQ ID NO: 52	Npuc(A33T)_(FSK)eGFP_His6x		MIKIATRKYLGKQNVYGIGVERDH NFALKNGFITHN**FSK**GEELFTGVD HNFALKNGFINHNVSKGEELFTGV VPILVELDGDVNGHKFSVRGEGEG DATNGKLTLKFICTTGKLPVPWPT LVTTLTYGVQCFSRYPDHMKQHD FFKSAMPEGYVQERTISFKDDGTY KTRAEVKFEGDTLVNRIELKGIDF KEDGNILGHKLEYNFNSHNVYITA DKQENGIKANFKIRHNVEDGSVQL ADHYQQNTPIGDGPVLLPDNHYLS TQSKLSKDPNEKRDHMVLLEFVT AAGITLGMDELYKLEHHHHHH*
SEQ ID NO: 53	Npuc(A33)_(MSK)eGFP_His6x		MIKIATRKYLGKQNVYGIGVERDH NFALKNGFI**A**HN**MSK**GEELFTGV DHNFALKNGFINHNVSKGEELFTG VVPILVELDGDVNGHKFSVRGEGE GDATNGKLTLKFICTTGKLPVPWP TLVTTLTYGVQCFSRYPDHMKQH DFFKSAMPEGYVQERTISFKDDGT YKTRAEVKFEGDTLVNRIELKGID FKEDGNILGHKLEYNFNSHNVYIT ADKQENGIKANFKIRHNVEDGSVQ LADHYQQNTPIGDGPVLLPDNHYL STQSKLSKDPNEKRDHMVLLEFVT AAGITLGMDELYKLEHHHHHH*
SEQ ID NO: 54	Npuc(A33T)_(MSK)eGFP_His6x		MIKIATRKYLGKQNVYGIGVERDH NFALKNGFITHN**MSK**GEELFTGVD HNFALKNGFINHNVSKGEELFTGV VPILVELDGDVNGHKFSVRGEGEG DATNGKLTLKFICTTGKLPVPWPT LVTTLTYGVQCFSRYPDHMKQHD FFKSAMPEGYVQERTISFKDDGTY KTRAEVKFEGDTLVNRIELKGIDF KEDGNILGHKLEYNFNSHNVYITA DKQENGIKANFKIRHNVEDGSVQL ADHYQQNTPIGDGPVLLPDNHYLS TQSKLSKDPNEKRDHMVLLEFVT AAGITLGMDELYKLEHHHHHH*

SEQ ID NO: 55	Npuc(A33)_(ESK)eGFP_His6x		MIKIATRKYLGKQNVYGIGVERDH NFALKNGFIAHNESKGEELFTGVD HNFALKNGFINHNVSKGEELFTGV VPILVELDGDVNGHKFSVRGEGEG DATNGKLTLKFICTTGKLPVPWPT LVTTLTYGVQCFSRYPDHMKQHD FFKSAMPEGYVQERTISFKDDGTY KTRAEVKFEGDTLVNRIELKGIDF KEDGNILGHKLEYNFNSHNVYITA DKQENGIKANFKIRHNVEDGSVQL ADHYQQNTPIGDGPVLLPDNHYLS TQSKLSKDPNEKRDHMVLLEFVT AAGITLGMDELYKLEHHHHHH*
SEQ ID NO: 56	Npuc(A33T)_(ESK)eGFP_His6x		MIKIATRKYLGKQNVYGIGVERDH NFALKNGFITHNESKGEELFTGVD HNFALKNGFINHNVSKGEELFTGV VPILVELDGDVNGHKFSVRGEGEG DATNGKLTLKFICTTGKLPVPWPT LVTTLTYGVQCFSRYPDHMKQHD FFKSAMPEGYVQERTISFKDDGTY KTRAEVKFEGDTLVNRIELKGIDF KEDGNILGHKLEYNFNSHNVYITA DKQENGIKANFKIRHNVEDGSVQL ADHYQQNTPIGDGPVLLPDNHYLS TQSKLSKDPNEKRDHMVLLEFVT AAGITLGMDELYKLEHHHHHH*
SEQ ID NO: 57	Npuc(A33)_(KSK)eGFP_His6x		MIKIATRKYLGKQNVYGIGVERDH NFALKNGFIAHNVSKGEELFTGVD HNFALKNGFINHNVSKGEELFTGV VPILVELDGDVNGHKFSVRGEGEG DATNGKLTLKFICTTGKLPVPWPT LVTTLTYGVQCFSRYPDHMKQHD FFKSAMPEGYVQERTISFKDDGTY KTRAEVKFEGDTLVNRIELKGIDF KEDGNILGHKLEYNFNSHNVYITA DKQENGIKANFKIRHNVEDGSVQL ADHYQQNTPIGDGPVLLPDNHYLS TQSKLSKDPNEKRDHMVLLEFVT AAGITLGMDELYKLEHHHHHH*

SEQ ID NO: 58	Npuc(A33T)_(KSK)eGFP_His6x		MIKIATRKYLGKQNVYGIGVERDH NFALKNGFI**T**HN**KSK**GEELFTGVD HNFALKNGFINHNVSKGEELFTGV VPILVELDGDVNGHKFSVRGEGEG DATNGKLTLKFICTTGKLPVPWPT LVTTLTYGVQCFSRYPDHMKQHD FFKSAMPEGYVQERTISFKDDGTY KTRAEVKFEGDTLVNRIELKGIDF KEDGNILGHKLEYNFNSHNVYITA DKQENGIKANFKIRHNVEDGSVQL ADHYQQNTPIGDGPVLLPDNHYLS TQSKLSKDPNEKRDHMVLLEFVT AAGITLGMDELYKLEHHHHHH*
SEQ ID NO: 59	Npuc(A33)_(NSK)eGFP_His6x		MIKIATRKYLGKQNVYGIGVERDH NFALKNGFI**A**HN**NSK**GEELFTGVD HNFALKNGFINHNVSKGEELFTGV VPILVELDGDVNGHKFSVRGEGEG DATNGKLTLKFICTTGKLPVPWPT LVTTLTYGVQCFSRYPDHMKQHD FFKSAMPEGYVQERTISFKDDGTY KTRAEVKFEGDTLVNRIELKGIDF KEDGNILGHKLEYNFNSHNVYITA DKQENGIKANFKIRHNVEDGSVQL ADHYQQNTPIGDGPVLLPDNHYLS TQSKLSKDPNEKRDHMVLLEFVT AAGITLGMDELYKLEHHHHHH*
SEQ ID NO: 60	Npuc(A33T)_(NSK)eGFP_His6x		MIKIATRKYLGKQNVYGIGVERDH NFALKNGFI**T**HN**NSK**GEELFTGVD HNFALKNGFINHNVSKGEELFTGV VPILVELDGDVNGHKFSVRGEGEG DATNGKLTLKFICTTGKLPVPWPT LVTTLTYGVQCFSRYPDHMKQHD FFKSAMPEGYVQERTISFKDDGTY KTRAEVKFEGDTLVNRIELKGIDF KEDGNILGHKLEYNFNSHNVYITA DKQENGIKANFKIRHNVEDGSVQL ADHYQQNTPIGDGPVLLPDNHYLS TQSKLSKDPNEKRDHMVLLEFVT AAGITLGMDELYKLEHHHHHH*

SEQ ID NO: 61	Npuc(A33)_(SSK)eGFP_His6x		MIKIATRKYLGKQNVYGIGVERDH NFALKNGFIAHNSSKGEELFTGVD HNFALKNGFINHNVSKGEELFTGV VPILVELDGDVNGHKFSVRGEGEG DATNGKLTLKFICTTGKLPVPWPT LVTTLTYGVQCFSRYPDHMKQHD FFKSAMPEGYVQERTISFKDDGTY KTRAEVKFEGDTLVNRIELKGIDF KEDGNILGHKLEYNFNSHNVYITA DKQENGIKANFKIRHNVEDGSVQL ADHYQQNTPIGDGPVLLPDNHYLS TQSKLSKDPNEKRDHMVLLEFVT AAGITLGMDELYKLEHHHHHH*
SEQ ID NO: 62	Npuc(A33T)_(SSK)eGFP_His6x		MIKIATRKYLGKQNVYGIGVERDH NFALKNGFITHNSSKGEELFTGVD HNFALKNGFINHNVSKGEELFTGV VPILVELDGDVNGHKFSVRGEGEG DATNGKLTLKFICTTGKLPVPWPT LVTTLTYGVQCFSRYPDHMKQHD FFKSAMPEGYVQERTISFKDDGTY KTRAEVKFEGDTLVNRIELKGIDF KEDGNILGHKLEYNFNSHNVYITA DKQENGIKANFKIRHNVEDGSVQL ADHYQQNTPIGDGPVLLPDNHYLS TQSKLSKDPNEKRDHMVLLEFVT AAGITLGMDELYKLEHHHHHH*
SEQ ID NO: 63	Npuc(A33)_(CSK)eGFP_His6x		MIKIATRKYLGKQNVYGIGVERDH NFALKNGFIAHNCSKGEELFTGVD HNFALKNGFINHNVSKGEELFTGV VPILVELDGDVNGHKFSVRGEGEG DATNGKLTLKFICTTGKLPVPWPT LVTTLTYGVQCFSRYPDHMKQHD FFKSAMPEGYVQERTISFKDDGTY KTRAEVKFEGDTLVNRIELKGIDF KEDGNILGHKLEYNFNSHNVYITA DKQENGIKANFKIRHNVEDGSVQL ADHYQQNTPIGDGPVLLPDNHYLS TQSKLSKDPNEKRDHMVLLEFVT AAGITLGMDELYKLEHHHHHH*

SEQ ID NO: 64	Npuc(A33T)_(CSK)eGFP_His6x		MIKIATRKYLGKQNVYGIGVERDH NFALKNGFI**T**HN**CSK**GEELFTGVD HNFALKNGFINHNVSKGEELFTGV VPILVELDGDVNGHKFSVRGEGEG DATNGKLTLKFICTTGKLPVPWPT LVTTLTYGVQCFSRYPDHMKQHD FFKSAMPEGYVQERTISFKDDGTY KTRAEVKFEGDTLVNRIELKGIDF KEDGNILGHKLEYNFNSHNVYITA DKQENGIKANFKIRHNVEDGSVQL ADHYQQNTPIGDGPVLLPDNHYLS TQSKLSKDPNEKRDHMVLLEFVT AAGITLGMDELYKLEHHHHHH*
SEQ ID NO: 65	Npuc(A33)_(ASK)eGFP_His6x		MIKIATRKYLGKQNVYGIGVERDH NFALKNGFI**A**HN**ASK**GEELFTGVD HNFALKNGFINHNVSKGEELFTGV VPILVELDGDVNGHKFSVRGEGEG DATNGKLTLKFICTTGKLPVPWPT LVTTLTYGVQCFSRYPDHMKQHD FFKSAMPEGYVQERTISFKDDGTY KTRAEVKFEGDTLVNRIELKGIDF KEDGNILGHKLEYNFNSHNVYITA DKQENGIKANFKIRHNVEDGSVQL ADHYQQNTPIGDGPVLLPDNHYLS TQSKLSKDPNEKRDHMVLLEFVT AAGITLGMDELYKLEHHHHHH*
SEQ ID NO: 66	Npuc(A33T)_(ASK)eGFP_His6x		MIKIATRKYLGKQNVYGIGVERDH NFALKNGFI**T**HN**ASK**GEELFTGVD HNFALKNGFINHNVSKGEELFTGV VPILVELDGDVNGHKFSVRGEGEG DATNGKLTLKFICTTGKLPVPWPT LVTTLTYGVQCFSRYPDHMKQHD FFKSAMPEGYVQERTISFKDDGTY KTRAEVKFEGDTLVNRIELKGIDF KEDGNILGHKLEYNFNSHNVYITA DKQENGIKANFKIRHNVEDGSVQL ADHYQQNTPIGDGPVLLPDNHYLS TQSKLSKDPNEKRDHMVLLEFVT AAGITLGMDELYKLEHHHHHH*

SEQ ID NO: 67	Npuc(A33)_(GSK)eGFP_His6x		MIKIATRKYLGKQNVYGIGVERDH NFALKNGFI**A**HN**GSK**GEELFTGVD HNFALKNGFINHNVSKGEELFTGV VPILVELDGDVNGHKFSVRGEGEG DATNGKLTLKFICTTGKLPVPWPT LVTTLTYGVQCFSRYPDHMKQHD FFKSAMPEGYVQERTISFKDDGTY KTRAEVKFEGDTLVNRIELKGIDF KEDGNILGHKLEYNFNSHNVYITA DKQENGIKANFKIRHNVEDGSVQL ADHYQQNTPIGDGPVLLPDNHYLS TQSKLSKDPNEKRDHMVLLEFVT AAGITLGMDELYKLEHHHHHH*
SEQ ID NO: 68	Npuc(A33T)_(GSK)eGFP_His6x		MIKIATRKYLGKQNVYGIGVERDH NFALKNGFI**T**HN**GSK**GEELFTGVD HNFALKNGFINHNVSKGEELFTGV VPILVELDGDVNGHKFSVRGEGEG DATNGKLTLKFICTTGKLPVPWPT LVTTLTYGVQCFSRYPDHMKQHD FFKSAMPEGYVQERTISFKDDGTY KTRAEVKFEGDTLVNRIELKGIDF KEDGNILGHKLEYNFNSHNVYITA DKQENGIKANFKIRHNVEDGSVQL ADHYQQNTPIGDGPVLLPDNHYLS TQSKLSKDPNEKRDHMVLLEFVT AAGITLGMDELYKLEHHHHHH*

[00108] Note: by convention, residue numbering of INT_C segment excludes the formylmethionine translation of the start codon, then resumes numbering from the last residue of the INT_N segment.

[00109] Examples of linkers include but are not limited to: (1) poly-asparagine linker consisting of 4 to 15 asparagine residues, and (2) glycine-serine linker, consisting of various combinations and lengths of polypeptides consisting of glycine and serine. One of skill in the art can easily identify and use any linker that will successfully link the CIPS with an affinity tag.

[00110] In one example, the solid support can be a solid chromatographic resin backbone, such as a crosslinked agarose. The term "solid support matrix" or "solid matrix" refers to the solid backbone material of the resin which material contains reactive functionality permitting the covalent attachment of ligand (such as N-terminal intein segments) thereto. The backbone

material can be inorganic (e.g., silica) or organic. When the backbone material is organic, it is preferably a solid polymer and suitable organic polymers are well known in the art. Solid support matrices suitable for use in the resins described herein include, by way of example, cellulose, regenerated cellulose, agarose, silica, coated silica, dextran, polymers (such as polyacrylates, polystyrene, polyacrylamide, polymethacrylamide including commercially available polymers such as Fractogel, Enzacryl, and Azlactone), copolymers (such as copolymers of styrene and divinyl- benzene), mixtures thereof and the like. Also, co-, ter- and higher polymers can be used provided that at least one of the monomers contains or can be derivatized to contain a reactive functionality in the resulting polymer. In an additional embodiment, the solid support matrix can contain ionizable functionality incorporated into the backbone thereof.

[00111] Reactive functionalities of the solid support matrix permitting covalent attachment of the N-terminal intein segments are well known in the art. Such groups include hydroxyl (e.g., Si-OH), carboxyl, thiol, amino, and the like. Conventional chemistry permits use of these functional groups to covalently attach ligands, such as N-terminal intein segments, thereto. Additionally, conventional chemistry permits the inclusion of such groups on the solid support matrix. For example, carboxy groups can be incorporated directly by employing acrylic acid or an ester thereof in the polymerization process. Upon polymerization, carboxyl groups are present if acrylic acid is employed or the polymer can be derivatized to contain carboxyl groups if an acrylate ester is employed.

[00112] Affinity tags can be peptide or protein sequences cloned in frame with protein coding sequences that change the protein's behavior. Affinity tags can be appended to the N- or C-terminus of proteins which can be used in methods of purifying a protein from cells. Cells expressing a peptide comprising an affinity tag can be pelleted, lysed, and the cell lysate applied to a column, resin or other solid support that displays a ligand to the affinity tags. The affinity tag and any fused peptides are bound to the solid support, which can also be washed several times with buffer to eliminate unbound (contaminant) proteins. A protein of interest, if attached to an affinity tag, can be eluted from the solid support via a buffer that causes the affinity tag to dissociate from the ligand resulting in a purified protein, or can be cleaved from the bound affinity tag using a soluble protease. As disclosed herein, the affinity tag is cleaved through the self-cleaving action of the Npuc intein segment in the active intein complex.

[00113] Examples of affinity tags can be found in Kimple et al. Curr Protoc Protein Sci 2004 Sep; Arnau et al. Protein Expr Purif 2006 Jul; 48(1) 1-13; Azarkan et al. J Chromatogr

B Analyt Technol Biomed Life Sci 2007 Apr 15; 849(1-2) 81-90; and Waugh et al. Trends Biotechnol 2005 Jun; 23(6) 316-20, all hereby incorporated by reference in their entirety for their teaching of examples of affinity tags.

[00114] Examples of affinity include, but are not limited to, maltose binding protein, which can bind to immobilized maltose to facilitate purification of the fused target protein; Chitin binding protein, which can bind to immobilized chitin; Glutathione S transferase, which can bind to immobilized chitin; Poly-histidine, which can bind to immobilized chelated metals; FLAG octapeptide, which can bind to immobilized anti-FLAG antibodies.

[00115] Affinity tags can also be used to facilitate the purification of a protein of interest using the disclosed modified peptides through a variety of methods, including, but not limited to, selective precipitation, ion exchange chromatography, binding to precipitation-capable ligands, dialysis (by changing the size and/or charge of the target protein) and other highly selective separation methods.

[00116] In some aspects, affinity tags can be used that do not actually bind to a ligand, but instead either selectively precipitate or act as ligands for immobilized corresponding binding domains. In these instances, the tags are more generally referred to as purification tags. For example, the ELP tag selectively precipitates under specific salt and temperature conditions, allowing fused peptides to be purified by centrifugation. Another example is the antibody Fc domain, which serves as a ligand for immobilized protein A or Protein G-binding domains.

[00117] As disclosed herein, a protein of interest (POI), or target protein can attach to the C-terminal intein segment at its C-terminus. The C-terminal segment can be genetically fused to the protein of interest, for example. Methods of recombinant protein production are known to those of skill in the art. The C-terminal intein segment can comprise modifications when compared to a native *Npu* DnaE C-terminal intein segment. An example of such a modification includes, but is not limited to, a mutation of a highly conserved serine residue to a histidine residue. An example of such a mutation can be found in SEQ ID NO: 10, which also includes a mutation of a highly conserved aspartic acid to glycine. By "highly conserved" is meant identical amino acids or sequences that occur within aligned protein sequences across species.

[00118] The N-terminal intein segment can further comprise a sensitivity-enhancing motif (SEM), which renders the splicing or cleaving activity of the assembled intein complex highly sensitive to extrinsic conditions. This sensitivity-enhancing motif can render the split intein, when assembled (meaning the C-terminal intein segment comprising the protein of interest and the N-terminal intein segment are non-covalently associated, or covalently

linked), more likely to cleave under certain conditions. Therefore, the sensitivity-enhancing motif can render the split intein more sensitive to extrinsic conditions when compared to a native, or naturally occurring, intein. For example, the split intein disclosed herein can be 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100% or more sensitive to extrinsic conditions when compared to a native intein, where the sensitivity is defined as the percent cleavage under non-permissive conditions subtracted from the percent cleavage under permissive conditions. For example, if an intein shows no cleavage at pH 8.5 and 5 hours incubation, but 100% cleavage at pH 6.2 and 5 hours incubation, then the intein would show 100% sensitivity to this pH change under these conditions. Specifically, the *Npu* DnaE mutated intein disclosed herein can be more sensitive, and thus more likely to cleave the protein of interest, when certain conditions are present. These extrinsic conditions can be, for example, pH, temperature, or exposure to a certain compound or element, such as a chelating agent. Cleaving can occur at a greater rate, for example, at a pH of 6.2, when the sensitivity enhancing motif is present on the N-terminal intein segment, as compared to either an N-terminal intein segment which doesn't comprise the sensitivity enhancing motif, or a native or naturally occurring N-terminal intein segment. In another example, sensitivity can occur at a temperature of 0°C to 45°C, when the sensitivity enhancing motif is present on the N-terminal intein segment, as compared to either an N-terminal intein segment which doesn't comprise the sensitivity enhancing motif, or a native or naturally occurring N-terminal intein segment.

[00119] The sensitivity enhancing motif (SEM) can be on the N-terminus of the N-terminal intein segment, for example. The SEM can be reversible. By "reversible" it is meant that the cleaving behavior of the split intein can be altered under a given extrinsic condition. This behavior of the intein, however, is reversible when the extrinsic condition is removed. For example, an intein may cleave or splice when at a pH of 6.2, but may not cleave or splice at a pH of 8.5. The SEMs disclosed herein can be designed such that when appended to, for example, the N-terminus of the N-terminal intein segment, they introduce a sensitivity enhancing element that allows for more precise control of cleavage of the protein of interest. This sensitivity has enabled the successful use of the inteins under conditions that are compatible with commercially relevant cell culture expression platforms.

[00120] In one example, a SEM can comprise the structure: aa1 – aa2 – aa3 – aa4, wherein aa1 is a non-polar amino acid, aa2 is a negatively-charged amino acid, aa3 is a non-polar amino acid and aa4 is a positively charged amino acid. For example, aa1 can be glycine, alanine, isoleucine, leucine, methionine, phenylalanine, proline, tryptophan, or valine; aa2

can be aspartic acid or glutamic acid; aa3 can be glycine, alanine, isoleucine, leucine, methionine, phenylalanine, proline, tryptophan, or valine; and aa4 can be arginine or lysine.

[00121] Also disclosed herein is a SEM, wherein the SEM comprises the structure:aa1 – aa2 – aa3 - aa4 – aa5, wherein aa1 is a non-polar amino acid, aa2 is a negatively-charged amino acid, aa3 is a non-polar amino acid, aa4 is a positively charged amino acid and aa5 is a non-polar amino acid. For example, aa1 can be glycine, alanine, isoleucine, leucine, methionine, phenylalanine, proline, tryptophan, or valine; aa2 can be aspartic acid or glutamic acid or cysteine; aa3 can be glycine, alanine, isoleucine, leucine, methionine, phenylalanine, proline, tryptophan, or valine; aa4 can be arginine or lysine; and aa5 can be glycine, alanine, isoleucine, leucine, methionine, phenylalanine, proline, tryptophan, or valine.

[00122] Disclosed herein are SEMS, wherein the reversible SEM comprises the sequence G-E-G-H-H (SEQ ID NO: 20), G-E-G-H-G (SEQ ID NO: 21), G-D-G-H-H (SEQ ID NO: 22), or G-D-G-H-G (SEQ ID NO: 23).

[00123] Disclosed herein is a method of purifying a protein of interest, the method comprising: utilizing a split intein comprising two separate peptides: an N-terminal intein segment and a C-terminal intein segment; immobilizing the N-terminal intein segment to a solid support; attaching a protein of interest to the C-terminal intein segment, wherein cleaving of the C-terminal intein segment is highly sensitive to extrinsic conditions when compared to a native intein; exposing the N-terminal intein segment and the C-terminal intein segment to each other so that they associate; washing the solid support to remove non-bound material; placing the associated the N-terminal intein segment and the C-terminal intein segment under conditions that allow for the intein to self-cleave; and isolating the protein of interest.

[00124] Table 2 shows a list of amino acids, their abbreviations, polarity, and charge.

Table 2: Amino Acids

Amino Acid	3-Letter Code	1-Letter Code	Polarity	Charge
Alanine	Ala	A	nonpolar	neutral
Arginine	Arg	R	Basic polar	positive
Asparagine	Asn	N	polar	neutral
Aspartic acid	Asp	D	acidic polar	negative
Cysteine	Cys	C	nonpolar	neutral
Glutamic acid	Glu	E	acidic	negative

			polar	
Glutamine	Gln	Q	polar	neutral
Glycine	Gly	G	nonpolar	neutral
Histidine	His	H	Basic polar	Positive (10%) Neutral (90%)
Isoleucine	Ile	I	nonpolar	neutral
Leucine	Leu	L	nonpolar	neutral
Lysine	Lys	K	Basic polar	positive
Methionine	Met	M	nonpolar	neutral
Phenylalanine	Phe	F	nonpolar	neutral
Proline	Pro	P	nonpolar	neutral
Serine	Ser	S	polar	neutral
Threonine	Thr	T	polar	neutral
Tryptophan	Trp	W	nonpolar	neutral
Tyrosine	Tyr	Y	polar	neutral
Valine	Val	V	nonpolar	neutral

[00125] Disclosed herein are vectors comprising nucleic acids encoding the C-terminal intein segment as disclosed herein, as well as cell lines comprising said vectors. As used herein, plasmid or viral vectors are agents that transport the disclosed nucleic acids, such as those encoding a C-terminal intein segment and a peptide of interest, into a cell without degradation and include a promoter yielding expression of the gene in the cells into which they can be delivered. In one example, a C-terminal intein segment and peptide of interest are derived from either a virus or a retrovirus. Retroviral vectors are able to carry a larger genetic payload, i.e., a transgene or marker gene, than other viral vectors and for this reason are a commonly used vector. However, they are not as useful in non-proliferating cells. Adenovirus vectors are relatively stable and easy to work with, have high titers, and can be delivered in aerosol formulation, and can transfect non-dividing cells. Pox viral vectors are large and have several sites for inserting genes; they are thermostable and can be stored at room temperature. Disclosed herein is a viral vector which has been engineered so as to suppress the immune response of a host organism, elicited by the viral antigens. Vectors of this type can carry coding regions for Interleukin 8 or 10.

[00126] Viral vectors can have higher transfection (ability to introduce genes) abilities than chemical or physical methods to introduce genes into cells. Typically, viral vectors contain nonstructural early genes, structural late genes, an RNA polymerase III transcript, inverted terminal repeats necessary for replication and encapsidation, and promoters to

control the transcription and replication of the viral genome. When engineered as vectors, viruses typically have one or more of the early genes removed and a gene or gene/promoter cassette is inserted into the viral genome in place of the removed viral DNA. Constructs of this type can carry up to about 8 kb of foreign genetic material. The necessary functions of the removed early genes are typically supplied by cell lines which have been engineered to express the gene products of the early genes in trans.

[00127] The fusion DNA encoding a modified peptide can be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted protein-coding sequence. Depending on the host-vector system utilized, any one of a number of suitable transcription and translation elements can be used. For instance, when expressing a modified eukaryotic protein, it can be advantageous to use appropriate eukaryotic vectors and host cells. Expression of the fusion DNA results in the production of a modified protein.

[00128] Also disclosed herein are cell lines comprising the vectors or peptides disclosed herein. A variety of cells can be used with the vectors and plasmids disclosed herein. Non-limiting examples of such cells include somatic cells such as blood cells (erythrocytes and leukocytes), endothelial cells, epithelial cells, neuronal cells (from the central or peripheral nervous systems), muscle cells (including myocytes and myoblasts from skeletal, smooth or cardiac muscle), connective tissue cells (including fibroblasts, adipocytes, chondrocytes, chondroblasts, osteocytes and osteoblasts) and other stromal cells (e.g., macrophages, dendritic cells, thymic nurse cells, Schwann cells, etc.). Eukaryotic germ cells (spermatocytes and oocytes) can also be used, as can the progenitors, precursors and stem cells that give rise to the above-described somatic and germ cells. These cells, tissues and organs can be normal, or they can be pathological such as those involved in diseases or physical disorders, including, but not limited to, infectious diseases (caused by bacteria, fungi yeast, viruses (including HIV, or parasites); in genetic or biochemical pathologies (e.g., cystic fibrosis, hemophilia, Alzheimer's disease, schizophrenia, muscular dystrophy, multiple sclerosis, etc.); or in carcinogenesis and other cancer-related processes.

[00129] The eukaryotic cell lines disclosed herein can be animal cells, plant cells (monocot or dicot plants) or fungal cells, such as yeast. Animal cells include those of vertebrate or invertebrate origin. Vertebrate cells, especially mammalian cells (including, but not limited to, cells obtained or derived from human, simian or other non-human primate, mouse, rat, avian, bovine, porcine, ovine, canine, feline and the like), avian cells, fish cells (including zebrafish cells), insect cells (including, but not limited to, cells obtained or derived

from Drosophila species, from Spodoptera species (e.g., Sf9 obtained or derived from S. frugiperida, or HIGH FIVE™ cells) or from Trichoplusa species (e.g., MG1, derived from T. ni)), worm cells (e.g., those obtained or derived from C. elegans), and the like. It will be appreciated by one of skill in the art, however, that cells from any species besides those specifically disclosed herein can be advantageously used in accordance with the vectors, plasmids, and methods disclosed herein, without the need for undue experimentation.

[00130] Examples of useful cell lines include, but are not limited to, HT1080 cells (ATCC CCL 121), HeLa cells and derivatives of HeLa cells (ATCC CCL 2, 2.1 and 2.2), MCF-7 breast cancer cells (ATCC BTH 22), K-562 leukemia cells (ATCC CCL 243), KB carcinoma cells (ATCC CCL 17), 2780AD ovarian carcinoma cells (see Van der Buick, A. M. et al., Cancer Res. 48:5927-5932 (1988), Raji cells (ATCC CCL 86), Jurkat cells (ATCC TIB 152), Namalwa cells (ATCC CRL 1432), HL-60 cells (ATCC CCL 240), Daudi cells (ATCC CCL 213), RPMI 8226 cells (ATCC CCL 155), U-937 cells (ATCC CRL 1593), Bowes Melanoma cells (ATCC CRL 9607), WI-38VA13 subline 2R4 cells (ATCC CLL 75.1), and MOLT-4 cells (ATCC CRL 1582), as well as heterohybridoma cells produced by fusion of human cells and cells of another species. Secondary human fibroblast strains, such as WI-38 (ATCC CCL 75) and MRC-5 (ATCC CCL 171 can also be used. Other mammalian cells and cell lines can be used in accordance with the present invention, including, but not limited to CHO cells, COS cells, VERO cells, 293 cells, PER-C6 cells, M1 cells, NS-1 cells, COS-7 cells, MDBK cells, MDCK cells, MRC-5 cells, WI-38 cells, WEHI cells, SP2/0 cells, BHK cells (including BHK-21 cells); these and other cells and cell lines are available commercially, for example from the American Type Culture Collection (P.O. Box 1549, Manassas, Va. 20108 USA). Many other cell lines are known in the art and will be familiar to the ordinarily skilled artisan; such cell lines therefore can be used equally well.

[00131] Once obtained, the proteins of interest can be separated and purified by appropriate combination of known techniques. These methods include, for example, methods utilizing solubility such as salt precipitation and solvent precipitation; methods utilizing the difference in molecular weight such as dialysis, ultra-filtration, gel-filtration, and SDS-polyacrylamide gel electrophoresis; methods utilizing a difference in electrical charge such as ion-exchange column chromatography; methods utilizing specific affinity such as affinity chromatography; methods utilizing a difference in hydrophobicity such as reverse-phase high performance liquid chromatography; and methods utilizing a difference in isoelectric point, such as isoelectric focusing electrophoresis. These are discussed in more detail below.

[00132] Disclosed is a method of purifying a protein of interest, the method comprising: utilizing a split intein comprising two separate peptides: an N-terminal intein segment and a C-terminal intein segment; immobilizing the N-terminal intein segment to a solid support; attaching a protein of interest to the C-terminal intein segment, wherein cleaving of the C-terminal intein segment is highly sensitive to extrinsic conditions when compared to a native intein. exposing the N-terminal intein segment and the C-terminal intein segment to each other so that they associate; washing the solid support to remove non-bound material; placing the associated the N-terminal intein segment and the C-terminal intein segment under conditions that allow for the intein to self-cleave; and isolating the protein of interest.

[00133] Also disclosed herein are kits. A kit, for example, can include the split intein disclosed herein (an N-terminal intein segment and a C-terminal intein segment) as well as a protein of interest, optionally. The kit can also include instructions for use.

C. Experimental

[00134] The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how the compounds, compositions, articles, devices and/or methods claimed herein are made and evaluated, and are intended to be purely exemplary of the invention and are not intended to limit the scope of what the inventors regard as their invention. Efforts have been made to ensure accuracy with respect to numbers (*e.g.*, amounts, temperature, etc.), but some errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, temperature is in °C or is at ambient temperature, and pressure is at or near atmospheric.

Example 1: Modifications of the Npu split intein for accelerated C-terminal cleavage

[00135] To guide the modification of the active site of the intein, the Npu split intein protein sequence and structure (pdb 4QFQ) were aligned with *Mycobacterium smegmatis* DnaBi class 3 intein structure (pdb 6OWN) in Pymol software. Emphasis was put on the highly conserved WCT motif observed in class 3 inteins, but not in class 1 inteins (Figures 5-6). Intein structures are highly conserved across all classes of inteins, which was further confirmed by aligning other structures from the class 1 inteins like the *Mycobacterium tuberculosis* RecA contiguous mini intein (pdb 2L8L) and the cyanophage P-SSM2 gp41-1 split intein (pdb 6QAZ). Thus, the alignment showed overlapping of the conserved W-C-T motif residues from class 3 inteins over the class 1 Npu split intein amino acid positions. In this case, the W overlaps with F74 in the Npu_N intein fragment, while C and T overlap with positions D16 and A33 in Npuc intein fragment (Figure 7), respectively. Because the C

residue overlaps with previous D16G modification in the Npuc segment for accelerated C-terminal cleavage, and since C may introduce a new nucleophile which can result in unwanted N-terminal cleavage, the C mutation was removed from further experimentation and left as G. To show the influence of W-T residues, the native F74 and A33 of the Npu intein were modified in a factorial design to show the individual and combined effects of the Npu intein variants (Figure 9). Both F74 and A33 are in the vicinity with catalytic H72 of Npu, and thus modifying these residues may have a direct impact on the intein's cleaving activity.

[00136] To assess if the amino acid modifications could have a major impact on Npu intein structure, the amino acid sequences were modeled using the AlphaFold2 tool (M. M. Schütze et al., Nature Methods, 2022). The predicted model showed structure conservation with over 90% confidence (Figure 20). This prediction implies that no major folding changes of global structure would be generated by the proposed modifications and thus support the conservation of intein activity. Additionally, the amino acids in the active site of the generated intein structure were modeled and evaluated in Pymol software to predict possible intermolecular interactions with other amino acids in proximity. Figure 21 shows possible pi-cation interactions between F74W and the Npu_N catalytic residue H72, while A33T is predicted to have hydrogen bonding with the Npu_N H72 and K73 residues. This suggests that these modifications have a higher chance of affecting the rate of the self-cleaving reaction of Npu.

[00137] Results in Figure 11 suggest that individual mutants would decrease the rate of intein self-cleavage except for the single mutant A33T, which shows ~1.5x faster cleavage than the control (Figure 14). Furthermore, in the factorial design, F74W was combined with different amino acids of similar size such as A33T, A33V, and A33S. Valine is a small hydrophobic amino acid just like the native alanine residue in Npuc. Threonine is a beta-branch such as valine but has the hydroxyl polar group (-OH), whereas serine has a similar size but lacks a methyl group (-CH3). Results show that the F74W, A33V double mutant behaves similarly to F74W, A33, but the F74W, A33S double mutant has a ~1.5x slower self-cleaving rate compared to the control (Figures 12-13). On the contrary, the F74W, A33T variant displayed ~2x-faster cleaving rates compared to the control (Figure 12-14). Therefore, the dual mutant F74W, A33T was selected for further characterization of pH controllability. The experiment was then repeated, but with the elution buffer held at pH 8.5 to determine the rate of intein cleavage at the higher pH. The results in Figure 15 show that the dual mutant continues to display significant pH sensitivity, with no noticeable intein cleavage in the first

five hours. The observed cleavage is comparable to that of the control (F74, A33) under these conditions, indicating that pH controllability is preserved in the F74W, A33T dual mutant.

[00138] Experiments were repeated with eGFP as the model target protein. In this case, more mutants were tested to expand the understanding of size and polarity in these positions: F74W/Y/F/V and A33V/S/C/T/N/L (See Figure 23). On one hand, the mutations on the F74 with tyrosine was used to emulate size and properties with an isosteric residue, while the valine was used as a contrast in size better to understand the effect of size in this position. On the other hand, the A33 had more mutations to substitute this position with polar properties such as cysteine and serine. At the same time, the asparagine and leucine residues were used to probe the combinatorial effect of sidechain size and polarity. The results in Figure 2 show that single mutants F74Y and F74V drastically decreased cleaving kinetics, while F74W had a comparable performance to native F74. This suggests that although isosteric residues like tyrosine can fill the cavity, a polar hydroxyl group can disrupt the catalytic activity. Similarly, the A33N and A33L, regardless of their respective polarity, both displayed loss in activity, indicating that this position requires a smaller size residue to form the active site. On the contrary, the A33V/T/S/C all showed faster cleavage, particularly the small polar residues. This implies that the polar groups can participate in hydrogen bonding to better coordinate catalytic residues for the cleavage reaction. Thus, this evidence supports that Npuc A33 can be substituted with small polar residues like serine, cysteine, and threonine for accelerated cleavage phenotypes.

[00139] When evaluating the double mutants, F74W,A33N and F74W,A33L both showed complete activity loss, supporting the evidence observed on the single mutants. The F74W,A33C and F74,A33S both performed comparably to their respective single mutant A33C and A33S. The double mutant F74W,A33T once again displayed faster cleaving kinetics (see Figure 23). This confirms the previously observed trend with the experiments using TrxA model protein (shown in Figures 12-14). Threonine has an additional methyl group in the beta carbon, which could allow for more van der Waal interactions with neighbor residue F74W. The methyl could potentially help to maximize the hydrophobic interactions in the core for better structural organization of the polar OH group. Residues like cysteine and serine do not have that methyl group, reducing the odds of steering the polar group in the optimal orientation as threonine would in combination with F74W as shown in Figure 20.

Methods:

DNA plasmid construction for Npu split intein

[00140] A C-terminal (GGGS) flexible linker followed by a chitin-binding domain (CBD) tag was fused to the C-terminus of the MGDGHG-Npu$_N$ fragment via Gibson assembly in a pET Npu$_N$ expression vector. The CBD tag allows affinity capture of the Npu$_N$ intein fragment and allows evaluation of on-column intein self-cleavage. Thioredoxin A (TrxA, with the N-terminal +1,+2,+3 sequence MSD) protein was attached to the C-terminus of Npu$_C$ through Gibson assembly in a pET vector and was used as a model protein to assess the intein self-cleaving reaction. The generated clones were subsequently modified through site-directed mutagenesis with inverse PCR ligations to generate split intein variants: pET_M-GDGHG-Npu$_N$(F74W)-GGGS-CBD (SEQ ID NO: 30), and pET_M-Npu$_C$(A33T, S, or V)_TrxA_strepII_His6x (SEQ ID NOS: 22-24). The constructs with eGFP model protein were also cloned through Gibson Assembly to generate pET_M-Npu$_C$(A33T, S, C, N, or L)_(VSK)eGFP_His6x and pET_M-GDGHG-Npu$_N$(F74W, Y, or V)-GGGS-CBD (see sequences) to confirm and expand more mutants in the Npu positions mentioned above (see Figure 21).

Recombinant protein expression in *E. coli*

[00141] The individual Npu split intein DNA clones were transformed into competent *Escherichia coli* strain BLR (DE3) (F-ompT hsdSB(rB-,mB-) gal dcm (DE3)) bacteria. Transformants were cultured in 5 mL of Luria broth media (10 g NaCl, 5 g Yeast extract, 10 g Tryptone per liter) supplemented with 100 µg/mL ampicillin, shaking vigorously at 200 rpm in 37°C for 16-18 hrs. Overnight cultures were then diluted at a ratio of 1:100 (vol./vol.) into 2x Luria Broth media (10 g NaCl, 10 g Yeast extract, 20 g Tryptone per liter) supplemented with 100 µL ampicillin in 500 mL Ultrayield™ flasks and shaken at 320 rpm in 37°C until reaching OD_{600} of 2.0-4.0. The cultures were then induced with 0.5 mM (final concentration) of isopropyl ß-D-1-thiogactopyranoside (IPTG) for 20 hrs. at 16°C for protein expression.

Recombinant protein purification and self-cleaving kinetic rate characterization

[00142] The cell culture volumes were harvested in a 1:1 volume ratio with the respective variant combinations. The resulting combinations of all pellets were resuspended and lysed together in 1/10 of the original culture vol. with column buffer (20 mM AMPD + 20 mM PIPES + 200 mM NaCl at pH 8.5) to promote intein association. Lysates were centrifuged at

max velocity for 10 min at 4°C to remove major host cell insoluble impurities. The assembled intein complexes in clarified lysates were captured on 500 µL of NEB chitin affinity resin 50% slurry and washed with pH 8.5 buffer (20 mM AMPD + 20 mM PIPES + 500 mM NaCl + 0.2% vol./vol. Tween20 at pH 8.5) in a 10 mL Poly-Prep® chromatography column. The wash was followed with a quick wash (W2) with elution buffer (20 mM AMPD + 20 mM PIPES + 200 mM NaCl at pH 6.2) to shift the buffer conditions to a mildly acidic pH (Figure 10). All the capture and washing steps were carried out at 4°C to suppress premature intein activity. The resin was evenly resuspended in 1 Column volume (CV) of elution buffer, and a sample was taken and mixed at a 1:1 vol./vol. ratio with of 2x-SDS loading dye and boiled at 95° C for 5 min to stop the reaction. This sample is considered the time point zero sample of the reaction. The columns were then brought to room temperature to proceed with the self-cleavage, and resin samples were taken at 1 hr., 3 hrs., 5 hrs., and 24 hrs. as previously described to monitor the intein self-cleaving reaction (Figure 10). The experiment to test the eGFP model protein with more mutants was performed as described in this section (See Figure 22).

Example 2: Characterization of the effects of the target protein leader sequence over the intein cleaving rate.

[00143] To understand the impact of the F74W,A33T dual mutant on the previously described dependency of intein cleavage rate on target protein leader sequence (Figure 16), the first three amino acids of TrxA were modified. The Npu intein has exceptional cleavage performance, particularly when the first two amino acids are bulkier and hydrophobic, while smaller and polar/charged residues are correlated with poor cleavage (Figure 16). To study the effects of polar/charged and smaller amino acid leader sequences, the first three native amino acids of TrxA (MSD) were modified to DIG and EVQ through site-directed mutagenesis. Cleavage rates of the resulting variants were analyzed and compared with the control Npu intein (F74,A33) variant, as described in Example 1. The goal of this experiment was to confirm if the observed intein self-cleaving acceleration extends to known slow-cleaving amino acid leader sequences. This experiment was repeated as described below but with eGFP as the model protein with different first amino acid (+1) identities to explore the impact of the single mutant A33T and double mutant F74W,A33T. The selected amino acids represent different properties: E as acidic, K as basic, A as small hydrophobic, G as small flexible, P as rigid, and S, T, and C as small polars. N, M, and F were used respectfully as larger amino acids for bulky polar, bulky aliphatic, and bulky aromatic.

[00144] The results of this experiment showed that the control intein variant was cleaving slowly as expected, with less than 20% cleavage observed at pH 6.2 in the first five hours of the reaction for both the EVQ and DIG leader sequences. On the contrary, the F74W,A33T dual mutant exhibited dramatically accelerated self-cleavage. Analysis of the cleaving rate indicated that cleavage was accelerated by ~19x and ~25x in EVQ and DIG, respectively, when compared to the control (Figure 18). Furthermore, the EVQ, DIG, and MSD leader sequences all display more than 50% of precursor protein cleaved at the 3 hr time point (Figure 19), which supports the previously observed accelerated cleavage trend. This confirms a lower dependency on the target protein leader sequence and offers better performance for more proteins with slow-cleaving amino acids.

[00145] In the +1 amino acid study, the A33T single mutant consistently showed faster cleaving kinetics for all the different +1 residues when compared to their respective control (See Figure 20). However, the overall property trend was comparable to the control, where bulkier non-polar residues tend to cleave faster than smaller and polar residues. One interesting observation was made with glutamic acid, which reached over 90% cleavage within the first hour, while its control reached about 40% cleavage, confirming rate acceleration. Another remarkable residue was alanine that, within the first hour, reached about 50% cleavage, while the control reached about 20% cleavage in the first four hours. Surprisingly, glycine reached over 90% cleavage within four while its control reported approximately 50% cleavage in the same timepoint. Lysine demonstrated noticeable improvement as it reached about 55% in the first four hours, while its control only reached about 15% by that time. The polar neutral residues like N, S, and C all had comparable cleavage to their respective control. In contrast, the double mutant F74W,A33T displayed variable results. Phenylalanine was too fast to appreciate cleavage by the zero hour time point, and methionine also showed faster cleavage, reaching almost 100% cleavage in the first hour. Polar resides neutral and polar charged behave very similarly to their respective and controls. One exception was alanine, which reached about 50% cleaved within the first hour, emulating the cleavage rate for single mutant A33 with +1 methionine. Overall, the single mutant A33T showed more consistent improvement across different +1 residues (Figure 8).

Methods:

DNA plasmid construction for Npu split intein

[00146] The first three amino acids of native TrxA (MSD) in pET_M-Npuc(A33T)_(MSD)TrxA_strepII_His6x (SEQ ID NO: 33) were modified through site-directed mutagenesis with inverse PCR ligations to generate split intein variants: pET_M-Npuc(A33T)_(EVQ)TrxA_strepII_His6x (SEQ ID NO: 38), and pET_M-Npuc(A33T)_(DIG)TrxA_strepII_His6x (see new sequences below). For the +1 amino acid study, the pET_M-NpuC(A33T)_X^{+1}-eGFP_His6x and pET_M-NpuC(A33)_X^{+1}-eGFP_His6x were also generated through inverse PCR cloning (X^{+1}: F, M, E, K, N, C, S, A, or G).

Recombinant protein expression in *E. coli*

[00147] The individual Npu split intein DNA clones were transformed in competent *Escherichia coli* strain BLR (DE3) (F-ompT hsdSB(rB-,mB-) gal dcm (DE3)) bacteria. Transformants were cultured in 5 mL of Luria broth media (10 g NaCl, 5 g Yeast extract, 10 g Tryptone per liter) supplemented with 100 µg/mL ampicillin, shaking vigorously at 200 rpm in 37°C for 16-18 hr. Overnight cultures were then diluted at a ratio of 1:100 (vol./vol.) into 2x Luria Broth media (10 g NaCl, 10 g Yeast extract, 20 g Tryptone per liter) supplemented with 100 µg/mL ampicillin in 500 mL Ultrayield™ flasks and shaken at 320 rpm in 37°C until reaching an OD_{600} of 2.0-4.0. The cultures were then induced with 0.5 mM (final concentration) of isopropyl ß-D-1-thiogactopyranoside (IPTG) for 20 hrs. at 16°C for protein expression.

Recombinant protein purification and self-cleaving kinetic rate characterization

[00148] The cell culture volumes were harvested in a 1:1 volume ratio with the respective variant combinations. The resulting combinations of cell pellets were resuspended and lysed together in 1/10 of the original culture volume with column buffer (20 mM AMPD + 20 mM PIPES + 200 mM NaCl at pH 8.5) to promote intein association. Lysates were centrifuged at maximum velocity for 10 min at 4° C to remove insoluble impurities. The assembled intein complexes in clarified lysates were captured and washed (20 mM AMPD + 20 mM PIPES + 500 mM NaCl + 0.2% vol./vol. Tween20 at pH 8.5) on 500 µL of NEB chitin affinity resin 50% slurry at pH 8.5 buffer in a 10 mL Poly-Prep® chromatography columns. Then, a second quick wash (W2) was performed with elution buffer (20 mM AMPD + 20 mM PIPES + 200 mM NaCl at pH 6.2) to provide a mildly acidic pH (Figure 10). All the capture and washing steps were carried out at 4°C to suppress premature intein activity. The resin was then evenly resuspended in 1 CV of elution buffer, and a sample was taken and mixed at a 1:1 vol./vol. ratio with 2x-SDS loading dye and boiled at 95°C for 5 min to stop the reaction. This sample is considered the time point zero of the reaction. The columns were brought to

room temperature to proceed with the self-cleavage, and resin samples were taken at 1 hr., 3 hrs., 5 hrs., and 24 hrs (for the +1 amino acid study 0 hr., 0.5 hr., 1 hr., 2 hrs., 4 hrs. resin timepoint samples were taken). as previously described to monitor the intein self-cleaving reaction (Figure 10).

Example 3: Covalent immobilization of Npu_N variants (cysteine-free)

[00149] To evaluate the self-cleaving performance of the Npu dual mutant in the context of a self-cleaving affinity chromatography method, the Npu_N(F74W) variant was immobilized on a solid stationary phase. To ensure that Npu_N intein has efficient and single-point thiol conjugation on a unique immobilization point, all cysteine residues (C1, C29, and C60) were modified to alanine or serine, and a unique C residue was introduced at the Npu_N C-termini. In this case, a bicistronic pET DNA vector was used to maximize the stability and titer expression of Npu_N ligand variants (PCT/U52021/030161, herein incorporated by reference in its entirety).

[00150] The Npu_N(F74W) variant was conjugated onto an Avitide proprietary bromoacetic agarose resin through a thiol-ether bond (see methods section below, Figure 25). The NpuN ligand was quantified using a spectrophotometer and the concentration was adjusted to ~20 mg/mL. To evaluate the efficacy of conjugation, a sample of control Npu_N ligand (F74) was immobilized at a similar concentration. The ligand immobilization efficiency was determined by subtracting the final flowthrough and washing the calculated protein from the initially calculated protein. The average amount of immobilized Npu_N on bromoacetyl activated resin control F74 and F74W mutant was 13.99 mg/mL and 14.04 mg/mL, respectively (Figure 26). The ANOVA Tukey-Krammer all means comparison analysis shows that there is no significant difference between the control and the variant F74W, thus confirming that the F74W mutation does not impact the coupling reaction efficiency (See Figure 26).

[00151] To further test and compare the self-cleaving performance for single mutant A33T and double mutant F74W,A33T against the control with the generated resin variants, these were tested with eGFP model protein to under pH 6.2 and pH 8.5 (as described in the methods section below). Figure 27 showed that the single mutant F74W performed comparably with the control F74 in both acidic and basic pH. This confirms that F74W does not change much the activity as observed in the previous example. The single mutant A33T at pH 8.5 showed negligible cleavage by the first four hours of the reaction, similar to the control. In this case, it was concluded that the single mutant A33T preserves pH controllability, but it can accelerate the cleavage when compared to the control. Furthermore, the double mutant F74W,A33T again

shows about twice more cleavage in the first four hours, however it remains in negligible margins as only 8% of captured protein cleaved. The pH experiment proves that the single mutant A33T and double mutants F74W, A33T preserve the overall pH controllability feature, as it notably suppresses activity in the first four hours at pH 8.5 at room temperature. This enables the new variants to be used in downstream applications as it can protein capture and separate impurity without significant losses in the process.

[00152] Moreover, to evaluate the effects of temperature over intein's self-cleavage activity, the experiments were repeated in pH 6.2 in 4 °C versus room temperature as reference. Figure 28 shows a comparison of samples at 6.2 in cold temperature (4°C). The single mutant F74W cleaves slower than the control, as the control was able to cleave about 28% in 24 hrs., while F74W gets less than 5% cleaved. In contrast, single mutant A33T and double mutant F74W,A33 both showed accelerated cleavage when compared to the control. While it seems that both have more than 80% cleavage in the course of 24 hrs., the single mutant A33T reached more than 95% cleavage, thus suggesting that the single mutant A33T outperforms all the other variants in cold temperature. This experiment shows the first example of cleaving mutant intein that can cleave in colder temperatures, potentially enabling downstream applications for temperature-sensitive proteins. This additional evidence supports the combination of colder temperature and more alkaline pH conditions to effectively suppress premature self-cleavage for single mutant A33T, further providing controllability over the intein system, and thereby the purification process.

Methods:

DNA plasmid construction for Npu split intein

[00153] The bicistronic construct pET_MGDGHG-Npu$_N$-GGGS-HHHHHH-C_stop_RBS_M-Npu$_C$(HA) (SEQ ID NO: 20; PCT/U52021/030161) was modified through site-directed mutagenesis with inverse PCR ligations to generate variant: pET_MGDGHG-Npu$_N$(F74W)-GGGS-HHHHHH-C_stop_RBS_M-Npu$_C$(AHA) (SEQ ID NO: 6). For pET_Npu$_C$(A33T)_(VSK)eGFP_His6x (SEQ ID NO: 19), the variant was generated with inverse PCR ligation from pET_Npu$_C$(A33)_(VSK)eGFP_His6x (SEQ ID NO: 49)

Recombinant protein expression in *E. coli*

[00154] The individual Npu split intein DNA clones were transformed in competent *Escherichia coli* strain BLR (DE3) (F-ompT hsdSB(rB-,mB-) gal dcm (DE3)) bacteria. Transformants were cultured in 5 mL of Luria broth media (24 g Yeast extract, 20 g Tryptone, 4 mL glycerol, 17 mM KH_2PO_4, 72 mM K_2HPO_4 per liter) supplemented with 100

ug/mL ampicillin, shaking vigorously at 200 rpm in 37° C for 16-18 hrs. Overnight cultures were then diluted at a ratio of 1:100 (vol./vol.) into 500 mL of Terrific Broth media (10 g NaCl, 10 g Yeast extract, 20 g Tryptone per liter) supplemented with 100 ug/mL ampicillin in 2.5 L Ultrayield™ flasks and shaken at 320 rpm in 37° C until reaching OD_{600} of 2.0-4.0. The cultures were then induced with 0.5 mM (final concentration) of isopropyl β-D-1-thiogactopyranoside (IPTG) for 20 hrs. at 16° C for protein expression.

Npu Ligand purification and sample preparation

[00155] The cell culture volumes were harvested, resuspended, and lysed in 1/5 of the original culture vol. with column buffer (25 mM MOPS + 10 mM imidazole + 15 mM NaCl at pH 7.5) in a homogenizer at 4° C. Lysates were centrifuged at max velocity for 15 min at 4° C to remove insoluble impurities. The Npu variants were purified using two Cytiva Ni-NTA prepacked HisTrap® FF 5 mL FPLC cartridges (cat. No. 17525501) connected in tandem on an Äkta pure 25 FPLC system. Samples were loaded at a flow rate of 2.0 mL/min., washed with 5 CVs of IMAC column buffer, and then eluted in a linear gradient mode with 6 CVs of elution buffer (25 mM MOPS + 200 mM imidazole + 15 mM NaCl at pH 8.0). Elution fractions were pooled and loaded on a pre-packed DEAE Q-Sepharose 3.15 mL anion exchange (AEX) column and washed with AEX column buffer (25 mM MOPS, 15 mM NaCl at pH 7.5) as a polishing purification step. The Npu_N ligands were then eluted with a linear gradient of AEX elution buffer (25 mM MOPS, 500 mM NaCl at pH 7.5). Elutions were then buffer exchanged into the coupling buffer (100 mM Boric acid, 25 mM sodium tetraborate, 150 mM NaCl at pH 8.5) and concentrated in a Sartorius Vivaspin® Turbo 10kDa MWCO concentrator tube.

Immobilization of Npu ligands

[00156] The Npu was conjugated through reaction with a bromoacetyl functional group on an agarose bead (proprietary bromoacetic activated coupling resin, Avitide) with the unique thiol on the Npu_N C-terminal cysteine. The bromoacetyl group is very reactive with reduced thiols. To synthesize 1 mL of Npu_N resin variants, concentrated Npu_N ligands were diluted into 1 mL of coupling buffer at a concentration of ~20 mg/mL. The diluted ligand samples were supplemented with a final concentration of 25mM Tris(2-carboxyethyl) phosphate (TCEP, Thermo product, cat. No. 77720) to reduce all the terminal cysteine thiols on Npu_N. Then, 2mL of bromoacetic resin 50% slurry was transferred into 10 mL Poly-Prep® chromatography columns and allowed to settle. The resin bead was then equilibrated with 10 CVs of coupling buffer, then 1 mL of Npu_N(F74) (seq. Id no. 20) and Npu_N(F74W) (seq Id no. 6) ligand variants were added respectably in each resin bead. The coupling reaction was

performed in the sealed gravity columns, mixing gently at 30 °C covered from the light with tin foil for 16 hrs. After the conjugation, sample flowthrough was collected, and the resin was washed with 10 CVs of coupling buffer to remove residual unreacted Npu_N. The conjugated resins were then incubated in a blocking buffer (50 mM L-Cystein-HCl at pH 8.5) to block the unreacted bromoacetic groups on the agarose bead. Lastly, 1 M NaCl was used to wash out any residual contaminants on the resin, then was stored in a 20% ethanol solution at 4° C. The conjugation for each ligand variant was performed in triplicates.

pH and temperature effects study

[00157] The Npu_C_(VSK)eGFP and Npu_C(A33T)_(VSK)eGFP cell pellets were resuspended and lysed in 1/10 of the original culture volume with column buffer (20 mM AMPD + 20 mM PIPES + 200 mM NaCl at pH 8.5) to promote intein association. Lysates were centrifuged at maximum velocity for 10 min at 4 °C to remove insoluble impurities. The assembled intein complexes in clarified lysates were captured and washed (20 mM AMPD + 20 mM PIPES + 500 mM NaCl + 0.2% vol./vol. Tween20 at pH 8.5) on 200 µL of generated resin at pH 8.5 buffer in 10 mL Poly-Prep® chromatography columns. Then, a second quick wash (W2) was performed with elution buffer (20 mM AMPD + 20 mM PIPES + 200 mM NaCl at pH 6.2) to provide a mildly acidic pH for control and for higher pH 8.5 (Figure 27). These were prepared in All the capture and washing steps were carried out at 4°C to suppress premature intein activity. The resin was then evenly resuspended in 1 CV of elution buffer, and a sample was taken and mixed at a 1:1 vol./vol. ratio with 2x-SDS loading dye and boiled at 95°C for 5 min to stop the reaction. This sample is considered the time point zero of the reaction. The pH 6.2 and 8.5 were brought to room temperature to proceed with the self-cleavage, one duplicate set of the pH 6.2 columns was kept in at 4 °C, and resin samples were taken at 0.5 hr., 1 hr., 2 hrs., 4 hrs. and 24 hrs. (for cold temperature only 0 hr., 2 hr., 4 hrs., and 24 hrs.) as previously described to monitor the intein self-cleaving reaction (Figure 10).

[00158] It will be apparent to those skilled in the art that various modifications and variations can be made in the present invention without departing from the scope or spirit of the invention. Other embodiments of the invention will be apparent to those skilled in the art from consideration of the specification and practice of the invention disclosed herein. It is intended that the specification and examples be considered as exemplary only, with a true scope and spirit of the invention being indicated by the following claims.

REFERENCES

1. Iwai H, Zuger S, Jin J, Tam PH (2006) Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostoc punctiforme. Febs Letters 580: 1853-1858.
2. Mootz HD, Dhar T (2011) Modification of transmembrane and GPI-anchored proteins on living cells by efficient protein trans-splicing using the Npu DnaE intein. Chemical Communications 47: 3063-3065.
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: 215-223.
4. Guan D, Ramirez M, Chen Z (2013) Split intein mediated ultra-rapid purification of tagless protein (SIRP). Biotechnol Bioeng 110: 2471-2481.
5. Wood DW, Wu W, Belfort G, Derbyshire V, Belfort M (1999) A genetic system yields self-cleaving inteins for bioseparations. Nature Biotechnology 17: 889-892.
6. Shi C, Tarimala A, Meng Q, Wood DW (2014) A general purification platform for toxic proteins based on intein trans-splicing. Appl Microbiol Biotechnol 98: 9425-9435.
7. Vila-Perello M, Liu Z, Shah NH, Willis JA, Idoyaga J, et al. (2013) Streamlined expressed protein ligation using split inteins. J Am Chem Soc 135: 286-292.
8. Perler, F. B., et al., (1994) Nucleic Acids Res 22, 1125-1127
9. Perler, F. B. (2002) Nucleic Acids Res 30, 383-384. [19] Saleh, L., et al. (2006) Chem Rec 6, 183-193.
10. Liu, X. Q., et al. (2003) J Biol Chem 278, 26315-26318
11. Yang, J., et al. (2004) Mol Microbiol 51 , 1185-1192.
12. Telenti, A., et al. (1997) J Bacterid 179, 6378-6382.
13. Wu, H., Xu, M. Q., et al. (1998) Biochim Biophys Acta 1387, 422-432.
14. Derbyshire, V., et al. (1997) Proc Natl Acad Sci U S A 94, 11466-11471.
15. Duan, X., et al. (1997) Cell 89, 555-564.
16. Ichiyanagi, K., et al. (2000) J Mol Biol 300, 889-901.
17. Klabunde, T., et al. (1998) Nat Struct Biol 5, 31-36.
18. Ding, Y., et al. (2003) J Biol Chem 278, 39133-39142. [29] Xu, M. Q., et al. (1996) Embo J 15, 5146-5153.
19. Scott, C.P., et al., (1999) Proc Natl Acad Sci U.S.A. 96, 13638-13643.
20. Xu, M.Q., et al., (2001) Methods 24, 257-277.
21. Evans, T. C1 Jr., et al. (2000) J Biol Chem 275, 9091-9094.

22. Kwon, Y. et al., (2006) Angew Chem Int Ed 45, 1726-1729.

23. Tori, K., Dassa, B., Johnson, M. A., Southworth, M. W., Brace, L. E., Ishino, Y., ... & 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*♦. *Journal of Biological Chemistry*, *285*(4), 2515-2526.)

24. Prabhala, S. V., Gierach, I. and Wood, D. W., "The Evolution of Intein-Based Affinity Methods as Reflected in 30 years of Patent History," *Frontiers in Molecular Biosciences*, Vol 9, Article 857566, (2022).

CLAIMS

What is claimed is:

1. An amino acid sequence comprising 90% or more identity to SEQ ID NO: 2, wherein SEQ ID NO: 2 comprises a tryptophan at position 74.
2. An amino acid sequence comprising 90% or more identity to SEQ ID NO: 12, wherein SEQ ID NO: 12 comprises a threonine at position 34.
3. An amino acid sequence consisting of SEQ ID NO: 2.
4. An amino acid sequence consisting of SEQ ID NO: 12.
5. A protein purification system, wherein the system comprises a split intein comprising two separate peptides: an N-terminal intein segment and a C-terminal intein segment, wherein the N-terminal intein segment comprises 90% or more identity to SEQ ID NO: 2, wherein SEQ ID NO: 2 comprises a tryptophan at position 74.
6. The protein purification system of claim 5, wherein the N-terminal intein segment is linked to a solid support.
7. The protein purification system of claim 5, wherein cleaving of the C-terminal intein segment is highly sensitive to extrinsic conditions when compared to a native intein.
8. The protein purification system of claim 5, wherein the C-terminal intein segment comprises 90% identity to SEQ ID NO: 12, wherein SEQ ID NO: 12 comprises a threonine at position 34.
9. The protein purification system of claim 5, wherein the N-terminal intein segment has been further modified so as not to comprise any internal cysteine residues.
10. The protein purification system of claim 5, wherein a purification tag is attached to the N-terminal intein segment at its C-terminus.
11. The protein purification system of claim 10, wherein the purification tag comprises one or more histidine residues.
12. The protein purification system of claim 11, wherein the N-terminal intein segment comprises amino acids at its C-terminus which allow for covalent immobilization.
13. The protein purification system of claim 12 wherein the one or more amino acids at the C-terminus are cysteine residues.
14. The protein purification system of claim 6 wherein the N-terminal intein segment is immobilized on a solid chromatographic resin backbone.

15. The protein purification system of claim 5, wherein a protein of interest (POI) is attached to the C-terminal intein segment at its C-terminus.
16. The protein purification system of claim 5, wherein the N-terminal intein segment further comprises a sensitivity-enhancing motif, which renders it highly sensitive to extrinsic conditions.
17. The protein purification system of claim 16, wherein the sensitivity-enhancing motif is on the N-terminus of the N-terminal intein segment.
18. The protein purification system of claim 16, wherein the extrinsic condition is pH, temperature, or both.
19. A protein purification system, wherein the system comprises a split intein comprising two separate peptides: an N-terminal intein segment and a C-terminal intein segment, wherein the C-terminal intein segment comprises 90% or more identity to SEQ ID NO: 11.
20. The protein purification system of claim 19, wherein the N-terminal intein segment is linked to a solid support.
21. The protein purification system of claim 19, wherein cleaving of the C-terminal intein segment is highly sensitive to extrinsic conditions when compared to a native intein.
22. The protein purification system of claim 19, wherein the N-terminal intein segment comprises 90% identity to SEQ ID NO: 2.
23. The protein purification system of claim 19 wherein the N-terminal intein segment has been further modified so as not to comprise any internal cysteine residues.
24. The protein purification system of claim 19, wherein a purification tag is attached to the N-terminal intein segment at its C-terminus.
25. The protein purification system of claim 24, wherein the purification tag comprises one or more histidine residues.
26. The protein purification system of claim 25, wherein the N-terminal intein segment comprises amino acids at its C-terminus which allow for covalent immobilization.
27. The protein purification system of claim 26, wherein the one or more amino acids at the C-terminus are cysteine residues.
28. The protein purification system of claim 19, wherein the N-terminal intein segment is immobilized on a solid chromatographic resin backbone.
29. The protein purification system of claim 19, wherein a protein of interest (POI) is attached to the C-terminal intein segment at its C-terminus.

30. The protein purification system of claim 19, wherein the N-terminal intein segment further comprises a sensitivity-enhancing motif, which renders it highly sensitive to extrinsic conditions.
31. The protein purification system of claim 30, wherein the sensitivity-enhancing motif is on the N-terminus of the N-terminal intein segment.
32. The protein purification system of claim 30, wherein the extrinsic condition is pH, temperature, or both.
33. A method of purifying a protein of interest, the method comprising:
    a. utilizing a split intein comprising two separate peptides: an N-terminal intein segment and a C-terminal intein segment, wherein the N-terminal intein segment comprises 90% or more identity to SEQ ID NO: 2;
    b. immobilizing the N-terminal intein segment to a solid support;
    c. attaching a protein of interest to the C-terminal intein segment, wherein cleaving of the C-terminal intein segment is highly sensitive to extrinsic conditions when compared to a native intein.
    d. exposing the N-terminal intein segment and the C-terminal intein segment to each other so that they associate;
    e. washing the solid support to remove non-bound material;
    f. placing the associated the N-terminal intein segment and the C-terminal intein segment under conditions that allow for the intein to self-cleave;
    g. isolating the protein of interest.
34. The method of claim 33, wherein the C-terminal intein segment comprises 90% or more identity to SEQ ID NO: 12.
35. A method of purifying a protein of interest, the method comprising:
    a. utilizing a split intein comprising two separate peptides: an N-terminal intein segment and a C-terminal intein segment, wherein the C-terminal intein segment comprises 90% or more identity to SEQ ID NO: 12;
    b. immobilizing the N-terminal intein segment to a solid support;
    c. attaching a protein of interest to the C-terminal intein segment, wherein cleaving of the C-terminal intein segment is highly sensitive to extrinsic conditions when compared to a native intein.
    d. exposing the N-terminal intein segment and the C-terminal intein segment to each other so that they associate;
    e. washing the solid support to remove non-bound material;

f. placing the associated the N-terminal intein segment and the C-terminal intein segment under conditions that allow for the intein to self-cleave;

g. isolating the protein of interest.

36. The method of claim 35, wherein the N-terminal intein segment comprises 90% or more identity to SEQ ID NO: 2.