REVIEW

THE PROTEIN SOCIETY

New tools for recombinant protein production in *Escherichia coli*: A 5-year update

Germán L. Rosano ,* Enrique S. Morales, and Eduardo A. Ceccarelli

Instituto de Biología Molecular y Celular de Rosario (IBR), CONICET. Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Rosario, Argentina

Abstract: The production of proteins in sufficient amounts is key for their study or use as biotherapeutic agents. *Escherichia coli* is the host of choice for recombinant protein production given its fast growth, easy manipulation, and cost-effectiveness. As such, its protein production capabilities are continuously being improved. Also, the associated tools (such as plasmids and cultivation conditions) are subject of ongoing research to optimize product yield. In this work, we review the latest advances in recombinant protein production in *E. coli*.

Keywords: *Escherichia coli*; *E. coli* strains; recombinant protein production; expression plasmid; affinity tags; fusion partners; inclusion bodies

Introduction

The study of proteins or their use in biotechnological applications often requires their isolation from other cellular components. Purification can be performed from the natural source of the protein; however, this approach is usually cumbersome and inefficient for most of them. The coding sequence for the protein of interest can be inserted into an appropriate expression vector and transformed into a prokaryotic host, such as the bacterium *Escherichia coli*. Using *E. coli* as a microbial cell factory for producing recombinant proteins lowers the costs of production and improves the yield. Nowadays, many proteins of commercial interest are produced in *E. coli*.[1] In the lab, the recombinant production of proteins in *E. coli* is the method of choice for their structural and functional study.

From gene cloning to protein purification, the cellular and molecular tools needed in all steps of the process are widely accessible, and many alternatives are available. Still, failure to obtain a functional recombinant protein is not uncommon, due to protein toxicity to the host or aggregation in inclusion bodies. That is why there is continuous interest in novel approaches that optimize recombinant protein production in *E. coli*. Numerous reviews have covered different aspects of the topic in detail.[2–5] In this review, we cover advances reported in the last 5 years, in the areas of host engineering,

expression vector design, and culture conditions. The newly developed tools show much promise in the field, and we expect them to disseminate in the scientific community rapidly. Lastly, for those who are about to embark on the fascinating world of the heterologous expression of proteins, we advise not only to read this review, but also refer to our previous one, given that both are complementary.[4]

The BL21(DE3) strain: all hail the king

Amid the biotechnological revolution that occurred in the last decades of the 20th century, different *E. coli* lines were tested for their characteristics in the production of recombinant proteins. The B line emerged as the winner given its salient features. One derivative, BL21(DE3), has become the preferred host for recombinant protein production. Genome sequencing of strains of the B line has helped understand the molecular basis of useful phenotypes for heterologous protein synthesis.[6–8]

Fast cell growth and nonmotility

In rich media, common *E. coli* strains have a doubling time of about 20 minutes. Differences in growth time are more notorious in minimal media: B cells typically grow faster than other lines (such as the cloning strain K-12 and its derivatives) under these conditions.[9] Many researchers rediscover that the B cells are nonmotile: When culture vessels are left unattended on the bench, B cells sink to the bottom while cultures with cloning strains remain turbid. Due to a large deletion in *fli* genes required for flagellar proteins biosynthesis, B cells do not have flagella.[6] Some authors have proposed that this in part explains the fast cell growth of B cells as flagellar biosynthesis and assembly is an energy-intensive process.[9]

Diminished acetate production

Compared to K-12, *E. coli* B produces less acetate during cultivation with glucose as the sole carbon source.[10] The pH of cultivation media is typically around 7.0, so acetate is in equilibrium with acetic acid, which in turn diffuses into the cell, altering internal pH control and impairing cell viability.[11] Also, during cultivation, acid production further lowers pH causing more acetic acid to accumulate. In line with this, Wang et al. showed that cultivation at pH 7.5–8.5 improves the production of recombinant proteins by lowering acetate stress.[12]

Protease deficiency

Accumulation of any given protein is a fine balance between its biosynthesis and its degradation. Deficiency in key proteolytic systems in *E. coli* B can extend the lifetime of recombinant proteins in certain cases. An IS186 insertion in the promoter of the *lon* gene eliminates this major protease.[13] Also, a deletion in the gene for the outer membrane protease OmpT can lead to less proteolysis during purification, as cell disruption causes a massive release of this protease.[14]

Superior secretion capabilities

In some occasions, secretion of the foreign proteins yields better results than production in the cytoplasm.[15] B cells secrete more proteins than other laboratory strains, partly due to the existence of a second T2S secretion system.[9]

The features above tipped the scale in favor of the B line for the production of recombinant proteins. A major breakthrough came with the generation of the derivative BL21(DE3) by Studier and Moffat.[16] The BL21(DE3) strain carries a copy of phage T7 RNA polymerase (T7RNAP) under control of the *lac*UV5 promoter. Genes of interest are cloned under control of a T7 promoter in expression plasmids, and protein production begins upon addition of the gratuitous inducer isopropyl β-D-1-thiogalactopyranoside (IPTG). This system provides the user with full control of the induction of protein synthesis with high selectivity and activity. All of these attributes established BL21(DE3) as the preferred host. Its genome has been sequenced and recently curated (GenBank entry CP001509.3).[8] Later derivatives improved other aspects inherent to the production of foreign proteins in *E. coli*, such as rare codon usage and disulfide bond formation. However, the system based on the T7RNAP and its promoter has remained virtually the same since its inception.

Advances in Strain Engineering

As explained above for BL21(DE3) and derivatives, strain engineering has advanced the capabilities of *E. coli* as a cell factory (Table I). Genome manipulation for maximizing heterologous protein production can be undertaken by two different approaches: hypothesis-driven and discovery-driven. A hypothesis-driven strategy aims to manipulate molecular components of a known pathway or process and directly tackling the problem that hampers protein production. In a discovery-driven strategy, cells are mutagenized and then screened or selected for increased protein production.

Hypothesis-driven strain design

Popular strains such as BL21(DE3) pLysS (for control of basal expression), CodonPlus or Rosetta (for codon bias correction), Origami, SHuffle or CyDisCo (for correct disulfide bond formation), Tuner or Lemo21 (for tunable induction), and many others were rationally constructed using this approach.[4] In the last 5 years, several strains were described in which known molecular systems were exploited to enhance protein production. For example, protein secretion via the Tat secretion pathway allows for export of fully folded proteins (fused to a TorA signal peptide) up to 150 kDa in molecular mass. This system can be an excellent alternative to the Sec pathway; however, its low abundance may result in poor yields.[17] Browning et al. engineered

Table I. *Escherichia coli Strains Frequently Used for Recombinant Protein Production*

Strain	Supplier	Application/key features
Strain		
BL21(DE3)	Multiple companies	Routine protein production from coding sequences under control of the T7 promoter/protease deficiency
BL21(DE3) pLysS	Multiple companies	Reduction of basal expression/pLysS plasmid confers chloramphenicol resistance
BL21-CodonPlus(DE3)-RI(PL	Agilent	Codon bias correction/pRI(PL) plasmid provides extra copies of rare tRNAs genes
Rosetta™(DE3) and derivatives	Novagen (Merck)	Codon bias correction/pRARE plasmid provides extra copies of rare tRNAs genes
Origami and derivatives	Novagen (Merck)	Disulfide-bonded protein production/double mutant in reductases (*trxB* and *gor*) resulting in more oxidant conditions in the cytoplasm
SHuffle® T7 Express and derivatives	NEB	Disulfide-bonded protein production/double mutant in reductases (*trxB* and *gor*) resulting in more oxidant conditions in the cytoplasm. Contains the disulfide bond isomerase DsbC for improved fidelity in disulfide bond formation in the cytoplasm
CyDisCo	Ruddock laboratory (University of Oulu)	Disulfide-bonded protein production/expression of mitochondrial thiol oxidase Erv1p (yeast) and protein disulfide isomerase (human)
Tuner(DE3) and derivatives	Novagen (Merck)	Tunability of protein production levels/deletion in lacy permease allows uniform entry of IPTG
Lemo21(DE3)	NEB	Tunability of protein production levels/adjustable levels of T7 lysozyme (inhibitor of T7RNAP) by rhamnose supplementation
BL21-AI	ThermoFisher	Tunability of protein production levels/T7RNAP gene under control of the araBAD promoter
OverExpress™ C41(DE3) and C43(DE3)	Lucigen	Production of membrane/toxic proteins/T7RNAP gene under control of p*lac*Weak promoter. In addition, C43(DE3) has a mutated LacI repressor that binds longer to the *lac* operator, thus delaying T7RNAP transcription. It is also a Lon protease revertant
ArcticExpress (DE3)	Agilent	Production of aggregation-prone proteins/constitutive expression of chaperones Cpn10 and Cpn60 from the psychrophilic bacterium *Oleispira antarctica*, which show high refolding activities at 4–12°C
Newly developed strains		
SixPack	Pósfai laboratory[23]	Codon bias correction/insertion of extra copies of genes encoding six rare tRNAs in a ribosomal RNA operon
Mutant56(DE3)	de Gier laboratory[32]	Production of membrane/toxic proteins/mutant T7RNAP with weakened binding to the T7 promoter
TatExpress BL21	Robinson laboratory[18]	Enhanced protein secretion/p*tac* promoter upstream of *tatABCD operon for increased levels of Tat secretion pathway*
Marionette	Addgene	Protein coexpression/independent control of expression using 12 different inducers
RiboTite	Dixon laboratory[20,21]	Control of basal expression and tunability of protein production levels/introduction of riboswitches upstream of the T7RNAP gene and the coding sequence of interest

BL21 by placing the strong inducible promoter p*tac* upstream of the chromosomal *tatABCD* operon. The resulting strain, dubbed TatExpress BL21, was shown to secrete 30 mg L^{-1} of recombinant human growth hormone into the periplasm.[18]

Another recent development pertains tighter control of basal expression and tunability of protein production by dual transcription-translational control aided by riboswitches. For extremely toxic proteins, leaky expression can lead to cell death. The BL21(DE3) pLysS strain contains a plasmid (pLysS) with the gene for the T7 lysozyme, a natural inhibitor of T7RNAP.[19] This provides an efficient mechanism to inhibit the small amount of T7RNAP synthesized in the absence of inducer, due to its stochastic transcription from the *lac*UV5 promoter. However, even with this tighter control, leaky expression is still known to occur. Moreover, IPTG is a potent inducer of gene expression. Fine-tuning its concentration to reduce the expression of a toxic gene product is a laborious task. In this case, tunable systems allowing for easy and predictable optimization of inducer concentration are highly desirable. The Dixon lab devised the RiboTite system, consisting of the BL21(IL3) strain (or more recently, the BL21[LV2] strain[20]) and the pETORS plasmid.[21] BL21(IL3) possesses the T7RNAP gene with a very similar configuration as in BL21(DE3). However, an orthogonal riboswitch sequence is contained in the 5′ untranslated region of the T7RNAP gene and also before the gene of interest cloned in the pETORS expression plasmid. A riboswitch is a segment in a messenger RNA that folds into intricate structures that block gene expression by interfering with translation. Binding of an effector molecule induces a change in conformation permitting the regulation of expression post-transcriptionally. In RiboTite, the riboswitch is a modified version of the adenine-sensing *add* A-riboswitch from *Vibrio vulnificus* engineered to bind the effector pyrimido-pyrimidine-2,4-diamine (PPDA).[22] So, the expression of the foreign coding sequence can only occur in the presence of both IPTG and PPDA, which effectively reduces leaky expression to almost undetectable levels. Moreover, the amount of recombinant protein can be modulated by tuning PPDA concentration.

The problem of codon bias in recombinant protein production can be addressed by supplementing extra copies of genes coding for rare tRNAs. These extra copies were included in the pRIL and pRARE plasmids, leading to the strains BL21(DE3) CodonPlus and Rosetta, respectively. Interestingly, Lipinszki et al. integrated the six least abundant tRNA species into a ribosomal operon in the chromosome of *E. coli* BL21(DE3).[23] In this way, expression of rare tRNAs is coupled to the actual needs for translational capacity. Additionally, the burden of a second plasmid and supplementation of its corresponding selection antibiotic are avoided. The resulting strain, named SixPack, outperformed (or performed as well as) both BL21(DE3) and Rosetta2(DE3) in the production of recombinant proteins.

Coexpression of proteins requires cloning the corresponding genes in compatible plasmids. Alternatively, the Duet series of plasmids from Novagen allows for cloning two coding sequences under separate T7 promoters in the same plasmid. Up to eight proteins can be coexpressed using compatible Duet plasmids. However, the stoichiometric control of protein level is almost impossible. In the Duet system, all coding sequences are under the influence of the same promoter, which can pose a problem if the levels of the components of a heterologous metabolic pathway need to be fine-tuned for optimum yield. In an impressive work combining genetic engineering, rational design, and directed evolution, Meyer et al. created the Marionette strains.[24] In these cell lines (available in MG1655, DH10B, and BL21 backgrounds), the genes for 12 different genetic regulators were inserted in the genome. The chosen transcription factors control gene expression by binding (positive regulator) or detaching (negative regulator) from their cognate promoter and respond to chemical inducers added to the medium. Theoretically, up to 12 genes of interest can be cloned (the authors tested a five-enzyme lycopene biosynthetic pathway) in plasmids, with each coding sequence under the control of a different inducible promoter of the sensor array. In this way, the level of each recombinant protein can be manipulated at will by adding the proper inducer, with little cross-reactivity, high specificity, low leakiness, and ample dynamic range.

Discovery-driven strain selection and screening

In a discovery-driven approach, mutagenesis with chemical agents or transposons, directed evolution methods or spontaneous mutations under strong selection pressures are used to alter the bacterial genome. Then, selections and screening assays are performed to pick strains displaying better protein production than the wild-type. In this strategy, previous knowledge of mechanisms that hamper protein production is not necessary. Discovery-driven approaches work well for optimizing the synthesis of toxic proteins. For example, the expression of toxic recombinant proteins causes cell death or poor growth. So, growth of IPTG-resistant colonies may indicate that genome modification in that strain allowed for gene expression of the offending protein. The popular C41(DE3) and C43(DE3) strains were isolated in this way.[25] Protein production was tolerated because homologous recombination of the *lac*UV5 promoter of the T7RNAP gene in the original BL21(DE3) strain with the *lac* promoter of the *lac* operon resulted in a weaker promoter (p*lac*Weak). Thus, T7RNAP levels are reduced in comparison to BL21(DE3) and then, sublethal amounts of recombinant protein can be obtained.[26]

The process is greatly facilitated if additional strong selection or screening methods are used, most importantly in cases where the expression of the protein does not cause growth impairment. For example, the recombinant protein can be fused to an antibiotic selection marker or a fluorescent protein.[27–29] The underlying

assumption is that if the marker protein is functional, then the heterologous protein must be correctly folded as well. Although some examples can be found (reviewed in Schlegel et al.[30]), alas, mutations in the isolated over-producing strains are typically not characterized. Also, successful outcomes can sometimes be protein-dependent, that is, when the production of other proteins (even homologues) is tested, the isolated strain may fail to produce significant amounts.[31] This limits the applicability of the isolated strains and explains why, apart from C41(DE3) and C43(DE3), there have not been major advances in strain isolation of the B line with superior expression capabilities using mutagenesis approaches. Of importance, the de Gier lab reported the isolation and characterization of the *E. coli* membrane protein production strain Mutant56(DE3), which showed better yields for the production of membrane proteins than C41(DE3) and C43(DE3).[32] In this strain, the mutation that allowed better protein production changed one amino acid in T7RNAP, weakening its binding to the T7 promoter.

Advances in Plasmid Design

The sequence encoding for the protein of interest is generally cloned in an expression plasmid. The plasmid must contain at least a promoter and a translation initiation region to direct the expression of the coding sequence, a selectable marker, and replication elements. Additionally, the vector may contain other genetic elements to facilitate the detection, purification, or solubilization of the protein, such as sequences encoding for affinity tags and fusion partners. In the last few years, there have been many advances in all of these features, which are described later.

Promoters and translation initiation regions

Promoters influence protein yield by modulating two key aspects in the expression of heterologous genes: stringency of repression before induction (low stringency leading to high levels of leaky expression) and rate of transcription after inducer addition. An "ideal" promoter should not allow for basal expression and should permit the synthesis of high amounts of messenger RNA ("strong" promoter) after induction. Also, manipulation of RNA levels by adjusting inducer concentration (tunability) is another desirable trait.

As already explained, the T7 promoter is the most widely used for recombinant protein production. It is present in the pET series of vectors and is probably one of the strongest promoters known. However, when using the BL21(DE3)/pET vectors system, the levels of recombinant proteins are difficult to manipulate, as IPTG is a potent inducer even at very low concentrations. For this reason, other alternatives such as the tunable araBAD promoter may be more suitable.[33] The list of promoters used for recombinant protein production is long and includes more than 10 different options, where the user can select for strong/weak, tunable/constitutive, or chemically inducible/thermally inducible promoters.[4,34]

In the last few years, tools for the determination of the best promoter for a given coding sequence for the protein of interest were generated. Yang et al. have designed a vector suite for the screening of 10 IPTG-inducible promoters (T7, A3, *lpp*, *tac*, *pac*, Sp6, *lac*, *npr*, *trc*, and *syn*). These promoters are contained in plasmids with a PLICing position, so that a target previously amplified with phosphorothioated primers can be cloned into 10 vectors in a single step without using restriction enzymes.[35] Similarly, Cheng et al. devised a method for rapid promoter replacement, called ReToAd ("*re*treat *to ad*vance").[36] Seven promoters were cloned in a specific region of the vector containing the gene of interest by whole-plasmid amplification and touchdown polymerase chain reaction (PCR) in a single reaction. Then, colonies were screened for optimized protein production. PLICable promoters and ReToAd are interesting methods to select the best promoter for any given protein. Undoubtedly, their dissemination and inclusion of other promoters are warranted. Of note, Anilionyte et al. discovered that variants of the promoter PthrC are self-inducible by growth phase transition (specifically, when the culture reaches an optical density around 0.5, a common value used for inducer addition).[37] The use of PthrC-derived promoters in expression vectors eliminates monitoring cell density and manual induction, which would be useful in high-throughput trials of protein production.

Translation initiation regions contain a Shine-Dalgarno sequence and a linker region to the start codon. These sequences are optimized for protein production in expression vectors. However, due to cloning procedures, new suboptimal sequences may be generated between the promoter and the start of the gene of interest ("cloning scars"). Mirzadeh et al. proposed a PCR-based method to generate a library of the vector–coding sequence junction.[38] Briefly, the cloning scar sequence was changed in all possible combinations by PCR with degenerate primers. The library was transformed in *E. coli*, and protein production was screened via cell sorting as the coding sequence of interest was fused to green fluorescent protein (or more recently, by translationally coupling the coding sequence to an antibiotic resistance gene[39]). The authors reported a 1000-fold difference between low and high expression vectors, which highlights the importance of optimizing the translation initiation region. Nevertheless, new restriction-free cloning methods are becoming increasingly popular and permit cloning of the gene without generating cloning scars.[40]

Selection markers

Plasmid maintenance is ensured by including a selection marker. Under the selection pressure, only cells containing the plasmid survive. In protein overproducing strains, accumulation of the heterologous protein causes

Recombinant protein production in *Escherichia coli*

a metabolic burden and nonproducing cells eventually overtake the culture, resulting in yield decline over time.[41] Commonly, plasmids contain genes conferring antibiotic resistance. Media supplementation with antibiotics is simple, cost-effective, and convenient and is by far the most common strategy for protein synthesis at a lab scale. However, at larger scales, the use of antibiotics is frowned upon due to its associated costs, environmental pollution, and regulatory restrictions. Much progress has been made to develop antibiotic-free selection systems, most importantly, in the area of plasmid-addiction and nutrient prototrophies. Recently, Ali et al. designed expression vectors containing the gene encoding the enzyme enoyl-acyl carrier protein reductase from *Vibrio cholerae* (*fab*V), which confers resistance to Triclosan, a nonantibiotic biocide polychloro phenoxy phenol.[42] Protein production levels were the same when compared to cells carrying the same expression plasmid containing a β-lactamase gene for selection. Also, the construction of high-copy number expression plasmids with increased stability has been described. Primelles Eguia et al. used a *par* locus, a *cis*-acting locus that allows stable plasmid inheritance, to assure retention of the plasmid pAR-KanI.[43] After 8 hours of induction, almost all cells maintained the vector in the absence of antibiotic. In contrast, if cells bearing the commercial pET28 vector are grown in the absence of antibiotic, only 5% retain the plasmid.

The systems mentioned earlier select for plasmid-containing cells, but they do not select for protein-overproducing ones. Even if cells contain the expression vector, protein production can diminish over time due to accumulated mutations and insertion of mobile sequences in the coding sequences of interest.[41,44] Rubjberg et al. reported a very interesting approach that rewards over-producing cells. The system relies on product-addiction for mevalonate production, so only over-producing cells can survive.[45] This feat was achieved by placing two non-conditionally essential genes under the control of the pBAD promoter in a mevalonate-producing *E. coli* strain. Then, an engineered mevalonate-responsive AraC variant was provided so that cells became product-addicted. The resulting strain showed high-yield and stable mevalonate production over 95 generations. Another advantage is that the system does not rely on external inputs, that is, supplementation of media. The system works for this particular case (mevalonate production) but this strategy and others relying on similar principles[46,47] pave the way for future advances.

Affinity tags and fusion partners

Overproducing the protein of interest in *E. coli* is just the first of many challenges in the process. Subsequent goals include the detection and purification of the protein to homogeneity (at least >95% purity as judged by gel electrophoresis in denaturing conditions). Fusion of the protein of interest to affinity tags and fusion partners markedly facilitate downstream processing.

Commercial antibodies are available for many of the tags (for example, anti-His tag, anti-FLAG, anti-StrepII tag, to name a few) allowing for easy detection of the tagged protein by western blot. Also, the tags can bind to their cognate substrates or cofactors when immobilized in resins, thus permitting protein purification by affinity chromatography. Moreover, some fusion partners have solubility-enhancing properties that increase the yield of soluble protein. Typical examples include glutathione S-transferase (GST), maltose binding protein (MBP), and members of the small ubiquitin-like modifier (SUMO) family of proteins.[4] Choosing one of the many options is not trivial, as fusion tags may work in one case but not in another. To easily screen among different fusion partners, Correa et al. designed a T7 vector suite (based on the pET32 plasmid) that includes six histidine-tagged proteins with solubility-enhancing properties.[48,49] Cloning of the coding sequence of interest is achieved by restriction-free cloning using only a pair of primers for all vectors.

In the last 5 years, there have been many reports describing new fusion partners and affinity tags, including in-depth reviews.[50,51] In the area of small peptide tags, Ojima-Kato et al. reported that the addition of the sequence coding for serine-lysine-isoleucine-lysine (SKIK tag) after the initial N-terminal methionine codon markedly improves the expression of recombinant proteins.[52] Nguyen et al. devised the NT-11 tag, derived from the first 11 amino acid residues of a duplicated carbonic anhydrase from *Dunaliella*.[53] The NT-11 tag enhanced the soluble production of all tested proteins, which were also His-tagged. Moving up to fusion tags of higher molecular mass, the Kumar lab designed a 34-amino acid heparin-binding affinity tag (HB-tag).[54,55] HB-tagged proteins can be purified using a heparin-agarose resin under naturing or denaturing conditions. Tagged proteins can be eluted by a simple NaCl gradient. Han et al. used a truncated maltotriose-binding protein from *Pyrococcus furiosus* (MBP-Pyr) as a solubility enhancing tag. They also included a modified histidine tag, consisting of intercalating histidine and glutamate residues (HE tag). MBP-Pyr is closely related to the widely used MBP tag. However, higher expression and solubility of HE-MBP(Pyr)-fusion proteins relative to His-tagged MBP(Pyr)-fusion proteins was achieved.[56] The Zarate lab used small (<10 kDa) metal-binding proteins as fusion tags (SmbP and CusF), which have the dual function of solubility enhancers and binding to resins with immobilized metal ions.[57,58]

Even though heterologous protein aggregation in inclusion bodies (IBs) is typically undesired, there has been increased advocacy for IBs.[59,60] IB formation allows high product yield and purity and the production of toxic proteins in denaturing conditions. In some cases, this strategy may be the only option to produce a recombinant protein. In that venue, Jong et al. discovered that a 39-amino acid signal sequence of the protein TorA from *E. coli* (ssTorA) promoted the formation of

IBs of even highly soluble proteins such as thioredoxin and MBP.[61] The ssTorA tag is an interesting tool for producing proteins where other strategies have failed or may even become the first choice in particular cases.

Culture Conditions and Growth Monitoring

Changing the conditions of the culture is the easiest way to alter the growth of *E. coli* and directly impacts on the volumetric yield of the recombinant protein. Researchers often overlook the benefits of optimizing external factors (except the temperature of the culture after induction). Previously, we alerted that Luria-Bertani (LB) medium (arguably, the most popular media for growing *E. coli*) is far from being the most suitable media for protein over-production.[4] Simply switching to richer broths (like Terrific Broth) leads to higher final cell densities. Also, autoinduction media has gained popularity, and many examples of successful protein production by growing *E. coli* in autoinduction media abound in the literature. Autoinduction media consists of at least two carbon sources: glucose and lactose (glycerol is also added to increase yields).[62] Glucose is the preferred carbon source, and once it is depleted (typically during exponential growth), lactose starts being consumed. This, in turn, induces protein production from systems based on the *lac* promoter. There are several advantages of using auto-induction media: higher bacterial densities are achieved, the time point of induction is highly reproducible, and stopping the culture for IPTG addition is no longer needed. Briand et al. have shown that autoinduction can also be achieved by internally altering cell metabolism. They serendipitously discovered that producing human Hsp70 in *E. coli* somehow interferes with the function of the endogenous enzyme glyceraldehyde 3-phosphate dehydrogenase, leading indirectly and consequently to LacI removal and subsequent expression of the gene of interest, if under the influence of lac-based promoter.[63] This platform (named SILEX, *Self-InducibLe Ex*pression) allows autoinduction without any medium adaptation and works under many culture conditions.

Complex media provide abundant availability of nutrients but little control over the metabolic state of cells in culture. Protein production is benefited by lowering the growth rate so that the equilibrium between synthesis and folding can be reached. Culture temperature is the easiest way to manipulate growth rate. However, in large bioreactors, temperature control can increment the costs of production exponentially, so other means of growth rate control are utilized. We agree with the view of Krause et al. that there is a disconnection between principles applied in recombinant protein production in fed-batch cultures and shake-flask cultures.[64] In the former, degree of aeration and rate of glucose feeding are closely monitored, as these parameters are of utmost importance to attain high yields. However, at lab scale, this is rarely taken into account. Luckily, advances in this area show promising results that will lead to their adoption. For example, controlled carbon source supply

in shake flasks has been achieved by controlled release of glucose by its diffusion from silicone beads[65] or in situ degradation of polysaccharides.[66] Oxygen availability can be modulated by proper flask design[67] or by adding immiscible oxygenated oils.[68] Lastly, devices that monitor these parameters and others have been adapted for cultures in microplates and shake flasks so that the best conditions for recombinant protein production can be found.[69] However, because culture condition monitoring at shake-flask scale involves specialized hardware, it is still not widely used.

Lastly, light may soon become another parameter to adjust during cultivation with the introduction of blue light-inducible T7RNAPs (Opto-T7RNAPs).[70] To create Opto-T7RNAP, the polymerase was split into two parts, which were then each fused to light-inducible dimerizers. Upon illumination of the culture with LEDs emitting blue light, the dimerizers interact, and a functional T7RNAP is generated. Thus, blue light acts as an inducer of expression. Moreover, Opto-T7RNAP dissociates in the dark, thereby allowing for dynamic and precise control of protein production. The blue light/Opto-T7RNAP systems can replace IPTG as an inducer, which is an attractive feature, especially at high-scale fermentations.

Toolkits for the Expressionist

The reviews that cover the field of recombinant protein production in *E. coli* describe the many molecular tools available for a successful outcome. Yet, many researchers outside the field who require a recombinant protein are overwhelmed by the amount of information. In fact, many have confessed to us that this is one of the reasons for the abandonment of the project. In this section, three toolkits equipped with different amounts of resources, depending on the research group's needs, are described (Fig. 1).

The starter kit

At this point, it should be clear that every lab aiming at obtaining a recombinant protein in *E. coli* should have at least a stock of the BL21(DE3) strain (including a pLys stock for tighter control of basal expression) and vectors of the pET series. We strongly recommend the T7 vector suite from Correa et al.[48] Cloning into these six different plasmids is quite easy and requires one pair of primers for PCR as employed in restriction-free cloning. These plasmids will also allow for surveying the effect of five (or six in the latest update[49]) fusion tags on solubility. Other reagents needed in this kit include nickel-NTA-agarose beads for purification and the Tobacco Etch Virus (TEV) protease (which can be prepared *in house*[71]) for tag removal.

The intermediate combo

In addition to the tools listed earlier, strains that correct for codon usage usually improve the levels of production of recombinant protein dramatically.[72] If the gene of interest has not been codon-optimized, then the BL21

DO YOU WANT TO PRODUCE A RECOMBINANT PROTEIN IN *E. coli*?

	STRAINS	VECTORS	CULTIVATION
STARTER KIT	BL21(DE3) BL21(DE3) pLysS	pET plasmids T7 vector suite	LB medium Temperature control
INTERMEDIATE COMBO	TUNABILITY Lemo21/Tuner/AI CODON BIAS CodonPlus/Rosetta	TUNABILITY pBAD series of vectors and derivatives	Autoinduction medium
PROFESSIONAL PLATFORM	TOXIC/MB PROTEINS C41/C43 (DE3) DISULFIDE BONDS Origami/SHuffle CyDisCo	CO-EXPRESSION pDuet vectors FOLDING ASSISTANCE Chaperone co-expression	Online monitoring Controlled glucose release systems Special flasks

Figure 1. Toolkits available for recombinant protein production. The items in the figure are those thoroughly tested over many years of widespread use, so novel discoveries described in the text were not included. MB, membrane.

(DE3) CodonPlus or Rosetta strains are must-haves. Strains [like Lemo21(DE3) and Tuner™ (DE3)][73] and vectors (for example, containing the araBAD promoter)[74] that allow for tunable levels of synthesis are good choices for better controlling the amount of recombinant protein produced, which can be particularly helpful in cases of protein toxicity. This intermediate combo is suitable for meeting the needs of protein-intensive methodologies, such as nuclear magnetic resonance or X-ray crystallography. In fact, the components enumerated in this section are the most widely used combination in structural biology projects.[75]

The professional platform

Finally, in addition to the tools listed earlier, a complete platform encompasses others needed for challenging projects. In cases of membrane, very toxic, disulfide-rich or aggregation-prone proteins, then specialty strains and vectors come in handy. For membrane proteins, the C41(DE3) and C43(DE3) (and the newly developed Mutant56) strains give better results than BL21(DE3). Correct disulfide-bridges configuration can be achieved by using the Origami and SHuffle strains and derivatives. In our hands, the plasmid suite from Takara encoding for different sets of molecular chaperones is suitable in cases of aggregation-prone proteins. If protein coproduction is needed, the Duet series of vectors (Novagen) has been used with much success. Also, vectors for screening the effect of different promoters of gene expression are good additions to the platform. The novel tools described in this article are included in this subsection, mainly due to the lack of dissemination in the scientific community. However, we have no doubt they will become more popular in the future once researchers start including more and more of them in their recombinant protein production schemes.

Where the Field Is Heading

It is exciting that novel strategies and molecular tools for recombinant protein production in *E. coli* are being reported regularly. We especially praise those that allow for testing multiple variants at once easily. This increases not only the chance of success but also the chances of adoption by newcomers that appreciate easy setups and efficient use of resources. We believe that the use of strains and vectors suites for fast optimization will become the norm. Also, innovations in protein design and engineering must be adopted to maximize protein production. In general, researchers are reluctant to alter the sequence of the protein, as most amino acid substitutions can be destabilizing or disrupt protein function.[76] Nevertheless, tweaking the protein sequence by directed evolution methods can improve the production of recombinant proteins. Depending on the introduced mutation, the overall structure and function can be retained.[76,77] Also, tools allowing for post-translational modifications are still scarce. Even though glycosylation patterns can be obtained using engineered *E. coli* strains,[78,79] more research is warranted in this area. Finally, even though the popularity of *E. coli* as a host is high, other microorganisms should be considered as options, such

as *Bacilli* strains, *Pseudomonas fluorescens*, *Corynebacterium glutamicum*, and many others. In this regard, recently described shuttle vectors for easy transfer between bacteria are nice additions to the arsenal of tools available for recombinant protein production.[80]

Acknowledgments

References

1. Baeshen MN, Al-Hejin AM, Bora RS, Ahmed MM, Ramadan HA, Saini KS, Baeshen NA, Redwan EM (2015) Production of biopharmaceuticals in *E. coli*: current scenario and future perspectives. J Microbiol Biotechnol 25:953–962.

2. Lebendiker M, Danieli T (2014) Production of prone-to-aggregate proteins. FEBS Lett 588:236–246.

3. Jia B, Jeon CO (2016) High-throughput recombinant protein expression in *Escherichia coli*: current status and future perspectives. Open Biol 6:160196.

4. Rosano GL, Ceccarelli EA (2014) Recombinant protein expression in *Escherichia coli*: advances and challenges. Front Microbiol 5:172.

5. Gileadi O (2017) Recombinant protein expression in *E. coli*: a historical perspective. Methods Mol Biol 1586:3–10.

6. Jeong H, Barbe V, Lee CH, Vallenet D, Yu DS, Choi SH, Couloux A, Lee SW, Yoon SH, Cattolico L, Hur CG, Park HS, Ségurens B, Kim SC, Oh TK, Re L, Studier FW, Daegelen P, Kim JF (2009) Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). J Mol Biol 394:644–652.

7. Studier FW, Daegelen P, Lenski RE, Maslov S, Kim JF (2009) Understanding the differences between genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3) and comparison of the *E. coli* B and K-12 genomes. J Mol Biol 394:653–680.

8. Kim S, Jeong H, Kim EY, Kim JF, Lee SY, Yoon SH (2017) Genomic and transcriptomic landscape of *Escherichia coli* BL21(DE3). Nucleic Acids Res 45:5285–5293.

9. Yoon SH, Jeong H, Kwon S-K, Kim JF. Genomics, biological features, and biotechnological applications of *Escherichia coli* B: "is B for better?!". In: Lee SY, Ed. , 2009Systems biology and biotechnology of Escherichia coli. Dordrecht: Springer Netherlands; p. 1–17.

10. Son YJ, Phue JN, Trinh LB, Lee SJ, Shiloach J (2011) The role of Cra in regulating acetate excretion and osmotic tolerance in *E. coli* K-12 and *E. coli* B at high density growth. Microb Cell Fact 10:52.

11. Leone S, Sannino F, Tutino ML, Parrilli E, Picone D (2015) Acetate: friend or foe? Efficient production of a sweet protein in *Escherichia coli* BL21 using acetate as a carbon source. Microb Cell Fact 14:106.

12. Wang H, Wang F, Wang W, Yao X, Wei D, Cheng H, Deng Z (2014) Improving the expression of recombinant proteins in *E. coli* BL21 (DE3) under acetate stress: an alkaline pH shift approach. PloS one 9:e112777.

13. Sai Sree L, Reddy M, Gowrishankar J (2001) IS186 insertion at a hot spot in the lon promoter as a basis for lon protease deficiency of *Escherichia coli* B: identification of a consensus target sequence for IS186 transposition. J Bacteriol 183:6943–6946.

14. Grodberg J, Dunn JJ (1988) ompT encodes the *Escherichia coli* outer membrane protease that cleaves T7 RNA polymerase during purification. J Bacteriol 170:1245–1253.

15. Burdette LA, Leach SA, Wong HT, Tullman-Ercek D (2018) Developing gram-negative bacteria for the secretion of heterologous proteins. Microb Cell Fact 17:196.

16. Studier FW, Moffatt BA (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J Mol Biol 189:113–130.

17. DeLisa MP, Tullman D, Georgiou G (2003) Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. Proc Natl Acad Sci U S A 100:6115–6120.

18. Browning DF, Richards KL, Peswani AR, Roobol J, Busby SJW, Robinson C (2017) *Escherichia coli* "TatExpress" strains super-secrete human growth hormone into the bacterial periplasm by the tat pathway. Biotechnol Bioeng 114:2828–2836.

19. Studier FW (1991) Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. J Mol Biol 219:37–44.

20. Horga LG, Halliwell S, Castineiras TS, Wyre C, Matos C, Yovcheva DS, Kent R, Morra R, Williams SG, Smith DC, Dixon N (2018) Tuning recombinant protein expression to match secretion capacity. Microb Cell Fact 17:199.

21. Morra R, Shankar J, Robinson CJ, Halliwell S, Butler L, Upton M, Hay S, Micklefield J, Dixon N (2016) Dual transcriptional-translational cascade permits cellular level tuneable expression control. Nucleic Acids Res 44:e21.

22. Robinson CJ, Vincent HA, Wu MC, Lowe PT, Dunstan MS, Leys D, Micklefield J (2014) Modular riboswitch toolsets for synthetic genetic control in diverse bacterial species. J Am Chem Soc 136:10615–10624.

23. Lipinszki Z, Vernyik V, Farago N, Sari T, Puskas LG, Blattner FR, Posfai G, Gyorfy Z (2018) Enhancing the translational capacity of *E. coli* by resolving the codon bias. ACS Synth Biol 7:2656–2664.

24. Meyer AJ, Segall-Shapiro TH, Glassey E, Zhang J, Voigt CA (2019) *Escherichia coli* "marionette" strains with 12 highly optimized small-molecule sensors. Nat Chem Biol 15:196–204.

25. 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.

26. Schlegel S, Genevaux P, de Gier JW (2015) Deconvoluting the genetic adaptations of *E. coli* C41(DE3) in real time reveals how alleviating protein production stress improves yields. Cell Rep 10:1758–1766.

27. Gul N, Linares DM, Ho FY, Poolman B (2014) Evolved *Escherichia coli* strains for amplified, functional expression of membrane proteins. J Mol Biol 426:136–149.

28. Skretas G, Georgiou G (2010) Simple genetic selection protocol for isolation of overexpressed genes that enhance accumulation of membrane-integrated human G protein-coupled receptors in *Escherichia coli*. Appl Environ Microbiol 76:5852–5859.

29. Drew D, Lerch M, Kunji E, Slotboom DJ, de Gier JW (2006) Optimization of membrane protein overexpression and purification using GFP fusions. Nat Methods 3:303–313.

30. Schlegel S, Genevaux P, de Gier JW (2017) Isolating *Escherichia coli* strains for recombinant protein production. Cell Mol Life Sci 74:891–908.

31. Skretas G, Georgiou G (2009) Genetic analysis of G protein-coupled receptor expression in *Escherichia coli*: inhibitory role of DnaJ on the membrane integration of the human central cannabinoid receptor. Biotechnol Bioeng 102:357–367.

32. Baumgarten T, Schlegel S, Wagner S, Low M, Eriksson J, Bonde I, Herrgard MJ, Heipieper HJ, Norholm MH,

Slotboom DJ, de Gier JW (2017) Isolation and characterization of the *E. coli* membrane protein production strain Mutant56(DE3). Sci Rep 7:45089.

33. Marschall L, Sagmeister P, Herwig C (2017) Tunable recombinant protein expression in *E. coli*: promoter systems and genetic constraints. Appl Microbiol Biotechnol 101:501–512.

34. Terpe K (2006) Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotechnol 72:211–222.

35. Yang J, Ruff AJ, Hamer SN, Cheng F, Schwaneberg U (2016) Screening through the PLICable promoter toolbox enhances protein production in *Escherichia coli*. Biotechnol J 11:1639–1647.

36. Cheng F, Xiang C, Zhang XJ, Liu ZQ, Zheng YG (2018) ReToAd: simple method for the rapid replacement of promoters to improve protein production. Biotechnol Lett 40:957–964.

37. Anilionyte O, Liang H, Ma X, Yang L, Zhou K (2018) Short, auto-inducible promoters for well-controlled protein expression in *Escherichia coli*. Appl Microbiol Biotechnol 102:7007–7015.

38. Mirzadeh K, Martinez V, Toddo S, Guntur S, Herrgard MJ, Elofsson A, Norholm MH, Daley DO (2015) Enhanced protein production in *Escherichia coli* by optimization of cloning scars at the vector-coding sequence junction. ACS Synth Biol 4:959–965.

39. Rennig M, Martinez V, Mirzadeh K, Dunas F, Rojsater B, Daley DO, Norholm MHH (2018) TARSyn: Tunable antibiotic resistance devices enabling bacterial synthetic evolution and protein production. ACS Synth Biol 7:432–442.

40. van den Ent F, Lowe J (2006) RF cloning: a restriction-free method for inserting target genes into plasmids. J Biochem Biophys Methods 67:67–74.

41. Corchero JL, Villaverde A (1998) Plasmid maintenance in *Escherichia coli* recombinant cultures is dramatically, steadily, and specifically influenced by features of the encoded proteins. Biotechnol Bioeng 58:625–632.

42. Ali SA, Chew YW, Omar TC, Azman N (2015) Use of FabV-Triclosan plasmid selection system for efficient expression and production of recombinant proteins in *Escherichia coli*. PloS One 10:e0144189.

43. Eguia FAP, Ramos HR, Kraschowetz S, Omote D, Ramos CRR, Ho PL, Carvalho E, Goncalves VM (2018) A new vector for heterologous gene expression in *Escherichia coli* with increased stability in the absence of antibiotic. Plasmid 98:22–30.

44. Rugbjerg P, Myling-Petersen N, Porse A, Sarup-Lytzen K, Sommer MOA (2018) Diverse genetic error modes constrain large-scale bio-based production. Nat Commun 9:787.

45. Rugbjerg P, Sarup-Lytzen K, Nagy M, Sommer MOA (2018) Synthetic addiction extends the productive life time of engineered *Escherichia coli* populations. Proc Natl Acad Sci U S A 115:2347–2352.

46. Lv Y, Qian S, Du G, Chen J, Zhou J, Xu P (2019) Coupling feedback genetic circuits with growth phenotype for dynamic population control and intelligent bioproduction. Metab Eng 54:109–116.

47. Enrico Bena C, Grob A, Isalan M, Bosia C, Ceroni F (2018) Commentary: synthetic addiction extends the productive life time of engineered *Escherichia coli* populations. Front Bioengin Biotechnol 6:77.

48. Correa A, Ortega C, Obal G, Alzari P, Vincentelli R, Oppezzo P (2014) Generation of a vector suite for protein solubility screening. Front Microbiol 5:67.

49. Ortega C, Prieto D, Abreu C, Oppezzo P, Correa A (2018) Multi-compartment and multi-host vector suite for recombinant protein expression and purification. Front Microbiol 9:1384.

50. Paraskevopoulou V, Falcone FH (2018) Polyionic tags as enhancers of protein solubility in recombinant protein expression. Microorganisms 6:47.

51. Vandemoortele G, Eyckerman S, Gevaert K (2019) Pick a tag and explore the functions of your pet protein. Trends Biotechnol. https://doi.org/10.1016/j.tibtech.2019.03.016.

52. Ojima-Kato T, Nagai S, Nakano H (2017) N-terminal SKIK peptide tag markedly improves expression of difficult-to-express proteins in *Escherichia coli* and *Saccharomyces cerevisiae*. J Biosci Bioeng 123:540–546.

53. Nguyen TKM, Ki MR, Son RG, Pack SP (2019) The NT11, a novel fusion tag for enhancing protein expression in *Escherichia coli*. Appl Microbiol Biotechnol 103: 2205–2216.

54. Jayanthi S, Gundampati RK, Kumar TKS (2017) Simple and efficient purification of recombinant proteins using the heparin-binding affinity tag. Curr Protoc Prot Sci 90:6 16 11–16 16 13.

55. Morris J, Jayanthi S, Langston R, Daily A, Kight A, McNabb DS, Henry R, Kumar TKS (2016) Heparin-binding peptide as a novel affinity tag for purification of recombinant proteins. Prot Express Purif 126:93–103.

56. Han Y, Guo W, Su B, Guo Y, Wang J, Chu B, Yang G (2018) High-level expression of soluble recombinant proteins in *Escherichia coli* using an HE-maltotriose-binding protein fusion tag. Prot Express Purif 142:25–31.

57. Vargas-Cortez T, Morones-Ramirez JR, Balderas-Renteria I, Zarate X (2016) Expression and purification of recombinant proteins in *Escherichia coli* tagged with a small metal-binding protein from *Nitrosomonas europaea*. Prot Express Purif 118:49–54.

58. Vargas-Cortez T, Morones-Ramirez JR, Balderas-Renteria I, Zarate X (2017) Production of recombinant proteins in *Escherichia coli* tagged with the fusion protein CusF3H. Prot Express Purif 132:44–49.

59. Wurm DJ, Quehenberger J, Mildner J, Eggenreich B, Slouka C, Schwaighofer A, Wieland K, Lendl B, Rajamanickam V, Herwig C, Spadiut O (2018) Teaching an old pET new tricks: tuning of inclusion body formation and properties by a mixed feed system in *E. coli*. Appl Microbiol Biotechnol 102:667–676.

60. Ramon A, Senorale-Pose M, Marin M (2014) Inclusion bodies: not that bad. Front Microbiol 5:56.

61. Jong WSP, Vikström D, Houben D, van den Berg van Saparoea HB, de Gier J-W, Luirink J (2017) Application of an *E. coli* signal sequence as a versatile inclusion body tag. Microb Cell Fact 16:50.

62. Studier FW (2005) Protein production by auto-induction in high density shaking cultures. Prot Express Purif 41: 207–234.

63. Briand L, Marcion G, Kriznik A, Heydel JM, Artur Y, Garrido C, Seigneuric R, Neiers F (2016) A self-inducible heterologous protein expression system in *Escherichia coli*. Sci Rep 6:33037.

64. Krause M, Neubauer A, Neubauer P (2016) The fed-batch principle for the molecular biology lab: controlled nutrient diets in ready-made media improve production of recombinant proteins in *Escherichia coli*. Microb Cell Fact 15:110.

65. Jeude M, Dittrich B, Niederschulte H, Anderlei T, Knocke C, Klee D, Buchs J (2006) Fed-batch mode in shake flasks by slow-release technique. Biotechnol Bioeng 95: 433–445.

66. Krause M, Ukkonen K, Haataja T, Ruottinen M, Glumoff T, Neubauer A, Neubauer P, Vasala A (2010) A novel fed-batch based cultivation method provides high

cell-density and improves yield of soluble recombinant proteins in shaken cultures. Microb Cell Fact 9:11.

67. Bruecher D (2015) Evolution of shake flask technology. Genet Engin Biotechnol News 35:32–33.

68. Sklodowska K, Jakiela S (2017) Enhancement of bacterial growth with the help of immiscible oxygenated oils. RSC Advan 7:40990–40995.

69. Klockner W, Buchs J (2012) Advances in shaking technologies. Trends Biotechnol 30:307–314.

70. Baumschlager A, Aoki SK, Khammash M (2017) Dynamic blue light-inducible T7 RNA polymerases (Opto-T7RNAPs) for precise spatiotemporal gene expression control. ACS Synth Biol 6:2157–2167.

71. Tropea JE, Cherry S, Waugh DS (2009) Expression and purification of soluble his(6)-tagged TEV protease. Methods Mol Biol 498:297–307.

72. Rosano GL, Ceccarelli EA (2009) Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted *Escherichia coli* strain. Microb Cell Fact 8:41.

73. Wagner S, Klepsch MM, Schlegel S, Appel A, Draheim R, Tarry M, Hogbom M, van Wijk KJ, Slotboom DJ, Persson JO, de Gier J-W (2008) Tuning *Escherichia coli* for membrane protein overexpression. Proc Natl Acad Sci U S A 105:14371–14376.

74. Guzman LM, Belin D, Carson MJ, Beckwith J (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol 177:4121–4130.

75. Structural Genomics C; China Structural Genomics C; Northeast Structural Genomics C, Graslund S, Nordlund P, Weigelt J, Hallberg BM, Bray J, Gileadi O, Knapp S, Oppermann U, Arrowsmith C, Hui R, Ming J, dhe-Paganon S, Park HW, Savchenko A, Yee A, Edwards A, Vincentelli R, Cambillau C, Kim R, Kim SH, Rao Z, Shi Y, Terwilliger TC, Kim CY, Hung LW, Waldo GS, Peleg Y, Albeck S, Unger T, Dym O, Prilusky J, Sussman JL, Stevens RC, Lesley SA, Wilson IA, Joachimiak A, Collart F, Dementieva I, Donnelly MI, Eschenfeldt WH, Kim Y, Stols L, Wu R, Zhou M, Burley SK, Emtage JS, Sauder JM, Thompson D, Bain K, Luz J, Gheyi T, Zhang F, Atwell S, Almo SC, Bonanno JB, Fiser A, Swaminathan S, Studier FW, Chance MR, Sali A, Acton TB, Xiao R, Zhao L, Ma LC, Hunt JF, Tong L, Cunningham K, Inouye M, Anderson S, Janjua H, Shastry R, Ho CK, Wang D, Wang H, Jiang M, Montelione GT, Stuart DI, Owens RJ, Daenke S, Schütz A, Heinemann U, Yokoyama S, Büssow K, Gunsalus KC (2008) Protein production and purification. Nat Methods 5:135–146.

76. Sachsenhauser V, Bardwell JC (2018) Directed evolution to improve protein folding in vivo. Curr Opin Struct Biol 48:117–123.

77. Wang T, Badran AH, Huang TP, Liu DR (2018) Continuous directed evolution of proteins with improved soluble expression. Nat Chem Biol 14:972–980.

78. Mueller P, Gauttam R, Raab N, Handrick R, Wahl C, Leptihn S, Zorn M, Kussmaul M, Scheffold M, Eikmanns B, Elling L, Gaisser S (2018) High level in vivo mucin-type glycosylation in *Escherichia coli*. Microb Cell Fact 17:168.

79. Nothaft H, Szymanski CM (2013) Bacterial protein N-glycosylation: new perspectives and applications. J Biol Chem 288:6912–6920.

80. Nakata PA (2017) Construction of pDUO: a bicistronic shuttle vector series for dual expression of recombinant proteins. Plasmid 89:16–21.