Mini-Review Article

In vivo cleavage of solubility tags as a tool to enhance the levels of soluble recombinant proteins in *Escherichia coli*

Filipe S. R. Silva[1], Sara P. O. Santos[1], Roberto Meyer[1,2], Eduardo S. Silva[1,2], Carina S. Pinheiro[1,2], Neuza M. Alcantara-Neves[1,2], Luis G. C. Pacheco [1,3,*]

[1] Post-Graduate Program in Biotechnology, Institute of Health Sciences, Federal University of Bahia, Salvador-BA, Brazil. [2] Post-Graduate Program in Immunology, Institute of Health Sciences, Federal University of Bahia, Salvador-BA, Brazil. [3] Department of Biotechnology, Institute of Health Sciences, Federal University of Bahia, Salvador-BA, Brazil.

* Correspondence: L. Pacheco (E-mail: luis.pacheco@ufba.br).

Running title: Controlled intracellular processing of recombinant proteins

Word count: ~2,669 words (main text)

16 **Abstract**

17 Recombinant proteins are generally fused with solubility enhancer tags to improve target
18 protein folding and solubility. However, the fusion protein strategy usually requires the use of
19 expensive proteases to perform *in vitro* proteolysis and additional chromatography steps to
20 obtain tag-free recombinant proteins. Expression systems based on intracellular processing of
21 solubility tags in *Escherichia coli*, through co-expression of a site-specific protease, are useful
22 for simplifying the recombinant protein purification process, for screening molecules that fail
23 to remain soluble after tag removal, and to promote higher yields of soluble target protein.
24 Herein, we review controlled intracellular processing (CIP) systems, tailored to produce
25 soluble untagged proteins in *E. coli*. We discuss the different genetic systems available for
26 intracellular protein processing regarding system design features, significant advantages and
27 limitations of the various strategies.

28

29 **Keywords:** Recombinant proteins; controlled intracellular processing; *Escherichia coli;* protein
30 solubility; site-specific protease.

31

Introduction

Escherichia coli is the microorganism of choice for the production of recombinant proteins (Feng et al. 2014). It is estimated that *ca.* 88% of protein structures deposited in Protein Data Bank derived from proteins produced in this host organism (Nettleship et al. 2010). Besides, this bacterial chassis produces more than 30% of FDA-approved biopharmaceuticals. The advantages of the *E. coli* expression systems include minimal requirements of laboratory structure and sterile procedures, short doubling time, high dry-weight yields in recombinant proteins, and straightforward process scale-up (Sezonov et al. 2007). However, a major bottleneck of this bacterial system involves the poor recovery of recombinant proteins in their soluble forms, particularly when working with proteins of eukaryotic origin (Costa et al. 2014).

To overcome solubility problems, a widely used approach is the fusion-tag technology, in which the gene encoding the target protein is fused with the coding sequence of a highly soluble protein. This technology often leads to improvements in solubility and stability of a given recombinant protein of interest, then contributing to a streamlined purification process (Kosobokova et al. 2016). The most commonly used solubility enhancers include Maltose-Binding Protein (MBP), Glutathione-S-Transferase (GST), Thioredoxin A (TrxA), and N Utilization Substance Protein A (NusA). The reasons why these proteins improve the solubilities of their partners are not completely understood to date; however, it is already known that polypeptide chains rich in positively charged amino acids can increase electrostatic repulsion among residues during translation, then avoiding aggregation (Kang et al. 2015). Additionally, the tag-fusion technology can prevent defective mRNA structures when fused to the N-terminal portion of the target sequence, and the solubility enhancer proteins are more likely to be highly soluble than shorter peptides (Waugh 2005).

Nevertheless, removing fusion protein tags requires expensive site-specific proteases, such as Tobacco Etch Virus Protease (TEVp), Human Rhinovirus Protease 3C (HRV 3C), Enterokinase, and Factor Xa. Besides, additional chromatography steps are entailed to remove both fusion tags and protease from target protein eluate (Li 2011). Likewise, some proteins may aggregate after the solubility tag is removed. Therefore, *in vivo* intracellular processing can be useful as a tool to verify molecules that will fail to remain soluble after cleavage is performed (Cesaratto et al. 2016; Kapust and Waugh 2000; Lu and Aon 2014).

In vivo solubility tag cleavage systems

Even though *in vivo* experiments involving TEV protease had been reported back in the 90s (Parks et al. 1995), it was only in the 2000s that co-expression of proteases in *E. coli* was described as a system for controlled intracellular processing (CIP) (Kapust and Waugh 2000) (Fig. 1; Table 1). This system resulted from the co-expression of green-fluorescent protein (GFP) fused to MBP (MBP -TEV cleavage site- GFP) with a modified version of TEVp. The fusion protein-encoding cassette was physically segregated from the TEVp-encoding cassette by cloning in different plasmid vectors (Fig. 1a-b). The fusion protein substrate was inserted into an IPTG inducible vector, while TEVp was cloned in a tetracycline-inducible vector. It was proposed to be more suitable to use two different chemically induced promoters for each gene because they are more versatile since it is possible to control each gene expression by changing inducer concentration. Besides, it is also possible to control the timing that each gene will be expressed (Fig. 1b). The study established 2 hours of delayed expression of TEVp after fusion protein substrate induction by IPTG, leading to significant improvement of passenger protein solubility. It can be crucial to reduce host metabolism burden, which is a concern for protein yield and solubility. Another strategy is the use of a low copy plasmid to drive protease expression, avoiding deleterious

early metabolism burden (Fig. 1b-c). In other cases, insoluble passenger proteins produced in an unfused form in *E. coli,* became soluble when fused with MBP and processed *in vivo* by TEVp. An additional finding revealed CIP systems could be used as a diagnostic tool to determine if the passenger protein will remain soluble when separated from the solubility tag (Kapust and Waugh 2000).

Although double induction may be the key to avoid the metabolic burden, it may complicate large scale processes and make it more expensive and time-consuming. The opposite approach was used in order to build another expression system capable of cleaving solubility tags *in vivo* and produce Diaminopropionate Ammonia-Lyase (DAL) in its soluble form, by co-expressing mutant variants of TEVp (Wei et al. 2012). Instead of using double induction, the same promoter sequence was used for both expression cassettes to be induced at the same time with IPTG (Fig. 1a). This procedure can balance the side effects of using multiple inductions at different times, leading to a more straightforward process. In fact, bacteria co-expressing the TEVpM2 variant showed the highest DAL activity, with TEVpM2 splitting GST-DAL more effectively, and a higher DAL amount was obtained in *E. coli.* On the other hand, a significant decrease in cell growth was reported after only 5-6 hours post-induction, due to metabolic burden arising from overexpression of the two recombinant proteins (TEVp and DAL). Additionally, it is also possible that using multiple *lac* operator sites will require more *LacI* repressor protein, which may lead to leakiness and anticipate metabolic stress.

Systems induced by physical rather than chemical stimuli may be another way to tackle the metabolic burden and keep the recombinant process simple. Temperature-sensitive pHsh promoter was used together with a pT7-*lacO* promoter in order to produce a controlled intracellular processing system (Feng et al. 2014) (Fig. 1e). This physically inducing system displayed successful results during the co-expression of human rhinovirus protease 3C

(HRV3C) by removing solubility tags from target proteins in *E. coli*. The reporter gene Enhanced GFP (EGFP) was fused with Trx, DsbA, GST, Nus, TF, and MBP. IPTG was used to activate fusion substrate expression. pHsh promoter, which is activated by alternative sigma factor 32 (σ^{32}) when the temperature is changed to 42°C, was used to control the expression of HRV3C protease. HRV3C was also fused to the GST tag in order to guarantee protease solubility *in vivo*. GST-HRV3C expression was induced by increasing temperature to 42° C for 1h, and then the target protein substrate was induced for 5 hours. The target protein was obtained without solubility tag and was purified by a single step of nickel-NTA affinity chromatography. Results showed high levels of purified native EGFP when fused with the Trx tag. They repeated their protocol using Bluetongue virus (BTV) protein, which is known as difficult to express, accumulated in inclusion bodies when directly expressed without fusion tag. However, a soluble pattern was noticed when BTV protein was produced fused to a solubility tag. Although 42 °C is not the ideal temperature for *E. coli* growth, this temperature activating-promoter can make double induction cheaper than dual chemical induction. Therefore, these physically induced systems allow for easy identification of a more suitable solubility enhancer partner and can be a cheaper option to remove tags *in vivo* prior purification.

Despite CIP systems displayed effectiveness by enhancing native protein solubility, it generally requires two plasmids to both fusion protein substrate and protease. Consequently, steps of cloning, plasmid transformation, and antibiotic selection generally have to be performed at least twice, increasing method complexity. As an attempt to simplify cloning steps, a TEVp expression cassette was inserted into *the E. coli* chromosome by λ-Red recombineering (Luo et al. 2015) (Fig. 1d). To accomplish that, overlap extension PCR was used to insert downstream and upstream homology arms (HA) of the chromosomal *malE* gene from *E. coli* to the cassette T7-TEV-*aacC1*. Then, red-competent BL21 cells were

electroporated with HA flanked T7-TEV-aacC1. The knock-in strain was termed *E. coli* LS2416 and the expression of two target proteins in LS2416 cells, GFP and N-TIMP (N-terminal inhibitory domain of human tissue inhibitor of metalloproteases-2), confirmed the functionality of this chromosome-based system. Without solubility tag, GFP was expressed slightly soluble, while N-TIMP was produced almost entirely insoluble. When MBP was fused upstream of each gene with the TEVp recognition site (ENLYFQ↓G) between the MBP and target protein, GFP was totally soluble, whereas N-TIMP exhibited high soluble percentage. The main advantage of this chromosome-based system is the requirement of only one vector to be cloned. Besides, a lower metabolic load is also expected due to the use of resistance genes for plasmid maintenance.

Optimizing proteases for *in vivo* proteolysis

Highly specific proteases are essential for building CIP systems. The purpose of using these enzymes is the reduction of nonspecific cleavage occurrence. Despite low stability and solubility of wild type TEV protease, this is the most co-expressed protease, due to the high specificity of its target cleavage site ENLYFQG/S (Kapust et al. 2001; Parks et al. 1995). However, wild type TEVp may be problematic for *in vivo* processing giving rise to low yield and solubility because of autoproteolysis that generates a truncated enzyme form with reduced activity (Kapust et al. 2001; Wei et al. 2012). Kapust et al. (2001) (Kapust et al. 2001) solved the autoproteolysis problem by mutating internal cleavage site amino acids and obtained TEVp S219V not only as a more stable protease but also a more efficient catalyst. Alternative TEVp variant (TEV_{sh}) containing mutations T17S/N68D/ I77V was described as more soluble *in vitro* (van den Berg et al. 2006). Another modified TEVp with mutations L56V/S135G remained soluble at higher concentrations and displayed improved catalytic activity compared with TEVp S219V (Cabrita et al. 2007). In order to obtain optimized *in vivo* proteolysis, Wei

et al. (2012) (Wei et al. 2012) combined all these mutations in a new TEVp variant, the TEVpM2. In their study, the coding gene for Emerald GFP (EmGFP) was fused downstream to TEVp variants. Fluorescence analysis revealed that *E. coli* expressing variant TEVpM2 (T17S/L56V/N68D/I77V/S135G) had higher fluorescence than other variants, suggesting that mutations in TEVp sequence resulted in higher *in vivo* solubility.

In contrast, instead of using TEVp, Nallamsetty et al. (2004) (Nallamsetty et al. 2004) performed cleavage of fusion proteins with the Tobacco Vein Mottling Virus protease (TVMVp), both *in vivo* and *in vitro*. Similar to TEVp, TVMVp is active in a wide range of ionic strength, highly active at low temperatures, and even has comparable catalytic efficiency to TEVp. Moreover, both enzymes have high proteolytic stringency, so they do not cleave in nonspecific sites. On the other hand, TEV and TVMV proteases display different sequence specificities, not cleaving each other's recognition site. While TEVp favorite cleaving site is ENLYFQS, the canonical target site for TVMV is ETVRFQS. Therefore, TVMVp can be useful to replace TEVp when fusion substrate has a peptide sequence that resembles the TEVp recognition site. Likewise, both proteases can be used together, allowing for the removal of two distinct tags. Additionally, TVMVp has the advantage of not cleaving itself into inactive fragments, as wild type TEVp does.

CIP systems can make protein purification easier

There are several well-established, straightforward semi-automated protocols for high throughput protein purification. However, most of them are adapted for unfused proteins. In order to purify fused proteins, two consecutive immobilized metal affinity chromatographic (IMAC) steps are often necessary. The first IMAC purifies the fusion protein, and following site-specific proteolysis, a second subtractive IMAC is generally implemented to remove the cleaved tag and the site-specific protease (Wang et al. 2015). Donnelly et al. (2006) (Donnelly

et al. 2006) used the CIP approach to obtain untagged proteins faster. They took advantage of TEVp and TVMVp distinct specificities and modified target fusion substrate to contain two distinct protease recognition sites (MBP–TVMVsite–his6-tag–TEVsite-target protein). That strategy is slightly different from those used by most CIP systems, where his-tag is not separated from the target protein at the end of the purification process. They co-expressed TVMVp with 16 protein substrates enhanced by fusion with MBP, in which the yield of pure protein failed after the second IMAC when they tried to apply traditional protocols. TVMVp was produced constitutively in BL21, and the released MBP was observed in all SDS-PAGE lanes, showing that all fusion protein substrates were cleaved by TVMVp *in vivo*. Based on the abundance of proteins on soluble fraction, 10 out of 16 proteins were sufficiently soluble, and two of them were found in both soluble and insoluble fractions. This finding corroborates that CIP systems can be used to eliminate false positives without using *in vitro* cleavage step for screening for molecules that become insoluble when separated from MBP. The remaining proteins were purified by standard protocols, highly efficient for his-tagged proteins. This dual tag approach demonstrated that using *in vivo* proteolysis can highly improve purity and yield in semi-automated protocols.

Improving the solubility of target proteins using synthetic genetic circuits

In a recent study, our group demonstrated that a genetic regulatory cascade could be used to control the *in vivo* removal of a solubility tag from a fusion recombinant protein using a single plasmid (Silva et al. 2019) (Fig. 1f). To assemble this expression vector, termed pSOLC, the genetic modules were built as follow: (i) first module contained the sequence coding for a fusion target protein consisting of the solubility tag KDPG aldolase (EDA) as well as a Gly-Ser-Gly-Ser flexible linker, and a canonical TEVp cleavage recognition site; (ii) the second module encoded the target protein EGFP and was placed under the control of

module #1; (iii) the third module was designed to express the TetR repressor and TEVp, permitting not only the release of the target protein but also preventing the collapse of the genetic circuit due to the accumulation of regulators. When all synthetic biological components were cloned into a single plasmid, our genetic system's functionality was superior to the use of two different plasmids, highlighting a total soluble recombinant protein yield of 272.0 ± 60.1 μg/mL of culture. In addition, free EGFP composed 46.5% of the total purified protein fraction, separating easily from remaining fusion EDA-EGFP. The advantage of this regulatory cascade is the intrinsic interaction of the genetic elements following a single chemical input, leading to the simultaneous production of a fusion recombinant protein and a site-specific protease, which then cleaves the solubility tag from the target protein. However, a significant part of the recombinant protein remained in its fusion form after intracellular processing, showing a limitation of the approach. The strategy merits additional developments due to the requirement of a single induction with only one inducer as well as the possibility of use in different cell lineages (Silva et al. 2019).

Conclusion

It is unlikely that CIP systems will replace standard expression systems for the production of fused recombinant proteins, but it is worth considering them as an alternative to solve problems related to protein solubility and purification methodology. Three main obstacles may limit the application of CIP systems in recombinant protein production: (i) co-expression of proteases may lead to metabolic burden and, consequently, poor soluble protein yield; (ii) the use of multiple inductions could make target protein production more expensive and complicated; and (iii) cloning steps can be time and resource-consuming when various genes have to be inserted in an *E. coli* strain. As discussed herein, the enhanced metabolic load is boosted when cells are forced to overexpress at the same level, two heterologous genes

using the same promoter and similar plasmid copy numbers. However, stress-responsive promoters might be an alternative to avoid the metabolic burden and make the process cheaper and less complicated.

Authors' contributions

FSRS and LGCP conceived the manuscript idea. FSRS led the writing and construction of figure. SPOS and ESS contributed with literature search and manuscript writing. RM, CSP, NMAN and LGCP reviewed the draft versions, read, and approved the final manuscript.

Funding

FSRS is the recipient of a scholarship from FAPESB. LGCP and NMAN are recipients of research fellowships from the National Council for Scientific and Technological Development of Brazil (CNPq). Work at our group is partially funded through the FAPESB/CNPq PRONEM-2014 grant awarded to LGCP.

Conflicts of interest

The authors report no conflict of interest.

References

Afanador GA, Matthews KA, Bartee D, Gisselberg JE, Walters MS, Freel Meyers CL, Prigge ST. 2014. Redox-dependent lipoylation of mitochondrial proteins in Plasmodium falciparum. *Mol. Microbiol.* **94**:156–171.

Berg S Van Den, Ake P-, Berglund H. 2006. Improved solubility of TEV protease by directed evolution **121**:291–298.

Bujnicki JM, Prigge ST, Caridha D, Chiang PK. 2003. Structure, evolution, and inhibitor

255 interaction of S-adenosyl-L-homocysteine hydrolase from Plasmodium falciparum.
256 *Proteins Struct. Funct. Genet.* **52**:624–632.

257 Cabrita LD, Gilis D, Robertson AMYL, Dehouck Y, Rooman M, Bottomley SP. 2007.
258 Enhancing the stability and solubility of TEV protease using in silico design:2360–2367.

259 Cesaratto F, Burrone OR, Petris G. 2016. Tobacco Etch Virus protease: A shortcut across
260 biotechnologies. *J. Biotechnol.*

261 Costa S, Almeida A, Castro A, Domingues L. 2014. Fusion tags for protein solubility,
262 purification, and immunogenicity in Escherichia coli: The novel Fh8 system. *Front.*
263 *Microbiol.* **5**:1–20.

264 Didovyk A, Borek B, Hasty J, Tsimring L. 2016. Orthogonal Modular Gene Repression in
265 Escherichia coli Using Engineered CRISPR/Cas9. *ACS Synth. Biol.* **5**:81–88.

266 Donnelly MI, Zhou M, Millard CS, Clancy S, Stols L, Eschenfeldt WH, Collart FR,
267 Joachimiak A. 2006. An expression vector tailored for large-scale, high-throughput
268 purification of recombinant proteins. *Protein Expr. Purif.* **47**:446–454.

269 Du Y, Gisselberg JE, Johnson JD, Lee PJ, Prigge ST, Bachmann BO. 2010. Lactococcus
270 lactis fabH, encoding ??-ketoacyl-acyl carrier protein synthase, can be functionally
271 replaced by the plasmodium falciparum congener. *Appl. Environ. Microbiol.* **76**:3959–
272 3966.

273 Eisenmesser EZ, Kapust RB, Nawrocki JP, Mazzulla MJ, Pannell LK, Waugh DS, Byrd RA.
274 2000. Expression, purification, refolding, and characterization of recombinant human
275 interleukin-13: Utilization of intracellular processing. *Protein Expr. Purif.* **20**:186–195.

276 Feng Y, Xu Q, Yang T, Sun E, Li J, Shi D, Wu D. 2014. A novel self-cleavage system for

277 production of soluble recombinant protein in Escherichia coli. *Protein Expr. Purif.*
278 **99C**:64–69. http://www.ncbi.nlm.nih.gov/pubmed/24727155.

279 Kang YS, Song JA, Han KY, Lee J. 2015. Escherichia coli EDA is a novel fusion expression
280 partner to improve solubility of aggregation-prone heterologous proteins. *J. Biotechnol.*
281 **194**:39–47. http://dx.doi.org/10.1016/j.jbiotec.2014.11.025.

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

285 Kapust RB, Waugh DS. 2000. Controlled Intracellular Processing of Fusion Proteins by TEV
286 Protease **318**:312–318.

287 Kosobokova EN, Skrypnik KA, Kosorukov VS. 2016. Overview of Fusion Tags for
288 Recombinant Proteins **81**.

289 Li Y. 2011. Self-cleaving fusion tags for recombinant protein production. *Biotechnol. Lett.*
290 **33**:869–881.

291 Lu JZ, Muench SP, Allary M, Campbell S, Roberts CW, Mui E, McLeod RL, Rice DW,
292 Prigge ST. 2007. Type I and type II fatty acid biosynthesis in Eimeria tenella: Enoyl
293 reductase activity and structure. *Parasitology* **134**:1949–1962.

294 Lu Q, Aon JC. 2014. Co-expression for intracellular processing in microbial protein
295 production. *Biotechnol. Lett.* **36**:427–441.

296 Luo X, Li L, Chai M, Zhang Q, Shang G. 2015. Escherichia coli BL21 (DE3) chromosome-
297 based controlled intracellular processing system for fusion protein separation. *J.*
298 *Microbiol. Methods* **114**:35–37. http://dx.doi.org/10.1016/j.mimet.2015.04.013.

McFeeters RL, Altieri AS, Cherry S, Tropea JE, Waugh DS, Byrd RA. 2007. The high-precision solution structure of Yersinia modulating protein YmoA provides insight into interaction with H-NS. *Biochemistry* **46**:13975–13982.

Muench SP, Prigge ST, Zhu L, Kirisits MJ, Roberts CW, Wernimont S, McLeod R, Rice DW. 2006. Expression, purification and preliminary crystallographic analysis of the Toxoplasma gondii enoyl reductase. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **62**:604–606.

Nallamsetty S, Kapust RB, Tözsér J, Cherry S, Tropea JE, Copeland TD, Waugh DS. 2004. Efficient site-specific processing of fusion proteins by tobacco vein mottling virus protease in vivo and in vitro. *Protein Expr. Purif.* **38**:108–115.

Nallamsetty S, Waugh DS. 2006. Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. *Protein Expr. Purif.* **45**:175–182.

Nettleship JE, Assenberg R, Diprose JM, Rahman-huq N, Owens RJ. 2010. Recent advances in the production of proteins in insect and mammalian cells for structural biology. *J. Struct. Biol.* **172**:55–65. http://dx.doi.org/10.1016/j.jsb.2010.02.006.

Parks TD, Howard ED, Wolpert TJ, Arp DJ, Dougherty WG. 1995. Expression and Purification of a Recombinant Tobacco Etch Virus NIa Proteinase: Biochemical Analyses of the Full-Length and a Naturally Occurring Truncated Proteinase Form. *Virology*.

Raran-Kurussi S, Waugh DS. 2016. A dual protease approach for expression and affinity purification of recombinant proteins. *Anal. Biochem.* **504**:30–37.

Schubot FD, Cherry S, Austin BP, Tropea JE, Waugh DS. 2005. Crystal structure of the protease-resistant core domain of Yersinia pestis virulence factor YopR. *Protein Sci.*

14:1679–1683. http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=15930010&retmode=ref&cmd=prlinks%5Cnpapers3://publication/doi/10.1110/ps.051446405.

Sezonov G, Joseleau-Petit D, D'Ari R. 2007. Escherichia coli physiology in Luria-Bertani broth. *J. Bacteriol.*

Silva F, Santos S, Meyer R, Alcantara-Neves N, Pinheiro C, Pacheco LGC. 2019. Single-Input Regulatory Cascade for in vivo Removal of the Solubility Tag in Fusion Recombinant Proteins Produced by Escherichia coli. *Front. Bioeng. Biotechnol* **7**.

Wang HZ, Chu ZZ, Chen CC, Cao AC, Tong X, Ouyang C Bin, Yuan QH, Wang MN, Wu ZK, Wang HH, Wang S Bin. 2015. Recombinant passenger proteins can be conveniently purified by one-step affinity chromatography. *PLoS One* **10**:1–17.

Waugh DS. 2005. Making the most of affinity tags **23**.

Wei L, Cai X, Qi Z, Rong L, Cheng B, Fan J. 2012. In vivo and in vitro characterization of TEV protease mutants. *Protein Expr. Purif.* **83**:157–163. http://dx.doi.org/10.1016/j.pep.2012.03.011.

Figure legend

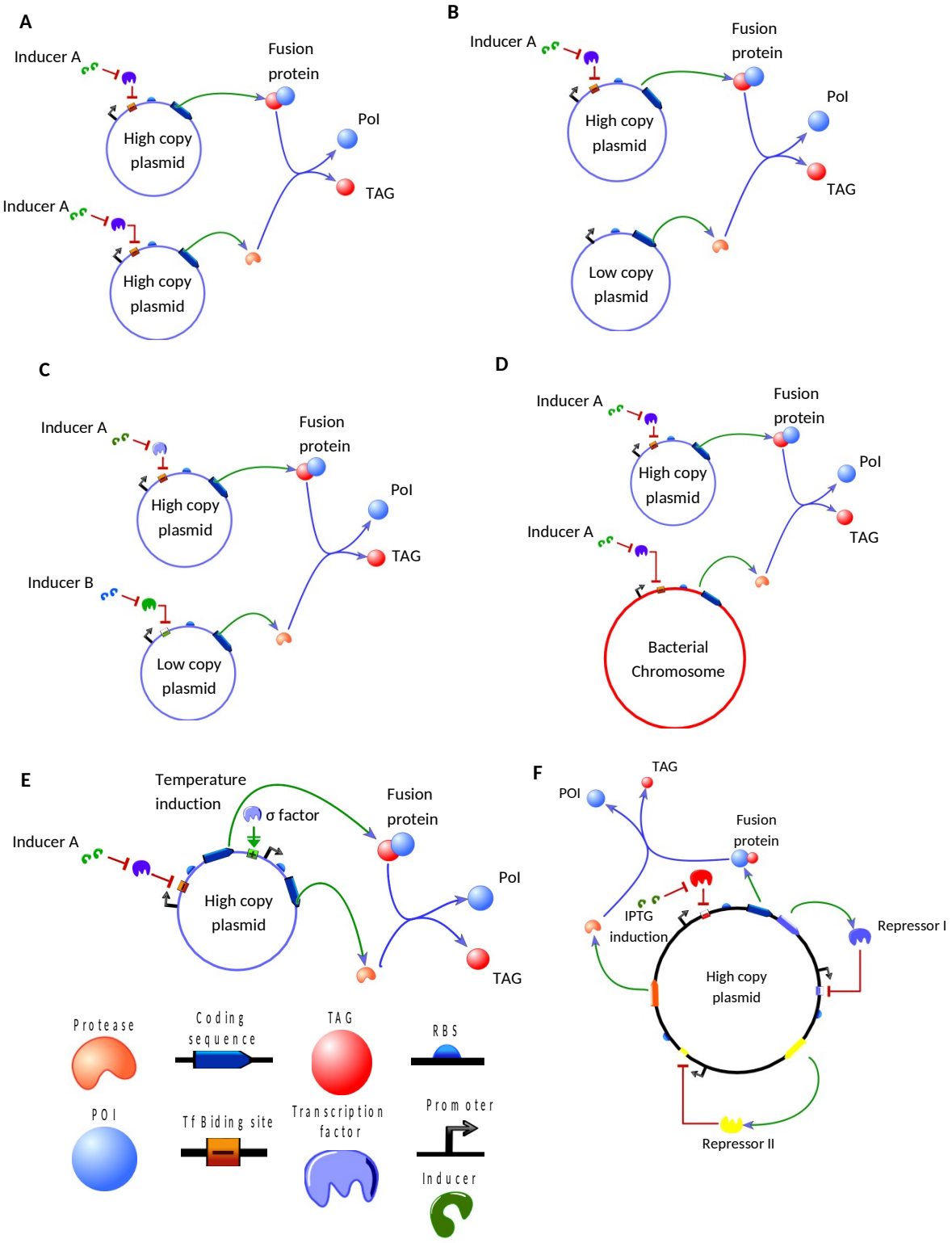
Fig.1 Schematic of controlled intracellular processing (CIP) systems in *E. coli*. A) Simultaneous double induction of fusion protein and site-specific protease using the same inducer molecule (*e.g*., IPTG). **B)** Expression of the fusion protein of interest (PoI) is regulated by a chemically inducible promoter, whereas the site-specific protease is constitutively expressed from a low copy number plasmid. **C)** Double induction at different times using an inducer molecule (such as IPTG) for activating fusion protein production and a second inducer (*e.g.*, Anhydrotetracycline or Arabinose) for stimulating protease expression. **D)** The bacterial strain contains a protease expression unit in its chromosome to produce the protease, and the fusion protein is expressed from a plasmid. **E)** Target protein fusion is chemically induced (*e.g*., IPTG), and protease expression is activated under stress condition by sigma (σ) transcription factor (*e.g.*, turning the temperature to 42ªC). **F)** The production of both PoI and protease are activated by the induction of a single promoter by IPTG. IPTG addition to the media generates the translation of the fusion protein and the first repressor protein. The first repressor inhibits production of the second repressor protein through binding to the promoter's operator site. This releases expression of the site-specific protease to perform cleavage of the fusion protein.

Table 1. Overview of studies that have used controlled intracellular processing for recombinant protein production

Protease/ Tags	Target protein	Protease induction	Refs.
TEV protease			
MBP	Green Fluorescent Protein (GFP), TIMP N-terminal inhibitory domain of human tissue inhibitor of metalloproteinases-2 (TIMP), Human cyclindependent kinase 4 inhibitor (p16), and oncoprotein encoded by human papillomavirus (E6).	Anhydrotetracycline	(Kapust and Waugh, 2000)
	human IL-13 (hIL-13)	Constitutive	(Eisenmesser et al., 2000)
	S-adenosylho- mocysteine hydrolase	Constitutive	(Bujnicki et al., 2003)
	Enoyl reductase enzyme (ENR)	Constitutive	(Muench et al., 2006)
	Type III secretion system effector (YopR)	Anhydrotetracycline	(Schubot et al., 2005)
	Enoyl reductase (ENR)	Constitutive	(Lu et al., 2007)
	Yersinia modulating protein (YmoA)	Constitutive	(McFeeters et al., 2007)
	B-Ketoacyl-acyl carrier protein (ACP)	Constitutive	(Du et al., 2010)
	E. coli lipoate ligase (EcLplA); lipoate ligase 1 (LipL1); - lipoate ligase 2 (LipL2).	Constitutive	(Afanador et al., 2014)
	dCAS9	Isopropyl β-D-1-thiogalactopyranoside	(Didovyk et al., 2016)
	Green Fluorescent Protein (GFP), N-terminal inhibitory domain of human tissue inhibitor of metalloproteases-2 (TIMP).	Isopropyl β-D-1-thiogalactopyranoside	(Luo et al., 2015)
MBP and NusA	Green Fluorescent Protein (GFP), glyceraldehyde 3-phosphate dehydrogenase (G3PDH), dihydrofolate reductase (DHFR), rhodanese, luciferase, tissue inhibitor of metalloproteinases-1 (TIMP), YopN, YopJ, YopT, YscK, YscL, and YscO.	Constitutive	(Nallamsetty and Waugh, 2006)
acyl carrier protein (ACP)	Glucokinase (GlcK), α-Amylase (Amy) and GFP	Arabinose (0.2%) or IPTG (0.4 mM)	(Wang et al., 2015)
GST	Diaminopropionate ammonia-lyase (DAL), maize 2-Cys peroxiredoxin A (Prx).	Isopropyl β-D-1-thiogalactopyranoside	(Wei et al., 2012)

KDPG Aldolase (EDA)	Enhanced Green Fluorescent Protein (EGFP)	Isopropyl β-D-1-thiogalactopyranoside	(Silva et al., 2019)
TVMV protease			
MBP, GST, and TRX.	Transcription termination/antitermination protein NusG	Anhydrotetracycline	(Nallamsetty et al., 2004)
MBP	Hypothetical (*B. cereus*), Cytoplasmic protein, Regulatory protein, RNA ligase, Cytoplasmic protein Hypothetical, Inner membrane proteinC, Inner membrane protein, Cytoplasmic protein C, Hydrophilic protein, Regulatory proteinC, SAM methyltransferaseC, Galactitol enzyme IIA, Transport protein, Hypothetical, Urease accessory protein	Constitutive	(Donnelly et al., 2006)
Human rhinovirus 3C (HRV3C) protease			
MBP	Super folder Green fluorescent protein sf-GFP	L-arabinose (0.2%)	(Raran-Kurussi and Waugh, 2016)
Trx, DsbA, GST, Nus, TF, and MBP	EGFP, Bluetongue virus (BTV) protein	Temperature (42 °C)	(Feng et al., 2014)

Fig 1


A
Inducer A
Fusion protein
High copy plasmid
PoI
TAG
Inducer A
High copy plasmid
B
Inducer A
Fusion protein
High copy plasmid
PoI
TAG
Low copy plasmid
C
Inducer A
Fusion protein
High copy plasmid
PoI
TAG
Inducer B
Low copy plasmid
D
Inducer A
Fusion protein
High copy plasmid
PoI
TAG
Inducer A
Bacterial Chromosome
E
Temperature induction
σ factor
Fusion protein
Inducer A
High copy plasmid
PoI
TAG
Protease
Coding sequence
TAG
RBS
POI
Tf Biding site
Transcription factor
Promoter
Inducer
F
TAG
POI
Fusion protein
IPTG induction
Repressor I
High copy plasmid
Repressor II