Strategies for efficient production of recombinant proteins in *Escherichia coli*: alleviating the host burden and enhancing protein activity

Zi-Xu Zhang[†], Fang-Tong Nong[†], Yu-Zhou Wang, Chun-Xiao Yan, Yang Gu, Ping Song[*] and Xiao-Man Sun[*]

Abstract

Escherichia coli, one of the most efficient expression hosts for recombinant proteins (RPs), is widely used in chemical, medical, food and other industries. However, conventional expression strains are unable to effectively express proteins with complex structures or toxicity. The key to solving this problem is to alleviate the host burden associated with protein overproduction and to enhance the ability to accurately fold and modify RPs at high expression levels. Here, we summarize the recently developed optimization strategies for the high-level production of RPs from the two aspects of host burden and protein activity. The aim is to maximize the ability of researchers to quickly select an appropriate optimization strategy for improving the production of RPs.

Keywords: *Escherichia coli*, Recombinant protein, Host burden, Post-translational modification, Inclusion bodies

Introduction

Since the last century, the emergence of recombinant protein (RP) expression systems has revolutionized biotechnology. Excitingly, with the advancement of biotechnology, the yield of RPs has increased from the gram to the kilogram scale, and the range of applications has expanded from traditional food and chemical industries to biopharmaceuticals [1, 2]. For example, it is projected that the industrial enzyme market will grow from USD 6.6 billion in 2021 to USD 9.1 billion by 2026 [3], illustrating the enormous market value and growth potential of RPs. Similarly, a variety of protein drugs have been successfully marketed, including monoclonal antibodies (mAbs), recombinant vaccines, and hormones,

demonstrating that RPs already play a significant role in the biopharmaceutical field [4].

Due to its inexpensive fermentation requirements, rapid proliferation ability and stable high-level expression, *Escherichia coli* (hereafter *E. coli*) has become the mainstay of RP expression among prokaryotic expression hosts [5]. As early as the 1970s, *E. coli* was applied in the production of clinical drugs, such as the hormones somatostatin [6] and insulin [7], which were commercialized early on. As a gold standard for expressing RPs, *E. coli* BL21(DE3) and the pET expression system are widely used in research and commercial production. This is primarily attributed to the T7 RNA polymerase (RNAP) from λ prophage in the genome of BL21(DE3), which can specifically recognize the T7 promoter (P_{T7}) on the pET plasmid and transcribe at eightfold the speed of the *E. coli* native RNAP [8, 9]. In recent years, several BL21(DE3)-derived strains have been widely used to produce various types of RPs, including C41/C43(DE3) (for the production

[†]Zi-Xu Zhang and Fang-Tong Nong contributed equally to this work

*Correspondence: songping@njnu.edu.cn; xiaomansun@njnu.edu.cn

School of Food Science and Pharmaceutical Engineering, Nanjing Normal University, Xuelin Road, Qixia District, Nanjing, People's Republic of China

Zhang *et al. Microbial Cell Factories*   (2022) 21:191

Page 2 of 18

of membrane proteins) [10], BL21(DE3)-pLysS (for reduction of T7 RNAP expression intensity) [11], BL21Star(DE3) (for improvement of mRNA stability) [12], and SixPack (for codon bias correction) [13]. Such efficient production capacity has given it an unassailable position in structural research, new enzyme mining and industrial production [14, 15].

Despite the availability of so many alternative expression systems, there is no guarantee that every type of protein will have a high yield or catalytic/functional activity. The occurrence of these phenomena can be attributed to two main aspects: (i) the host burden caused by the massive production of RPs [16] and (ii) the limited post-translational modification (PTM) capacity and generation of inclusion bodies (IBs) [17]. In fact, any production of RPs, especially toxic proteins, will inevitably compete with the host for resources, which are mainly reflected in the additional DNA replication burden, competition for transcription- and translation-related elements (RNAP, ribosomes, tRNA, and amino acids), and the additional energy and substrates consumed by PTMs [18]. For instance, high-level expression of membrane proteins can lead to the saturation of the Sec translocator-dependent transport pathway, affecting electron transport in the respiratory chain and inhibiting the expression of key enzymes of the tricarboxylic acid cycle [19]. Similarly, glucose dehydrogenase (GDH, an industrial enzyme) leads to significant autolysis of the bacterial cell during the later stages of fermentation [20]. To solve this problem, various means of genetic engineering and synthetic biology have been applied to alleviate host burden, including optimization of the expression intensity of T7 RNAP and pET expression systems (Fig. 1A) [21, 22], as well as balancing or decoupling the cell growth and RP production [23–25]. These optimization strategies effectively relieve or even remove the metabolic burden and increase the capacity of unit cell production. However, when proteins are synthesized at high rates, limited PTMs and molecular chaperones can lead to protein misfolding and the formation of a large number of IBs, affecting the functional activity and solubility of certain proteins. Therefore, the production of highly active RPs is also an important optimization aim, which can be achieved by strengthening or supplementing PTMs, increasing proteolysis and overexpressing suitable molecular chaperones [26]. This review summarizes different classes of optimization strategies developed in recent years from the two main aspects of alleviating host burden and optimizing protein activity, providing a reference for increasing the production of different RPs and discusses the future development direction of related optimization strategies.

Optimization of target protein expression rate based on the gold standard T7 RNAP platform

When T7 RNAP is sufficiently induced, its powerful transcriptional capacity enables the rapid production of large amounts of mRNA, bringing the yield of RPs to 50% of the total cellular protein in just a few hours [27]. However, a strong production capacity is a double-edged sword, especially in the expression of toxic proteins. Numerous studies have shown that growth inhibition during RP production is mainly attributed to excessively strong gene transcription, and translation further exacerbates the host burden [21, 28, 29]. Therefore, the ability to precisely balance the intensity of RP transcription and translation levels is key to reducing host burden and increasing production. This is usually optimised in two aspects as follows: T7 RNAP and pET plasmid.

Regulation of the target protein expression rate-T7 RNAP

The easiest way to control the expression intensity of RPs is to regulate the amount and activity of T7 RNAP, which is often achieved by optimizing transcription or translation levels. In the BL21(DE3) genome, the T7 RNAP gene is controlled by the lacUV5 promoter (P_{lacUV5}), which is a strongly inducible promoter that ensures rapid expression and accumulation after induction (induced by Isopropyl-beta-D-thiogalactopyranoside (IPTG)) [30]. However, high levels of expression are not compatible with some RPs, especially toxic proteins. Accordingly, many studies increased the production of toxic proteins by reducing the transcript level of T7 RNAP. For example, the membrane protein expression host C41(DE3) was obtained by stress screening, while the autolysin expression host BL21(DE3-lac1G) was constructed by recombining P_{lacUV5} with P_{lac} sequences [10, 20, 31]. Furthermore, the P_{lacUV5} is independent of CRP, which makes it leakier than P_{lac} [32]. Replacing the promoter of T7 RNAP with other kinds of inducible promoters is an effective way to regulate transcription levels and reduce leakage (Fig. 1B). Du et al. [32] tested the effects of three inducible promoters (P_{araBAD}, P_{rhaBAD} and P_{tet}) on the transcriptional intensity and leaky expression of T7 RNAP, respectively. It was found that all three promoters were suitable for prolonged fermentation of toxic proteins, whereby P_{rhaBAD} and P_{tet} were able to regulate T7 RNAP transcription more rigorously, providing additional options for the expression of various RPs, especially toxic proteins. Similarly, enhancing the ability to block proteins is also an effective way to reduce leaky expression. In addition to the conversion of PlacUV5 to Plac, the study found that the lac repressor gene (lacI) was also mutated (V192F, referred to as mLacI hereafter) in the membrane protein expression host (C41/

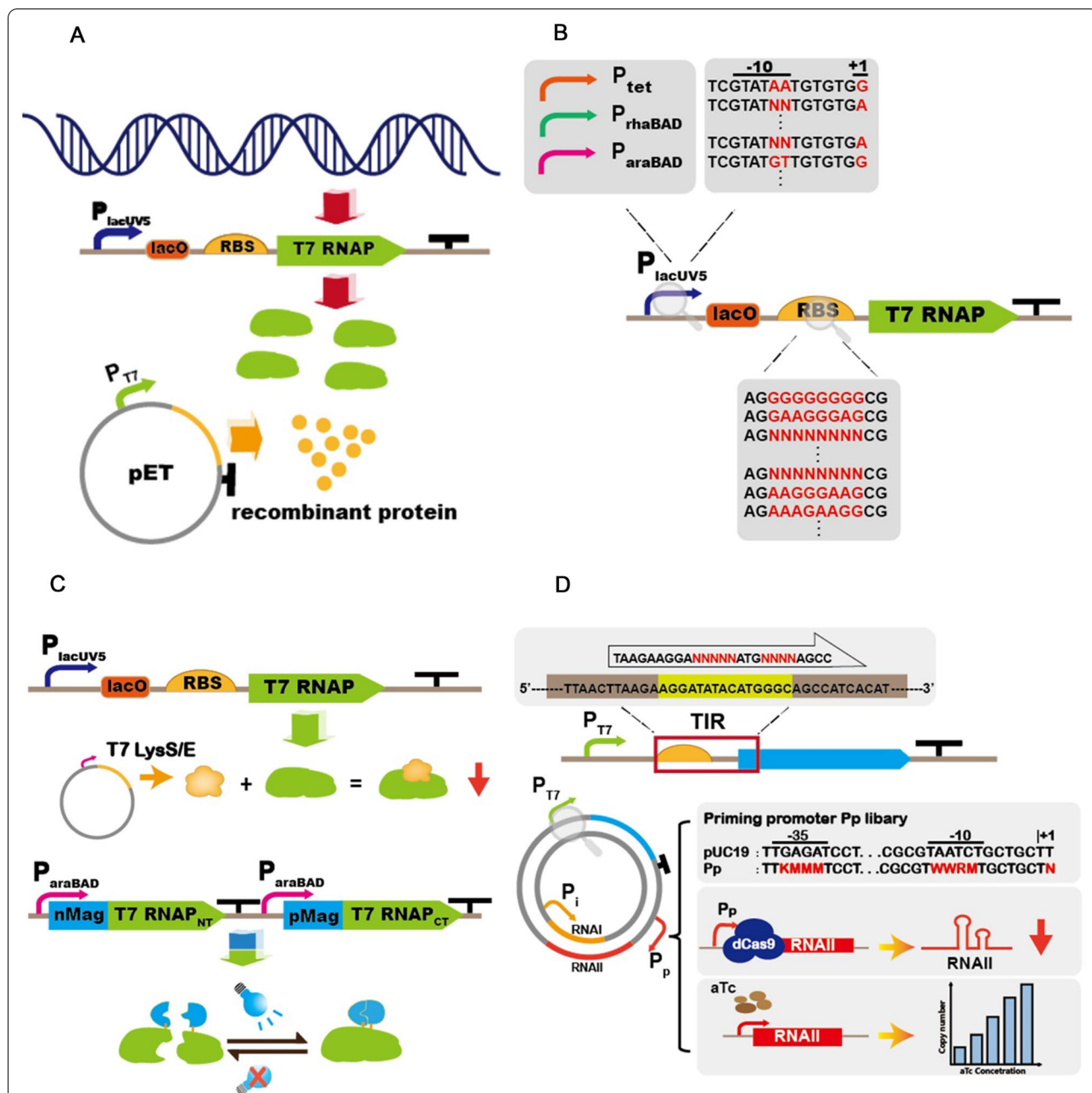Zhang *et al. Microbial Cell Factories*   (2022) 21:191

Page 3 of 18

Fig. 1 The optimization expression strategies for T7 RNAP and pET plasmids. **A** Illustration of protein expression of recombinant protein genes on pET plasmids. **B** Optimization of T7 RNAP transcription and translation level, including substitutions of different promoters, and mutations in promoter functional region and RBS sequence. **C** regulation of T7 RNAP activity. The conventional approach is to utilize lysozyme or light-induction to regulate. **D** Optimization of pET plasmids based on expression intensity and copy numbers. Among them, the expression intensity was optimized by constructing an ITR library to screen for optimal expression results. The degree of binding of RNA-i to RNA-p determines the replication intensity of the plasmid to control the copy numbers. By constructing a promoter library for RNA-p, replacing the inducible promoter, and using dCas9 to regulate expression intensity, the copy numbers can be controlled

C43(DE3)) [33]. Excitingly, mLacI can specifically bind to the lac operator site, but the blocking effect cannot be removed by the addition of IPTG. Based on this phenomenon, Kim et al. [31] developed an anti-leakage

expression system for the overproduction of membrane proteins. Among them, mLacI expression is regulated by the rhamnose inducible promoter P_{rhaBAD}. When trace amounts of L-rhamnose were added, T7 RNAP leakage

Zhang *et al. Microbial Cell Factories*   (2022) 21:191

Page 4 of 18

expression could be inhibited during host growth, reducing growth burden. With the increasing concentration of L-rhamnose, mLacI is abundantly produced and thus reduces the transcription intensity of T7 RNAP, even in the presence of IPTG. This approach makes it possible to control the rate of protein production.

Unlike the transcriptional level, which is controlled by the promoter and RNAP, the strength of translation is mainly determined by the nucleotide sequence and arrangement of the ribosome binding site (RBS) (Fig. 1B). Liang et al. [34] designed 10 RBS sequences with different expression intensities for expressing T7RNAP using an RBS calculator, which was successfully implemented in five Gram-negative and one Gram-positive bacteria. To further extend the regulatory range, Li et al. [35] constructed a more extensive RBS library of T7 RNAP using CRISPR/Cas9 and cytosine base editor, with expression levels ranging from 28 to 220% of the wild-type strain. Using this library, the authors obtained customized hosts for eight difficult-to-express proteins in just three days. The tested model RPs included an autolytic protein, membrane protein, antimicrobial peptide, and insoluble protein, while the production of the industrial enzyme GDH was increased 298-fold. These results show that optimizing the expression intensity of T7 RNAP can effectively improve the RP production, and regulation of the translational level makes it easier to construct screening libraries and rapidly obtain optimized hosts for individual RPs.

Since it is an enzyme, the catalytic activity of T7 RNAP is also a key factor affecting the rate and efficiency of transcription. Mutations of key amino acid residues in T7 RNAP are one of the most effective methods to tune its activity, whose mechanisms are divided into two categories: weakening the binding ability to P_{T7} or generating code-shifting mutations to reduce the catalytic activity [36–38]. For example, Baumgarten et al. [37] found a single amino acid mutation (A102D) of T7 RNAP in the membrane protein expression host Mt56(DE3), which reduced the ability to bind to the P_{T7} and decreased the RP production rate. In addition, the addition of T7 RNAP inhibitors is also a way to effectively regulate T7 RNAP activity, and various derivative hosts including BL21(DE3)-pLysS, BL21(DE3)-pLysE, and Lemo21(DE3) have been developed based on this principle [39–41] (Fig. 1C). With the development of synthetic biology, researchers hope to change the strength of T7 RNAP activity in logic gates to precisely and dynamically regulate the process of growth and production. A variety of T7 RNAP expression systems regulated by light induction have been developed successively, achieving dynamic regulation of RP production [42–44]. For example, the Opto-T7RNAPs system splits the T7RNAP into two

fragments and expresses them in tandem with a light-sensitive dimerization domain. When the fragments are expressed and irradiated by the light of a specific wavelength, T7 RNAP can resume its transcriptional activity, with up to 80-fold change in activity between blue light and darkness [43]. Regrettably, these studies have only been validated with fluorescent proteins or lycopene, and have not been applied to RP production.

Regulation of the target protein expression rate-pET plasmid

Another key factor affecting the expression rate of RPs depends on the combination of different elements on the pET plasmid, including sequences of relevant functional regions near P_{T7} (-35/-10 region, translation initiation region (TIR) and operator sequence) and replicon [45]. As the core region of the pET plasmid, various functional regions near the P_{T7} determine the rigor of basal expression before induction and the appropriate transcription rate after induction.

To reduce the host burden of leaky expression, several more rigorous inducible systems have been combined with P_{T7} to increase the yield of toxic or structurally complex proteins, such as the cumate operator [46], inducible translational ON orthogonal riboswitch [47], and temperature-regulated self-induction [48]. After solving the leaky expression problem, an urgent task is to quickly screen the appropriate expression intensity of various RPs. In contrast to complex genomic manipulations, the combination of degenerate primers and MEGAWHOP PCR or enzymatic digestion and ligation allows rapid access to very large libraries of various functional sequences, including promoter mutation and TIR libraries [22, 49–51]. It is worth noting that the optimal promoter-TIR combination will not necessarily give the best results (Fig. 1D). For example, the optimal combination yielded a 131-fold increase in the expression of superfolder green fluorescent protein (sfGFP), while the highest yield was achieved after single-factor optimization (TIR) of the expression of DNA glycosylase Neil3, with a threefold increase, and combinatorial optimization produced only a twofold increase [22]. Therefore, the use of resistance markers to flexibly screen the expression levels of RPs is expected to become a faster and more accurate library screening tool, especially when multiple libraries are combined [52].

Replicons, genetic elements that replicate as autonomous units, determine the copy numbers of vectors and compatibility with other plasmids. As many expression units reside in each cell, it is logical to assume that a high plasmid dosage results in higher production of RPs [45]. However, this view does not apply to all RPs, as high copy numbers can contribute to rapid accumulation of large

Zhang *et al. Microbial Cell Factories*   (2022) 21:191

Page 5 of 18

amounts of mRNA and RPs, resulting in increased host burden. It was found that each additional plasmid molecule in the host cell increases the metabolic burden by 0.063% [53]. Therefore, an appropriate copy number can provide a balance between growth and production. Generally, replicon replacement is a preferred method for regulating copy numbers [38, 54], with choices ranging from high-copy-number replicons (pUC series, 500–700 copies [55]) to low-copy-number replicons (pSC101, <5 copies [56]). However, this permanent adjustment of copy numbers makes it difficult to balance the host burden of high copy numbers or low production due to insufficient plasmid copies. Recently, this challenge has been overcome by the dynamic copy number regulation system, which works by regulating key genes of the plasmid replication machinery (priming RNA (RNA-p) and inhibitory RNA (RNA-i)). The degree of binding of RNA-i to RNA-p determines the replication intensity of the plasmid to control the copy numbers. Using inducer-based RNA-p/i promoter libraries, CRISPRi and inducer regulation (Fig. 1D), multiple replicons based on ColE1 can achieve controlled regulation of copy numbers during RP production [53, 57]. For example, Rouches et al. constructed a pUC19 plasmid library spanning 1194 mutants to achieve copy number variations between 1 and 800, thereby optimizing the violacein synthesis pathway and the efficiency of CRISPRi [53]. The appearance of dynamic copy number regulation systems has changed the traditional handling of gene copy numbers, providing a powerful tool to reduce the host burden and improve RP production.

Dual optimization of growth and production—balancing and decoupling

During the exponential growth phase, the content of RNAPs, ribosomes and various essential proteins is generally constant [58]. Coincidentally, induction of RP expression is usually done in the mid-exponential phase, but rapid transcription and translation can lead to an uneven distribution of host resources and thus affect growth [59]. Ceroni et al. [60] developed a burden monitor that allows real-time detection of the host burden through changes in green fluorescence intensity (GFP integrated into the λ locus). It was found that the expression intensity of RPs and the molecular weight was proportional to the host burden in MG1655 and DH10β, with the highest reduction of fluorescence intensity reaching more than 90%. At the same time, there was a significant decrease in RP production under high burden conditions. Therefore, another key to improving RP production is to achieve the dual optimization of growth and production, which is best solved by balancing

the allocation of resources or removing the interference between the two fermentation stages.

Balancing cell growth and recombinant protein production

No matter how the production rate is optimized, the RPs will compete for the host nutritional resources, affecting normal growth. Exogenous supplementation can effectively compensate for the nutrients consumed during RP production. Depending on the consumption during the production of pramlintide, some amino acids are categorized as growth-promoting (GP1, including serine, aspartic acid, glutamic acid, threonine and proline) and protein production promoting (GP2, including cysteine, methionine, leucine and alanine) [24]. The combination of 5 mM GP1 at inoculation with 2.5 mM GP1 and GP2 after 6 h in fermentation was the most economical and effective, resulting in a 40% increase of pramlintide production (protein concentration of 3.09 ± 0.12 g/L). In addition, this strategy was also applied to the production of granulocyte colony-stimulating factors.

For the host, reducing unnecessary energy expenditure or blocking byproduct formation can effectively alleviate the burden associated with RP production. The accumulation of acetate is an important factor in the RP production, since it inhibits cell growth and protein synthesis [59]. Blocking the phosphotransferase system (PTS) can effectively reduce the rate of glucose uptake and decrease the production of acetate, which has been applied to increase the production of enhanced GFP (eGFP) [61], vaccines [62], and glutamate dehydrogenase [63]. In addition, knocking out flagellar formation-related genes can reduce energy consumption in *E. coli*. Jae et al. [55] further knocked down the major flagellar regulator (FlhC) in a PTS-blocked strain, which increased the ATP pool and NADPH/NADP$^+$ ratio. These strategies demonstrate that it is feasible to redistribute energy metabolism and reduce by-product formation for the increased RP production.

In addition to the host burden caused by competition for resources, the RP production often triggers a cellular stress response (CSR). Therefore, blocking the emergence of CSR can prevent the down-regulation of a large number of growth-related genes and alleviate the negative effects of CSR on the host [64]. Sharma et al. [64] compared the transcriptomes of cultures of different RPs and selected a series of up-regulated genes for knockout. The results showed that the double knockout mutant BW25113ΔelaA + ΔcysW (DKO) had the highest activity in asparaginase production with 70.3 units/ml. To further unravel the mechanisms involved in CSR mitigation by the DKO strain, Guleria et al. [65] used the strain to overexpress the Rubella E1 gene and performed a transcriptome analysis. Compared to the

Zhang *et al. Microbial Cell Factories*   (2022) 21:191

Page 6 of 18

wild type, down-regulation of multiple genes related to growth-critical processes was suppressed in the DKO strain, including translation, transcription, RNA and ribosome biogenesis, transport, energy metabolism and other catabolic processes. It suggests that the host burden caused by RPs can be effectively mitigated by blocking CSR, which has the potential to serve as a chassis cell to develop an efficient platform for recombinant protein production.

In general, the native genes encoding most heterologous RPs have rare codons, which often affect their translation and folding rate [66]. Two strategies can be applied to alleviate the host burden: heterologous gene codon optimization and supplementation of rare tRNAs. The former not only requires significant experimental resources, but also results in heavy competition for the internal tRNA pool, placing a heavier burden on the host [67]. Conversely, the appropriate introduction of rare codons can improve the yield and solubility of RPs and reduce the host burden [68, 69]. Accordingly, the over-expression of rare tRNAs is a more economical means of optimization. A variety of commercial expression strains, including the Rosetta™(DE3) series and BL21-CodonPlus(DE3), have been developed based on this principle [45]. Unlike the two commercial strains, the newly developed expression host SixPack [13] integrates six of the least abundant tRNA genes into the BL21(DE3) chromosome behind a ribosomal manipulator for expression. This not only relieves the burden of plasmid-based tRNA expression, but also regulates the expression intensity of rare tRNAs through ribosomes, avoiding the waste of resources. This host has been demonstrated to outperform BL21(DE3) and Rosetta2(DE3) in the expression of RPs from eight different origins.

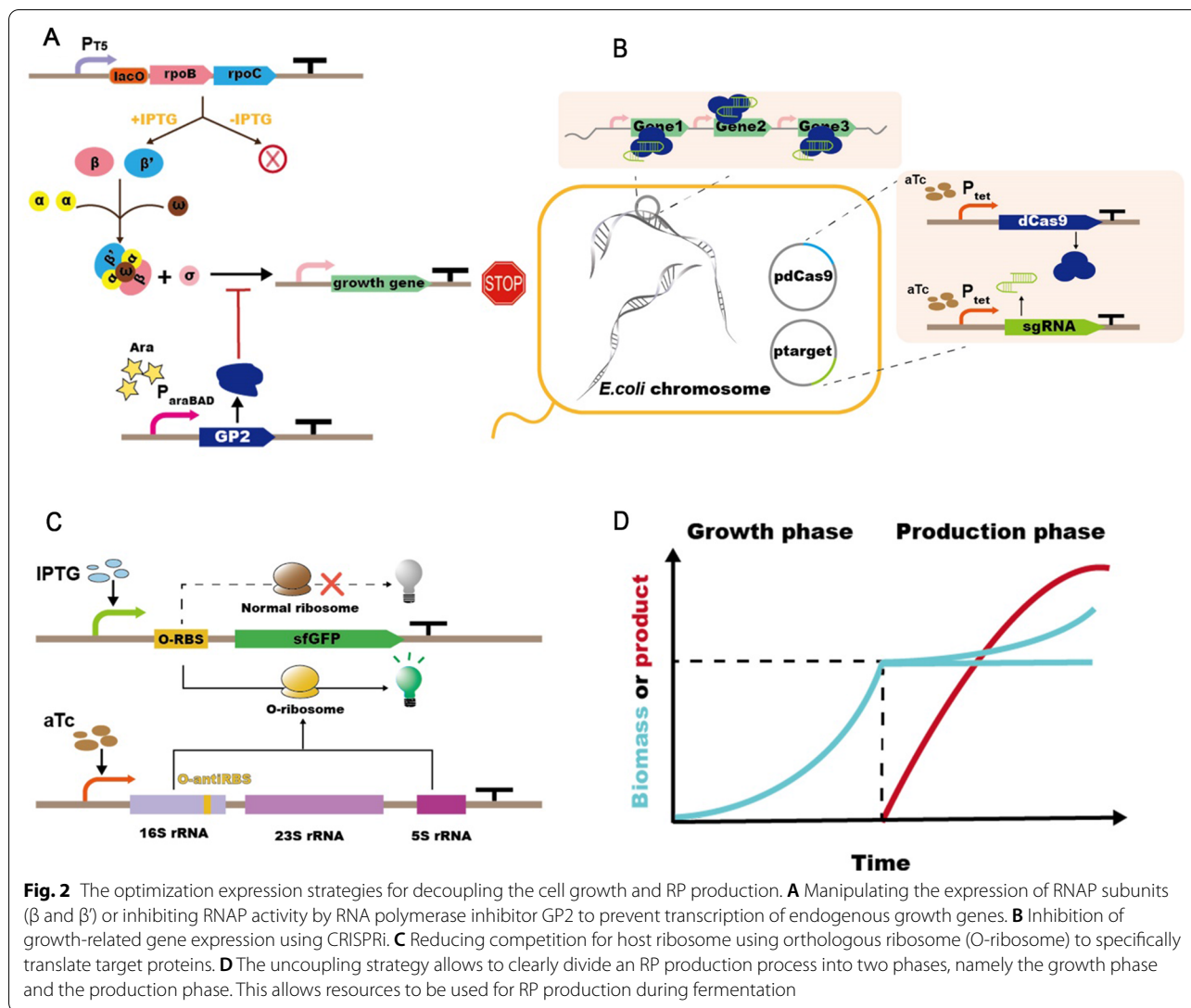## Decoupling cell growth and recombinant protein production

The mechanisms inducing host burdens vary depending on the class of RPs, and a more simplistic approach would be to decouple the cell growth from RP production, effectively reducing the difficulty of resource allocation. In the first stage, the host cells are cultured at a normal growth rate without competition from RP production. Once the culture has reached the stable stage, growth will be stopped and RP production induced so that most of the resources are used for product synthesis. This two-stage fermentation process has been successfully applied to RP production [70].

The auto-induction system is a decoupling method often applied in industrial production. Traditional auto-induction media are usually supplemented with glucose, lactose, or glycerol. When glucose is present, it inhibits the uptake of lactose by the bacterium and prevents RP production. After glucose is exhausted, lactose is transported into the cells to induce RP production [71]. To further expand the range of applications and reduce leaky expression, several types of auto-induction systems have been developed, based on principles such as quorum sensing [72], phosphate induction [73], or molecular chaperones that unblock catabolite repression [74, 75]. Notably, the phosphate-based auto-induction system can be used under different culture conditions, including 384-well plates, shake flasks and bioreactors [69]. Melgar et al. [76] combined this system with lysozyme and DNA/RNA endonuclease to achieve auto-induction and autolysis, allowing the release of more than 90% of the protein and facilitating its application in industrial production.

However, auto-induction systems cannot achieve growth arrest during production, and interrupting cell growth can more efficiently allocate resources to RP production, which is often achieved by inhibiting or blocking the expression of growth-critical genes. A variety of decoupling strategies have been applied to RP production by controlling or inhibiting the expression of endogenous RNAP (Fig. 2A) [25, 77, 78]. Excitingly, blocking the expression of endogenous RNAP improves the efficiency of the insertion of non-canonical amino acids (ncAA) at specific sites, expanding the application range of this strategy [79]. Similarly, blocking the normal replication of chromosomes can also achieve growth arrest. Kasari et al. [80] added serine recombinase recognition sites at both ends of the replication start (oriC) of the chromosome and blocked normal DNA replication by temperature-induced expression of serine recombinase, which resulted in a fivefold increase in the product yield. However, this approach completely blocks the normal growth of the host and cannot achieve a dynamic balance between growth and production. By contrast, inhibition of growth-related proteins (DNA replication, or nucleotide synthesis-related proteins) using CRISPRi can dynamically regulate the growth state (Fig. 2B) [81]. Li et al. [82] constructed a sgRNA library targeting growth-related genes, and 1332 different sgRNAs were screened to reduce host growth and increase GFP accumulation. Among them, GFP production increased more than fivefold when *sibB*/*ibsB* was inhibited.

In fact, the fundamental purpose of decoupling growth and production is to make the best use of the host resources. If a series of orthologous elements are utilized to prevent RP production from depleting key growth resources, the goal of alleviating the host burden can be achieved. Because of the universality and complexity of the cellular translation machinery, there is no unique ribosome in *E. coli* that recognizes specific mRNAs to achieve orthogonal translation [83]. Interaction between RBS and 16S rRNA in the ribosomal

Zhang *et al. Microbial Cell Factories*   (2022) 21:191

Page 7 of 18

Fig. 2 The optimization expression strategies for decoupling the cell growth and RP production. **A** Manipulating the expression of RNAP subunits (β and β′) or inhibiting RNAP activity by RNA polymerase inhibitor GP2 to prevent transcription of endogenous growth genes. **B** Inhibition of growth-related gene expression using CRISPRi. **C** Reducing competition for host ribosome using orthologous ribosome (O-ribosome) to specifically translate target proteins. **D** The uncoupling strategy allows to clearly divide an RP production process into two phases, namely the growth phase and the production phase. This allows resources to be used for RP production during fermentation

subunit is a key regulatory step in the recognition and initiation of translation (Fig. 2C). Darlington et al. [83] evaluated the feasibility of developing orthogonal translation systems development by modeling, further customizing 16S rRNA to successfully develop a more efficient orthologous ribosome (o-ribosome). When no orthologous mRNA is present in the host, the o-ribosome can still translate the endogenous mRNA. With increasing expression of the orthologous mRNA, the o-ribosome recognizes and translates it, preventing this mRNA from occupying the host ribosome and interfering with normal metabolism, which is especially useful in the expression of toxic proteins. However, the o-ribosome is defective and produces proteins with a tenfold lower capacity than that of the natural ribosome. To

solve this problem, various optimization strategies have been applied to improve the orthogonal translation system in recent years [84, 85]. Among them, Liu et al. [84] utilized phage-assisted continuous evolution technology for rapid optimization of 16S rRNA by screening pressure. After multiple rounds of directed evolution, the mutant o-ribosome achieved faster translation, resulting in 6.3-fold higher RP production than the wild-type. Most importantly, this ribosome can introduce ncAAs into the protein with high efficiency, which is 9.08-fold higher than that of the native ribosome, improving the application of orthogonal translation systems in RP production. In brief, whether it is to inhibit or block the expression of growth-essential genes or to use o-ribosomes to express RPs, the aim is to ensure normal growth of the host during the growth phase (Fig. 2D).

Zhang *et al. Microbial Cell Factories*   (2022) 21:191

Page 8 of 18

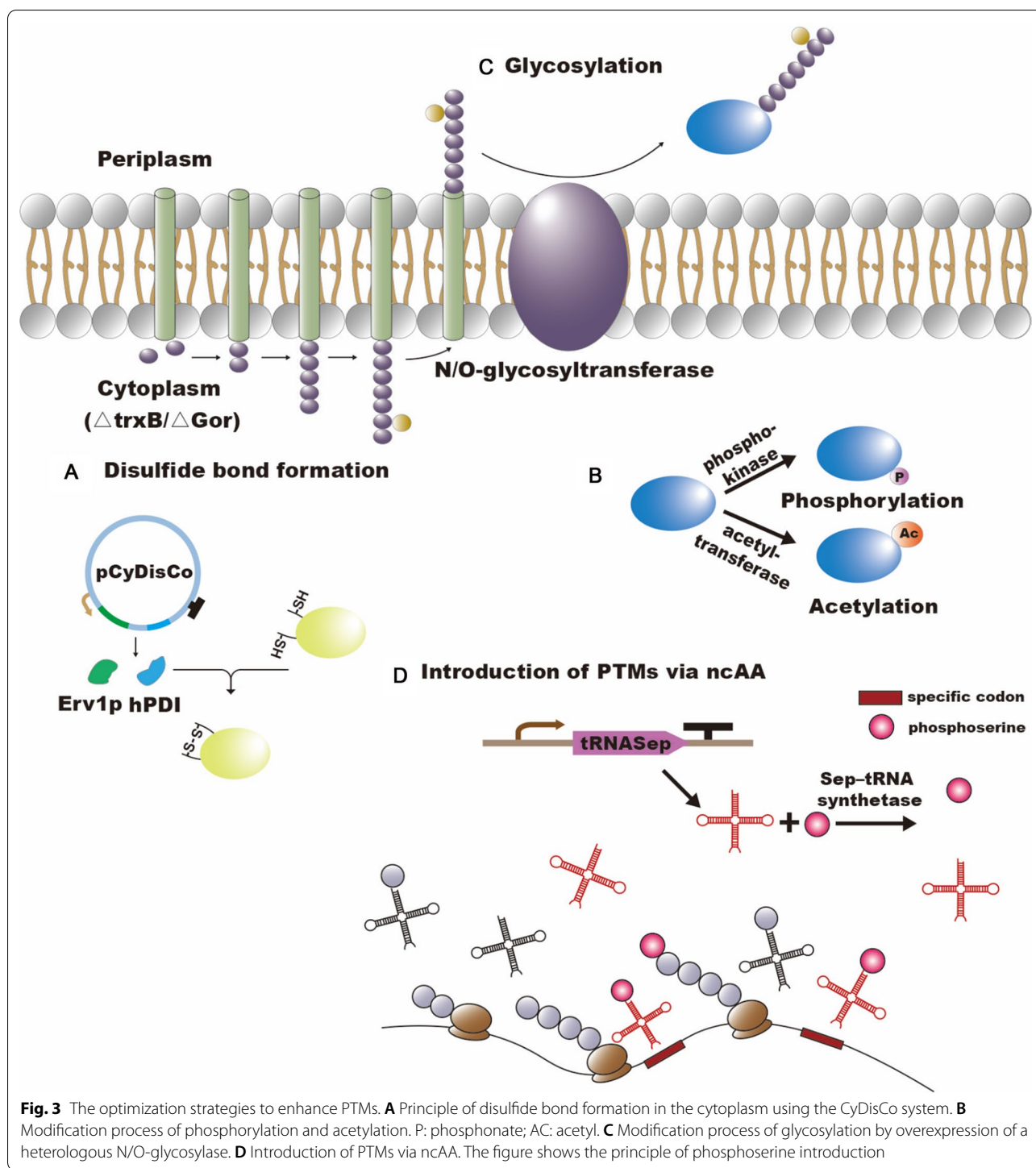## Optimizing protein activity—another key to the production

In addition to ensuring the quantity of RPs, the functional activity of the protein at high yields is also a key focus of RP production. When the expression rate or quantity of RPs exceeds the capacity of the host cell, it will result in a large number of proteins that misfold and aggregate, eventually producing IBs [17]. This phenomenon has greatly hindered the use of *E. coli* in various fields, especially the expression of protein-based drugs. The key reason for the generation of IBs is the limited PTM capacity and folding efficiency, which are the top priorities for optimizing the functional activity of RPs.

Enhancement of post-translational modifications

Most proteins with complex structures contain multiple disulfide bonds (DSBs) that maintain their normal conformation, including insulin [7] and epidermal growth factor [86]. As an oxidative process, the natural DSB formation is completed in the periplasmic space of *E. coli* and not in the reductive environment of the cytoplasm, which requires the protein to be localized and translocated to the appropriate location for modification [87]. The common protein translocation pathways are divided into three main categories: SecB-dependent, SRP-mediated and TAT translocation pathways [88]. Among them, SecB-dependent and SRP-mediated pathways both complete the translocation process by binding to SecA, and genetic fusion of signal peptides to RPs can enable them to utilize these pathways to translocate. Commonly used signal peptides include pelB, OmpA and DsbA [89, 90], but each signal peptide triggers a different mechanism that greatly affects the effectiveness of RP transport. In contrast to SRP-mediated DsbA, SecB-dependent OmpA drives the synthesis of endogenous secreted and membrane proteins, preventing Sec translocator saturation [89]. In recent years, the TAT translocation pathway has attracted the interest of researchers due to its natural "quality control" system, which can prioritize the output of correctly folded proteins [91]. The "TatExpress" strain was successfully developed and applied for the gram-level production of human growth hormone, proving its great potential [92]. In addition to the above translocation pathways, a signal peptide based on the N-terminal sequence of penicillin-binding protein 2 (PBP2) was shown to anchor the fusion protein to the cytoplasmic membrane. Interestingly, the high expression of PBP2 affects morphological changes in *E. coli* (rods to spheres) and interacts with lysis transglycosylase leading to host lysis [93]. This phenomenon has the potential to be developed into a self-cleaving transport system for rapidly accumulating RPs production.

Compared to the narrow periplasmic space, the cytoplasm has enough space to accomplish more protein folding and increase productivity. By blocking the natural reduction pathway in a $\Delta gor/\Delta trxB$ strain, the reductive cytoplasmic environment becomes oxidative, which facilitates the formation of DSBs [94]. The earliest commercial DSB-forming *E. coli* strain, Origami from Novagen, was developed based on this principle. By overexpressing sulfhydryl oxidase from the yeast mitochondria and disulfide bond isomerase from human cells, a host called CyDisCo was developed for the production of RPs with high DSB content, and was able to produce even perlecan with 44 DSBs (Fig. 3A) [95, 96]. Apart from the above, other means of optimization, including replacement of sulfhydryl oxidases from other sources [97], inversion or development of a periplasmic transmembrane disulfide bond-forming enzyme DsbB [98, 99], were also used to improve the efficiency and capacity of DSB formation.

In addition to the formation of DSBs, the efficiency of other PTMs also affects the functional activity of RPs, such as phosphorylation, acetylation (Fig. 3B), glycosylation and many other modifications that are often found in mAbs and functional proteins [100–102]. Among them, glycosylation is one of the most abundant and complex PTMs [103]. By linking monosaccharides, oligosaccharides or polysaccharides to proteins, the variety of protein functional activities is greatly expanded. Currently, over 70% of therapeutic proteins are modified by glycosylation, and precision glycosylation can effectively enhance the use of glycoproteins in the medical industry [102]. Compared to eukaryotes, *E. coli* does not have a natural mechanism for glycosylation of encoded proteins. Therefore, it can be used as a suitable chassis cell to develop bottom-up glycoengineering for different types of glycoproteins [104]. The first N-glycosylation expression system was successfully developed in *E. coli* by introducing genes related to N-glycosylation of *Campylobacter jejuni*, opening the curtain on the glycoprotein synthesis in *E. coli* [105] (Fig. 3C). Over the last two decades, many efforts have conferred the potential to produce a wide range of N/O-glycoproteins from *E. coli* or cell-free extracts, including optimization of glycosyltransferase substrate identification and orthogonality [102, 106–108], exploration of glycosylase function from multiple sources [107–109] and optimization of host environment, metabolic pathways and culture conditions [110–113]. Based on these studies, a variety of medically relevant products are in production and in the clinical phase, such as recombinant vaccine exotoxin A [114], therapeutic protein O-glycosylated interferon-α2b [115] and N-glycosylated mannose3-N-acetylglucosamine2 [116]. In a similar way to DSB, the glycosylation process in the above systems is mostly completed in the

Fig. 3 The optimization strategies to enhance PTMs. **A** Principle of disulfide bond formation in the cytoplasm using the CyDisCo system. **B** Modification process of phosphorylation and acetylation. P: phosphonate; AC: acetyl. **C** Modification process of glycosylation by overexpression of a heterologous N/O-glycosylase. **D** Introduction of PTMs via ncAA. The figure shows the principle of phosphoserine introduction

periplasmic space. In recent years, several studies have identified cytoplasmic glycosylation systems in various bacteria, laying the foundation for the development of novel glycosylation systems in *E. coli* [117–119]. Among them, the asparagine (N)-glucosyltransferase from Actinobacillus pleuropneumoniae (ApNGT) can be actively

expressed in the *E. coli* cytoplasm and transfer glucose residues to the naturally N-terminal glycosylation site of the protein (e.g. recombinant human EPO) [117]. Based on this discovery, Tytgat et al. [120] developed an N-glycosylation system in *E. coli* cytoplasm. Using ApNGT in combination with various oligosaccharide synthesis

Zhang *et al. Microbial Cell Factories*   (2022) 21:191

Page 10 of 18

pathways (e.g. human milk oligosaccharides and gly-cosphingolipids), glycosylation modifications of various glycoproteins (glycoconjugate vaccines and multivalent glycopolymers) have been achieved. Surprisingly, the system can complete the glycosylation of megadalton protein assemblies, which can be used as customized car-riers for delivery of drugs and vaccines.

It is worth mentioning that the orthogonality of ncAAs with specific codons can be used to introduce various types of modified amino acids more directly and pre-cisely. Park et al. [121] successfully introduced phos-phorylated serine residues into RPs at specific sites by orthogonal pairing of SepRS/tRNASep (Fig. 3D). Simi-larly, phosphor-threonine [122] and phospho-tyrosine [123] were utilized for RP modification. In addition to phosphorylation, acetylation, methylation and ubiquit-ination have been successfully introduced into various RPs [124]. In conclusion, the introduction of PTMs using ncAAs has the potential to once again make *E. coli* a "star host" for biopharmaceuticals.

Elimination of inclusion bodies

In addition to limited PTMs, a variety of factors such as misfolding, low solubility, and host burden also contrib-ute to IB formation. Three strategies are usually used to solve the problems: (i) enhancing solubility; (ii) improv-ing correct folding efficiency; (iii) optimizing the appro-priate expression intensity. Among them, the relevant aspects of (iii) have been described above.

The use of peptide tags is the most direct and effective means to enhance the solubility of RPs. Common tags include maltose binding protein (MBP), glutathione-S-transferase (GST), carbohydrate-binding module (CBM), thioredoxin, and NusA, which have been reviewed by Ki et al. [125]. Notably, a novel CBM (CBM66) was shown to have a pro-solubilizing effect on several types of RPs and to increase production titer [126]. For example, the combination of poly (ethylene terephthalate) hydrolase and CBM resulted in a 3.7-fold improvement compared to the other commercial labels (MBP and GST), without affecting protein bioactivity. However, if the molecular weight of the peptide tag is close to or larger than that of the RP, it will override the solubility of the RP itself. Furthermore, the subsequent label removal can nega-tively affect the solubility and stability of RPs. Conversely, the use of peptide tags with smaller molecular weights allows more reliable evaluation and optimization of the solubility of RPs. In recent years, a variety of low-molec-ular-weight protein tags have contributed to the solubili-zation and yield enhancement of various RPs, including the NEXT tag [127], low-molecular-weight protamine [128], and 6HFh8 [129]. Kim et al. utilized 6HFh8 [129] to express a variety of growth factor proteins. Among

them, 6HFh8-aFGF and 6HFh8-VEGF165 obtained high respective yields of 9.7 and 3.4 g/L in a 5-L batch supple-ment fermentation, with a purity of more than 99%. The removal of the small peptide tags does not significantly affect the solubility and functional activity, which is suit-able for the purification of small RPs.

Molecular chaperones are a class of auxiliary pro-teins that facilitate the folding and assembly of peptide structures, ensuring proper folding and preventing the aggregation of newly translated peptides [130]. *E. coli* possesses several molecular chaperone systems, such as GroES/EL and DnaK-DnaJ-GrpE, all with different func-tions [131]. Among them, DnaK-DnaJ-GrpE not only helps correctly fold newly translated peptides, but also functions during co- and post-translational modifica-tion. By contrast, the GroES/EL system associates with peptides only post-translationally, powering the repair of misfolded proteins [127]. It is easy to understand that the folding efficiency can be effectively enhanced by over-expression of molecular chaperones, which is usually done in three combinations: GroES/GroEL, DnaK-DnaJ-GrpE, and co-expression. However, co-expression is usu-ally not better than expressing a single factor, and only some chaperones can have a beneficial effect on protein folding [132]. Huang et al. [133] expressed distinct com-binations of molecular chaperones to enhance the solu-bility and activity of polyunsaturated fatty acid isomerase (PAI). The results showed that overexpression of GroES/EL increased the solubility of PAI from 29 to 97% and improved its specific activity by 57.8%. By contrast, the co-expression of DnaK-DnaJ-GrpE or GroES/EL had a weakening effect, resulting in only an 11.9% increase in activity.

Conclusion and outlook

Different types of RPs from different origins have highly specific characteristics, and there can be no single opti-mization strategy that applies to all proteins. This review summarizes the recently developed optimization strate-gies from the two major aspects of alleviating the host burden and optimizing functional activity, which helps researchers quickly select an appropriate expression strategy for their protein of interest (Table 1, Fig. 4). Encouragingly, with the continued development of syn-thetic biology, systems biology, and various gene editing tools, it is becoming less difficult to rapidly develop a customized host. Multiple in vivo mutagenesis strategies facilitate adaptive laboratory evolution for rapid screen-ing of strongly tolerant expression hosts, including DNA replication proteins, RNAP and T7 RNAP fused with base deaminases [134–137]. Construction of artificial organelles allows for *E. coli* compartmentalization, which has the potential to accomplish precise PTMs [138, 139].

Zhang *et al. Microbial Cell Factories*   (2022) 21:191

Page 11 of 18

Table 1 Application of strategies to enhance recombinant protein production in *E.Coli*

Type of optimization	Optimization strategy	Specific optimization method	Application	References
Regulating the RPs expression level	T7 RNAP transcriptional regulation	The promoter lacUV5 and lac were recombined to yield the promoter variants lac-1G. (A at the + 1 position was changed to G)	Over ten autolysis proteins were increased in yield. GDH activity was increased from 37.5 to 452.0 U/ml at 43 h	[20]
	T7 RNAP transcriptional regulation	Mutant LacI (V192F, does not bind IPTG) prevents leaky expression of T7 RNAP and dynamically regulates transcript levels	Three membrane proteins were increased in yield. *E. coli* cytosine transporter protein increased 4.5-fold at 6 h	[31]
	T7 RNAP translational regulation	Rapid screening of suitable expression hosts using the base editor and CRISPR/Cas9 to construct an T7 RNAP RBS library	The expression level of a target gene in the variant strain library ranged from 28 to 220% of the parental strain. The GDH expression exhibited a 298-fold increase	[35]
	T7 RNAP activity regulation	Single amino acid mutation in T7 RNAP (A102D), which reduced the ability to bind to the P_{T7} and decreased the RP production rate	Seven membrane proteins with varying degrees of production improvement (data not mentioned in the article)	[37]
	T7 RNAP activity regulation	T7 RNAP was split into two fragments and expressed in tandem with a light-sensitive dimerization domain, which are functional activity under blue light	The expression intensity of mCheery was increased 80-fold under blue light irradiation compared to dark conditions	[43]
	Plasmid expression regulation	Construction of an ITR library for pET-28a using degenerate primers and MEGAWHOP PCR, and rapid screening of mutants for strong transporter-competent signal peptides by β-lactamase	The multiple types of RPs (mAbs and human growth hormone (hGH)) were increased in yield. The yield of hGH reached 2.56 mg/L, a more than threefold increase	[51]
	Plasmid expression regulation	Regulation of ColE1 plasmid replication-associated gene expression intensity using CRISPRi and the inducible promoter (P_{tet})	The Plasmid libraries containing 1194 different copy numbers, increasing the yield of violacein	[53]

Zhang *et al. Microbial Cell Factories*   (2022) 21:191

Page 12 of 18

Table 1 (continued)

Type of optimization	Optimization strategy	Specific optimization method	Application	References
Balancing or decoupling the growth and production	Nutrient supplementation	Amino acids were supplemented according to the level of demand (refer to Sect. 3.1 for specific types and added content)	The production of pramlintide increased by 40% (protein concentration of 3.09 ± 0.12 g/L)	[24]
	Blocking the phosphotransferase system	Blocking the phosphotransferase system (PTS) by knockout ptsG or integration ptsG mutants. can effectively reduce the rate of glucose uptake and decrease the production of acetate	A variety of RPs were increased in yield, including eGFP (increased by 282%), vaccines (increased by 3.5-fold) and glutamate dehydrogenase (increased by 14.84%)	[61–63]
	Blocking the cellular stress response	Transcriptome analysis to identify genes associated with cellular stress response and knockout them. The blocking of CSR alleviates the downregulated expression of a variety of growth-essential genes	The double knockout mutant BW25113ΔelaA + ΔcysW (DKO) had the highest activity in asparaginase production with 70.3 units/ml	[64]
	Rare codon supplementation	Integrating six of the least abundant tRNA genes into the BL21(DE3) chromosome behind a ribosomal manipulator for expression	Increased yields of eight proteins, which have different lengths and rare codon contents	[13]
	Blocking host growth	Induction of serine recombinase expression and knockout oriC (replication start gene) at late growth stages, blocking host growth	fivefold increase in RFP production	[80]
	Inhibiting host growth	constructing a sgRNA library targeting growth-related genes, and 1332 different sgRNAs were screened to reduce host growth and increase GFP accumulation	GFP production increased more than fivefold when sibB/IbsB was inhibited	[82]
	RPs Expression using o-ribosome	Construction of a specific 16S rRNA recognizing the RPs RBS site for translation	The yield of RPs is 6.3-fold higher than that of the wild type	[84]
Enhancing protein activity	Enhancing transit capacity to the periplasmic space	Enhanced TAT translocation pathway by overexpression of TatABC membrane protein. Meanwhile, the TAT translocation pathway was exploited by the signal peptide TorA fusion RPs	The yield of purified periplasmic hGH are 5.4 g/L	[92]

Table 1 (continued)

| Type of optimization	Optimization strategy	Specific optimization method	Application	References
	Enhancing disulfide bond formation	a host called CyDisCo was developed by over-expressing sulfhydryl oxidase from the yeast mitochondria and disulfide bond isomerase from human cells	Efficient expression of DSB-rich RPs, including antibodies and therapeutic proteins, perlecan, a protein with 44 DSBs, can also be efficiently generated using this host	[95, 96]
	Efficient glycosylation in the cytoplasm	Combination of ApNGT overexpression and various oligosaccharide synthesis pathways for cytoplasmic N-glycosylation	The various glycoproteins have been achieved, including glycoconjugate vaccines, multivalent glycopolymers and megadalton protein assemblies	[120]
	Fusion solubilisation tag improves protein solubility	a variety of low-molecular-weight protein tags have contributed to the solubilization and yield enhancement of various RPs, requiring only fusion expression with recombinant proteins	The two growth factor fusions 6HFh8 tag can substantially increase the yield, reaching 9.7 and 3.4 g/L respectively	[129]
	Overexpression of molecular chaperones	Select appropriate molecular chaperones for overexpression to improve folding efficiency, including GroES/GroEL, DnaK-DnaJ-GrpE, and co-expression	The overexpression of GroES/EL increased the solubility of polyunsaturated fatty acid isomerase from 29 to 97% and improved its specific activity by 57.8%	[133]
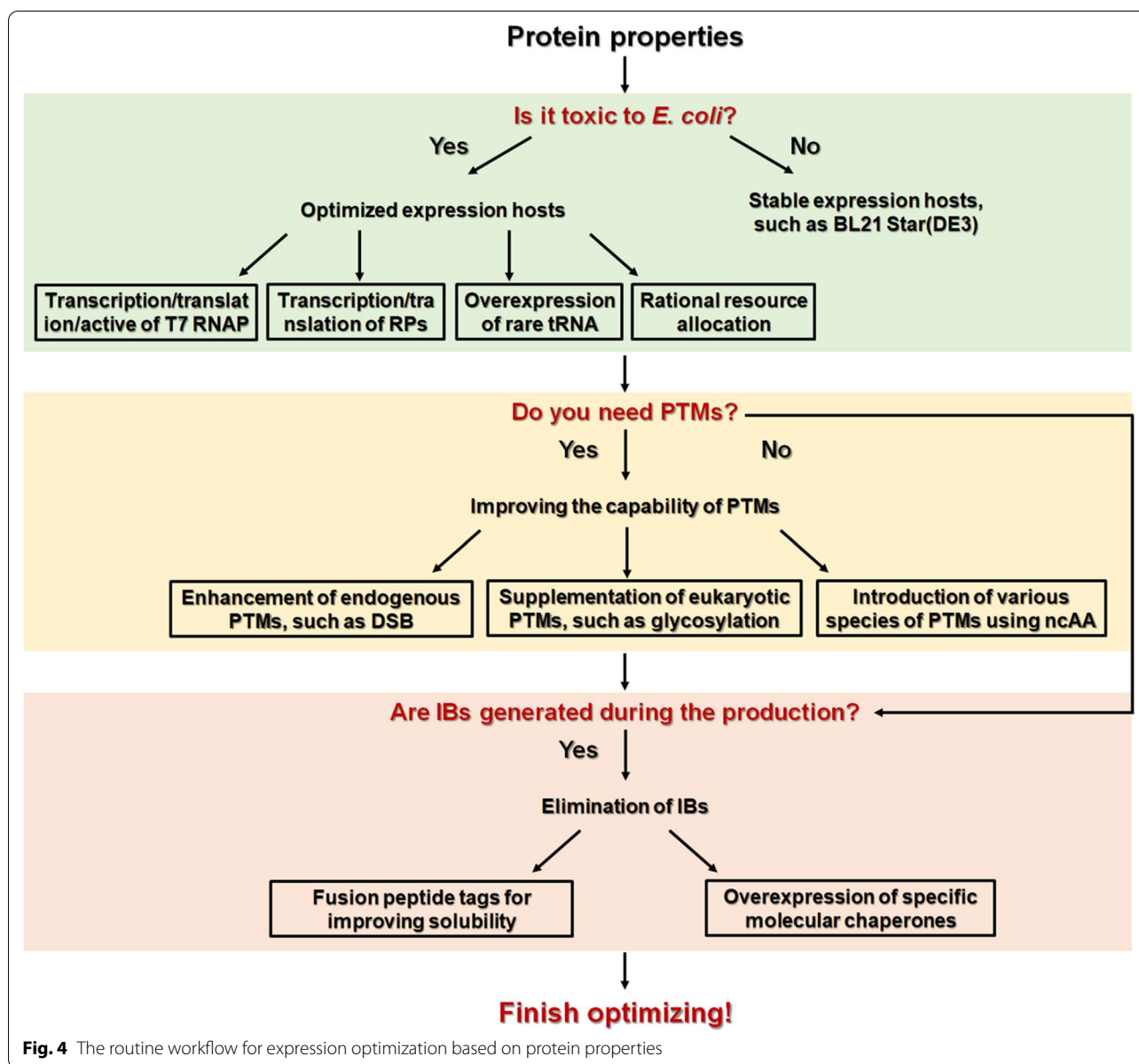
Fig. 4 The routine workflow for expression optimization based on protein properties

In addition, researchers are updating the BL21(DE3) genome annotation, as well as combining mathematical modeling, statistical analysis, and computer aided design to achieve precise optimization [140, 141]. In conclusion, we have reason to believe that *E. coli* will remain one of the brightest stars among RP production hosts.

Author contributions
ZXZ wrote the manuscript. FTN and YZW helped with preparation of the manuscript. CXY reviewed and edited the manuscript. PS and XMS conceptualized, reviewed and edited the manuscript. All authors read and approved the final manuscript.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Zhang *et al. Microbial Cell Factories*   (2022) 21:191

Page 15 of 18

References

1. Puetz J, Wurm FM. Recombinant proteins for industrial versus pharmaceutical purposes: a review of process and pricing. Processes. 2019;7:476–84.
2. Deckers M, Deforce D, Fraiture M-A, Roosens NH. Genetically modified micro-organisms for industrial food enzyme production: an overview. Foods. 2020;9:326–45.
3. Industrial Enzymes Market. https://www.marketsandmarkets.com/Market-Reports/industrial-enzymes-market-237327836.html. Accessed Jan 2022.
4. Walsh G. Biopharmaceutical benchmarks 2018. Nat Biotechnol. 2018;36:1136–45.
5. Deo S, Turton KL, Kainth T, Kumar A, Wieden H-J. Strategies for improving antimicrobial peptide production. Biotechnol Adv. 2022**:**107968–84.
6. Itakura K, Hirose T, Crea R, Riggs AD, Heyneker HL, Bolivar F, Boyer HW. Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. Science. 1977;198:1056–63.
7. Williams DC, Van Frank RM, Muth WL, Burnett JP. Cytoplasmic inclusion bodies in *Escherichia coli* producing biosynthetic human insulin proteins. Science. 1982;215:687–9.
8. Studier FW, Moffatt BA. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J Mol Biol. 1986;189:113–30.
9. Iost I, Guillerez J, Dreyfus M. Bacteriophage T7 RNA polymerase travels far ahead of ribosomes in vivo. J Bacteriol. 1992;174:619–22.
10. Miroux B, Walker JE. 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. 1996;260:289–98.
11. Studier FW. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. J Mol Biol. 1991;219:37–44.
12. Lopez PJ, Marchand I, Joyce SA, Dreyfus M. The C-terminal half of RNase E, which organizes the *Escherichia coli* degradosome, participates in mRNA degradation but not rRNA processing in vivo. Mol Microbiol. 1999;33:188–99.
13. Lipinszki Z, Vernyik V, Farago N, Sari T, Puskas LG, Blattner FR, Posfai G, Gyorfy Z. Enhancing the translational capacity of E. coli by resolving the codon bias. ACS Synth Biol. 2018;7:2656–64.
14. Chapman J, Ismail AE, Dinu CZ. Industrial applications of enzymes: recent advances, techniques, and outlooks. Catalysts. 2018;8:238–63.
15. Sørensen HP, Mortensen KK. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. J Biotechnol. 2005;115:113–28.
16. Weber J, Li Z, Rinas U. Recombinant protein production provoked accumulation of ATP, fructose-1, 6-bisphosphate and pyruvate in E coli K12 strain TG1. Microb Cell Fact. 2021;20:1–8.
17. Tripathi NK, Shrivastava A. Recent developments in bioprocessing of recombinant proteins: expression hosts and process development. Front Bioeng Biotechnol. 2019;7:420–54.
18. Rugbjerg P, Sommer MO. Overcoming genetic heterogeneity in industrial fermentations. Nat Biotechnol. 2019;37:869–76.
19. Wagner S, Baars L, Ytterberg AJ, Klussmeier A, Wagner CS, Nord O, Nygren P-A, van Wijk KJ, de Gier J-W. Consequences of membrane protein overexpression in *Escherichia coli*. Mol Cell Proteomics. 2007;6:1527–50.
20. Sun XM, Zhang ZX, Wang LR, Wang JG, Liang Y, Yang HF, Tao RS, Jiang Y, Yang JJ, Yang S. Downregulation of T7 RNA polymerase transcription enhances pET-based recombinant protein production in *Escherichia coli* BL21(DE3) by suppressing autolysis. Biotechnol Bioeng. 2021;118:153–63.
21. Tan S-I, Ng I-S. New insight into plasmid-driven T7 RNA polymerase in *Escherichia coli* and use as a genetic amplifier for a biosensor. ACS Synth Biol. 2020;9:613–22.
22. Shilling PJ, Mirzadeh K, Cumming AJ, Widesheim M, Köck Z, Daley DO. Improved designs for pET expression plasmids increase protein production yield in *Escherichia coli*. Commun Biol. 2020;3:1–8.
23. Lozano Terol G, Gallego-Jara J, Sola Martínez RA, Cánovas Díaz M, de Diego PT. Engineering protein production by rationally choosing a carbon and nitrogen source using *E. coli* BL21 acetate metabolism knockout strains. Microb Cell Fact. 2019;18:1–19.
24. Kumar J, Chauhan AS, Shah RL, Gupta JA, Rathore AS. Amino acid supplementation for enhancing recombinant protein production in *E. coli*. Biotechnol Bioeng. 2020;117:2420–33.
25. Stargardt P, Feuchtenhofer L, Cserjan-Puschmann M, Striedner G, Mairhofer J. Bacteriophage inspired growth-decoupled recombinant protein production in *Escherichia coli*. ACS Synth Biol. 2020;9:1336–48.
26. Mital S, Christie G, Dikicioglu D. Recombinant expression of insoluble enzymes in *Escherichia coli*: a systematic review of experimental design and its manufacturing implications. Microb Cell Fact. 2021;20:1–20.
27. Graumann K, Premstaller A. Manufacturing of recombinant therapeutic proteins in microbial systems. Biotechnol J. 2006;1:164–86.
28. Mittal P, Brindle J, Stephen J, Plotkin JB, Kudla G. Codon usage influences fitness through RNA toxicity. Proc Natl Acad Sci USA. 2018;115:8639–44.
29. Li Z, Rinas U. Recombinant protein production associated growth inhibition results mainly from transcription and not from translation. Microb Cell Fact. 2020;19:1–11.
30. Jeong H, Barbe V, Lee CH, Vallenet D, Yu DS, Choi S-H, Couloux A, Lee S-W, Yoon SH, Cattolico L. Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). J Mol Biol. 2009;394:644–52.
31. Kim SK, Lee D-H, Kim OC, Kim JF, Yoon SH. Tunable control of an Escherichia coli expression system for the overproduction of membrane proteins by titrated expression of a mutant lac repressor. ACS Synth Biol. 2017;6:1766–73.
32. Du F, Liu Y-Q, Xu YS, Li ZJ, Wang YZ, Zhang ZX, Sun XM. Regulating the T7 RNA polymerase expression in E coli BL21(DE3) to provide more host options for recombinant protein production. Microb Cell Fact. 2021;20:1–10.
33. Kwon S-K, Kim SK, Lee D-H, Kim JF. Comparative genomics and experimental evolution of Escherichia coli BL21 (DE3) strains reveal the landscape of toxicity escape from membrane protein overproduction. Sci Rep. 2015;5:1–13.
34. Liang X, Li C, Wang W, Li Q. Integrating T7 RNA polymerase and its cognate transcriptional units for a host-independent and stable expression system in single plasmid. ACS Synth Biol. 2018;7:1424–35.
35. Li ZJ, Zhang ZX, Xu Y, Shi TQ, Ye C, Sun XM, Huang H. CRISPR-Based Construction of a BL21 (DE3)-derived variant strain library to rapidly improve recombinant protein production. ACS Synth Biol. 2022;11:343–52.
36. Temme K, Hill R, Segall-Shapiro TH, Moser F, Voigt CA. Modular control of multiple pathways using engineered orthogonal T7 polymerases. Nucleic Acids Res. 2012;40:8773–81.
37. Baumgarten T, Schlegel S, Wagner S, Löw M, Eriksson J, Bonde I, Herrgård MJ, Heipieper HJ, Nørholm MH, Slotboom DJ. Isolation and characterization of the E coli membrane protein production strain Mutant56 (DE3). Sci Rep. 2017;7:1–14.
38. Tan S-I, Hsiang C-C, Ng I-S. Tailoring genetic elements of the plasmid-driven T7 system for stable and robust one-step cloning and protein expression in broad *Escherichia coli*. ACS Synth Biol. 2021;10:2753–62.
39. Huang J, Villemain J, Padilla R, Sousa R. Mechanisms by which T7 lysozyme specifically regulates T7 RNA polymerase during different phases of transcription. J Mol Biol. 1999;293:457–75.
40. Zhang X, Studier FW. Mechanism of inhibition of bacteriophage T7 RNA polymerase by T7 lysozyme. J Mol Biol. 1997;269:10–27.
41. Schlegel S, Löfblom J, Lee C, Hjelm A, Klepsch M, Strous M, Drew D, Slotboom DJ, de Gier J-W. Optimizing membrane protein overexpression in the *Escherichia coli* strain Lemo21(DE3). J Mol Biol. 2012;423:648–59.
42. Han T, Chen Q, Liu H. Engineered photoactivatable genetic switches based on the bacterium phage T7 RNA polymerase. ACS Synth Biol. 2017;6:357–66.
43. Baumschlager A, Aoki SK, Khammash M. Dynamic blue light-inducible T7 RNA polymerases (Opto-T7RNAPs) for precise spatiotemporal gene expression control. ACS Synth Biol. 2017;6:2157–67.
44. Raghavan AR, Salim K, Yadav VG. Optogenetic control of heterologous metabolism in E. coli. ACS Synth Biol. 2020;9:2291–300.
45. Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: advances and challenges. Front Microbiol. 2014;5:172–88.
46. Chaudhary AK, Lee EY. Tightly regulated and high level expression vector construction for *Escherichia coli* BL21(DE3). J Ind Eng Chem. 2015;31:367–73.

Zhang *et al. Microbial Cell Factories*   (2022) 21:191

Page 16 of 18

47. Horga LG, Halliwell S, Castiñeiras TS, Wyre C, Matos CF, Yovcheva DS, Kent R, Morra R, Williams SG, Smith DC. Tuning recombinant protein expression to match secretion capacity. Microb Cell Fact. 2018;17:1–18.

48. Anilionyte O, Liang H, Ma X, Yang L, Zhou K. Short, auto-inducible promoters for well-controlled protein expression in *Escherichia coli*. Appl Microbiol Biotechnol. 2018;102:7007–15.

49. Nie Z, Luo H, Li J, Sun H, Xiao Y, Jia R, Liu T, Chang Y, Yu H, Shen Z. High-throughput screening of T7 promoter mutants for soluble expression of cephalosporin C acylase in *E. coli*. Appl Biochem Biotechnol. 2020;190:293–304.

50. Mirzadeh K, Martinez V, Toddo S, Guntur S, Herrgard MJ, Elofsson A, Nørholm MH, Daley DO. Enhanced protein production in *Escherichia coli* by optimization of cloning scars at the vector–coding sequence junction. ACS Synth Biol. 2015;4:959–65.

51. Mirzadeh K, Shilling PJ, Elfageih R, Cumming AJ, Cui HL, Rennig M, Nørholm MH, Daley DO. Increased production of periplasmic proteins in *Escherichia coli* by directed evolution of the translation initiation region. Microb Cell Fact. 2020;19:1–12.

52. Rennig M, Martinez V, Mirzadeh K, Dunas F, Rojsater B, Daley DO, Nørholm MH. TARSyn: tunable antibiotic resistance devices enabling bacterial synthetic evolution and protein production. ACS Synth Biol. 2018;7:432–42.

53. Rouches MV, Xu Y, Cortes LBG, Lambert G. A plasmid system with tunable copy number. Nat Commun. 2022;13:1–12.

54. Ajikumar PK, Xiao W-H, Tyo KE, Wang Y, Simeon F, Leonard E, Mucha O, Phon TH, Pfeifer B, Stephanopoulos G. Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. Science. 2010;330:70–4.

55. Minton NP. Improved plasmid vectors for the isolation of translational lac gene fusions. Gene. 1984;31:269–73.

56. Nordström K. Plasmid R1—replication and its control. Plasmid. 2006;55:1–26.

57. Li C, Zou Y, Jiang T, Zhang J, Yan Y. Harnessing plasmid replication mechanism to enable dynamic control of gene copy in bacteria. Metab Eng. 2022;70:67–78.

58. Segall-Shapiro TH, Meyer AJ, Ellington AD, Sontag ED, Voigt CA. A 'resource allocator' for transcription based on a highly fragmented T7 RNA polymerase. Mol Syst Biol. 2014;10:742–56.

59. Yari K, Fatemi SS-A, Tavallaei M. High level expression of recombinant BoNT/A-Hc by high cell density cultivation of *Escherichia coli*. Bioproc Biosyst Eng. 2012; 35**:**407–14.

60. Ceroni F, Algar R, Stan G-B, Ellis T. Quantifying cellular capacity identifies gene expression designs with reduced burden. Nat Methods. 2015;12:415–8.

61. Jung H-M, Im D-K, Lim JH, Jung GY, Oh M-K. Metabolic perturbations in mutants of glucose transporters and their applications in metabolite production in *Escherichia coli*. Microb Cell Fact. 2019;18:1–14.

62. Fuentes LG, Lara AR, Martínez LM, Ramírez OT, Martínez A, Bolívar F, Gosset G. Modification of glucose import capacity in *Escherichia coli*: physiologic consequences and utility for improving DNA vaccine production. Microb Cell Fact. 2013;12:1–11.

63. Cheng L, Yang X, Li S, Fu Q, Fu S, Wang J, Li F, Lei L, Shen Z. Impact of gene modification of phosphotransferase system on expression of glutamate dehydrogenase protein of Streptococcus suis in *Escherichia coli*. Biotechnol Biotec EQ. 2017;31:612–8.

64. Sharma AK, Shukla E, Janoti DS, Mukherjee KJ, Shiloach J. A novel knock out strategy to enhance recombinant protein expression in *Escherichia coli*. Microb Cell Fact. 2020;19:1–10.

65. Guleria R, Jain P, Verma M, Mukherjee KJ. Designing next generation recombinant protein expression platforms by modulating the cellular stress response in *Escherichia coli*. Microb Cell Fact. 2020;19:1–17.

66. Gustafsson C, Govindarajan S, Minshull J. Codon bias and heterologous protein expression. Trends Biotechnol. 2004;22:346–53.

67. Lipońska A, Ousalem F, Aalberts DP, Hunt JF, Boël G. The new strategies to overcome challenges in protein production in bacteria. Microb Biotechnol. 2019;12:44–7.

68. Rahmen N, Schlupp CD, Mitsunaga H, Fulton A, Aryani T, Esch L, Schaffrath U, Fukuzaki E, Jaeger K-E, Büchs J. A particular silent codon exchange in a recombinant gene greatly influences host cell metabolic activity. Microb Cell Fact. 2015;14:1–14.

69. Zhong C, Wei P, Zhang YHP. Enhancing functional expression of codon-optimized heterologous enzymes in *Escherichia coli* BL21(DE3) by selective introduction of synonymous rare codons. Biotechnol Bioeng. 2017;114:1054–64.

70. Ye Z, Li S, Hennigan JN, Lebeau J, Moreb EA, Wolf J, Lynch MD. Two-stage dynamic deregulation of metabolism improves process robustness & scalability in engineered *E. coli*. Metab Eng. 2021;68:106–18.

71. Faust G, Stand A, Weuster-Botz D. IPTG can replace lactose in auto-induction media to enhance protein expression in batch-cultured *Escherichia coli*. Eng Life Sci. 2015;15:824–9.

72. Nocadello S, Swennen EF. The new pLAI (lux regulon based auto-inducible) expression system for recombinant protein production in *Escherichia coli*. Microb Cell Fact. 2012;11:1–10.

73. Menacho-Melgar R, Ye Z, Moreb EA, Yang T, Efromson JP, Decker JS, Wang R, Lynch MD. Scalable, two-stage, autoinduction of recombinant protein expression in E coli utilizing phosphate depletion. Biotechnol Bioeng. 2020;117:2715–27.

74. Briand L, Marcion G, Kriznik A, Heydel J-M, Artur Y, Garrido C, Seigneuric R, Neiers F. A self-inducible heterologous protein expression system in *Escherichia coli*. Sci Rep. 2016;6:1–11.

75. Shariati FS, Keramati M, Valizadeh V, Cohan RA, Norouzian D. Comparison of *E. coli* based self-inducible expression systems containing different human heat shock proteins. Sci Rep. 2021; 11**:**1–10.

76. Menacho-Melgar R, Moreb EA, Efromson JP, Yang T, Hennigan JN, Wang R, Lynch MD. Improved two-stage protein expression and purification via autoinduction of both autolysis and auto DNA/RNA hydrolysis conferred by phage lysozyme and DNA/RNA endonuclease. Biotechnol Bioeng. 2020;117:2852–60.

77. Izard J, Gomez Balderas CD, Ropers D, Lacour S, Song X, Yang Y, Lindner AB, Geiselmann J, de Jong H. A synthetic growth switch based on controlled expression of RNA polymerase. Mol Syst Biol. 2015;11:840–55.

78. Stargardt P, Striedner G, Mairhofer J. Tunable expression rate control of a growth-decoupled T7 expression system by L-arabinose only. Microb Cell Fact. 2021;20:1–17.

79. Galindo Casas M, Stargardt P, Mairhofer J, Wiltschi B. Decoupling protein production from cell growth enhances the site-specific incorporation of noncanonical amino acids in E coli. ACS Synth Biol. 2020;9:3052–66.

80. Kasari M, Kasari V, Kärmas M, Jõers A. Decoupling growth and production by removing the origin of replication from a bacterial chromosome. ACS Synth Biol. 2022. doi: https://doi.org/10.1021/acssynbio. 1c00618.

81. Li S, Jendresen CB, Grünberger A, Ronda C, Jensen SI, Noack S, Nielsen AT. Enhanced protein and biochemical production using CRISPRi-based growth switches. Metab Eng. 2016;38:274–84.

82. Li S, Jendresen CB, Landberg J, Pedersen LE, Sonnenschein N, Jensen SI, Nielsen AT. Genome-wide CRISPRi-based identification of targets for decoupling growth from production. ACS Synth Biol. 2020;9:1030–40.

83. Wan X, Pinto F, Yu L, Wang B. Synthetic protein-binding DNA sponge as a tool to tune gene expression and mitigate protein toxicity. Nat Commun. 2020;11:1–12.

84. Liu F, Bratulić S, Costello A, Miettinen TP, Badran AH. Directed evolution of rRNA improves translation kinetics and recombinant protein yield. Nat Commun. 2021;12:1–14.

85. Kolber NS, Fattal R, Bratulic S, Carver GD, Badran AH. Orthogonal translation enables heterologous ribosome engineering in Ecoli. Nat Commun. 2021;12:1–12.

86. Yadwad V, Wilson S, Ward O. Production of human epidermal growth factor by an ampicillin resistant recombinant *Escherichia coli* strain. Biotechnol Lett. 1994;16:885–90.

87. De Marco A. Strategies for successful recombinant expression of disulfide bond-dependent proteins in *Escherichia coli*. Microb Cell Fact. 2009;8:1–18.

88. McElwain L, Phair K, Kealey C, Brady D. Current trends in biopharmaceuticals production in *Escherichia coli*. Biotechnol Lett. 2022;89:1–15.

89. Ytterberg AJ, Zubarev RA, Baumgarten T. Posttranslational targeting of a recombinant protein promotes its efficient secretion into the *Escherichia coli* periplasm. Appl Environ Microbiol. 2019;85:e00671-e719.

90. Gawin A, Ertesvåg H, Hansen SAH, Malmo J, Brautaset T. Translational regulation of periplasmic folding assistants and proteases as a valuable strategy to improve production of translocated recombinant proteins in *Escherichia coli*. BMC Biotechnol. 2020;20:1–11.

Zhang *et al. Microbial Cell Factories*   (2022) 21:191

Page 17 of 18

91. Alanen HI, Walker KL, Suberbie MLV, Matos CF, Bönisch S, Freedman RB, Keshavarz-Moore E, Ruddock LW, Robinson C. Efficient export of human growth hormone, interferon α2b and antibody fragments to the periplasm by the *Escherichia coli* Tat pathway in the absence of prior disulfide bond formation. BBA-Mol Cell Res. 2015;1853:756–63.

92. Guerrero Montero I, Richards KL, Jawara C, Browning DF, Peswani AR, Labrit M, Allen M, Aubry C, Davé E, Humphreys DP. *Escherichia coli* "TatExpress" strains export several g/L human growth hormone to the periplasm by the Tat pathway. Biotechnol Bioeng. 2019;116:3282–91.

93. Legaree BA, Adams CB, Clarke AJ. Overproduction of penicillin-binding protein 2 and its inactive variants causes morphological changes and lysis in *Escherichia coli*. J Bacteriol. 2007;189:4975–83.

94. Derman AI, Prinz WA, Belin D, Beckwith J. Mutations that allow disulfide bond formation in the cytoplasm of *Escherichia coli*. Science. 1993;262:1744–7.

95. Matos CF, Robinson C, Alanen HI, Prus P, Uchida Y, Ruddock LW, Freedman RB, Keshavarz-Moore E. Efficient export of prefolded, disulfide-bonded recombinant proteins to the periplasm by the Tat pathway in *Escherichia coli* CyDisCo strains. Biotechnol Prog. 2014;30:281–90.

96. Sohail AA, Gaikwad M, Khadka P, Saaranen MJ, Ruddock LW. Production of extracellular matrix proteins in the cytoplasm of E coli: making giants in tiny factories. Int J Mol Sci. 2020;21:688–702.

97. Zhang W, Zheng W, Mao M, Yang Y. Highly efficient folding of multi-disulfide proteins in superoxidizing *Escherichia coli* cytoplasm. Biotechnol Bioeng. 2014;111:2520–7.

98. Hatahet F, Ruddock LW. Topological plasticity of enzymes involved in disulfide bond formation allows catalysis in either the periplasm or the cytoplasm. J Mol Biol. 2013;425:3268–76.

99. Mizrachi D, Robinson M-P, Ren G, Ke N, Berkmen M, DeLisa MP. A water-soluble DsbB variant that catalyzes disulfide-bond formation in vivo. Nat Chem Biol. 2017;13:1022–8.

100. Walsh G, Jefferis R. Post-translational modifications in the context of therapeutic proteins. Nat Biotechnol. 2006;24:1241–52.

101. Lapteva YS, Vologzhannikova AA, Sokolov AS, Ismailov RG, Uversky VN, Permyakov SE. In Vitro N-Terminal Acetylation of Bacterially Expressed Parvalbumins by N-Terminal Acetyltransferases from *Escherichia coli*. Appl Biochem Biotechnol. 2021;193:1365–78.

102. Natarajan A, Jaroentomeechai T, Cabrera-Sánchez M, Mohammed JC, Cox EC, Young O, Shajahan A, Vilkhovoy M, Vadhin S, Varner JD. Engineering orthogonal human O-linked glycoprotein biosynthesis in bacteria. Nat Chem Biol. 2020;16:1062–70.

103. Eichler J, Koomey M. Sweet new roles for protein glycosylation in prokaryotes. Trends Microbiol. 2017;25:662–72.

104. Harding CM, Feldman MF. Glycoengineering bioconjugate vaccines, therapeutics, and diagnostics in E coli. Glycobiology. 2019;29:519–29.

105. Wacker M, Linton D, Hitchen PG, Nita-Lazar M, Haslam SM, North SJ, Panico M, Morris HR, Dell A, Wren BW. N-linked glycosylation in Campylobacter jejuni and its functional transfer into E coli. Science. 2002;298:1790–3.

106. Silverman JM, Imperiali B. Bacterial N-glycosylation efficiency is dependent on the structural context of target sequons. J Biol Chem. 2016;291:22001–10.

107. Ollis AA, Zhang S, Fisher AC, DeLisa MP. Engineered oligosaccharyltransferases with greatly relaxed acceptor-site specificity. Nat Chem Biol. 2014;10:816–22.

108. Kightlinger W, Warfel KF, DeLisa MP, Jewett MC. Synthetic glycobiology: parts, systems, and applications. ACS Synth Biol. 2020;9:1534–62.

109. Keys TG, Wetter M, Hang I, Rutschmann C, Russo S, Mally M, Steffen M, Zuppiger M, Müller F, Schneider J. A biosynthetic route for polysialylating proteins in *Escherichia coli*. Metab Eng. 2017;44:293–301.

110. Yates LE, Mills DC, DeLisa MP. Bacterial glycoengineering as a biosynthetic route to customized glycomolecules. Adv Biochem Eng Biot. 2018;175:167–200.

111. Yates LE, Natarajan A, Li M, Hale ME, Mills DC, DeLisa MP. Glyco-recoded *Escherichia coli*: Recombineering-based genome editing of native polysaccharide biosynthesis gene clusters. Metab Eng. 2019;53:59–68.

112. Strutton B, Jaffé SR, Pandhal J, Wright PC. Producing a glycosylating *Escherichia coli* cell factory: the placement of the bacterial oligosaccharyl transferase pglB onto the genome. Biochem Biophys Res Commun. 2018;495:686–92.

113. Pratama F, Linton D, Dixon N. Genetic and process engineering strategies for enhanced recombinant N-glycoprotein production in bacteria. Microb Cell Fact. 2021;20:1–25.

114. Ihssen J, Kowarik M, Dilettoso S, Tanner C, Wacker M, Thöny-Meyer L. Production of glycoprotein vaccines in *Escherichia coli*. Microb Cell Fact. 2010;9:1–13.

115. Valderrama-Rincon JD, Fisher AC, Merritt JH, Fan Y-Y, Reading CA, Chhiba K, Heiss C, Azadi P, Aebi M, DeLisa MP. An engineered eukaryotic protein glycosylation pathway in *Escherichia coli*. Nat Chem Biol. 2012;8:434–6.

116. Du T, Buenbrazo N, Kell L, Rahmani S, Sim L, Withers SG, DeFrees S, Wakarchuk W. A bacterial expression platform for production of therapeutic proteins containing human-like O-linked glycans. Cell Chem Biol. 2019;26:203–12.

117. Naegeli A, Neupert C, Fan Y-Y, Lin C-W, Poljak K, Papini AM, Schwarz F, Aebi M. Molecular analysis of an alternative N-glycosylation machinery by functional transfer from *Actinobacillus pleuropneumoniae* to *Escherichia coli*. J Biol Chem. 2014;289:2170–9.

118. Rempe KA, Spruce LA, Porsch EA, Seeholzer SH, Nørskov-Lauritsen N, Geme JWS. Unconventional N-Linked Glycosylation Promotes Trimeric Autotransporter Function in *Kingella kingae* and *Aggregatibacter aphrophilus*. MBio. 2015;6:e01206-e1215.

119. Keys TG, Aebi M. Engineering protein glycosylation in prokaryotes. Curr Opin Syst Biol. 2017;5:23–31.

120. Tytgat HL, Lin C-W, Levasseur MD, Tomek MB, Rutschmann C, Mock J, Liebscher N, Terasaka N, Azuma Y, Wetter M. Cytoplasmic glycoengineering enables biosynthesis of nanoscale glycoprotein assemblies. Nat Commun. 2019;10:1–10.

121. Park H-S, Hohn MJ, Umehara T, Guo L-T, Osborne EM, Benner J, Noren CJ, Rinehart J, Söll D. Expanding the genetic code of *Escherichia coli* with phosphoserine. Science. 2011;333:1151–4.

122. Zhang MS, Brunner SF, Huguenin-Dezot N, Liang AD, Schmied WH, Rogerson DT, Chin JW. Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing. Nat Methods. 2017;14:729–36.

123. Hoppmann C, Wong A, Yang B, Li S, Hunter T, Shokat KM, Wang L. Site-specific incorporation of phosphotyrosine using an expanded genetic code. Nat Chem Biol. 2017;13:842–4.

124. Davis L, Chin JW. Designer proteins: applications of genetic code expansion in cell biology. Nat Rev Mol Cell Bio. 2012;13:168–82.

125. Ki M-R, Pack SP. Fusion tags to enhance heterologous protein expression. Appl Microbiol Biotechnol. 2020;104:2411–25.

126. Ko H, Kang M, Kim M-J, Yi J, Kang J, Bae J-H, Sohn J-H, Sung BH. A novel protein fusion partner, carbohydrate-binding module family 66, to enhance heterologous protein expression in *Escherichia coli*. Microb Cell Fact. 2021;20:1–12.

127. Jo BH. An intrinsically disordered peptide tag that confers an unusual solubility to aggregation-prone proteins. Appl Environ Microbiol. 2022;88:e00097-e122.

128. Choi SW, Pangeni R, Jung DH, Kim SJ, Park JW. Construction and characterization of cell-penetrating peptide-fused fibroblast growth factor and vascular endothelial growth factor for an enhanced percutaneous delivery system. J Nanosci Nanotechno. 2018;18:842–7.

129. Kim YS, Lee H-J, Han M-H, Yoon N-K, Kim Y-C, Ahn J. Effective production of human growth factors in *Escherichia coli* by fusing with small protein 6HFh8. Microb Cell Fact. 2021;20:1–16.

130. Schlieker C, Bukau B, Mogk A. Prevention and reversion of protein aggregation by molecular chaperones in the E coli cytosol: implications for their applicability in biotechnology. J Biotechnol. 2002;96:13–21.

131. Fatima K, Naqvi F, Younas H. A review: Molecular chaperone-mediated folding, unfolding and disaggregation of expressed recombinant proteins. Cell Biochem Biophys. 2021;79:153–74.

132. Yao D, Fan J, Han R, Xiao J, Li Q, Xu G, Dong J, Ni Y. Enhancing soluble expression of sucrose phosphorylase in *Escherichia coli* by molecular chaperones. Protein Expression Purif. 2020;169:105571–80.

133. Huang MN, Lu XY, Zong H, Bin ZG, Shen W. Bioproduction of trans-10, cis-12-Conjugated Linoleic Acid by a Highly Soluble and Conveniently Extracted Linoleic Acid Isomerase and an Extracellularly Expressed Lipase from Recombinant *Escherichia coli* Strains. J Microbiol Biotechnol. 2018;28:739–47.

Zhang *et al. Microbial Cell Factories*   (2022) 21:191

Page 18 of 18

134. Eom G, Lee H, Kim S. Development of a genome-targeting muta-tor for the adaptive evolution of microbial cells. Nucleic Acids Res. 2022;50:e38.

135. Moore CL, Papa LJ III, Shoulders MD. A processive protein chimera introduces mutations across defined DNA regions in vivo. J Am Chem Soc. 2018;140:11560–4.

136. Álvarez B, Mencía M, de Lorenzo V, Fernández LÁ. In vivo diversifica-tion of target genomic sites using processive base deaminase fusions blocked by dCas9. Nat Commun. 2020;11:1–14.

137. Pan Y, Xia S, Dong C, Pan H, Cai J, Huang L, Xu Z, Lian J. Random base editing for genome evolution in *Saccharomyces cerevisiae*. ACS Synth Biol. 2021;10:2440–6.

138. Wei S-P, Qian Z-G, Hu C-F, Pan F, Chen M-T, Lee SY, Xia X-X. Formation and functionalization of membraneless compartments in *Escherichia coli*. Nat Chem Biol. 2020;16:1143–8.

139. Wang Y, Liu M, Wei Q, Wu W, He Y, Gao J, Zhou R, Jiang L, Qu J, Xia J. Phase-Separated Multienzyme Compartmentalization for Terpene Biosynthesis in a Prokaryote. Angew Chem Int Ed. 2022;8:61–9.

140. Kim S, Jeong H, Kim E-Y, Kim JF, Lee SY, Yoon SH. Genomic and tran-scriptomic landscape of *Escherichia coli* BL21(DE3). Nucleic Acids Res. 2017;45:5285–93.

141. Packiam KAR, Ramanan RN, Ooi CW, Krishnaswamy L, Tey BT. Stepwise optimization of recombinant protein production in *Escherichia coli* utilizing computational and experimental approaches. Appl Microbiol Biotechnol. 2020;104:3253–66.

Publisher's Note