

(51) International Patent Classification:
C12P 21/06 (2006.01)(21) International Application Number:
PCT/US2010/002571(22) International Filing Date:
21 September 2010 (21.09.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/272,401 21 September 2009 (21.09.2009) US(71) Applicant (for all designated States except US): THE
BOARD OF TRUSTEES OF THE LELAND STAN-
FORD JUNIOR UNIVERSITY [US/US]; . (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **BOGYO, Matthew**
[US/US]; 1705 El Camino Real, Palo Alto, CA
94306-1106 (US). **GARCIA, K., Christopher** [US/US];
1705 El Camino Real, Palo Alto, CA 94306-1106 (US).
SHEN, Aimee [US/US]; 1705 El Camino Real, Palo
Alto, CA 94306-1106 (US). **LUPARDUS, Patrick, J.**
[US/US]; 1705 El Camino Real, Palo Alto, CA
94306-1106 (US).(74) Agent: **BEAUMONT, William, E.**; Juneau Partners,
Pllc, 333 N. Fairfax St., Suite 305, Alexandria, VA 22314
(US).(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ,
CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO,
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP,
KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD,
ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI,
NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD,
SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR,
TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG,
ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ,
TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,
EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU,
LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK,
SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

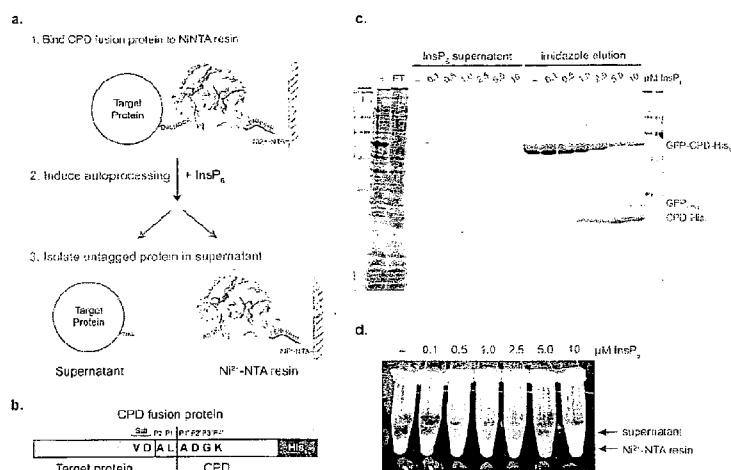
(54) Title: INDUCIBLE SELF-CLEAVING PROTEASE TAG AND METHOD OF PURIFYING RECOMBINANT PROTEINS
USING THE SAME

Fig. 1

(57) Abstract: A method of purifying a protein is disclosed which entails: a) fusing a site-specific affinity-tagged cysteine protease domain to a target protein to form a tagged fusion protein; b) activating the site-specific cysteine protease domain of the tagged fusion protein by subjecting the site-specific affinity-tagged cysteine protease domain to an inducer, which induces auto-processing at a cleavage site; thereby releasing untagged target protein; and c) isolating the untagged target protein.



Published:

- *without international search report and to be republished
upon receipt of that report (Rule 48.2(g))*

TITLE OF THE INVENTION**INDUCIBLE SELF-CLEAVING PROTEASE TAG AND METHOD OF PURIFYING
RECOMBINANT PROTEINS USING THE SAME****BACKGROUND OF THE INVENTION****Field of the Invention**

The present invention provides an inducible, self cleaving protease tag, and a method of purifying recombinant proteins using the same.

Description of the Background

The availability of simple, reliable, and cost-effective methods for recombinant protein purification is critical for the work of high throughput structural and proteomic centers and many individual researchers alike. While the addition of affinity tags such as poly-His and glutathione transferase (GST) to target proteins has greatly simplified purification strategies, it is often difficult to obtain soluble recombinant protein. As a result, affinity-tagged target proteins are often additionally fused to small proteins such as NusA and SUMO to improve their solubility, expression, and stability. However, these tags can alter the biological activity of target proteins and interfere with protein crystallization studies. Therefore many biological and biomedical applications require tag removal from the target protein. Most commonly used methods require the addition of exogenous site-specific proteases to cleave the affinity tag off the target protein at engineered sites. Unfortunately, the need for high levels of endoprotease for extended periods of time can result in unwanted cleavage events within the target protein. Furthermore, these endoproteases are costly, often exhibit poor solubility, and require the inclusion of additional chromatography steps to remove the exogenous protease.

SUMMARY OF THE INVENTION

Accordingly, the present invention provides an inducible, self-cleaving or autoprocessing protease tag, which may be used advantageously in the purification of recombinant proteins.

It is also an object of the present invention to provide a method of purifying recombinant proteins using the inducible, self-cleaving protease tag.

It is, moreover, an object of the present invention to provide a method of purifying recombinant proteins, wherein the purification, cleavage and separation of untagged protein from an endoprotease is condensed into a single step.

More particularly, it is an object of the present invention to provide a method of purifying a protein, which entails: a) fusing a site-specific affinity-tagged cysteine protease domain (CPD) protease to a target protein b) activating the site-specific CPD by subjecting the site-specific affinity-tagged CPD to an inducer, which induces autoprocessing at a cleavage site; thereby releasing untagged target protein; and c) isolating the untagged target protein.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a CPD fusion protein purification system. **(a)** Schematic of target protein purification using the CPD, described in detail in the text. **(b)** Schematic of CPD fusion protein. The Sall restriction site, which encodes the P4 and P3 residues Val and Asp, respectively, and the remaining P2-P4' residues contained within the CPD are shown. Prime positions refer to residues C-terminal to the autocleavage site, which is demarcated as a black vertical line. The composition of the residues added to the C-terminus of target proteins following autoprocessing can vary between one and four residues as described in Figure 3. Currently, the CPD system functions as a C-terminal fusion to target proteins and thus complements existing methods in which the affinity tag can only be applied as an N-terminal fusion¹³. **(c)** Purification of GFP using the CPD-His6 tag. GFP-CPD His6 bound to Ni²⁺-NTA resin was incubated with increasing amounts of InsP6

for 2 hrs at 4°C. GFP released into the supernatant was collected; Ni²⁺-bound proteins were then eluted from the resin by the addition of 200 mM imidazole. Collected fractions were analyzed by SDS-PAGE. **(d)** Visual analysis of GFP released into the supernatant fraction upon InsP6 addition to immobilized GFP-CPD-His6 fusion protein.

Figure 2 illustrates purification of test proteins using the CPD system. SDS-PAGE analysis using Coomassie stain of purifications of **(a)** biotin ligase (BirA, 35 kDa) using either the CPDHis6 or GST-His6 fusion tags, **(b)** CAD domain of STIM1 (14 kDa) using either the CPDHis6 or GST-His6 fusion tags. Asterisks indicate GST-STIM1(CAD)-His6 derived degradation products, and **(c)** mouse macrophage metalloelastin (MMP12) using CPDHis6 or His6-affinity tags. The diagonal arrows indicate a His6-tagged truncated MMIP12 product that is also observed during MMP12 purification from inclusion bodies. Large asterisk indicates a putative chaperone protein that co-purifies with MMIP12VDAL. In all cases, His6-tagged proteins bound to the Ni²⁺-NTA resin were incubated with 50 µM InsP6 for 1 hour at room temperature, and the resin was washed three times, followed by elution of Ni²⁺-bound proteins by 200 mM imidazole. CL, cleared lysate, +, IPTG induced culture, FT, flowthrough, IP6, elution from InsP6 incubation, E, imidazole elution prior to InsP6 addition.

Figure 3 is a schematic of pET-CPD expression vectors.

Figure 4 illustrates purification of the gp130 intracellular domain (ICD) using the CPD-His6 tag.

Figure 5 illustrates purification of PfSEN1 using the CPD system.

Figure 6 illustrates additional purification of CPD-cleaved MMP12vdal.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

To circumvent the above disadvantages, the present inventors have developed an on-bead cleavage purification system in which a site-specific affinity-tagged protease is fused directly to the target protein (**Fig. 1a and b**). A principal advantage of this approach is that affinity purification, cleavage, and separation of the untagged target protein from the endoprotease is condensed into a single step. This system combines the simplicity of onestep purification systems with many of the advantages of affinity tags, such as enhanced expression, integrity, and solubility of target proteins.

An important element of this purification method is the use of the *Vibrio cholerae* MARTX toxin cysteine protease domain (CPD). The CPD exhibits several properties that make it amenable to its development into an inducible, autocleaving protease tag. First, the CPD is a highly specific protease that cleaves exclusively after Leu residues⁴. In the native toxin, the CPD processes the MARTX toxin within interdomain junctions to release discrete effector domains. Secondly, the CPD is selectively activated by the eukaryotic-specific small molecule inositol hexakisphosphate (InsP6). Since InsP6 is absent from bacterial cells⁶, when the CPD-His6 tag is fused to the C-terminus of target proteins and expressed in *E. coli*, a CPD-His6 fusion protein can be purified from bacterial lysates in a protease-inactive form using imidazole affinity chromatography (IMAC, **Fig. 1a**). Addition of InsP6 to the immobilized fusion protein induces autoprocessing at the P1 Leu cleavage site (with P1 referring to the residue N-terminal to the scissile bond), which is located at the target protein-CPD junction. This processing event releases the untagged target protein into the supernatant, while the His6-tagged CPD remains immobilized on the Ni²⁺-NTA resin.

To demonstrate the feasibility of this system, the present inventors constructed pET expression vectors in which DNA encoding the CPD was cloned in to the Sall restriction site to generate pET-CPD vectors (**Fig. 3**). The fusion protein produced upon IPTG induction of *E. coli* harboring these vectors carries the P2-P1 residues of the native CPD (Ala-Leu, respectively) and

the P4-P3 residues encoded by the Sall site (Val-Asp, respectively) (**Fig. 1a and b**). Upon CPD-mediated autoprocessing of the fusion protein, the untagged target protein released from the resin carries four additional residues (Val-Asp-Als₂Leu) and the C-terminus. This C-terminal addition can be reduced to two amino acids (Glu-Leu) by cloning into the SacI site, or to a single amino acid (Leu) by cloning into the BamHI site and adding a Leu codon to the 3' cloning primer (**Fig. 3**).

As a proof-of-principle, the present inventors expressed and purified green fluorescent protein (GFP) as a fusion to CPD-His6 using IMAC; addition of increasing amounts of InsP6s stimulated the release of GFP from the Ni²⁺-NTA agarose beads (bead eluate, **Fig. 1c**). Since other site-specific proteases that are used to remove fusion tags have been observed to cleave target proteins at secondary sites, we sought to examine whether the CPD would spuriously cleave target proteins. To assess the fidelity of CPD-mediated processing of fusion proteins, we examined whether the CPD would cleave an intrinsically disordered protein after Leu residues within the target protein. We used the intracellular domain (ICD) of the cytokine receptor gp130 as a test substrate, since it is unstructured in solution by NMR⁷ and contains multiple Leu residues that might serve as cleavage substrates. The ICD-CPD-His6 fusion protein was expressed and purified from *E. coli* lysates using IMAC, and CPD-mediated cleavage of the immobilized fusion protein was activated by InsP6 addition. Autoprocessing occurred exclusively at the ICDCPD interdomain junction, with a single protein equivalent to the size of His6-tagged ICD being released into the supernatant fraction (**Fig. 4**). These results are consistent with our previous observation that CPD-mediated transcleavage is highly inefficient and strongly suggest that the CPD will not promiscuously cleave target proteins.

We noticed that the expression of the ICD-CPD-His6 fusion protein was approximately two-fold higher than the ICD-His6 protein in *E. coli* lysates (**Fig. 4**). This result suggested that the CPD might generally enhance target protein expression and/or solubility levels. To test this hypothesis, we compared the expression and solubility levels of CPD fusions to several other target proteins carrying either a His6-tag and/or GST-fusion tag (**Fig. 2, Fig. 5, Table 1**). In all cases, the presence of the CPD-His6 fusion tag increased the expression and solubility of target proteins. For example, fusion of the CPD-His6

tag to biotin ligase(BirA) from *E. coli* (BirA-CPD-His6_ raised BirA expression levels by three-fold over the GST-BirA construct8 (**Fig. 2a** and **Table 1**).

The CPD purification system also enhanced the expression, as well as purity, of a previously uncharacterized SUMO/Sentrin-specific peptidase 1 (SEN1) from the parasitic pathogen *Plasmodium falciparum*, the causative agent of malaria (**Fig. 5**). Although PfSenP1 carrying an N-terminal His6-tagged can be readily expressed and purified from *E. coli*, the N-terminal His7-tag must be removed by the addition of thrombin followed by multiple chromatography steps (**Table 2**). In contrast, when PfSEN1 is expressed as a fusion to CPD-His6 and released as untagged PfSEN1 upon InsP6 addition, only one minor contaminant copurifies with PfSEN1. This variant is easily removed using gel filtration chromatography, and the untagged PfSEN1 is of sufficient purity that we have used it to obtain diffraction-quality crystals. Although the heterologous expression of *P. falciparum* proteins in *E. coli* is typically challenging, we have observed that this system can enhance the expression and purification of other parasite proteins.

In addition to augmenting the expression of target proteins, the CPD-His6 fusions can protect target proteins from proteolytic degradation. This can be demonstrated by fusing the CRAC-activation domain (CAD) of the ER calcium sensor STIM1 to the CPD (**Fig. 2b**). CAD is a small 107 aa polypeptide that activates Ca²⁺ release-activated Ca²⁺ (CRAC) channels by binding to the CRAC channel protein Orai1. Until now, largescale expression and purification of this important regulatory domain has proven difficult due to its apparent instability even when fused to GST (**Fig. 2b**). Using the CPD system, we were able to obtain significant quantities of the intact CAD protein, which has subsequently been used in high-throughput screens for Orai1-CAD binding partners.

Moreover, the CPD purification system also increased the solubility of difficult-to-express proteins. Fusion of the mouse macrophage metalloelastase (MMP12) to CPDHis6 facilitated its purification from the soluble fraction of *E. coli* lysates, whereas His6-tagged MMP12 remained largely insoluble (**Fig.2c**). The currently used method for purification of His6-tagged MMP12 is a laborious procedure that requires solubilization of MMP12 inclusion bodies, refolding over

multiple days, followed by anion and cation exchange chromatography. The CPD purification system dramatically simplifies this purification procedure, allowing soluble, active MMP-12 to be isolated in approximately 7 hours (**Fig. 6** and **Table 3**).

Collectively, these results imply that the one-step purification systems such as the intein-chitinbinding-domain (CBD) and sortase-His6. While these systems simplify the purification of well-expressed proteins, the large size of the intein-CBD fusion tag can decrease target protein solubility, and sortase-His6 fusion tags do not increase target protein solubility.

Furthermore, unlike self-cleaving elastin-like polypeptide (ELP) tags, fusion proteins do not need to be subjected to the temperature cycles, pH shifts, or high salt concentrations, a feature that is critical for the purification of the intractable proteins. Based on the properties reported here, the CPD could replace the intein-tag in the self-cleaving-ELP system and potentially improve the solubility of ELP-tagged proteins while retaining their self-cleavability.

Indeed, a considerable strength of this method is that the CPD remains active over a wide range of conditions. CPD-mediated cleavage is complete within 1-2 hrs at temperatures between 4°C and 37°C, requires only micromolar of the small molecule InsP6 (an abundant and inexpensive reagent), and occurs efficiently both in the presence of standard protease inhibitor cocktails and in the absence of salt. This latter property carries the additional advantage of allowing the user to determine the buffer system in which to elute the target protein, which eliminates the need for desalting or buffer exchange steps that can reduce protein yields. Thus, the CPD system allows for considerable flexibility in optimizing purification procedures, as is often necessary for uncharacterized target proteins.

This versatility, combined with our observation that it can advantageously improve the solubility and integrity of difficult-to-express proteins (**Fig.2** and **Fig. 6**), indicates that it will have widespread utility in biological research. The simplicity of this system will also make it amenable for large-scale proteomic, structural genomic, and commercial applications by eliminating the cost and complexity associated with exogenous sitespecific proteases, potentially permitting its use in robotic systems for constructing protein arrays for screening purposes. Furthermore, it is also possible to generate mutants of the CPD that require high

concentrations of InsP6 for activation that would be suitable for use in Eukaryotic expression systems.

TERM DEFINITIONS

"CPD" means cysteine protease domain.

"MARTX" means multifunctional, autoprocessing RTX toxins produced by certain bacteria.

"Inducer" means a small molecule that induces autoprocessing at a cleavage site. In the present invention, that induced autocleavage releases an untagged target protein.

In more detail, **Fig. 3** is a schematic of pET-CPD expression vectors. Bent arrow, T7 promoter, Oval (RBS), ribosome binding site, green rectangle, target protein, grey rectangle, CPD, *V. cholera* MARTX (aa. 3440-3650), darker gray rectangle, Δ P1-CPD, *V. cholera* MARTX (aa. 3442-3650), darkest rectangle, Δ P2'-CPD, *V. cholera* MARTX (aa. 3444-3650), black rectangle, His6-tag, white rectangle, HA-tag. The dotted vertical line and arrow indicate the CPD cleavage site. Residues added onto the C-terminus of the target protein following CPD-mediated cleavage, and the relevant restriction site are shown. The composition of the amino acids added to the C-terminus of the target protein can be varied depending on the cloning site and pET-CPD vector used. It should be noted that the P1 Leu shown for pET22b-CPDBamHI-Leu must be encoded in the 3' cloning primer of the target gene (i.e. supplied by the target gene). Since the primary substrate specificity determinant for the CPD is a P1 Leu, and P2 and P3 residues are not strongly recognized by the CPD substrate binding pocket, the fusion protein produced from pET22b-CPDBamHI-LEU is efficiently autoprocessed by the CPD. Both pET22b and pET28a vector backbones were used to construct the CPD expression vectors.

Fig. 4 illustrates purification of the gp130 intracellular domain (ICD) using the CPDHis6 tag. Gp130 ICD(CPD)-His6 or gp130(ICD)-His6 bound to Ni²⁺-NTA resin was incubated with 100 μ M InsP6 for 2 hr at room temperature; the resin was washed four times, followed by elution of Ni²⁺-bound proteins by 200 mM imidazole. Purification fractions were analyzed by SDS-PAGE

followed by Coomassie staining. CL, cleared lysate, FT, flowthrough, IP6, elution from InsP6 incubation.

Fig. 5 illustrates purification of PfSEN1 using the CPD system. (a) SDS-PAGE analysis using Coomassie staining of *Plasmodium falciparum* SEN1 (PfSEN1, 25 kDa) purification using either the CPD-His6 or His6-affinity tags. PfSEN1-CPD-His6 or His6-PfSEN1 bound to the Ni²⁺-NTA resin was incubated with 100 μ M InsP6 for 2 hr at room temperature: the resin was washed three times, and wash fractions were collected. Ni²⁺-bound proteins were eluted by adding 200 mM imidazole, +, IPTG induced culture, CL, cleared lysate, E, imidazole elution prior to InsP6 addition, IP6, elution from InsP6, (b) UV trace PfSEN1 further purified by gel filtration chromatography following His6-tag removal. Inset, Coomassie staining of gel filtration fractions of PfSEN1 purifications. Thrombin refers to PfSEN1 purified thrombin-mediated removal of the N-terminal His6-tag, while InsP6 refers to InsP6-induced, CPD-mediated removal of the C-terminal CPD-His6-tag. The residues added to the resulting PfSEN1 protein are indicated; N-terminal GSHM for PfSEN1 (thrombin cleavage) and C-terminal VDAL for PfSEN1 (InsP6-activated CPD cleavage). (c) Coomassie staining of fractions taken during His6-PfSEN1 purification prior to thrombin incubation (-), following 12 hr thrombin incubation (+), and following subtractive IMAC to remove uncleaved His6-PfSEN1 (Ni²⁺-NTA). The yield of PfSEN1 diminished with each experimental manipulation.

Fig. 6 illustrates additional purification of CPD-cleaved MMP12VDAL. (a) Purification of MMP12VDAL by gel filtration chromatography. Inset, Coomassie staining of SDS-PAGE analysis of gel filtration fractions of MMP12VDAL. (b) MMP12 Fluorogenic substrate assay. The activity of MMP12 purified under denaturing conditions and refolded (MMP12 (Refolded)) and MMP12VDAL purified using the CPD system against a standard fluorogenic substrate were compared. Comparable rates of fluorogenic substrate cleavage are observed for MMP12 purified by the CPD method relative to the refolding method.

The results obtained and observed are summarized in the tables below.

Tables

Table 1 Target proteins expressed and purified by CPD purification method

Target protein	Yield ^a (mg/L culture)	Yield (nmol/L culture)	Activity
GFP _{VDAL} (CPD method)	3.3	105	Fluorescence at 511 nm
Gp130(ICD) _{VDAL} (CPD method)	5.9	188	n/a
Gp130(ICD)-His ₆	3.7	115	n/a
BirA _{VDAL} (CPD method)	10.9	202	Biotinylates LHHILDAQKMVWNHR BirA biotinylation site
GST-BirA-His ₆	12.0	90	Biotinylates LHHILDAQKMVWNHR BirA biotinylation site
PfSEN1 _{VDAL} (CPD method)	2.0	67	Cleaves PfSUMO
GSHM-PfSEN1	1.4	46	Cleaves PfSUMO
STIM1(CAD) _{VDAL}	2.1	148	Binds Orail
GST-STIM1(CAD)-His ₆	2.5 ^b	62	n/a
MMP12 _{VDAL} (CPD method)	1.4	47	Cleaves flourogenic peptide substrate Mca-PLGLDL(Dpa)AR
MMP12 (refold)	23	767	Cleaves flourogenic peptide substrate Mca-PLGLDL(Dpa)AR

^a Protein yield per litre of culture

^b Yield difficult to assess since GST-fusion protein degrades and falls out of solution over time

Table 2 Comparison of CPD-mediated purification of untagged PfSEN1 to thrombin-mediated purification

	PfSEN1-CPD-His₆	His₆-PfSEN1
Step 1	Prepare soluble lysate (1 hr)	Prepare soluble lysate (1 hr)
Step 2	IMAC purification (2 hr)	IMAC purification (2 hr)
Step 3	On-bead cleavage; collect supernatant (2 hr)	Imidazole elution
Step 4	Concentrate protein (0.5 hr)	Buffer exchange and concentrate protein (0.5 hr)
Step 5	Gel filtration chromatography (1 hr)	Thrombin cleavage overnight (>12 hr)
Step 6	Concentrate protein (0.5 hr)	Remove His ₆ -tag and uncleaved fusion with IMAC (1 hr)
Step 7		Concentrate protein an buffer exchange (0.5 hr)

Step 8		Gel filtration chromatography (1 hr)
		Concentrate protein (0.5 hr)
Total time	5 hr	>17.5 hr

Table 3 Comparison of CPD method to published method for purifying matrix metalloelastin (MMP12)

	MMP12-CPD-His₆	MMP12-His₆¹
Step 1	Prepare soluble lysate (1 hr)	Prepare and dissolve inclusion bodies (16 hr)
Step 2	IMAC purification (2 hr)	Protein refolding by dialysis in 1/20 volume (48 hr) a. 24 hr – 6M Urea, 4 L b. 24 hr – 3M Urea, 4 L c. 24 hr – 1M Urea
Step 3	On-bead cleavage; collect supernatant (2 hr)	Load partially refolded protein on anion and cation exchange tandem columns (2.5 hr)
Step 4	Concentrate protein (0.5 hr)	Wash columns with three buffers to fully refold protein on column (2 hr)
Step 5	Gel filtration chromatography (1 hr)	Elute protein from cation exchange column (0.5 hr)
Step 6	Concentrate protein (0.5 hr)	Concentrate protein (0.5 hr)
Total time	7 hr	3-4 days

Table 4. Primers used in examples.

#	Name	Sequence ^a	RE ^b
1	5' Sall A50CPD	CACGTCGACGCATTAGCGGATGGAAAAATACTACTCCAT	Sall
2	3' XhoI CPD	CACGTCGAGACCTTGCGCGTCCCAGCTTAG	XhoI
3	5' Scal CPD	TTCGAGCTCGCGGATGGAAAAATACTC	Scal
4	5' Sall HA-CPD	TCCGTCGACTACCCGTACGACGTCCCGGACTACGCGGCATTAGCGGATGGAAAA	Sall
5	5' BamHI CPD _{Low}	TCGGGATCCGGAAAAATACTCCATAATCAA	BamHI
6	5' BamHI CPD	TCGGGATCCGCATTAGCGGATGGAAAAATACTCCAT	BamHI
7	5' NdeI gfp	ATTCATATGGTGAGCAAGGGCGAG	NdeI
8	3' Sall gfp	CACGTCGACCTTGTACAGCTCGTCCAT	Sall
9	5' NdeI gp130(ICD)	ATTCATATGAATAAGCGAGACCTA	NdeI
10	3' Sall gp130(ICD)	TCCGTCGACCTGAGGCATGTAGCCGCC	Sall
11	5' NdeI BirA	GGCCATATGATGAAGGATAACACCGTGCCA	NdeI
12	3' Sall BirA	TCCGTCGACTTTTCTGCACTACGCAGGGA	Sall
13	5' NdeI CAD	AGCCATATGTATGCTCCAGAGGCCCTT	NdeI
14	3' Sall CAD	TCCGTCGACAGCAGGGTTGGGGCGTGT	Sall
15	5' NdeI MMP12	CTTCCATATGGCTCCCATG	NdeI
16	3' Sall MMP12	TAACGTCGACCTCGAGTCC	Sall

^a Restriction enzyme sequences are underlines, and HA tag is shown in italics. ^b Restriction site

Table 5: Strains used in examples

Strain	Genotype and relevant features	Reference
41	BL21(DE3)	Novagen
7	DH5 α	D.E. Cameron
269	pET22b in DH5 α	D.E. Higgins
195	pET28a in DH5 α	E. Ponder
330	pET22b-CPD _{Sall}	Present invention
329	pET28a-CPD _{Sall}	Present invention
331	pET22b-HA-CPD _{Sall}	Present invention
373	pET28a-HA-CPD _{Sall}	Present invention
374	pET22b-CPD _{BamHI-Leu}	Present invention
375	pET22b-CPD _{BamHI}	Present invention
197	pET22b-GFP-CPD	Present invention
371	pET22b-gp130(ICD)-CPD	Present invention
372	pET21a-gp130(ICD)	²
228	pET22b-BirA-CPD	Present invention
183	pGEX4T1-BirA	P.J. Lupardus
360	pET22b-PfSEN1-CPD	³
361	pET28a- PfSEN1	³
324	pET22b-STIM1(CAD)-CPD	Present invention
327	pGEX6-CAD128	⁴
359	pET22b-mMMP12	Present invention
358	pET41a-mMMP12	C. Overall

Methods

Bacterial growth conditions Overnight bacterial strains were grown at 37°C in Luria-Bertrani (LB) broth. Antibiotics were used at 100 µg/mL carbenicillin for pET22b vectors expressed in *E. coli*.

Strain construction Primers used are listed in **Table 4**; strains constructed are listed in **Table 5**. For construction of pET-CPD_{Sall} vectors, DNA encoding *Vibrio cholera* MARTX toxin amino acids 3440-3650 from *Vibrio cholerae* N16961 was PCR amplified from genomic DNA using primers #1 and #2. The resulting PCR fragment was cloned into the Sall and XhoI sites of the pET22b and pET28a expression vectors, respectively

(Novagen). For construction of the pET-CPDSacI vector, DNA encoding *Vibrio cholerae* MARTX toxin amino acids 3442-3650 from *Vibrio cholerae* N16961 was PCR amplified from genomic DNA using primers #3 and #2, and the resulting PCR fragment was cloned into the SacI and XhoI sites of pET22b. To construct the pET-HA-CPDSall vectors, DNA encoding the HA epitope tag was added to the 5' end of primer #4, and PCR amplification using primers #4 and #2 was used to fuse the HA tag directly to amino acid 3440 of *V. cholera* MARTX CPD. The resulting PCR fragment was cloned into the Sall and XhoI sites of the pET22b and pET28a expression vectors, respectively. For construction of the pET-CPDBamHI-Leu vector, DNA encoding *Vibrio cholerae* MARTX toxin amino acids 3440-3650 from *Vibrio cholerae* N16961 was PCR amplified from genomic DNA using primers #5 and #2, and the resulting PCR fragment was cloned into BamHI and XhoI sites of pET22b. For construction of the pETCPDBamHI vector, DNA *Vibrio cholerae* MARTX toxin amino acids 3440-3650 from *Vibrio cholerae* N16961 was PCR amplified from genomic DNA using primers #6 and #2 were used, and the resulting PCR fragment was cloned into the BamHI and XhoI sites of pET22b. The pET22b-GFP-CPD construct was cloned by PCR amplifying GFP from pEGFPN3(Clontech) using primers #7 and #8. To construct the pET22b-gp130(ICD)-CPD vector, amino acid 642-918 of gp130 corresponding to the intracellular domain were PCR amplified using primers #9 and #10 and pET21a-gp130(ICD) as a template. The pET22b-BirA-CPD vector was constructed by PCR amplifying the *birA* gene from the pGEX4T1-BirA template using primers #9 and #10. The pET22b-STIM1(CAD)-CPD plasmid was constructed by PCR amplifying DNA encoding amino acids 342-369 of STIM1 using pGEX6-CAD128 as a template and primers #13 and 14. The pET22b-nMMP12-CPD construct was constructed by PCR amplifying the catalytic domain of mouse MMP12 (aa 29-267) using pET41a-mMMP12 as a template using primers #15 and #16. In all cases, the resulting PCR products were cloned into the NdeI and Sall sites of pET22b-CPDSall.

Protein expression and purification. For purification of His6-tagged CPD fusion proteins, overnight cultures of the appropriate strain were diluted 1:500 into 1 L 2YT media and grown shaking at 37°C. When an OD600 of 0.6 was reached, IPTG was added to 250 µM, and cultures

were grown for 3-4 hrs at 30°C. Cultures were pelleted, resuspended in 25 mL lysis buffer (500 nM NaCl, 50 mM Tris-HCL, pH 7.5, 15mM imidazole, 10% glycerol) and flash frozen in liquid nitrogen. Lysates were thawed, then lysed by sonication and cleared by centrifugation at 15,000 x g for 30 minutes. His6-tagged CPD fusion proteins were affinity purified by incubating the lysates in batch with 0.5-1.0 mL Ni-NTA Agarose beads (Qiagen) with shaking for 2-3 hrs at 4°C. the binding reaction was pelleted at 1,500 x g, the supernatant was aside, and the pelleted Ni²⁺-NTA agarose beads were washed three times with lysis buffer. In some cases, 10% of the Ni²⁺-NTA beads containing immobilized CPD-His6 fusion proteins were removed, pelleted and then His6-tagged eluted using high imidazole buffer (500 nM NaCl, 50 mM Tris-HCL, pH 7.5, 175mM imidazole, 10% glycerol).

To liberate untagged target proteins into the supernatant fractions, 300-500 µL lysis buffer was added to the Ni²⁺-NTA beads containing CPD-His6 fusion proteins and the indicated amount of inositol hexakisphosphate (InsP6, Calbiochem) was added. In general, on-bead cleavage was allowed to proceed by nutating the beads in the presence of 50-100 µM InsP6 for 1-2 hr at room temperature or 4°C. The beads were pelleted at 1,500 x g, and the supernatant fraction was removed. The beads were then washed 3-4 times with 300-500 µL lysis buffer, and supernatant fractions retained. His6-tagged proteins remaining on the beads (i.e. cleaved CPD-His6) were eluted using high imidazole buffer (500 nM NaCl, 50 mM Tris-HCL, pH 7.5, 175mM imidazole, 10% glycerol) in 300-500 µL volumes. The elute was repeated 3-4 times, and elute fractions were collected. Purification of His6-tagged proteins lacking the CPD was performed in parallel. This general procedure was followed with the following exceptions: for purification of MMP12 constructs, the cultures were grown at 16°C overnight after IPTG induction, and 1 mM tris (2-carboxyethyl)phosphine (TCEP) was added to the lysis buffer to prevent misfolding of the protein. PfSEN1 and BirA protein purifications were performed exclusively at room temperature, since at 4°C, protein aggregation was observed. For removal of the His6-tag from His6-PfSEN1, thrombin beads (Calbiochem) that had been washed in PBS were added to the elute His6-PfSEN1, which had been buffer exchanged into PBS. Thrombin cleavage was allowed to proceed with shaking overnight for 12 hr at room temperature. Aliquots were taken before and after thrombin addition to monitor cleavage efficiency. Thrombin cleaved, untagged PfSEN1 was enriched by performing a subtractive Ni²⁺-NTA pull-down. Untagged PfSEN1

from both methods was then buffer-exchanged into gel filtration buffer (50 mM NaCl, 20 mM Tris pH 8.0). Protein purifications were analyzed by SDS-PAGE and Coomassie staining using GelCode Blue (Pierce). Purified protein concentrations of purified were determined by Bradford assay (Pierce).

Purification of MMP12-His6. MMP12-His6 was purified as previously described¹ with the following modifications. The cell pellet was resuspended in 100 mM NaCl, 100 mM Tris pH 8.0, 5.0 mM EDTA, 0.5 mM DTT, 100 µg/mL lysozyme and stirred for 2 hr. The cells were sonicated then centrifuged at 10,000 rpm for 10 min. The resulting inclusion bodies were washed two times and then resuspended in 50 mL 6M guanidine hydrochloride, 10 mM Tris pH 8.0 by stirring at 4°C overnight. The mixture was centrifuged at 15,000 rpm for 30 min, and 2 mL aliquots of supernatant were prepared. The supernatant was diluted 1:100 into denaturing buffer (6M Urea, 50 mM Tris pH 8.0, 10 mM CaCl₂, 30 mM NaCl, 5 mM DTT) to a final concentration of 0.1-0.2 mg/mL. The protein was then dialyzed for 24 hr in 2 L refolding buffer 1 (3 M Urea, 50 mM Tris pH 8.0, 10 mM CaCl₂, 30 mM NaCl, 5 mM DTT). The partiallyrefolded protein was then dialyzed in 4 L of refolding buffer 2 (1 M Urea, 50 mM HEPES pH 7.4, 10 mM CaCl₂, 5 mM DTT). The buffer exchanged protein was then purified using tandem 5 mL MonoQ and SP Sepharose (GE Healthcare) at 4°C. After loading the protein on the column, the column was washed with 50 mL of refolding buffer 2 without DTT at 1 M, 0.5, and 0 Murea, respectively. The protein was eluted from the SP column in 500 mM NaCl, 50 mM HEPES ph 7.4, 10 mM CaCl₂.

Gel filtration chromatography. Untagged PfSEN1 obtained from either thrombin or InsP6-mediated cleavage was concentrated using 10 kDa Centricon concentrator (Millipore) and buffer exchanged into 50 mM NaCl, 20 mM Tris pH 8.0 and purified on a Superdex 200 10/30 column (GE Healthcare) equilibrated in the same buffer. For MMP12, the gel filtration buffer contained 150 mM NaCl, 50 Mm Tris pH 7.4, 10 Mm TCEP. Gel filtrations were performed at 4°C.

Activity assays. Fluorescence of purified GFP at 511 nm was verified using a Molecular Devices fmax plate reader in black 96-well plates and 488 nm excitation. The activity of MMP12 was determined using the flourogenic substrate Mea-PLGLDL(Dpa)AR (Mea, (7-

methoxycoumarin-4-yl)acetyl, Dpa, *N*-3-(2,4-dinitrophenyl)-L-2,3-diaminopropionyl, Anaspec). Reactions were preformed in the assay buffer (50 Mm Tris pH 7, 150 mM NaCl, 10 Mm CaCl₂, 0.02% NaN₃, 5 mM TCEP) at 37°C. The substrate was used at 10 μM and the protein at 0.2 μM. The substrate hydrolysis was monitored continuously in a fluorescent plate reader (Molecular Devices) using an excitation wavelength of 325 nm and an emission wavelength of 395 nm.

In addition to the embodiments described above, other variations thereof may also be used in accordance with the present invention without departing from the spirit and scope thereof.

For example, aside from the *Vibrio cholera* MARTX cysteine protease domain (CPD), related MARTX CPDs from *Photobacterium luminescens* and *Vibrio vulnificus* also autoprocess in the presence of InsP6. Further, Clostridium toxin CPDs also work in the system.

Generally, MARTX toxins produced by *Vibrio sp.* and *Clostridium sp.* may be used. Exemplary species of *Vibrio* are *V. anguillarum*, *V. splendidus* and *V. vulnificus* in addition to *V. cholera*. Further, these toxins are specifically described as CPDs. For example, the Clostridium toxins are CPDs derived from the large glucosylating toxins produced by Clostridium sp. All of these bacterial species are commercially or readily available to the artisan.

Further, in addition to InsP6 (inositol hexakisphosphate), inositol pentakisphosphate (InsP5) may also be used but generally higher concentrations of the later are required to induce autocleavage.

Moreover, while *E. coli* is used as a well-known expression system the present invention, other bacterial hosts may also be used to produce the target recombinant protein. For example, *Bacillus systems* and *Lactobacillus lactis* systems may be used. Generally, any bacterial host system may be used provided that the expressed target protein is secreted into the media, and the media does not contain either InsP6 or InsP5, for example. Further, it is also possible to use the system in eukaryotic cells when the CPD is mutated to be less responsive to InsP6 and InsP5, for example. Basic cage mutations described in recent literature require higher than physiological

concentrations of InsP6 (cytosolic concentrations have been reported to be between 5-100 micromolar InsP6) to become activated.

Generally, the present method is conducted at a pH range of about 6.5 to 9.5. However, it is preferred that a pH range of about 7.5 to 8.5 be used. Further, the MARTX CPDs used in the present invention are preferably insensitive, i.e., no loss of protease activity, to salt concentration. For example, the MARTX CPDs used in the present invention generally exhibit little or no loss of activity in the presence of NaCl in a concentration of from 0 to 500 mM.

As noted above, the present method generally affords an increased expression of target proteins. In general, although the extent of increased expression varies from protein to protein, at least a two-fold increase in expression is commonly observed. Yet, greater increases are also observed. See the discussion above regarding a comparison of the expression level for (BirA-CPD-His6) in *E. coli* and (GST-BirA).

Additionally, various affinity tags may be used in accordance with the present invention being fused to the CPD. For example, streptavidin binding tags (SBP or Nanotag), CBP (Calmodulin binding tag) and Protein C-epitope tag may be mentioned. Further, it is also acceptable to combine the use of affinity-tagged CPD with other fusion or affinity tags. For example, the present inventors have constructed vectors in which the protein of interest can itself carry an affinity tag in addition to the affinity-tagged CPD (like the HA tag depicted in **Fig. 3**). Dually tagged proteins allow for tandem affinity purification of the target protein.

Finally, the CPD used may be modified to function as an N-terminal fusion. For example, the present inventors have observed that the CPD can cleave itself off a streptavidin support when a C-terminal biotinylated peptide sequence is fused to the CPD.

Having described the above embodiments, it will remain clear that various other changes and modifications may be made without departing from the spirit and scope of the present invention.

of PfSEN1 purifications. Thrombin refers to PfSEN1 purified thrombin-mediated removal of the N-terminal His₆-tag, while InsP₆ refers to InsP₆-induced, CPD-mediated removal of the C-terminal CPD-His₆-tag. The residues added to the resulting PfSEN1 protein are indicated; N-terminal GSHM for PfSEN1 (thrombin cleavage) and C-terminal VDAL for PfSEN1 (InsP₆-activated CPD cleavage). (c) Coomassie staining of fractions taken during His₆-PfSEN1 purification prior to thrombin incubation (-), following 12 hr thrombin incubation (+), and following subtractive IMAC to remove uncleaved His₆-PfSEN1 (Ni²⁺-NTA). The yield of PfSEN1 diminished with each experimental manipulation.

WHAT IS CLAIMED:

1. A method of purifying a protein, which comprises the steps of:
 - a) fusing a site-specific affinity-tagged cysteine protease domain to a target protein to form a tagged fusion protein;
 - b) activating the site-specific cysteine protease domain of the tagged fusion protein by subjecting the site-specific affinity-tagged cysteine protease domain to an inducer, which induces autoprocessing at a cleavage site; thereby releasing untagged target protein; and
 - c) isolating the untagged target protein.
2. The method of claim 1, wherein the cysteine protease domain is a *Vibrio sp* MARTX cysteine protease domain.
3. The method of claim 1, wherein the inducer is inositol hexakisphosphate, InsP₆.
4. The method of claim 1, wherein the cysteine protease domain in MARTX CPD of *Photobacterium luminescens*.
5. The method of claim 1, wherein the target protein is expressed in *E. coli*.
6. The method of claim 1, wherein the target protein is expressed in *Bacillus*.
7. The method of claim 1, wherein the target protein is expressed in *Lactobacillus lactis*.
8. The method of claim 1, wherein the target protein is expressed in a eukaryotic cell.

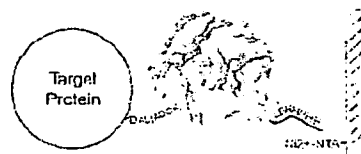
9. The method of claim 1, wherein said isolating comprises separating the target protein from bacterial lysates by affinity chromatography.
10. The method of claim 9, wherein the affinity chromatography is imidazole affinity chromatography.
11. The method of claim 1, wherein the site-specific affinity-tagged CPD is immobilized.
12. The method of claim 11, wherein the site-specific affinity-tagged CPD is immobilized on Ni^{+2} – NTA resin.
13. The method of claim 1, wherein the site-specific affinity-tagged CPD functions by C- or N-terminal fusion to the target protein.
14. The method of claim 1, wherein the site-specific affinity-tagged CPD enhances expression of the target protein.
15. The method of claim 1, wherein site-specific affinity-tagged CPD increases solubility of the target protein.
16. The method of claim 1, wherein the target protein is a parasite protein.
17. The method of claim 17, wherein the parasite is *P. falciparum*.
18. The method of claim 1, wherein the site-specific affinity-tagged CPD protects the target protein from proteolytic degradation.
19. The method of claim 1, wherein the target protein is MMP12.
20. The method of claim 1, which does not require temperature cycles.

21. The method of claim 1, which is conducted at a pH of 6.5 to 9.5.
22. The method of claim 1, wherein the cysteine protease domain is insensitive to salt.
23. The method of claim 1, which is effected at a temperature of between 4-37°C.
24. The method of claim 1, which is effected in about 1-2 hours.
25. The method of claim 1, wherein the untagged target protein is isolated from supernatant.
26. The method of claim 1, wherein the protein is purified is a recombinantly expressed protein.
27. The method of claim 1, wherein the inducer is inositol pentakisphosphate, InsP₅.
28. A recombinant and purified protein produced by the method of claim 1.

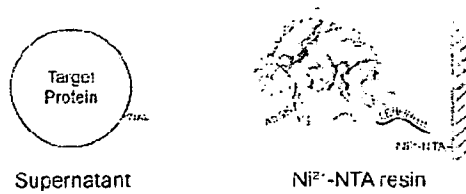
1/6

a.

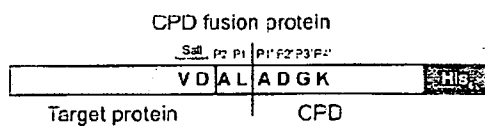
1. Bind CPD fusion protein to NiNTA resin

2. Induce autoproteolysis + InsP_6 

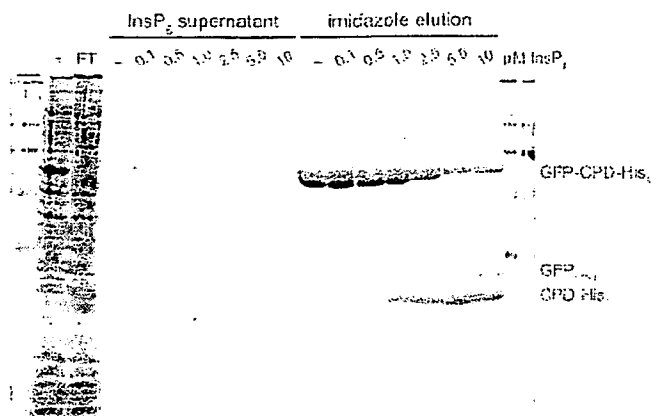
3. Isolate untagged protein in supernatant



b.



c.



d.



Fig. 1

2/6

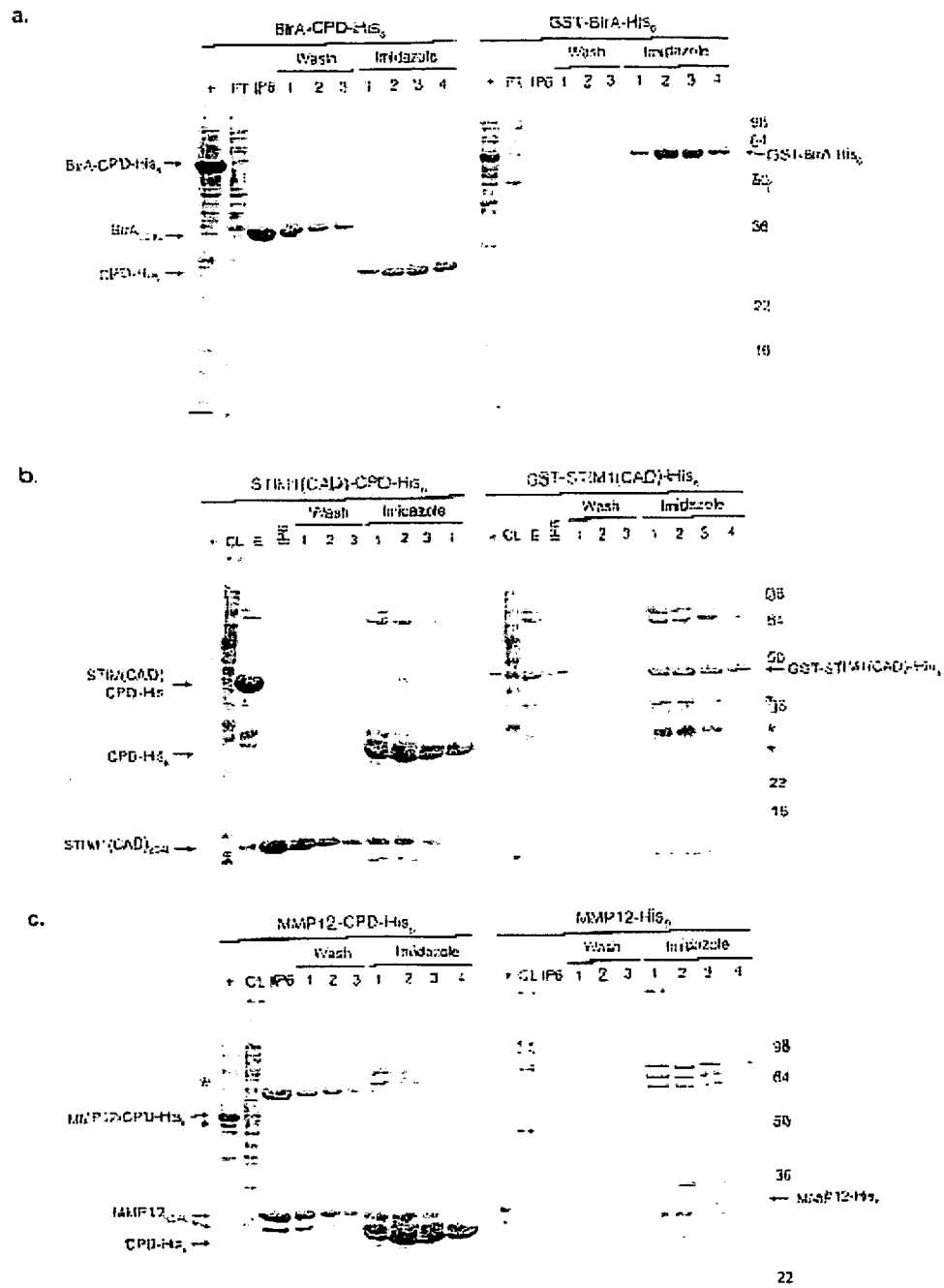


Fig. 2

3/6

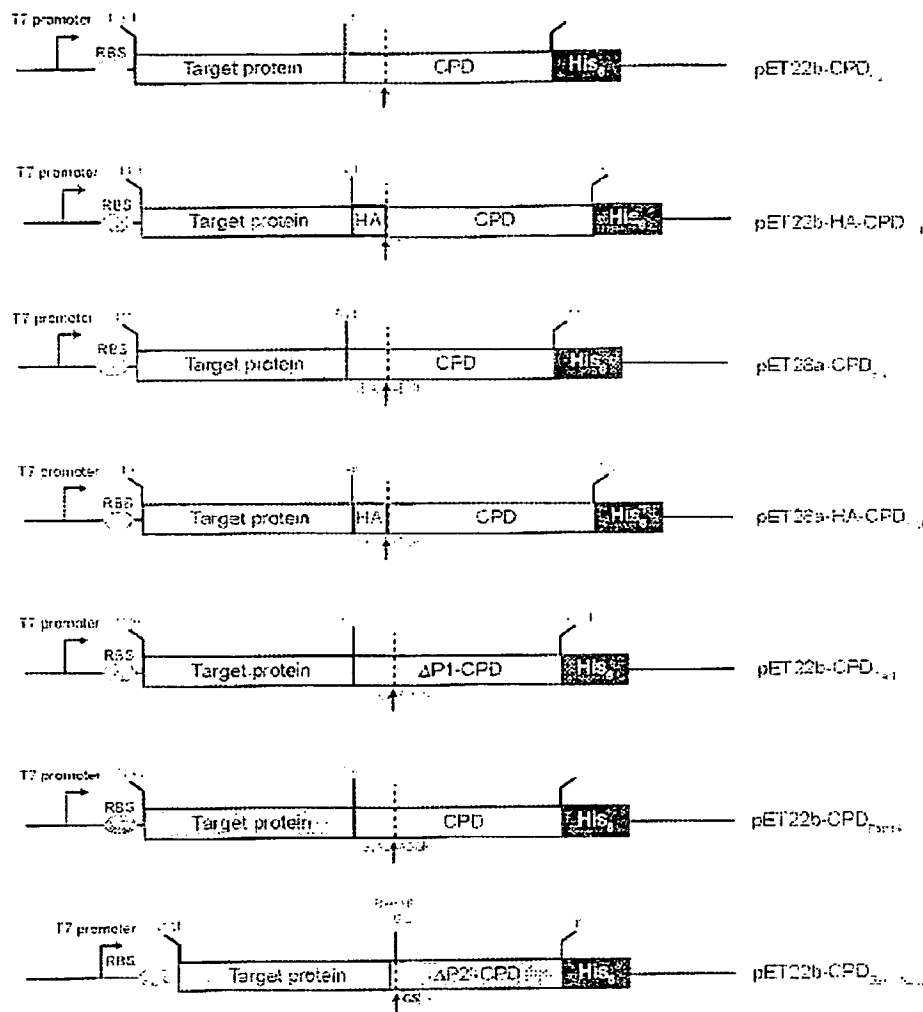


Fig. 3. Schematic of pET-CPD expression vectors. Bent arrow, T7 promoter, Oval (RBS), ribosome binding site, green rectangle, target protein, grey rectangle, CPD, *V. cholera* MARTX (aa. 3440-3650), darker grey rectangle, Δ P1-CPD, *V. cholera* MARTX (aa. 3442-3650), darkest grey rectangle, Δ P2'-CPD, *V. cholera* MARTIX (aa. 344-3650), black rectangle, His₆-tag, white rectangle, HA-tag. The dotted vertical line and arrow indicate the CPD cleavage site. Residues added onto the C-terminus of the target protein following CPD-mediated cleavage, and the relevant restriction site are shown. The composition of the amino acids added to the C-terminus of the target protein can be varied depending on the cloning site and pET-CPD vector used. It should be noted that the P1 Leu shown for pET22b-CPD_{BamHI-Leu} must be encoded in the 3' cloning primer of the target gene (i.e. supplied by the target gene). Since the primary substrate specificity determinant for the CPD is a P1 Leu, and P2 and P3 residues are not strongly recognized by the CPD substrate binding pocket, the fusion protein produced from pET22b-CPD_{BamHI-LEU} is efficiently autoprocessed by the CPD. Both pET22b and pET28a vector backbones were used to construct the CPD expression vectors.

4/6

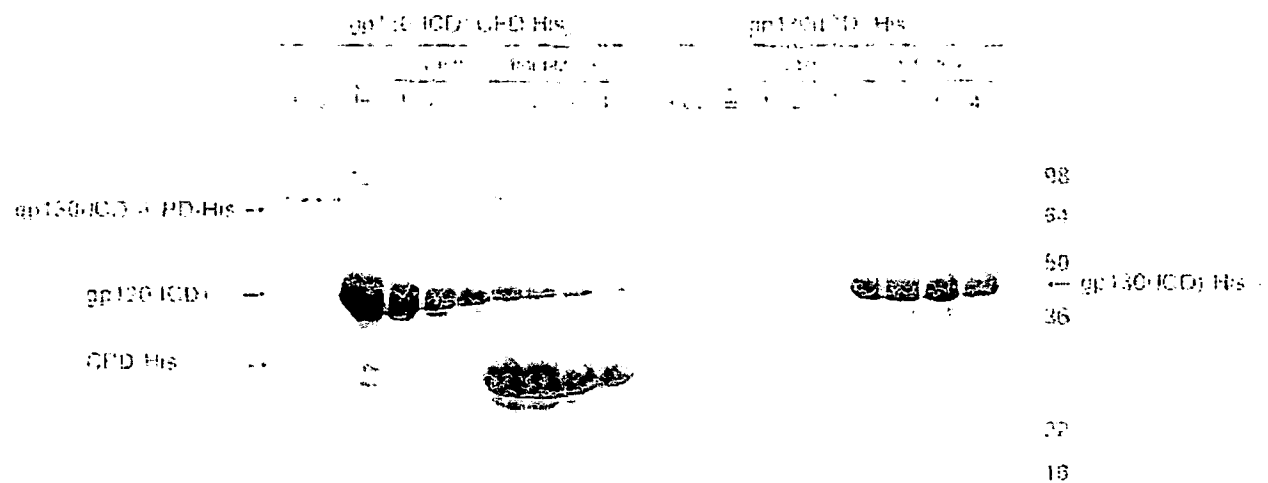


Fig. 4. Purification of the gp130 intracellular domain (ICD) using the CPDHis₆ tag. Gp130 ICD(CPD)-His₆ or gp130(ICD)-His₆ bound to Ni²⁺-NTA resin was incubated with 100 μM InsP₆ for 2 hr at room temperature; the resin was washed four times, followed by elution of Ni²⁺-bound proteins by 200 mM imidazole. Purification fractions were analyzed by SDS-PAGE followed by Coomassie staining. CL, cleared lysate, FT, flowthrough, IP6, elution from InsP₆ incubation.

5/6

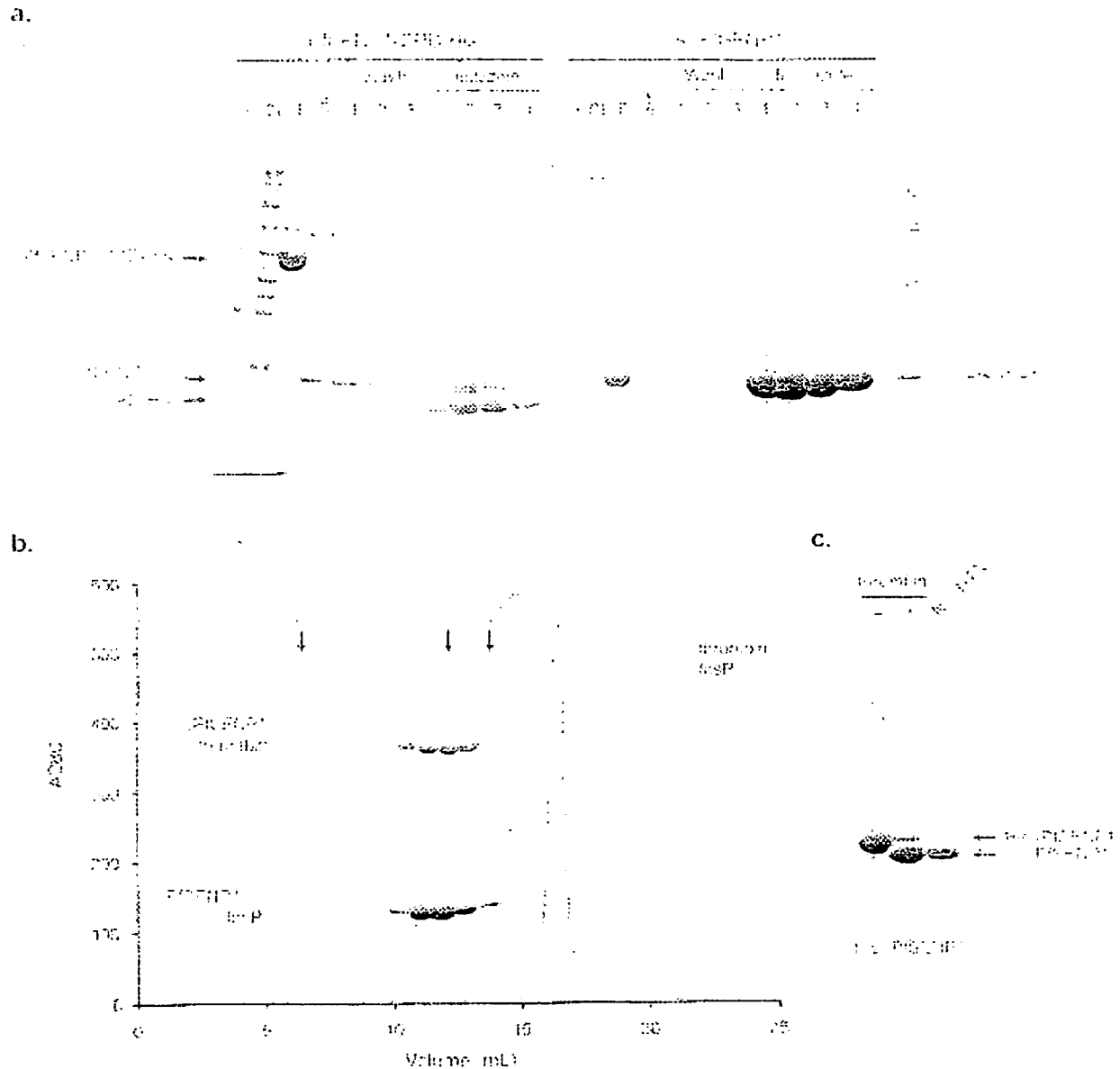


Fig. 5. Purification of PfSEN1 using the CPD system. **(a)** SDS-PAGE analysis using Coomassie staining of *Plasmodium falciparum* SEN1 (PfSEN1, 25 kDa) purification using either the CPD-His₆ or His₆-affinity tags. PfSEN1-CPD-His₆ or His₆-PfSEN1 bound to the Ni²⁺-NTA resin was incubated with 100 μ M InsP₆ for 2 hr at room temperature: the resin washed three times, and wash fractions were collected. Ni²⁺-bound proteins were eluted by adding 200 mM imidazole, +, IPTG induced culture, CL, cleared lysate, E, imidazole elution prior to InsP₆ addition, IP₆, elution from InsP₆, **(b)** UV trace PfSEN1 further purified by gel filtration chromatography following His₆-tag removal. Inset, Coomassie staining of gel filtration fractions

6/6

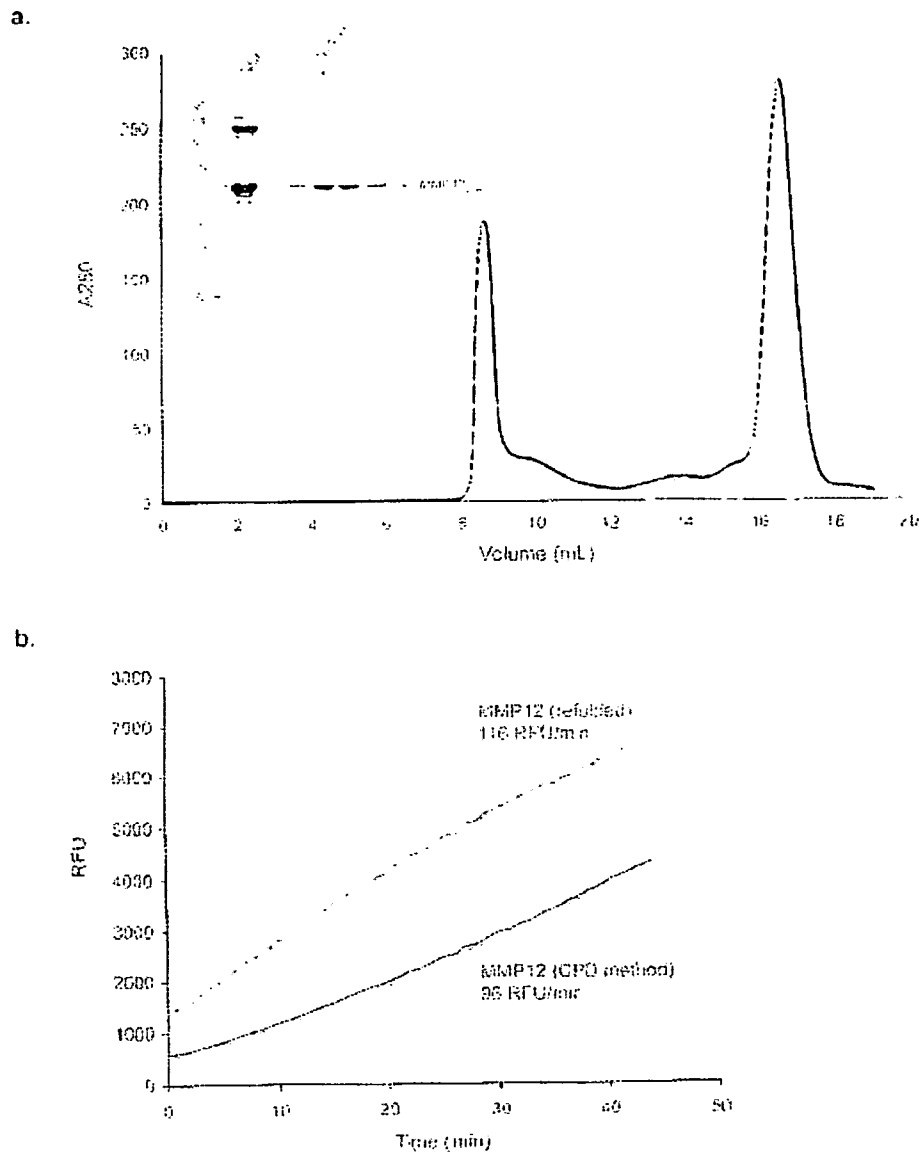


Fig. 6. Additional purification of CPD-cleaved MMP12_{VDAL}. **(a)** Purification of MMP12_{VDAL} by gel filtration chromatography. Inset, Coomassie staining of SDS-PAGE analysis of gel filtration fractions of MMP12_{VDAL}. **(b)** MMP12 Fluorogenic substrate assay. The activity of MMP12 purified under denaturing conditions and refolded (MMP12 (Refolded)) and MMP12_{VDAL} purified using the CPD system against a standard fluorogenic substrate were compared. Comparable rates of fluorogenic substrate cleavage are observed for MMP12 purified by the CPD method relative to the refolding method.