The pCri System: A Vector Collection for Recombinant Protein Expression and Purification

Theodoros Goulas1a, Anna Cuppari1a, Raquel Garcia-Castellanos1b, Scott Snipas2, Rudi Glockshuber3, Joan L. Arolas1, F. Xavier Gomis-Rüth1a

1 Proteolysis Lab, Molecular Biology Institute of Barcelona, CSIC, Barcelona Science Park, Helix Building, Barcelona, Spain, 2 Sanford-Burnham Medical Research Institute, La Jolla, California, United States of America, 3 Institute of Molecular Biology and Biophysics, Department of Biology, Zurich, Switzerland

Abstract

A major bottleneck in structural, biochemical and biophysical studies of proteins is the need for large amounts of pure homogeneous material, which is generally obtained by recombinant overexpression. Here we introduce a vector collection, the pCri System, for cytoplasmic and periplasmic/extracellular expression of heterologous proteins that allows the simultaneous assessment of prokaryotic and eukaryotic host cells (Escherichia coli, Bacillus subtilis, and Pichia pastoris). By using a single polymerase chain reaction product, genes of interest can be directionally cloned in all vectors within four different restriction sites at the 5’ end and multiple cloning sites at the 3’ end. In this way, a number of different fusion tags but also signal peptides can be incorporated at the N- and C-terminus of proteins, facilitating their expression, solubility and subsequent detection and purification. Fusion tags can be efficiently removed by treatment with site-specific peptidases, such as tobacco etch virus proteinase, thrombin, or sentrin specific peptidase 1, which leave only a few extra residues at the N-terminus of the protein. The combination of different expression systems in concert with the cloning approach in vectors that can fuse various tags makes the pCri System a valuable tool for high throughput studies.

Introduction

Researchers performing biochemical, biophysical and biological studies on proteins commonly require large amounts of pure homogeneous material, which cannot usually be purified from natural sources. Alternatively, proteins are over-expressed heterologously in various systems incorporating host cells of bacterial, yeast, insect, or mammalian origin [1–3]. A critical step in protein production, after target selection, is to examine as many parameters as possible and to identify the most promising strategy for protein expression and purification with a minimum of resources and time. Prior information on the protein of interest is crucial. An extensive search in databases such as NCBI (http://www.ncbi.nlm.nih.gov), UniProt (http://www.uniprot.org) and PDB (http://www.pdb.org) for known homologous proteins may identify possible problems and appropriate solutions for subsequent experiments. In addition, it is advisable to test protein orthologs of different origin, including distantly related or unrelated species (bacteria, archaea, and eukaryotes). At this point, analysis of the primary and secondary structure of both the encoding mRNA and the translated polypeptide may anticipate downstream problems.

There is a plethora of freely available software and databases for identifying protein families and sequence conservation patterns (PFAM) [4], putative signal peptides (SPs; SignalIP) [5], lipoboxes (DOLOP) [6], glycosylation, phosphorylation and other posttranslational modifications [7], transmembrane domains (TOPCONS, TMHMM, COUCHOPUS) [8–10], and unfolded/disordered regions (DisEMBL, PONDOR, PSIPRED Protein Sequence Analysis Workbench) [11–13]. Protein location within the cell, i.e. cytoplasmic, periplasmic, or extracellular (PSORT, http://psort.hgc.jp), provides an indication of the requirements of the protein for proper folding, including disulfide bond formation and the need for special chaperons in each cellular compartment [14–16]. Further prediction of the secondary structure content (JNET, LOMETS) [17,18] can give clues about possible protein domains and motifs, a characterisation which may prove useful for chopping full-length multi-domain proteins into globular moieties. In general, successful recombinant protein expression depends on the removal of wild-type SP, lipoboxes, postranslational signals, low-complexity regions, hydrophobic residues at the protein termini and membrane spanning regions, while conserving the boundaries of globular domains [19].
In parallel, eDNA characterisation is important in designing the cloning strategy and identifying potential problems at the transcriptional and translational levels. Although these processes are affected by a number of exo- and endo-nucleases, the stability of the resulting mRNA is critical in protein expression experiments [20]. mRNA can be protected by introducing sequences at the 5′ untranslated regions (UTRs) and stem loop structures at the 3′ UTRs [21]. The GC base content (>70%) may affect levels of expression and can be easily determined by sequence analysis software. Rare codons (GCUA 2.0) [22], especially consecutive ones, are frequently found in heterologous genes and may lead to translational errors due to ribosomal stalling [23,24]. Such codon bias can be remedied by replacing selected codons or, if necessary, by overall gene optimisation using appropriate software (OPTIMIZER) [25]. Once the above requirements are fulfilled, the gene can be inserted into the vector by directional cloning using restriction enzymes that do not cut within the gene sequence (NEBcutter) [26]. Efficiency of translation termination can be increased by introducing strong stop codons (UAA, especially in context when followed by a U base, or consecutive ones) at the end of the translated gene [27]. Although present in many expression vectors, transcription terminators can be included downstream of the transcribed gene if instability is predicted [28]. Finally, sources of cDNA can be found in the Mammalian Gene Collection (http://mgc.nci.nih.gov/) and at the home page of Culture Collection of the World (http://www.ecotao.com/holism/agric/hpcc.html).

No expression system is generic for all target proteins, so both bacterial and eukaryotic systems need to be explored. *Escherichia coli* provides the cheapest expression host, and it is the most widely used but its machinery is not as sophisticated as that of eukaryotic hosts, and it cannot always express well folded proteins of variable origin [15]. Other alternatives often need to be tested, including bacterial systems such as *Bacillus subtilis* [29] and more advanced eukaryotic systems such as the yeasts *Pichia pastoris* [1] and *Saccharomyces cerevisiae* [30], the baculovirus expression system in insect cells [3], mammalian cells [31], or cell-free systems using prokaryotic extracts [32], which have highly variable cost-efficiency ratios.

With *E. coli* alone, many variables can be tested in order to improve expression levels and achieve proper protein folding [2,33]. A number of specialised strains carrying mutations [34,35] or plasmids that co-express proteins favouring expression at the transcriptional or translational level (e.g. pRARE or plysE/plysS) are available [24,36]. Coupled expression of exogenous chaperones can assist in proper folding and prevent protein aggregation [37,38]. Expression can also be influenced by other parameters, such as the culture method (e.g. batch fermentation, fed batch and dialysis fermentation) [39], cell growth media composition (lysogeny broth (LB), the enriched terrific broth (TB), two times yeast and tryptone broth (2×YT), and auto-induction media) [40], and culture conditions like temperature (18–37°C), shaking, aeration and other physical variables. All these factors can affect production levels, secretion, protein folding, solubility and host proteolytic activity [41,42].

The many systems for introducing fusion tags currently available were originally developed to facilitate the detection and purification of recombinant proteins. Tags such as polyhistidine (His6-tag) and streptavidin-binding peptide (Strep-tag) allow purification by affinity chromatography and protein detection by Western blotting [43,44], and others such as C-terminally fused green fluorescent protein (GFP) are an indispensable tool for membrane protein biochemists [45]. Finally, several studies have shown that the introduction of tags at the N- or C-terminus of proteins can improve expression levels by providing an optimized environment for translation initiation and mRNA protection, protein solubility [46–48], and carrier-driven crystallisation experiments [49].

Here we present a collection of vectors with which various expression systems and fusion tags can be evaluated simply and effectively. We examine the applicability of this system and provide several test cases, which support its robustness and versatility. This vector collection, which has been extensively tested and modified, is freely available to the scientific community under Addgene (https://www.addgene.org).

**Materials and Methods**

**Genetic manipulations and vector preparation**

Three series of vectors were generated on the basis of vectors available from the European Molecular Biology Laboratory (pETMBP-1a, pETTRX-1a, and pETGST), Novagen (pET-26b, and pET-28a), MoBiTec (pHT-01, and pHJ-43), Invitrogen (pPICZA and pPIC2A), and from the Gockshuber laboratory (pRBI-DsbC) [50]. The inserted sequences for pCri-11, 13, and 14 were amplified from pET-15b-SUMO1 [51], pMIS3.0E [52], and pKLSLt [53], respectively. All vectors were prepared for directional cloning in *NeoI* or *NdeI* restriction sites at the 5′ end and in *XhoI* at the 3′ end. The gene coding for GFP (UniProt code: B6UPG7; 729 bp), including a multiple cloning site (MCS; from pETMBP-1a; 52 bp), was introduced into all vectors. The insert was cloned between the *NeoI* or *NdeI* and *XhoI* restriction sites and was modified to contain an *McsI* or *NheI* restriction site immediately after the *NeoI* and *NdeI* sites, respectively. Standard cloning techniques were used throughout [54]. Polymerase chain reaction (PCR) primers and DNA modifying enzymes were purchased from Sigma-Aldrich and Thermo-Scientific, respectively. PCR was performed using Phusion high-fidelity DNA polymerase (Thermo-Scientific) according to the manufacturer's instructions and following a standard optimisation step of a thermal gradient in each reaction. For vector preparation, a number of insertions and mutations introduced or eliminated nucleotide sequences. We followed a PCR-based strategy described elsewhere [55], including a *DpnI* digestion step to remove parental DNA. Digestion with restriction enzymes was carried out according to standard protocols. When necessary, a second round of digestion was performed before the final DNA purification step. DNA was purified from PCR reactions, enzymatic reactions, agarose gel band extractions, and vector extractions using OMEGA-Biotek purification kits. Chemically competent *E. coli* DH5α, BL21 (DE3), and Origami 2 (DE3) cells (Novagen) were prepared and transformed following Hanahan method [56]. Competent cells of P. pastoris KM71H (Invitrogen) and *B. subtilis* WB800N (MoBiTec) were prepared according to the manufacturer’s instructions.

**Protein expression and purification**

For expression trials, *mecR1* (UniProt code: P0A0Bl; an integral-membrane metallopeptidase) was cloned into vector pCri-9a and 13a; the gene coding for fragilysin (UniProt code: O86049; Ala212-Asp397; a soluble metalloendopeptidase) into pCri-1a, 4a, 6a and 8a; gfp into pCri-1a, 4a, 6a, 8a, 11a, and 14a; the gene coding for carboxypeptidase A2 (CPA2; UniProt code: P48802; Leu19-Tyr419; a soluble metalloendopeptidase) into pCri-8a, 9a, 16a, and 18a; and the gene coding for peptide-N-glycosidase F (PNGase F; UniProt code: P21163; Ala41-Asp354; a soluble glycosidase) into pCri-4a and 8a. The constructs were transformed in *E. coli* BL21 (DE3), Origami 2 (DE3), or *B. subtilis*...
cells and plated on LB plates supplemented with antibiotics (30 µg/mL kanamycin or 5 µg/mL chloramphenicol). A single colony was inoculated in 5 mL LB broth and incubated overnight at 30°C with stirring at 250 rpm. 1 mL of the pre-inoculum was used to inoculate 100 mL of LB broth and cells were left to grow at 37°C until OD$_{600nm}$ = 0.7–0.8. Subsequently, cells were incubated with 0.4–1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) to induce protein expression and kept for 5 h at 37°C or overnight at 20°C.

For expression trials in *P. pastoris* cells, vectors were linearized with *Pme* restriction enzyme and transformed using the *Pichia* EasyComp transformation kit (Invitrogen). Cells were inoculated in low salt yeast peptone dextrose (YPD) plates supplemented with EasyComp transformation kit (Invitrogen). Cells were inoculated in 100 mL of LB broth and cells were left to grow at 37°C until OD$_{600nm}$ = 0.7–0.8. Subsequently, cells were incubated with 0.4–1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) to induce protein expression and kept for 5 h at 37°C or overnight at 20°C.

For expression trials in *P. pastoris* cells, vectors were linearized with *Pme* restriction enzyme and transformed using the *Pichia* EasyComp transformation kit (Invitrogen). Cells were inoculated in low salt yeast peptone dextrose (YPD) plates supplemented with 100 µg/mL zeocin and incubated for 3–4 days at 28°C. Colonies were selected and grown in 100 mL buffered complex glycerol medium (BMGY) at 28°C until an OD$_{600nm}$ = 2. Cells were then harvested, resuspended in buffered complex methanol medium (BBMY), and protein expression was induced with 0.5% methanol.

Cells were separated from the growth media by centrifugation at 8,000×g for 30 min at 4°C. Secreted proteins were collected from the growth media and dialysed in buffer A (50 mM Tris-HCl, 250 mM NaCl, pH 7.5) and cytoplasmic proteins were extracted from the cells in the same buffer. For lysis, cells were sonicated with 3 pulses of 5 min each at 40% amplitude (Branson digital sonifier). Samples were collected before and after centrifugation (30,000×g for 30 min at 4°C) representing total and soluble protein fractions, respectively.

Selected samples were further purified by affinity chromatography using either nickel-nitrotriacetic acid- (Ni-NTA), maltose binding protein- (MBP) or glutathione S-transferase- (GST) HiTrap columns, or a Sepharose 4B matrix column (GE Healthcare Life Sciences). 10 mL of crude protein extract was applied to the columns, followed by three washes with buffer A. Proteins were eluted with buffer A supplemented with either 300 mM imidazole (Ni-NTA-affinity), 10 mM maltose (MBP-affinity), 10 mM reduced glutathione (GST-affinity) or 20 mM lactose (Sepharose-affinity). Finally, samples were buffer-exchanged to buffer B (20 mM Tris-HCl, 150 mM NaCl, pH 7.4) using a PD-10 desalting column (GE Healthcare Life Sciences). Samples were kept at 4°C at all times.

For expression and purification of MecR1, the cultures were scaled up to 6L, the collected cells were broken with a cell dismuter (Constant Cell Disruption Systems) at 2.4kBar and non-disrupted cells and cell debris were removed by centrifugation at 20,000×g for 45 min in a Sorvall centrifuge. Membranes were collected by ultracentrifugation at 150,000×g for 2 h at 4°C in a Beckman Optima L-90K using a 50.2 Ti rotor (Beckman) and 26.3-mL polycarbonate bottles with cap assembly (Beckman). Collected membranes were homogenized using a glass Potter and solubilized under gentle stirring by overnight incubation at 4°C in buffer C (50 mM Tris-HCl, 300 mM NaCl, 10 mM imidazole, 1 mM 1,4-dithio-D-threitol, pH 8.0) containing 100 mM lauryl-dimethylamine N-oxide (LDAO; Sigma) and EDTA-free proteinase inhibitor cocktail tablets (Roche). Non-solubilized proteins were removed by ultracentrifugation as described above. The sample was incubated overnight at 4°C with Ni-NTA resin (Invitrogen). The bound protein was batch purified in an open column (Bio-Rad), washed extensively, and the tagged protein eluted with buffer C plus 300 mM imidazole. The sample was desalted using a PD-10 column in buffer C containing 5 mM LDAO.

**Fusion-tag removal by proteinase cleavage**

Tobacco etch virus (TEV) proteinase and sentrin specific proteinase 1 (SENPI) were over-expressed in *E. coli* BL21 (DE3) pLysE cells using pET28-based vectors, which attach an N-terminal His$_6$-tag. Cultures (typically 4L) were grown in LB broth at 37°C until an OD$_{600nm}$ = 0.7–0.8, induced with 0.5 mM IPTG, and incubated either overnight at 20°C or for 5 h at 30°C for TEV proteinase or SENPI expression, respectively. Subsequently, cells were collected by centrifugation at 5,000×g for 30 min at 4°C and partially purified by Ni-NTA affinity chromatography as previously described [57,58]. Proteinases were stored at −80°C in buffer D (20 mM Tris-HCl, 50 mM NaCl, pH 7.5, 30% glycerol). Proteinase cleavage trials of tagged-proteins were performed overnight at 4°C in buffer B using various protein-proteinase ratios. For trials with thrombin (GE Healthcare Life Sciences), 2 units of proteinase were used to process 25 µg of protein in 100 µL of buffer C at room temperature and aliquots were taken at various time points.

**Enzymatic assays**

For hydrolytic activity measurements, PNGase F and fraglysin were partially purified by Ni-NTA-affinity chromatography as described above. Glycosidase activity of PNGase F was tested against the glycoprotein ribonuclease B (RNase B; New England Biolabs) at a w/w ratio of 1:5 PNGase F/RNase B and a final protein concentration of 0.5 mg/mL. Reactions were incubated overnight at 4°C and analysed by SDS-PAGE. Peptidase activity of fraglysin was tested against BODIPY FL-casein (Invitrogen) as previously described [59]. Crude protein extracts of CPA2 were used for assays after an initial activation with partial tryptic digestion in a w/w ratio of 1/100 of CPA2/trypsin at room temperature for 1 h. The activated protein was incubated with fucyl-acryloyl-L-phenylalanine-L-phenylalanine (0.05 mM; Sigma) in buffer B and the activity was monitored by measurement of the absorbance change at 330 nm.

**Western-blot analysis**

Protein samples were analyzed by Tricine-SDS-PAGE, transferred to Hybond ECL membranes (GE Healthcare Life Sciences), and finally blocked overnight at room temperature with 20 mL of blocking solution (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na$_2$HPO$_4$, 1.47 mM NaH$_2$PO$_4$, 0.05% Tween 20) containing 1.5% bovine serum albumin. MecR1 was detected by immunoblot analysis using custom polyclonal antibodies (Eurogentec) at dilution 1:1,000 and a secondary antibody (goat anti-rabbit IgG (H+L) peroxidase-conjugated antibody; Pierce) at dilution 1:1,000 (both in blocking solution). The immune complexes were detected using an enhanced chemiluminescence system (Super Signal West Pico Chemiluminescent; Pierce) according to the manufacturer’s instructions. Membranes were exposed to hyperfilm ECL films (GE Healthcare Life Sciences).

**Miscellaneous**

Denatured protein samples were analyzed by 10%–15% Tricine-SDS-PAGE [60] and stained with Coomassie-brilliant blue. Protein concentrations were routinely determined by absorbance at 280 nm, and, wherever necessary, corrected by the BCA protein assay method [Thermo Scientific] using bovine serum albumin as a standard. Protein identification by peptide mass fingerprinting was performed at the Protein Chemistry Facility of Centro de Investigaciones Biológicas (Madrid, Spain). Figures of vector maps were prepared with GENEIOUS (Biomatters).
Results and Discussion

Description of the pCri System

We generated a collection of vectors for recombinant protein overexpression in two bacterial (E. coli and B. subtilis) and one eukaryotic (P. pastoris) host strains. Vectors, available from commercial sources or laboratories, were initially modified by inserting new nucleotide sequences or point mutations, and finally evaluated for functionality. Most of the E. coli vectors are pET based [61] with the exception of pCri-12, which is based on pTrc99a [50]. The bacillus and yeast vectors are based on pHT [62] and pPICZ series [63], respectively, and can be stably propagated in E. coli cells when antibiotic resistance is conferred (Tables 1–3). In all vectors, protein expression is achieved by IPTG induction, except for the yeast vectors, for which methanol is required.

The collection consists of 29 vectors grouped into three main categories (Tables 1–3). Based on the available 5’end restriction sites for target gene cloning, the vectors are sorted into pCri-a and pCri-b series using either NeoI and MscI or NdeI and Nhel sites, respectively (Fig. 1 and Fig. S1). The pCri-a series is further separated into pCri-a and pCri-a-Strep based on the fusion tag that can be attached at the C-terminus of the target protein. Within each category, the vectors allow obtaining constructs with different fusion tags or expression in a particular host organism. Usage of the aforementioned 5’end restriction sites incorporates a methionine start codon, thus obviating the need to introduce it into the target gene during PCR amplification. An MCS universal for all vectors has been placed at the 3’end, which encodes seven rare restriction sites not found in most of the vectors (see vector maps for more details; Fig. S1). For convenience and tracking during vector preparation, a GFP insert is cloned within all vectors. The inserted genes can be sequenced from either terminus with specific primers as detailed in Table 4.

Preparation is greatly simplified, as only two restriction sites are used for directional cloning of a target gene into a large series of vectors. Although newer cloning techniques are now available (e.g. ligation independent cloning system [64]), this method was satisfactory. Cloning of target genes of variable size spanning from 150 to 7,000 base pairs was routinely performed with a success rate of more than seven out of ten positive clones when genes were cloned between an NeoI or Ndel and a XhoI site. To achieve reproducible results, it was essential to repeat double digestions of the vectors with all the restriction enzyme combinations.

Applications and main considerations of the pCri System

The choice and use of a suitable vector should be based on the properties of the target protein and the needs of the experiment in question. Here, in an effort to evaluate the functionality of the collection and to provide a rationale for the use of the experiment, we cloned and expressed several proteins of different origin and function:

Fusion tags assisting in protein purification. The pCri System allows the fusion of a His_6-tag at the N-terminus of the target protein, which can be in tandem with larger tags such MBP [43], GST [65], small ubiquitin-like modifier (SUMO) [66], and target protein, which can be in tandem with larger tags such MBP used as a first purification step through affinity chromatography tags add a functionality to the target protein, which is commonly indicated that the protein is produced not only in the extracellular environment of P. pastoris but also in the cytoplasm and periplasm of E. coli cells (Fig. 2D). In contrast, B. subtilis did not express the protein either extracellularly or intracellularly. In all cases, the protein was soluble and correctly processed after limited trypsin digestion, showing activity against small substrates. However, this is not always the case. Besides the oxidising conditions other posttranslational modifications. Correct folding and stabilization requires the formation of disulfide bonds in many proteins. These can be formed in oxidising environments as found in the periplasmic and extracellular environment of bacteria, or in specialised organelles of eukaryotes. B. subtilis has a large secretory capacity, whereas in E. coli secretion is mainly limited to the periplasm [68,69]. In P. pastoris, proteins are first driven to the endoplasmic reticulum and, after folding, they are secreted to the extracellular medium [70]. The pCri System includes vectors that fuse SP specialised for protein translocation to these cellular compartments. pCri-9 and 12 can be used with E. coli cells, whereas pCri-16 and 18 are suitable for expression in P. pastoris and B. subtilis, respectively. In the case of pCri-12, a disulfide-bond isomerase C (DsbC) is coexpressed with the target protein and provides additional support in the correct pairing of disulfide bonds in the periplasm [50].

As a test protein, we used human CPA2, which is commonly expressed in P. pastoris cells [71]. Unexpectedly, expression trials indicated that the protein is produced not only in the extracellular environment of P. pastoris but also in the cytoplasm and periplasm of E. coli cells (Fig. 2D). In contrast, B. subtilis did not express the protein either extracellularly or intracellularly. In all cases, the protein was soluble and correctly processed after limited trypsin digestion, showing activity against small substrates. However, this is not always the case. Besides the oxidising conditions other proteins may often participate in correct folding, including...
Table 1. pCri-a vectors.

| Name       | Vector size (→kbp) | Fusion tag⁵ | Extra residues after fusion removal | Fusion molecular mass (→kDa)⁶ | Resistance marker⁷ | Based on vector | Cells   |
|------------|-------------------|-------------|--------------------------------------|-------------------------------|---------------------|-----------------|---------|
| pCri-1a    | 7.2               | His₆-MBP-TEV-insert-His₆-tag | Gly-Ala | 43.1 | Kan⁺ | pETMBP-1a | E. coli |
| pCri-4a    | 6.4               | His₆-TRX-TEV-insert-His₆-tag | Gly-Ala | 14.3 | Kan⁺ | pETTrx-1a | E. coli |
| pCri-6a    | 6.7               | His₆-GST-TEV-insert-His₆-tag | Gly-Ala | 28.9 | Kan⁺ | pETGST-1a | E. coli |
| pCri-7a    | 6.0               | insert-His₆-tag | - | - | Kan⁺ | pET-28a | E. coli |
| pCri-8a    | 6.1               | His₆-TEV-insert-His₆-tag | Gly-Ala | 2.4 | Kan⁺ | pET-28a | E. coli |
| pCri-9a    | 6.1               | SP (pelB)-insert-His₆-tag | - | - | Kan⁺ | pET-26b | E. coli |
| pCri-11a   | 6.3               | His₆-SUMO-insert-His₆-tag | Ala | 12.6 | Kan⁺ | pET-26b | E. coli |
| pCri-12a   | 5.2               | SP (ompA)-insert-His₆-tag and DsbC coexpression | - | - | Amp⁺ | pRBI-DsbC | E. coli |
| pCri-13a   | 6.4               | His₆-MISTIC-THR⁸-insert-His₆-tag | Gly-Ser-Gly-Ala | 16.2 | Kan⁺ | pET-28a | E. coli |
| pCri-14a   | 6.5               | His₆-LSL-TEV-insert-His₆-tag | Gly-Ala | 19.6 | Kan⁺ | pET-28a | E. coli |
| pCri-15a   | 4.0               | His₆-TEV-insert | Gly-Ala | 2.4 | Zeo⁺ | pPICZA | P. pastoris |
| pCri-16a   | 4.4               | SP (n-factor)-insert-His₆-tag | - | - | Zeo⁺ | pPICZa | P. pastoris |
| pCri-17af  | 8.6               | His₆-TEV-insert-His₆-tag | Gly-Pro | 2.4 | Amp⁺, Cm⁺ | pH7-01 | B. subtilis |
| pCri-18af  | 8.7               | SP (samQ)-insert-His₆-tag | - | - | Amp⁺, Cm⁺ | pH7-43 | B. subtilis |

⁵Calculated vector sizes based on the electrophoretic mobility with a Lambda DNA marker as standard.
⁶For introducing a His₆-tag at the C-terminus of the target protein use a reverse primer without a stop codon.
⁷Molecular mass of C-terminal His₆-tag = 1 kDa.
⁸Kan⁺, 30–50 µg/mL; Amp⁺, 50–100 µg/mL; Cm⁺, 5 µg/mL; Zeo⁺, 25–100 µg/mL.
⁹Thrombin cleavage site (THR).
**Ampicillin resistance in E. coli and chloramphenicol resistance in B. subtilis.**

doi:10.1371/journal.pone.0112643.t001
Table 2. pCri-b vectors.

| Name   | Vector size (kbp)<sup>a</sup> | Fusion tag<sup>b</sup>                         | Extra residues after fusion removal | Fusion molecular mass (kDa)<sup>c</sup> | Resistance marker<sup>d</sup> | Based on vector | Cells     |
|--------|-------------------------------|-----------------------------------------------|-------------------------------------|-------------------------------------|-------------------------|----------------|-----------|
| pCri-1b | 7.2                           | His<sub>6</sub>-MBP-TEV-insert-His<sub>6</sub>-tag | Gly-His                            | 43                                  | Kan<sup>'</sup>           | pCri-1a         | E. coli   |
| pCri-4b | 6.4                           | His<sub>6</sub>-TRX-TEV-insert-His<sub>6</sub>-tag | Gly-Ala-His                        | 14.3                                | Kan<sup>'</sup>           | pCri-4a         | E. coli   |
| pCri-6b | 6.7                           | His<sub>6</sub>-GST-TEV-insert-His<sub>6</sub>-tag | Gly-His                            | 28.9                                | Kan<sup>'</sup>           | pCri-6a         | E. coli   |
| pCri-7b | 6.0                           | insert-His<sub>6</sub>-tag                    | -                                  | 2.4                                 | Kan<sup>'</sup>           | pCri-7a         | E. coli   |
| pCri-8b | 6.1                           | His<sub>6</sub>-tag-TEV-insert-His<sub>6</sub>-tag | Gly-His                            | -                                   | Kan<sup>'</sup>           | pCri-8a         | E. coli   |
| pCri-9b | 6.1                           | SP<sup>(pelB)</sup>-insert-His<sub>6</sub>-tag | -                                  | Kan<sup>'</sup>                      | pCri-9a                 | pCri-9a         | E. coli   |
| pCri-11b| 6.3                           | His<sub>6</sub>-SUMO-insert-His<sub>6</sub>-tag | His                                 | 12.6                                | Kan<sup>'</sup>           | pCri-11a        | E. coli   |
| pCri-12b| 5.2                           | SP<sup>(ompA)</sup>-insert-His<sub>6</sub>-tag and DsbC coexpression | -                                  | -                                   | Amp<sup>'</sup>          | pCri-12a        | E. coli   |
| pCri-14b| 6.5                           | His<sub>6</sub>-LSL-TEV-insert-His<sub>6</sub>-tag | Gly-His                            | 19.6                                | Kan<sup>'</sup>           | pCri-14a        | E. coli   |
| pCri-15b| 4.0                           | His<sub>6</sub>-TEV-insert-His<sub>6</sub>-tag | Gly-His                            | 2.4                                 | Zeo<sup>'</sup>          | pCri-15a        | P. pastoris |

<sup>a</sup>Calculated vector sizes based on the electrophoretic mobility with a Lambda DNA marker as standard.

<sup>b</sup>For introducing a His<sub>6</sub>-tag at the C-terminus of the target protein use a reverse primer without a stop codon.

<sup>c</sup>Molecular mass of C-terminal His<sub>6</sub>-tag ∼1 kDa.

<sup>d</sup>Kan<sup>'</sup>, 30–50 μg/mL; Amp<sup>'</sup>, 50–100 μg/mL; Cm<sup>'</sup>, 5 μg/mL; Zeo<sup>'</sup>, 25–100 μg/mL.

doi:10.1371/journal.pone.0112643.t002

Table 3. pCri-a-Strep vectors.

| Name   | Vector size (kbp)<sup>a</sup> | Fusion tag<sup>b</sup>                         | Extra residues after fusion removal | Fusion molecular weight (kDa)<sup>c</sup> | Resistance marker<sup>d</sup> | Based on vector | Cells     |
|--------|-------------------------------|-----------------------------------------------|-------------------------------------|-------------------------------------|-------------------------|----------------|-----------|
| pCri-1a-Strep | 7.2                           | His<sub>6</sub>-MBP-TEV-insert-Strep-tag | Gly-Ala                             | 43.1                                | Kan<sup>'</sup>           | pCri-1a         | E. coli   |
| pCri-4a-Strep | 6.4                           | His<sub>6</sub>-TRX-TEV-insert-Strep-tag | Gly-Ala                             | 14.3                                | Kan<sup>'</sup>           | pCri-4a         | E. coli   |
| pCri-7a-Strep | 6.0                           | insert-Strep-tag                              | -                                  | -                                   | Kan<sup>'</sup>           | pCri-7a         | E. coli   |
| pCri-8a-Strep | 6.1                           | His<sub>6</sub>-TEV-insert-Strep-tag          | Gly-Ala                             | 2.4                                 | Kan<sup>'</sup>           | pCri-8a         | E. coli   |
| pCri-9a-Strep | 6.1                           | SP<sup>(pelB)</sup>-insert-Strep-tag          | -                                  | -                                   | Kan<sup>'</sup>           | pCri-9a         | E. coli   |

<sup>a</sup>Calculated vector sizes based on the electrophoretic mobility with a Lambda DNA marker as standard.

<sup>b</sup>For introducing a Strep-tag at the C-terminus of the target protein use a reverse primer without a stop codon.

<sup>c</sup>Molecular mass of C-terminal Strep-tag ∼1.3 kDa.

<sup>d</sup>Kan<sup>'</sup>, 30–50 μg/mL.

doi:10.1371/journal.pone.0112643.t003
oxidases, foldases, isomerases and specialised chaperones [69]. Moreover, disulfide bond formation is not the only factor in proper protein folding and stability, and further posttranslational modifications (e.g. glycosylation) may be required, which can be provided by *P. pastoris* [70].

Another approach for disulfide bond formation exploits the oxidising cytoplasm of thioredoxin reductase B (trxB) and glutathione reductase (gor2) mutant *E. coli* cells (Origami 2) [34]. In contrast to the commonly used BL21 cells, Origami 2 efficiently expressed PNGase F, either with pCri-4a or 8a, soluble and catalytically active against RNase B (Fig. 2E and 2F). The protein contains disulfide bonds that require an oxidising environment, which is adequately formed in the cytoplasm of mutant cells. In addition, the combined use of thioredoxin A (TRX) as fusion protein in pCri-4a and expression in Origami 2 can lead to the overexpression of small multi-disulfide proteins, among others [69,72]. This system takes advantage of TRX, which acts as an oxidant when it operates in an oxidized milieu found in mutant cells [34], thus providing an additional mechanism for disulfide bond formation within the cytoplasm. TRX is subsequently removed by TEV protease cleavage in the presence of selected amounts of redox agents to assist in correct disulfide bond pairing [72,73].

**Expression of membrane proteins.** Membrane proteins are among the targets most requested and at the same time difficult to express and purify. To address this issue, a vector was prepared, which fuses a small protein from *B. subtilis* with target proteins (pCri-13a). This protein, known as the membrane-integrating sequence for translation of integral-membrane protein constructs (MISTIC), folds autonomously into membranes, simultaneously dragging the tagged-protein to the cell membrane [74]. Moreover, this vector contains a longer His8-tag in tandem with MISTIC in order to provide higher affinity for Ni-NTA affinity purification.

| Primer Name | Vector (pCri) | Sequence site | Primer sequence |
|-------------|---------------|---------------|-----------------|
| T7 promoter | 1-14a,b except 12a-b* | 5’ end | Universal |
| T7 terminator | 1-14a,b except 12a-b* | 3’ end | Universal |
| seq-pCri-1 | 1a and b internal* | 5’ end | GAAATCATGCCGAACATCCC |
| seq-pCri-4 | 4a and b internal* | 5’ end | GCGGCAACCAAAGTGGGTGCAC |
| seq-pCri-6 | 6a and b internal* | 5’ end | GACCATCTCCCAACTAGT |
| seq-pCri-11 | 11a and b internal* | 5’ end | CAAAAGAACTGGGAATG |
| 5’seq-pCri-12 | 12a** and b** | 5’ end | GATAACGAGGGCAAAAAATG |
| 3’seq-pCri-12 | 12a* and b* | 3’ end | CAAATGAAACACAAACAAAC |
| seq-pCri-13 | 13a internal* | 5’ end | CAGATTITATATCATC |
| seq-pCri-14 | 14a and b internal* | 5’ end | CTTCTGGAAATCCCT |
| 5’ AOX1 | 15a,b* and 16a** | 5’ end | Universal |
| 3’ AOX1 | 15a,b* and 16a* | 3’ end | Universal |
| 5’seq-pCri-17 | 17a* and 18a** | 5’ end | CTTATCATTTGAATTT |
| 3’seq-pCri-17 | 17a* and 18a* | 3’ end | GATTATATGTCAAAGGAC |

*Hybridises before Ncol, Ndel or Xhol restriction sites. **Hybridises before the SP.

doi:10.1371/journal.pone.0112643.t004
Figure 2. Protein expression and purification trials using the pCri System. (A) The GFP gene was cloned into pCri-1a, 4a, 6a, 8a, 11a, and 14a, the proteins expressed in E. coli BL21 cells, and subsequently purified by Ni-NTA-affinity chromatography except for MBP, GST, and LSL fusion products, which were purified by their respective specific affinity resins. (B) The gene coding for fragilysin was cloned into pCri-1a, 4a, 6a and 8a, and expressed in E. coli Origami 2 cells. Total (T) and soluble (S) fractions of crude protein extracts were further analysed by SDS-PAGE. All expression trials were performed at 20°C except for pCri-1a, which was also performed at 37°C. (C) Partially purified MBP-fragilysin before (−) and after (+) TEV protease cleavage. Arrows indicate the soluble fraction of fragilysin (white) and the MBP (black) after TEV protease cleavage. (D) Expression of CPA2 intracellularly (lanes 1 and 2) or periplasmatically (lanes 3 and 4) in E. coli cells, and extracellularly (lanes 5 and 6) in P. pastoris cells. Lanes indicate samples before (1, 3 and 5) and after (2, 4 and 6) tryptic digestion. Arrows indicate the pro-CPA2 (black), the mature form (grey) and the pro-peptide (white) after tryptic cleavage. (E) The PNGase F gene was cloned into pCri-4a and 8a and expressed overnight at 20°C in E. coli BL21 and Origami 2 cells. Total (T) and soluble (S) fractions of crude protein extracts were further analysed by SDS-PAGE. (F) Activity of affinity-purified TRX-PNGase F against glycosylated RNase B. (+) and (−) indicate presence and absence of PNGase F. Arrows indicate the PNGase F (black), native RNase B (grey) and deglycosylated RNase B (white). (G) MecR1 was expressed in E. coli BL21 using pCri-8a or 13a, and soluble fractions were analysed by Western blotting with specific antibodies as detailed in “Materials and Methods”. A black arrow indicates the detected MecR1. (H) Partially purified MISTIC-MecR1 after Ni-NTA-affinity purification. (I) Partially purified MBP-GFP, SUMO-GFP and MISTIC-MecR1 were digested with TEV protease,
For evaluation purposes, we cloned and expressed MecR1, a membrane metalloproteinase from *Staphylococcus aureus* implicated in methicillin resistance [75]. Detectable levels of expression were only achieved when the protein was fused with MISTIC, whereas mere fusion with an N-terminal His6-tag was unsuccessful (Fig. 2G). Moreover, expression yields of the protein were sufficient (0.4 mg of affinity purified protein per litre of culture) after solubilisation of the membranes with the zwitterionic detergent LDAO (Fig. 2H). Although further studies are required to assess the folding state of this protein, fusion with MISTIC allowed us to express it in milligram amounts. Many other membrane proteins were also expressed in fusion with various MISTIC constructs, indicating that this system could be an alternative approach for membrane proteins that are difficult to express [52].

**Removal of fusion tags.** In most cases, release of the target protein from any fused tag is desirable. In the pCri System, a TEV proteinase cleavage site is introduced immediately after the tag in all vectors except for pCri-11 and pCri-13, in which a SENP1 or thrombin site is found, respectively (Fig. S1). TEV proteinase is a highly specific enzyme that recognises an hexapeptide sequence [76], whereas SENP1 further offers robustness and high proteolytic activity in addition to high specificity, usually requiring only minute amounts for tag removal [77]. Moreover, use of a thrombin cleavage site in pCri-13a was necessary due to the low efficiency of TEV proteinase in the presence of detergents which are required during membrane-protein solubilisation [78]. In addition, linker sequences (Gly-Ser)7 and Gly-Tyr-Ala were introduced before and after the thrombin recognition site, respectively, to improve access for proteinase cleavage (Fig. S1) [79].

Tag removal was achieved with variable amounts of endopeptidases, different incubation times and temperatures (Fig. 2f). These studies indicated that optimisation trials are needed in each case to identify the best conditions for complete digestion (e.g. buffer, temperature, protease:substrate ratio). Proteinase cleavage and tag removal result in the incorporation of one or two extra residues at the N-terminus of the expressed protein except for pCri-4b and pCri-13a, which attach three and six residues, respectively (Tables 1–3).

## References

1. Gregg JM, Ceregino JL, Shi J, Higgins DR (2000) Recombinant protein expression in Pichia pastoris. Mol Biotechnol 16: 23–52.
2. Sorensen HP, Mørtensen KK (2005) Advanced genetic strategies for recombinant protein expression in Escherichia coli. J Biotechnol 115: 113–129.
3. Kost T, Condrea JP, Jarvis DL (2005) Baculovirus as versatile vectors for protein expression in insect and mammalian cells. Nat Biotechnol 23: 567–575.
4. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, et al. (2012) The Pfam protein families database. Nucleic Acids Res 40: D290–D301.
5. Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: Discriminating signal peptides from transmembrane regions. Nat Methods 8: 785–786.
6. Madan Babu M, Sankaran K (2002) DOLOP-Database of bacterial lipoproteins. Bioinformatics 18: 641–643.
7. Zhou F, Xue Y, Yao X, Xu Y (2006) A general user interface for prediction servers of proteins’ post-translational modification sites. Nat Protoc 1: 1318–1321.
8. Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J Mol Biol 305: 567–580.
9. Hayat S, Elofsson A (2012) BOCTOPOUS: Improved topology prediction of transmembrane β barrel proteins. Bioinformatics 28: 516–522.
10. Bernal A, Vihard H, Henein A, Elofsson A (2009) TOPCONS: consensus prediction of membrane protein topology. Nucleic Acids Res 37: W465–468.
11. Linding R, Jensen Lj, Diella F, Bork P, Gibson TJ, et al. (2005) Protein disorder prediction: Implications for structural proteomics. Structure 11: 1453–1459.
12. Li X, Romero P, Kani M, Dunker AK, Obradovic Z (1999) Predicting protein disorder for N-, C-, and internal regions. Genome Inf 10: 30–40.
13. McGuyer LJ, Bryson K, Jones DT (2000) The PSI-PRED protein structure prediction server. Bioinformatics 16: 404–405.
14. Land PA (2001) Microbial molecular chaperones. Adv Microb Physiol 44: 93–140.
15. Baneyx F, Mujacic M (2004) Recombinant protein folding and misfolding in Escherichia coli. Nat Biotechnol 22: 1399–1408.
16. Driessen AJM, Nouwen N (2008) Protein translocation across the bacterial cytoplasmic membrane. Annu Rev Biochem 77: 643–667.
17. Cole C, Barber JD, Barton GI (2008) The Jrrpd-3 secondary structure prediction server. Nucleic Acids Res 36: W197–W201.
