



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

Theodoros Goulas^{1*}, Anna Cuppari^{1#a}, Raquel Garcia-Castellanos^{1#b}, Scott Snipas², Rudi Glockshuber³, Joan L. Arolas¹, F. Xavier Gomis-Rüth^{1*}

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 homogenous 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 rare 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.

Citation: Goulas T, Cuppari A, Garcia-Castellanos R, Snipas S, Glockshuber R, et al. (2014) The pCri System: A Vector Collection for Recombinant Protein Expression and Purification. *PLoS ONE* 9(11): e112643. doi:10.1371/journal.pone.0112643

Editor: Dipankar Chatterji, Indian Institute of Science, India

Received: August 13, 2014; **Accepted:** October 9, 2014; **Published:** November 11, 2014

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files.

Funding: This study was supported in part by grants from European, Spanish, and Catalan agencies (FP7-HEALTH-2010-261460 "Gums&Joints"; FP7-PEOPLE-2011-ITN-290246 "RAPID"; FP7-HEALTH-2012-306029-2 "TRIGGER"; BFU2012-32862; CSD2006-00015; and 2014SGR9). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: fxgr@ibmb.csic.es (FXGR); thgcri@ibmb.csic.es (TG)

#a Current address: Structural MitoLab, Molecular Biology Institute of Barcelona, CSIC, Barcelona Science Park, Barcelona, Spain

#b Current address: Protein Expression Core Facility, Institute for Research in Biomedicine (IRB Barcelona), Barcelona Science Park, Barcelona, Spain

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, BOCTOPUS) [8–10], and unfolded/disordered regions (DisEMBL, PONDR, 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 (JPRED, 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, posttranslational signals, low-complexity regions, hydrophobic residues at the protein termini and membrane spanning regions, while conserving the boundaries of globular domains [19].

In parallel, cDNA 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 pHT-43), Invitrogen (pPICZA and pPICZαA), and from the Glockshuber 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 *Nco*I or *Nde*I restriction sites at the 5' end and in *Xho*I 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 *Nco*I or *Nde*I and *Xho*I restriction sites and was modified to contain an *Msc*I or *Nhe*I restriction site immediately after the *Nco*I and *Nde*I 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 *Dpn*I 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: P0A0B0; an integral-membrane metallopeptidase) was cloned into vector pCri-8a 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: P48052; Leu19-Tyr419; a soluble metalloexopeptidase) 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 $OD_{600\text{ nm}} \approx 0.7\text{--}0.8$. 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 *PmeI* 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 $OD_{600\text{ nm}} \approx 2$. Cells were then harvested, resuspended in buffered complex methanol medium (BMMY), 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-nitrilotriacetic 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 disrupter (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 lauryldimethylamine 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 His₆-tag. Cultures (typically 4L) were grown in LB broth at 37°C until an $OD_{600\text{ nm}} \approx 0.7\text{--}0.8$, 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 fragilysin 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 fragilysin 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 furyl-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 Na₂HPO₄, 1.47 mM NaH₂PO₄, 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 (HL) peroxidase-conjugated antibody; Pierce) at dilution 1:5,000 (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 pTrec99a [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 *NcoI* and *MscI* or *NdeI* and *NheI* 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 *NcoI* or *NdeI* 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 vectors, 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 His₆-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 the β -trefoil lectin module of protein LSL₁₅₀ from the mushroom *Laetiporus sulphureus* (LSL) [53] (Tables 1–3). The C-terminus of the target protein can likewise be furnished with a His₆-tag or a Strep-tag if the stop codon of the amplified gene is omitted. These tags add a functionality to the target protein, which is commonly used as a first purification step through affinity chromatography [43,53,65]. On this basis, we cloned and expressed GFP in pCri-

1a, 4a, 6a, 8a, 11a, and 14a. The proteins were purified by Ni-NTA affinity chromatography except for MBP, GST, and LSL fusion products, which were purified by their specific affinity resins (Fig. 2A). Nickel or cobalt affinity chromatography of His₆-tagged proteins are among the most commonly used methods for purification, but others using the affinity properties of MBP or GST, and the recently reported LSL₁₅₀, can provide better purification results under mild elution conditions. This choice among alternative affinity purification systems allows the best purification method to be used for each target protein. Moreover, many of those tags can be used to track poorly expressed proteins by Western-blot analysis, as they are otherwise undetectable by Coomassie-stained SDS-PAGE.

Fusion tags assisting in protein solubility. In addition, several studies showed that tags such as N-utilisation substrate A (NUSA), MBP, or the smaller GST and SUMO have positive effects on the cargo protein due to their solubility-enhancing or chaperoning properties [2,66,67]. Nevertheless, their working mechanism is still controversial, with several studies suggesting a more passive role due to their excellent solubility properties rather than a direct influence on the folding of their partner [47]. For example, fragilysin (Ala212-Asp397) [59], a bacterial enterotoxin metalloproteinase, was expressed in high amounts in fusion with MBP, TRX, GST, and His₆-tag, both at 37°C or 20°C (Fig. 2B). However, only MBP rendered the protein soluble during low temperature expression trials, whereas other fusions or expression at higher temperatures produced protein prone to aggregation. The protein remained in solution even after MBP removal (Fig. 2C) but catalytically inactive against fluorescent-labelled casein, indicating at least partial misfolding. Similar results were obtained when fragilysin was expressed with the smaller Z-tag (\approx 10 kDa) [59], indicating that fusion proteins may have a positive effect on target solubility without necessarily implying that it will be well folded and active. Nevertheless, these fusion tags can have an application in the expression of proteins with known solubility problems that need to be temporally stabilised until an adequate condition/solution is found [67].

Expression of proteins requiring disulfide bonds and 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 tryptic 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) ^a	Fusion tag ^b	Extra residues after fusion removal	Fusion molecular mass (~kDa) ^c	Resistance marker ^d	Based on vector	Cells
pCri-1a	7.2	His ₆ -MBP-TEV-insert-His ₆ -tag	Gly-Ala	43.1	Kan ^r	pETMBP-1a	<i>E. coli</i>
pCri-4a	6.4	His ₆ -TRX-TEV-insert-His ₆ -tag	Gly-Ala	14.3	Kan ^r	pETTrx-1a	<i>E. coli</i>
pCri-6a	6.7	His ₆ -GST-TEV-insert-His ₆ -tag	Gly-Ala	28.9	Kan ^r	pETGST-1a	<i>E. coli</i>
pCri-7a	6.0	insert-His ₆ -tag	-	-	Kan ^r	pET-28a	<i>E. coli</i>
pCri-8a	6.1	His ₆ -TEV-insert-His ₆ -tag	Gly-Ala	2.4	Kan ^r	pET-28a	<i>E. coli</i>
pCri-9a	6.1	SP (pelB)-insert-His ₆ -tag	-	-	Kan ^r	pET-26b	<i>E. coli</i>
pCri-11a	6.3	His ₆ -SUMO-insert-His ₆ -tag	Ala	12.6	Kan ^r	pET-26b	<i>E. coli</i>
pCri-12a	5.2	SP (ompA)-insert-His ₆ -tag and DsbC coexpression	-	-	Amp ^r	pRBI-DsbC	<i>E. coli</i>
pCri-13a	6.4	His ₆ -MISTIC-THR ^e -insert-His ₆ -tag	Gly-Ser-Gly ₃ -Ala	16.2	Kan ^r	pET-28a	<i>E. coli</i>
pCri-14a	6.5	His ₆ -LSL-TEV-insert-His ₆ -tag	Gly-Ala	19.6	Kan ^r	pET-28a	<i>E. coli</i>
pCri-15a	4.0	His ₆ -TEV-insert	Gly-Ala	2.4	Zeo ^r	pPICZA	<i>P. pastoris</i>
pCri-16a	4.4	SP (α-factor)-insert-His ₆ -tag	-	-	Zeo ^r	pPICZαA	<i>P. pastoris</i>
pCri-17a ^f	8.6	His ₆ -TEV-insert-His ₆ -tag	Gly-Pro	2.4	Amp ^r , Cm ^r	pHT-01	<i>B. subtilis</i>
pCri-18a ^f	8.7	SP (samYQ)-insert-His ₆ -tag	-	-	Amp ^r , Cm ^r	pHT-43	<i>B. subtilis</i>

^aCalculated vector sizes based on the electrophoretic mobility with a Lambda DNA marker as standard.

^bFor introducing a His₆-tag at the C-terminus of the target protein use a reverse primer without a stop codon.

^cMolecular mass of C-terminal His₆-tag ~1 kDa.

^dKan^r, 30–50 µg/mL; Amp^r, 50–100 µg/mL; Cm^r, 5 µg/mL; Zeo^r, 25–100 µg/mL.

^eThrombin cleavage site (THR).

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

^aCalculated vector sizes based on the electrophoretic mobility with a Lambda DNA marker as standard.

^bFor introducing a His₆-tag at the C-terminus of the target protein use a reverse primer without a stop codon.

^cMolecular mass of C-terminal His₆-tag ~1 kDa.

^dKan^r, 30–50 µg/mL; Amp^r, 50–100 µg/mL; Cm^r, 5 µg/mL; Zeo^r, 25–100 µg/mL.

doi:10.1371/journal.pone.0112643.t002

Table 3. pCri-a-Strep vectors.

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

^aCalculated vector sizes based on the electrophoretic mobility with a Lambda DNA marker as standard.

^bFor introducing a Strep-tag at the C-terminus of the target protein use a reverse primer without a stop codon.

^cMolecular mass of C-terminal Strep-tag ~1.3 kDa.

^dKan^r, 30–50 µg/mL.

doi:10.1371/journal.pone.0112643.t003

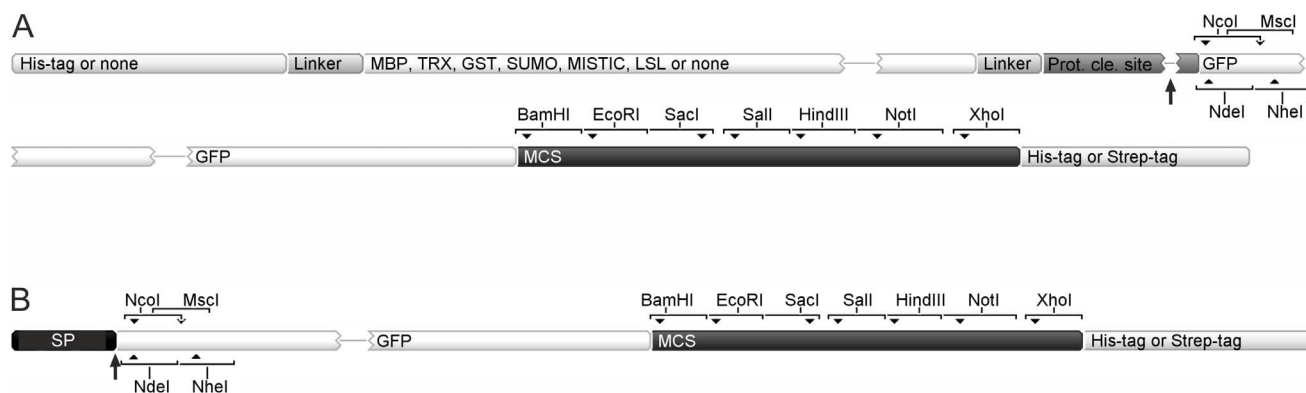


Figure 1. Vector overview of the pCri System. (A) Vectors for cytoplasmic protein expression. (B) Vectors for periplasmic and extracellular protein expression. An N-terminal His₆-tag can be fused in all vectors for intracellular expression except of pCri-7. Other tags can also be fused including MBP, TRX, GST, SUMO, MISTIC, and LSL (Table 1–3). In all vectors, a C-terminal His₆-tag or Strep-tag is attached if a stop codon is omitted within the target gene. Black arrows indicate the proteinase (i.e. TEV, SENP1 or thrombin) and signal peptide (SP) cleavage sites. Restriction sites allowing directional cloning are also shown. For more details regarding each vector, refer to Fig. S1.
doi:10.1371/journal.pone.0112643.g001

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 (*gor*⁻) 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 proteinase 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.

Table 4. Sequencing primers for the pCri System.

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	GACCATCTCCAACCTAGTG
seq-pCri-11	11a and b internal*	5'end	CAAAAGAAGTGGGAATG
5' seq-pCri-12	12a** and b**	5'end	GATAACGAGGGCAAAAAATG
3' seq-pCri-12	12a* and b*	3'end	CAAAGTAAACAACATAAAAC
seq-pCri-13	13a internal*	5'end	CAGATTTTATCCATCTC
seq-pCri-14	14a and b internal*	5'end	CTTCTGGAATCACCCCTC
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	CTTATCACTTGAAATTG
3' seq-pCri-17	17a* and 18a*	3'end	GATTTTATTAGTACAGGGAC

*Hybridises before *NcoI*, *NdeI* or *XhoI* restriction sites.

**Hybridises before the SP.

doi:10.1371/journal.pone.0112643.t004

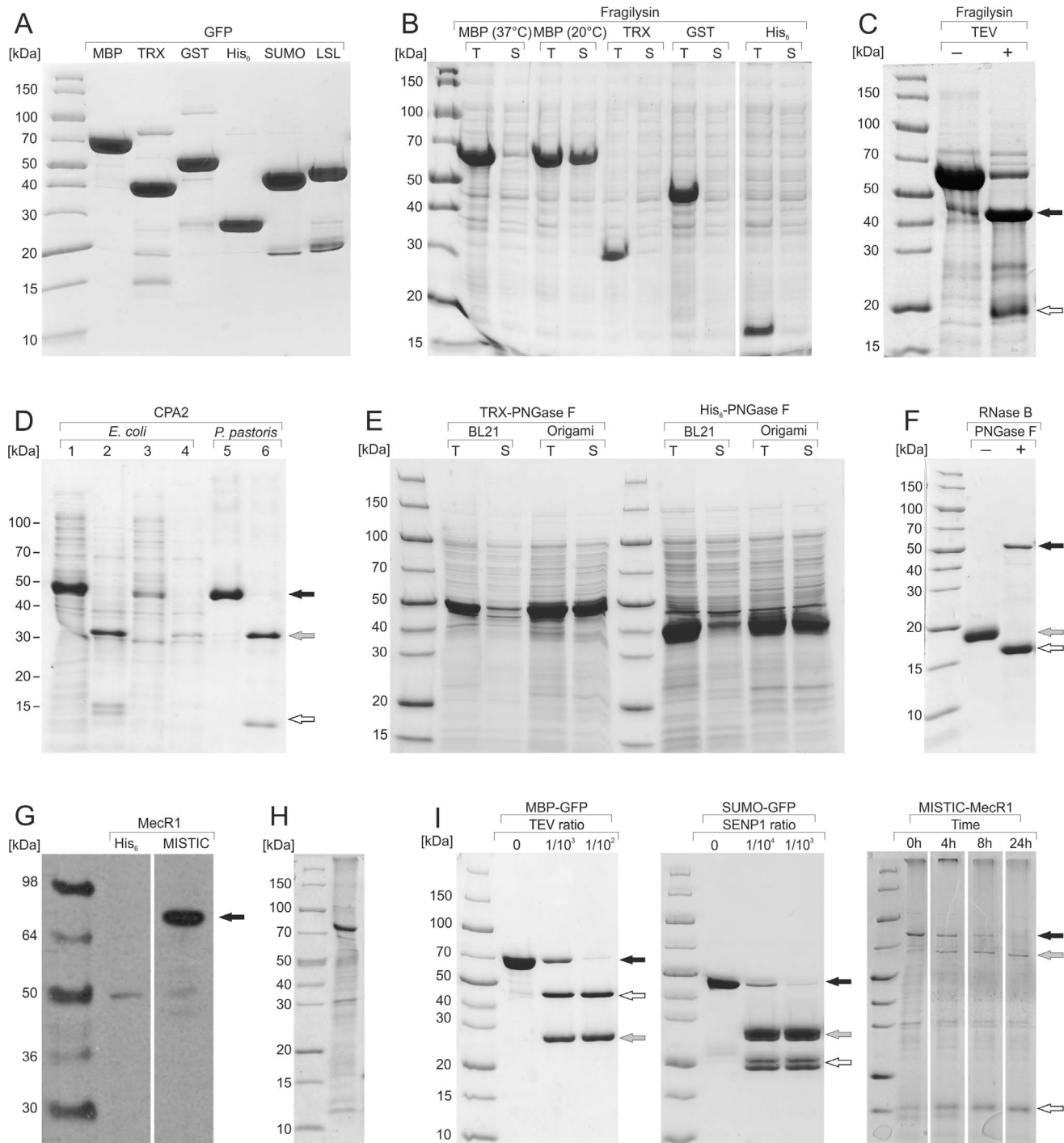


Figure 2. Protein expression and purification trials using the pCri System. (A) The GFP gene was cloned into pCri-1a, 4a, 6a, 8a, 11a, and 14a, the proteins expressed in *E. coli* BL21 cells, and subsequently purified by Ni-NTA-affinity chromatography except for MBP, GST, and LSL fusion products, which 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 proteinase cleavage. Arrows indicate the soluble fraction of fragilysin (white) and the MBP (black) after TEV proteinase 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 proteinase,

SENP1 or thrombin, respectively. For TEV proteinase and SENP1 digestions various ratios of proteinase:tagged-protein were tested in overnight incubations at 4°C, whereas for thrombin digestions 2 units of proteinase were used to digest 25 µg of protein for various times at room temperature. Arrows indicate tagged-protein (black), target protein (grey) and fused-tag (white) after proteinase cleavage. doi:10.1371/journal.pone.0112643.g002

For evaluation purposes, we cloned and expressed MecR1, a membrane metallopeptidase 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) to enable partial purification by Ni-NTA affinity chromatography 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)₅ and Gly₃-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. 2I). These studies indicated that optimisation trials are needed in each case to identify the best conditions for complete digestion (e.g. buffer, temperature, proteinase: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

- Cregg JM, Cereghino JL, Shi J, Higgins DR (2000) Recombinant protein expression in *Pichia pastoris*. *Mol Biotechnol* 16: 23–52.
- Sorensen HP, Mortensen KK (2005) Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J Biotechnol* 115: 113–128.
- Kost T, Condeary JP, Jarvis DL (2005) Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nat Biotechnol* 23: 567–575.
- Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, et al. (2012) The Pfam protein families database. *Nucleic Acids Res* 40: D290–D301.
- Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat Methods* 8: 785–786.
- Madan Babu M, Sankaran K (2002) DOLOP-Database of bacterial lipoproteins. *Bioinformatics* 18: 641–643.
- 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.
- 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.
- Hayat S, Elofsson A (2012) BOCTOPUS: Improved topology prediction of transmembrane β barrel proteins. *Bioinformatics* 28: 516–522.
- Bernsel A, Viklund H, Hennerdal A, Elofsson A (2009) TOPCONS: consensus prediction of membrane protein topology. *Nucleic Acids Res* 37: W465–468.
- Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, et al. (2003) Protein disorder prediction: Implications for structural proteomics. *Structure* 11: 1453–1459.
- Li X, Romero P, Rani M, Dunker AK, Obradovic Z (1999) Predicting protein disorder for N-, C-, and internal regions. *Genome Inf* 10: 30–40.
- McGuYn LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. *Bioinformatics* 16: 404–405.
- Lund PA (2001) Microbial molecular chaperones. *Adv Microb Physiol* 44: 93–140.
- Baneyx F, Mujacic M (2004) Recombinant protein folding and misfolding in *Escherichia coli*. *Nat Biotechnol* 22: 1399–1408.
- Driessen AJM, Nouwen N (2008) Protein translocation across the bacterial cytoplasmic membrane. *Annu Rev Biochem* 77: 643–667.
- Cole C, Barber JD, Barton GJ (2008) The Jpred-3 secondary structure prediction server. *Nucleic Acids Res* 36: W197–W201.

Conclusions

Here we introduce a vector collection designed for large-scale recombinant protein overexpression, and demonstrate its suitability in a series of test proteins. The choice of a suitable expression vector should be based on target and tag properties. The availability of a range of fusion tags allows the choice between different affinity purification methods. Moreover, some tags were included for specific use, such as MISTIC and TRX, which are intended for expression of membrane and disulfide rich proteins, respectively. In general, our common strategy first explores the effects of the presence or absence of N- or C- terminal tags (e.g. His₆-tag or Strep-tag) on each construct under different host cell growth conditions. Omission of the tag or alternation of the position can drastically influence the expression and solubility of the protein. If this approach is ineffective the chances of optimising the expression by testing other fusion combinations are reduced. Several reports showed the beneficial effects of the fusions on target solubility [2,67]. However, this is not always the case: the protein is often dragged into solution, rather than acting as a chaperone for the proper folding of its fusion partner. Removal of the fusion tag can revert the positive effect and cause precipitation [47,48]. If this occurs, then modified constructs, other homologous targets or even other expression systems need to be explored, including bacterial and eukaryotic cells that can be easily tested using the vector collection of the pCri System.

Supporting Information

Figure S1 Partial nucleotide sequence and translation of the pCri System vectors. (DOC)

Acknowledgments

We thank Prof. José M. Mancheño (Madrid, Spain) for providing pKLSL DNA, and Robin Rycroft for critical reading of the manuscript.

Author Contributions

Conceived and designed the experiments: TG FXGR. Performed the experiments: TG AC RGC. Analyzed the data: TG AC RGC FXGR. Contributed reagents/materials/analysis tools: SS RG FXGR. Wrote the paper: TG SS RGC RG JLA FXGR.

18. Wu S, Zhang Y (2007) LOMETS: A local meta-threading-server for protein structure prediction. *Nucleic Acids Res* 35: 3375–3382.
19. Dong A, Xu X, Edwards AM (2007) In situ proteolysis for protein crystallization and structure determination. *Nat Methods* 4: 1019–1021.
20. Hall MN, Gabay J, De'barboulle M, Schwartz M (1982) A role for mRNA secondary structure in the control of translation initiation. *Nature* 295: 616–618.
21. Massé E, Escorcia FE, Gottesman S (2003) Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. *Genes Dev* 17: 2374–2383.
22. Fuhrmann M, Hausherr A, Ferbitz L, Schödl T, Heitzer M, et al. (2004) Monitoring dynamic expression of nuclear genes in *Chlamydomonas reinhardtii* by using a synthetic luciferase reporter gene. *Plant Mol Biol* 55: 869–881.
23. McNulty DE, Claffee BA, Huddleston MJ, Kane JF (2003) Mistranslational errors associated with the rare arginine codon CGG in *Escherichia coli*. *Protein Expr Purif* 27: 365–374.
24. Gustafsson C, Govindarajan S, Minshall J (2004) Codon bias and heterologous protein expression. *Trends Biotechnol* 22: 346–353.
25. Puigbò P, Guzmán E, Romeu A, Garcia-Vallvé S (2007) OPTIMIZER: A web server for optimizing the codon usage of DNA sequences. *Nucleic Acids Res* 35: W126–W131.
26. Vincze T, Posfai J, Roberts RJ (2003) NEBcutter: A program to cleave DNA with restriction enzymes. *Nucleic Acids Res* 31: 3688–3691.
27. Poole ES, Brown CM, Tate WP (1995) The identity of the base following the stop codon determines the efficiency of in vivo translational termination in *Escherichia coli*. *EMBO J* 14: 151–158.
28. Newbury SF, Smith NH, Robinson EC, Hiles ID, Higgins CF (1987) Stabilization of translationally active mRNA by prokaryotic REP sequences. *Cell* 48: 297–310.
29. Westers L, Westers H, Quax WJ (2004) *Bacillus subtilis* as cell factory for pharmaceutical proteins: A biotechnological approach to optimize the host organism. *Biochim Biophys Acta* 1694: 299–310.
30. Porro D, Sauer M, Branduardi P, Mattanovich D (2005) Recombinant protein production in yeasts. *Mol Biotechnol* 31: 245–259.
31. Aricescu AR, Lu W, Jones EY (2006) A time- and cost-efficient system for high level protein production in mammalian cells. *Acta Crystallographica D62*: 1243–1250.
32. Schwarz D, Junge F, Durst F, Frölich N, Schneider B, et al. (2007) Preparative scale expression of membrane proteins in *Escherichia coli*-based continuous exchange cell-free systems. *Nat Protoc* 2: 2945–2957.
33. Makrides SC (1996) Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol Mol Biol Rev* 60: 512–538.
34. Stewart EJ, Aslund F, Beckwith J (1998) Disulfide bond formation in the *Escherichia coli* cytoplasm: An in vivo role reversal for the thioredoxins. *EMBO J* 17: 5543–5550.
35. Miroux B, Walker JE (1996) Over-production of proteins in *Escherichia coli*: Mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol* 260: 289–298.
36. Studier FW (1991) Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J Mol Biol* 219: 37–44.
37. Hartl FU, Hayer-Hartl M (2002) Molecular chaperones in the cytosol: From nascent chain to folded protein. *Science* 295: 1852–1858.
38. De Marco A (2007) Protocol for preparing proteins with improved solubility by co-expressing with molecular chaperones in *Escherichia coli*. *Nat Protoc* 2: 2632–2639.
39. Shiloach J, Fass R (2005) Growing *E. coli* to high cell density—a historical perspective on method development. *Biotechnol Adv* 23: 345–357.
40. Studier FW (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* 41: 207–234.
41. Baneyx F, Georgiou G (1992) Degradation of secreted proteins in *Escherichia coli*. *Ann NY Acad Sci* 665: 301–308.
42. Yee L, Blanch HW (1992) Recombinant protein expression in high cell density fed-batch cultures of *Escherichia coli*. *Biotechnol* 10: 1550–1556.
43. Nallamsetty S, Waugh DS (2007) A generic protocol for the expression and purification of recombinant proteins in *Escherichia coli* using a combinatorial His6-maltose binding protein fusion tag. *Nat Protoc* 2: 383–391.
44. Schmidt TG, Skerra A (2007) The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nat Protoc* 2: 1528–1535.
45. Drew D, Newstead S, Sonoda Y, Kim H, von Heijne G, et al. (2008) GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in *Saccharomyces cerevisiae*. *Nat Protoc* 3: 784–798.
46. Waugh DS (2005) Making the most of affinity tags. *Trends Biotechnol* 23: 316–320.
47. Nallamsetty S, Waugh D (2006) Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. *Protein Expr Purif* 45: 175–182.
48. Esposito D, Chatterjee DK (2006) Enhancement of soluble protein expression through the use of fusion tags. *Curr Opin Biotechnol* 17: 353–358.
49. Smyth DR, Mrozkiewicz MK, McGrath WJ, Listwan P, Kobe B (2003) Crystal structures of fusion proteins with large-affinity tags. *Protein Sci* 12: 1313–1322.
50. Maskos K, Huber-Wunderlich M, Glockshuber R (2003) DsbA and DsbC-catalyzed oxidative folding of proteins with complex disulfide bridge patterns in vitro and in vivo. *J Mol Biol* 325: 495–513.
51. Drag M, Salvesen GS (2008) DeSUMOylating enzymes-SENPs. *IUBMB Life* 60: 734–742.
52. Kefala G, Kwiatkowski W, Esquivies L, Maslennikov I, Choe S (2007) Application of Mystic to improving the expression and membrane integration of histidine kinase receptors from *Escherichia coli*. *J Struct Funct Genomics* 8: 167–172.
53. Angulo I, Acebrón I, de Las Rivas B, Muñoz R, Rodríguez-Crespo I, et al. (2011) High-resolution structural insights on the sugar-recognition and fusion tag properties of a versatile beta-trefoil lectin domain from the mushroom *Lactiporus sulphureus*. *Glycobiology* 21: 1349–1361.
54. Sambrook J, Russell WD (2001) *Molecular cloning: A laboratory manual*. Third Edit. Cold Spring Harbor (NY).
55. Hemsley A, Arnheim N, Toney MD, Cortopassi G, Galas DJ (1989) A simple method for site-directed mutagenesis using the polymerase chain reaction. *Nucleic Acids Res* 17: 6545–6551.
56. Hanahan D (1983) Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* 166: 557–580.
57. Kapust RB, Tózsér J, Fox JD, Anderson DE, Cherry S, et al. (2001) Tobacco etch virus protease: Mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng* 14: 993–1000.
58. Reverter D, Lima CD (2009) Preparation of SUMO proteases and kinetic analysis using endogenous substrates. *Methods Mol Biol* 497: 225–239.
59. Goulas T, Arolas JL, Gomis-Rüth FX (2011) Structure, function and latency regulation of a bacterial enterotoxin potentially derived from a mammalian adamalysin/ADAM xenolog. *Proc Natl Acad Sci* 108: 1856–1861.
60. Schagger H (2006) Tricine-SDS-PAGE. *Nat Protoc* 1: 16–22.
61. Studier WF, Rosenberg AH, Dunn JJ, Dubendorf JW (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol* 185: 60–89.
62. Phan TTP, Nguyen HD, Schumann W (2012) Development of a strong intracellular expression system for *Bacillus subtilis* by optimizing promoter elements. *J Biotechnol* 157: 167–172.
63. Daly R, Hearn MTW (2005) Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. *J Mol Recognit* 18: 119–138.
64. Aslanidis C, de Jong PJ (1990) Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res* 18: 6069–6074.
65. Harper S, Speicher DW (2011) Purification of proteins fused to glutathione S-transferase. *Methods Mol Biol* 681: 259–280.
66. Peroutka Iii RJ, Orcutt SJ, Strickler JE BT (2011) SUMO fusion technology for enhanced protein expression and purification in prokaryotes and eukaryotes. *Methods Mol Biol* 705: 15–30.
67. Kapust RB, Waugh DS (1999) *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci* 8: 1668–1674.
68. Simonen M, Palva I (1993) Protein secretion in *Bacillus* species. *Microbiol Rev* 57: 109–137.
69. De Marco A (2009) Strategies for successful recombinant expression of disulfide bond-dependent proteins in *Escherichia coli*. *Microb Cell Fact* 8: 26.
70. Macauley-Patrick S, Fazenda ML, McNeil B, Harvey LM (2005) Heterologous protein production using the *Pichia pastoris* expression system. *Yeast* 22: 249–270.
71. Reverter D, Garcia-Saez I, Catusas L, Vendrell J, Coll M, et al. (1997) Characterisation and preliminary X-ray diffraction analysis of human pancreatic procarboxypeptidase A2. *FEBS Lett* 420: 7–10.
72. Arolas JL, Botelho TO, Vilcinskas A, Gomis-Rüth FX (2011) Structural evidence for standard-mechanism inhibition in metallopeptidases from a complex poised to resynthesize a peptide bond. *Angew Chemie* 50: 10357–10360.
73. Sanglas L, Aviles FX, Huber R, Gomis-Rüth FX, Arolas JL (2009) Mammalian metallopeptidase inhibition at the defense barrier of *Ascaris* parasite. *Proc Natl Acad Sci U S A* 106: 1743–1747.
74. Roosild TP, Greenwald J, Vega M, Castronovo S, Riek R, et al. (2005) NMR structure of Mystic, a membrane-integrating protein for membrane protein expression. *Science* 307: 1317–1321.
75. Marrero A, Mallorqui-Fernández G, Guevara T, Garcia-Castellanos R, Gomis-Rüth FX (2006) Unbound and acylated structures of the MccR1 extracellular antibiotic-sensor domain provide insights into the signal-transduction system that triggers methicillin resistance. *J Mol Biol* 361: 506–521.
76. Kapust RB, Tózsér J, Copeland TD, Waugh DS (2002) The P1' specificity of tobacco etch virus protease. *Biochem Biophys Res Commun* 294: 949–955.
77. Butt TR, Edavettal SC, Hall JP, Mattern MR (2005) SUMO fusion technology for difficult-to-express proteins. *Protein Expr Purif* 43: 1–9.
78. Vergis JM, Wiener MC (2011) The variable detergent sensitivity of proteases that are utilized for recombinant protein affinity tag removal. *Protein Expr Purif* 78: 139–142.
79. Dvir H, Choe S (2009) Bacterial expression of a eukaryotic membrane protein in fusion to various Mystic orthologs. *Protein Expr Purif* 68: 28–33.