Recent advances in recombinant production of soluble proteins in *E. coli*

Ario de Marco[1*]

Abstract

Background *E. coli* still remains the most commonly used organism to produce recombinant proteins in research labs. This condition is mirrored by the attention that researchers dedicate to understanding the biology behind protein expression, which is then exploited to improve the effectiveness of the technology. This effort is witnessed by an impressive number of publications, and this review aims to organize the most relevant novelties proposed in recent years.

Results The examined contributions address several of the known bottlenecks related to recombinant expression in *E. coli*, such as improved glycosylation pathways, more reliable production of proteins whose folding depends on the formation of disulfide bonds, the possibility of controlling and even benefiting from the formation of aggregates or the need to overcome the dependence of bacteria on antibiotics during bacterial culture. Nevertheless, the majority of the published papers aimed at identifying the conditions for optimal control of the translation process to achieve maximal yields of functional exogenous proteins.

Conclusions Despite community commitment, the critical question of what really is the metabolic burden and how it affects both host metabolism and recombinant protein production remains elusive because some experimental results are contradictory. This contribution aims to offer researchers a tool to orient themselves in this complexity. The new capacities offered by artificial intelligence tools could help clarifying this issue, but the training phase will probably require more systematic experimental approaches to collect sufficiently uniform data.

Keywords Disulfide bonds, Glycosylation, Intrinsically disordered peptides, pET plasmids, T7 RNA polymerase

Background

The number of different organisms used for producing recombinant proteins is constantly increasing, but *E. coli* probably still remains the first option in most of research laboratories worldwide owing to the relative simplicity of this system. Surprisingly, such platform is still highly dynamic, and new theoretical knowledge and practical improvements are steadily integrated into the preexisting "technology backbone". Nevertheless, many of these contributions to innovation do not reach and affect the daily work of the lab; in most cases, there are no time and resources to implement and test all the proposed new approaches. The aim of this review is to present a preselected pool of proposals for technical advancement that appeared in the last 3–4 years to simplify the screening of solutions that could fit the specific problems affecting the different laboratories that struggle to produce their proteins of interest. In this sense, this overview will not recapitulate the historical background and standard approaches because these have been already covered

*Correspondence:
Ario de Marco
ario.demarco@ung.si
[1]Lab of Environmental and Life Sciences, University of Nova Gorica,
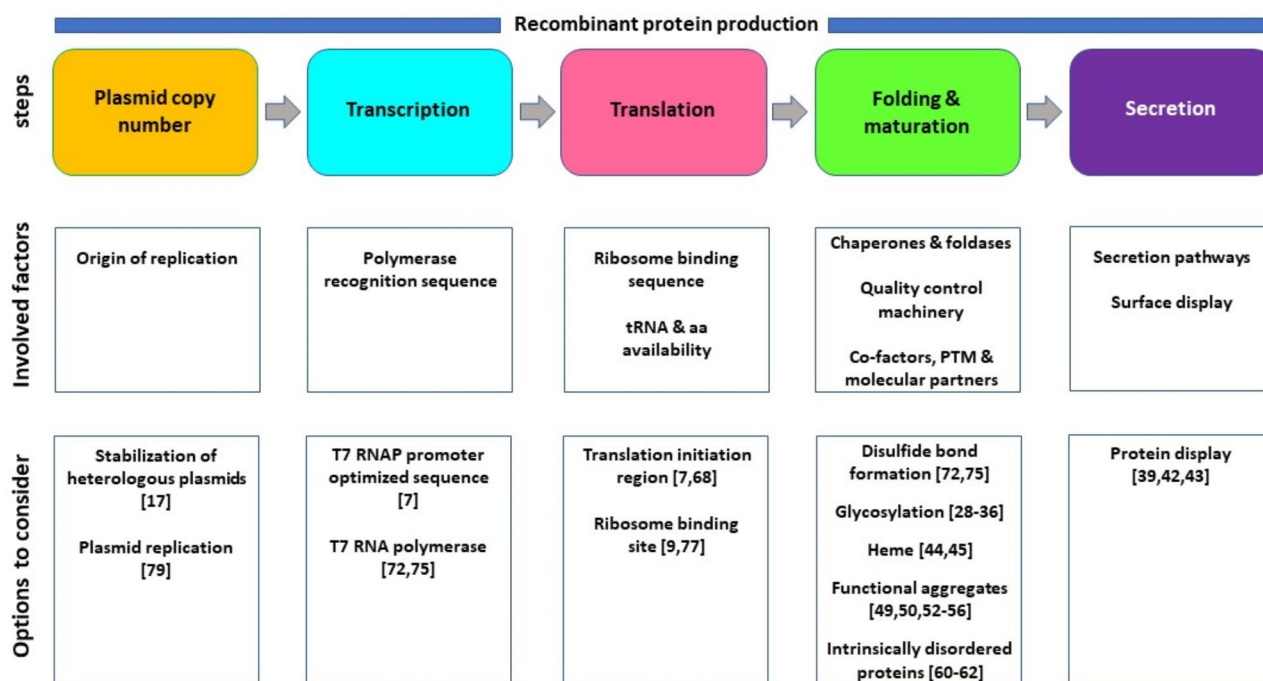Vipavska cesta 13, Nova Gorica 5000, Slovenia

Fig. 1 Schematic representation of successive phases controlling protein production and of the corresponding involved key factors, as inferred by the data reported in recent publications For each relevant step, the most promising approaches for improving yields and functionality of recombinant proteins that have been reported in the articles discussed in this review are listed. Numbers in brackets correspond to references. Of course, these options do not embrace the complete array of possibilities but rather mirror the reality described in the cited papers that preferentially focused on some topics, whereas others, such as the use of different promoters, the codon optimization, the overexpression of molecular chaperones or the addition of osmolytes, just to name a few, apparently were not investigated as actively in the last few years. This does not mean that such factors are less relevant for successful recombinant protein production, only that they were not further optimized lately

in other excellent reports [1–4]. Articles were chosen because either introduced new concepts or elaborated previous ideas into innovative perspectives. In this sense, the survey will not systematically address any aspect of recombinant protein production in *E. coli* because the research did not advance uniformly in the several sections that contributed to the whole technology. Furthermore, simple applications of already described protocols or incremental contributions are not listed, with the exception of those articles that confirm the effectiveness of recently proposed innovations. Finally, priority has been given to solutions that can be evaluated in, and profitable for, research labs rather than for industrial settings that possess resources and constraints that differ completely from what is usually available in academic conditions.

Particular attention will be dedicated to what is called "metabolic burden" because there is no agreement on what the conditions that generate it are. Indeed, starting from DNA transcription to finish with protein folding and secretion, we can imagine a long array of steps (Fig. 1), each of which might contribute to the competition between host and recombinant proteins and result in a stress factor leading to cell death and/or unproductive synthetic pathways, as thoroughly described in a recent review [5]. The principle "less is more", according to which decreasing the pace of exogenous protein synthesis would lead to higher yields [6], works in several cases, as well as, in other cases, the use of methods that accelerate the synthetic machinery results successfully [7], and even approaches that totally neglect transcription and translation issues to focus exclusively on elements affecting the final protein structure can substantially contribute to increasing the yields of functional recombinant proteins [8]. We know more, but not necessarily better, because several unexpected results contradict previous assumptions, as recently shown in the case of the positive effect on protein production induced by the introduction of rare codons in the sequences corresponding to the target protein [9]. A pragmatic and effective option is offered by the systemic, high-throughput approach that compares as many expression conditions as possible without a priori biases [10], but it is expensive and technically demanding. Unluckily, "rational" alternatives are still too unpredictable, probably because they are not supported by data sets sufficiently large to render them robust enough. This review will summarize the state-of-the-art and underline the missing information that will be necessary to fill these gaps.

Main text

Antibiotic-free plasmid selection

Antibiotic-based plasmid selection and maintenance are expensive and not completely effective in high-density cultures. This might already represent a significant metabolic burden for the cells since the constitutive expression of the plasmid-encoded genes that provide antibiotic resistance is needed, furthermore the system is costly and contributes to conditions favorable for the development of antimicrobial resistance and, for this reason, is not sustainable in the long term. Taken together, these arguments represent a valid motivation for the implementation of recombinant protein expression approaches independent of antibiotics [11–16]. Recently, a new complementation concept has been proposed that should overcome the limitations specific to the methods proposed in the past [17]. The genomic promoter of the essential gene *inf*A, encoding the small (72 amino acids) translation Initiation Factor 1 (IF1), was exchanged with an inducible arabinose promoter. Consequently, the strain can survive either in the presence of arabinose (provided in the growth medium) or because it has been transformed with a plasmid that encodes an *infA* copy. In the absence of arabinose, there is strong selection pressure for the preservation of the *inf*A-encoding plasmid, which can therefore be safely used to express any further protein of interest.

Proteins requiring the formation of disulfide bonds

Since antibiotic resistance is a major issue, there is also great interest in identifying alternative antimicrobial molecules that could be substitutes for older drugs in the future. Host defense peptides (HDPs) are part of the innate immune system and can contribute to controlling microbial infection. Since they are short and relatively rich in cysteines involved in disulfide bonds, they represent a class of constructs that are difficult to produce in *E. coli* and, consequently, thus far, have been preferentially synthesized. However, considerable research effort has been made to optimize their recombinant expression in bacteria as fusion proteins [18, 19]. The Origami strain has been preferred because of the oxidizing conditions of its cytoplasm, but in a recent publication [20], Lopez-Cano et al. reported that, surprisingly, HDPs expressed in wild-type BL21(DE3) accumulated at relatively high yields and possessed high activity, which correlated with increased disulfide bond formation efficiency. Since the reaction leading to disulfide bond formation can be ruled out in the strongly reducing cytoplasm of BL21(DE3), it is likely that oxidation occurs during protein purification. In this case, the role of GFP, which is used as the passenger protein for HDPs, might be relevant in preventing nonproductive aggregation of the fusion partner before the HDP can reach its native conformation.

It would be interesting to compare this procedure with combinations that exploit sulfhydryl oxidase and isomerase overexpression in the reducing cytoplasm [21, 22], which has already been proven to be highly effective in the production of recombinant proteins that require the formation of disulfide bonds to reach their native folding. From this perspective, the most recent version of this approach foresees a switchable system that exploits phosphate depletion to trigger the passage from reducing to oxidizing cytoplasm by turning on and turning off the accumulation of few key enzymes during the productive stationary phase [23]. First, the authors engineered a bacterial strain in which genes of the glutaredoxin pathway were deleted and the DNA sequence corresponding to thioredoxin B was fused to the DAS + 4 degradation tag for inducible removal, whereas the tunable disulfide bond isomerase DsbC and the sulfhydryl oxidase Erv1p were introduced. Next, the production of the model nanobody protein was evaluated by designing a bacterial culture such that the decrease in phosphate concentration during the bacterial exponential growth phase could induce a switch from reducing to oxidizing conditions and the accumulation of the foldases in the cytoplasm at the beginning of the stationary growth phase, when the recombinant expression of nanobodies starts. The authors summarized that the possibility of separating the metabolic requirements necessary for cell growth from the metabolic inputs necessary for recombinant protein expression allowed obtaining unmatched yields (100–800 mg/L in shake flasks, > 2 g/L in a bioreactor) of soluble and functional nanobodies. Interestingly, the optimized strain performed better at 37 °C than at 30 °C, suggesting that the described system did not reach its "metabolic burden", understood as the condition in which resources become limiting.

New versions of more conventional methods have been proposed, such as the expression of disulfide bond-dependent proteins in the oxidizing environment offered by the bacterial periplasm. From this perspective, promising advancements have been made with the exploitation of CASPON™ (CASPase-based fusiON), a sequence that comprises solubility-enhancing elements, a His tag for affinity purification and a recognition site for efficient cleavage by means of circularly permuted caspase-2 [24, 25], which has been effectively used to increase the yields of several peptides expressed recombinantly in *E. coli* [24, 26]. Despite the encouraging results, so far CASPON™ technology has been used only by researchers who originally contributed to its implementation. The reason for such apparent lack of interest is unknown, since the overall package has elevated technical requirements but possibly also intellectual property issues.

Another perspective is provided by research that used modeling to describe the variable capacity of *E.*

coli to form disulfide bonds in exogenous proteins [27]. The authors first analyzed how many endogenous proteins require oxidative folding during different bacterial growth phases and conditions to establish then the residual "physiological" capacity, at any step, suitable for oxidizing (and, in the case, isomerizing) recombinant proteins. Following this approach, the level of expression of recombinant proteins that require the formation of disulfide bonds for reaching their native folding should be regulated according to the residual host oxidative capacity, the amount of which varies over the culture time.

Protein glycosylation

In addition to the production of recombinant proteins possessing several trans disulfide bonds, the other major issue of *E. coli* that limits its use as a universal factory is that it lacks natural glycosylation pathways. This drawback prompted several researchers to engineer such bacteria to obtain mutants that are able to undergo eukaryotic-like glycosylation. A significant success in this field is represented by bacteria modified with an O-glycosylation machinery able to functionalize serine residues with different human cancer-associated glycans in vivo and ex vivo [28]. These bacteria have already demonstrated their reliability when used by other groups [29]. Furthermore, effective recombinant N-linked glycosylation has been obtained in *E. coli* transformed with the *Campylobacter*-derived PglB oligosaccharyltransferase [30] in combination with a modification of the secretion pathway. Specifically, the signal peptide cleavage site was mutated to expand the time of membrane residency and the oxidation parameters were tuned. However, the structural and functional characterization of the resulting recombinant proteins was too limited to judge the overall quality of the process. Recently, even more sophisticated platforms for the production of humanized N-glycosylated recombinant proteins have been proposed [31, 32] that wait for validation experiments outside the laboratories in which the innovations were originally developed.

Since glycosylation patterns can strongly affect the immunogenicity of proteins, particular attention has been given to glycoconjugated vaccines. The possibility of obtaining native glycosylation patterns is particularly important in the case of recombinant proteins used as vaccines because otherwise the antigenic potency can be significantly diminished or addressed toward irrelevant epitopes. Glycoconjugate vaccines are usually composed of a cell surface glycan covalently linked to an immunogenic carrier protein and provide broad protection but are cumbersome to produce. The introduction of the Protein Glycan Coupling Technology (PGCT) has transformed the production process by allowing the in vivo coupling of recombinant glycan antigens to carrier proteins in a process catalyzed by a coexpressed oligosaccharyltransferase enzyme [33]. This process was recently automated and optimized at each single protocol step after 24 different culture conditions were compared to obtain elevated high-quality vaccine production in *E. coli* [34]. In further attempts to improve the platform, the same authors engineered 11 *E. coli* strains with slightly different characteristics that might be relevant for specific applications [35] and increased the availability of the otherwise limiting glycan building block undecaprenyl phosphate [36]. To date, no other group has published results obtained using the material described above.

Protein surface display

The display of enzymes and antibodies at the bacterial membrane is a convenient means to obtain functional reagents that are optimally oriented outward and that can be used to decorate biosensor surfaces directly with activated bacteria, sparing any protein purification step [37, 38]. The efficacy of this system strongly depends on the actual membrane reagent density. Recently, the presence and quantification of proteins displayed on the bacterial surface have been assessed via an approach that should be less invasive than previous methods. These methods conventionally rely on the fusion of the target protein to a fluorescent protein or another relatively large reporter. The mass of such tags can affect the secretion and display efficiency of the target protein. Zhang et al. [39] proposed exploiting the GFP-split system in which only a minimal portion of GFP, corresponding to the 11th β-strand (1.8 kDa), is fused to the displayed protein, but its presence enables fluorescence recovery after complementation with the remaining GFP1−10 moiety, which can be produced inexpensively as a standard recombinant protein [40]. The resulting fluorescence intensity allows estimation of the degree of display. Of course, the reliability of the system depends on the quality of the complementary GFP1-10, and the optimization of its production protocol has been recently achieved by tuning the feeding conditions [41]. Successful display depends on the efficiency of the secretion process, and in the case of ABC transporters, it can be affected by the protein surface charge. This knowledge has led to efforts to identify secretion-optimized mutants, and the resulting PySupercharge algorithm is effective in proposing suitable amino acid substitutions that favor protein secretion [42]. Nevertheless, the functionality of the resulting mutants has not yet been validated. In the case of the hemolysin A secretion system, the critical parameter is the protein isoelectric point [43]. The introduction of positive charges and fusion with the S tag greatly promoted the secretion efficiency of the tested peptides and proteins, which were functional after tag release.

Proteins containing heme groups

Another class of proteins whose production can be challenging is those containing a heme group. An optimized mutant was identified among 52 recombinant *E. coli* strains differing in some genes involved in heme synthesis [44] because of its capacity to increase the yields of ten model heme proteins. The gain in terms of protein yield was 42–107%, but since the functionality was improved, the overall activity increased 6.5-fold. Interestingly, at least for cytochromes, their yields can also be significantly increased by truncating their transmembrane sequence, without detectable negative effects on protein functionality [45].

Functional protein aggregates

The discovery that proteins precipitated into inclusion bodies could maintain their functionality and catalytic activity radically changed the appreciation for these biological structures [46, 47]. Instead of designing trial–and-error, often ineffective, protocols aimed at obtaining the resuspension of the trapped proteins, the interest moved toward solutions that could directly exploit functionally active inclusion bodies. In parallel, researchers have started deciphering the factors involved in the formation of such aggregates with the aim of controlling the process and recovering "optimized inclusion bodies", for example, by exploiting the availability of the His-tag present in most recombinant proteins and tuning the concentration of bivalent cations [48]. Both aggregation-inducing tags and linkers fused to the target protein/enzyme seem relevant for recovering highly functional inclusion bodies. A recent paper reported the results of an automated approach for the preparation of libraries composed of assembled modular sequences and unbiased screening aimed at the identification of optimal tag/linker combinations [49]. The authors demonstrated that their selection modality could rationally restrict the number of constructs to be characterized in an advanced phase and inferred some general rules that were effective in recovering functional aggregates. Specifically, better results were obtained by introducing rigid proline/threonine linkers and exploiting C-terminal aggregation-inducing tags.

The aggregation of catalytically active proteins into inclusion bodies has also been successfully achieved via rational approaches [50]. The authors designed a three-component construct in which an N-terminal L_6KD peptide was directly fused to the SUMO protein, leaving the target proteins at the C-terminus. The goal was to exploit the SUMO chaperone activity to promote the correct folding of the target protein and the aggregation-prone N-terminal peptide [50] to precipitate the whole construct. Although the presence of the aggregation element seemed to be crucial for the stability of the constructs described in [49] and [51], the design reported in

[50] maintained, at least in the described examples, stable and functional inclusion bodies even after the tags were removed by means of the SUMO.

Other tags designed to favor protein aggregation and successive aggregate purification, such as the peptide HlyA60 derived from hemolysin A [52], have been proposed recently. Once purified, tagged protein aggregates with conserved functionality can be directly used to promote effective self-cleavage to simplify the separation of the target protein from the aggregating tag [53]. Specifically, an elastin-like polypeptide integrated with an engineered mini-intein sequence allowed the recovery of the fusion protein in basic buffers and effective target protein release under slightly acidic conditions, clearly outperforming the intein-based constructs proposed in the past. A different strategy aims at sequestering proteins into aggregates by fusing them with short cationic peptides [54]. Both fluorescent proteins and enzymes remain functional, but the initial data did not clarify how much this sequestration mechanism could impact host cell growth or whether it is suitable for any kind of protein. In a follow-up article [55], the same group demonstrated that both the length and charge density of the supercharged, disordered peptides significantly affected protein expression levels and subcellular localization, but the effects on bacterial metabolism were not further investigated. Understanding whether this type of protein condensate might represent a stabilizing or a perturbating factor for host cells in comparison with the equivalent expression rate of the same recombinant proteins that accumulate in soluble form would be meaningful.

Another strategy exploited ubiquitin-binding shuttle proteins (UBQLNs) modified with charge variants placed at the protein N-terminus to demonstrate that tuning the phase separation characteristics of the constructs was feasible [56]. Specifically, the insertion of negative charges correlated with a minor propensity for phase separation, but this effect could be counterbalanced by the addition of phase separation-promoting tyrosine residues. This feature might theoretically become relevant when choosing the tags to fuse at the N-terminus of target proteins. The authors compared FLAG, HA, Myc and ALFA tags, the net charge of which spans from −3 to 0, and concluded that the tyrosines present in the sequences compensated for charges, resulting in similar phase separation effects induced by all such tags. Although the observed results can be specific only for proteins of the UBQLN family, they suggest that the addition of amino acids at the N-termini of recombinant proteins should always require a specific evaluation, not only to verify the N-rule that determines the protein half-life but also to check their propensity to condense. More generally, it might be wise to perform a preliminary screening of constructs differing for N-terminal elements

(Tags or sequences designed for downstream modifications) before proceeding to large-scale production.

Innovative tags

In the last two decades, a large variety of tags (SpyTag, ALFAtag, VirD2, ybbr, APX2, LPTEG, etc.) fused to recombinant target proteins has been proposed, mostly with the aim of simplifying protein engineering, functionalization and downstream applications of the constructs [57] or simply monitoring the expression of the attached protein under different conditions [58]. Nevertheless, the previous generation of fusion partners (MBP, GST, thioredoxin, NusA, etc.) was mostly designed to improve recombinant protein solubility [59]. Following this original lane, some groups have recently investigated the possibility of exploiting intrinsically disordered peptides (IDPs) as tags to increase the yields of passenger proteins since native N-terminal disordered regions have a stabilizing effect on paired proteins [56]. The 53-amino acid NEXT tag has been isolated from the marine bacterium *Hydrogenovibrio marinus* and effectively improved the solubility of aggregation-prone proteins [60]. The NEXT tag scored better than MBP or GST, and since its presence did not affect passenger functionality, its removal is not necessary. The proposed mechanism of action is that NEXT exploits its high degree of solvent exposure and elevated dynamics to exclude neighboring macromolecules, thereby preventing protein aggregation. This capacity is apparently preserved in other IDPs, independent of their sequence [61]. The success of natural IDPs has stimulated the search for optimized synthetic IDP (SynIDP) sequences, which are recovered by screening large libraries to isolate candidates that are particularly effective in rescuing aggregation-prone proteins used as models [62]. The target moieties of the obtained fusion proteins were not simply soluble but also functional, and in the case of SynIDP-based constructs, they provided higher yields than fusions obtained by pairing the target protein with conventional tags such as MBP and SUMO.

Gene expression control

It seems that a relevant portion of laboratory-made plasmids contains mistakes [63], but even commercial vectors might present some flaws. An innovation with potentially high impact has been proposed by Shilling and coworkers [7], who identified some inaccuracies present in the popular pET vectors that result in less effective transcription and translation. Considering how extensively these plasmids are used, their substantial improvement would benefit a large community of researchers. The authors first reported that the insertion of the *lac* operator sequence (present in most of the pET vectors) caused the deletion of four bases belonging to the T7 promoter.

When restored, the presence of these bases resulted in statistically significant yield improvements in the tested recombinant proteins. Next, the optimization of the translation initiation region via the synthetic evolution approach enabled another relevant yield increase, and the effect of the two sequence modifications was often additive. Nevertheless, the optimization procedure imposed the presence of a fixed MQL sequence at the N term of the inserted target clone. According to the protein characteristics, the higher yields corresponded to soluble (and functional) as well as to insoluble proteins. The major limitation of this work is that only a few proteins were tested; therefore, it is difficult to assess how generally the extremely promising reported results are. Unfortunately, researchers [64–67] who tested partially or totally modified pET vectors for protein expression did not compare the resulting yields with those obtained via the original plasmids; therefore, the potential improvement remains undetermined. The same group that identified such inaccuracies also worked on the specific optimization of the translation initiation region of leader peptides fused to the protein to be secreted, also in this case, using pET vectors [68]. By applying directed evolution to conserve amino acid sequences, they managed to obtain base variants that allowed higher yields of the corresponding secreted and correctly cleaved proteins. It is true that in this case, it was probably not possible to permit the complete randomization of the N-terminal amino acids because these might have affected the efficiency of the leader peptide removal, but it is astonishing that the proposed sequences totally diverge from those reported in the other paper and that neither of the two cites the other article. Consequently, at present, it is not possible to evaluate the real advantage of alternative vector designs.

Another issue related to pET vectors is the actual capacity to preserve, by means of antibiotics, their theoretical production potential in host cells after the induction of recombinant expression. A recent report investigated the role of the host strains and concluded that, for those allowing the accumulation of high levels of T7 polymerase, such as BL21(DE3), it is beneficial, or even necessary, keeping short induction times and/or using kanamycin resistance instead of ampicillin [69]. This caution should prevent the effects of the toxicity induced by the overexpression of apparently any exogenous protein above "critical levels", the definition of which remains uncertain [70]. According to this view, such toxicity provides the selective pressure on *E. coli* cells necessary for inducing an adaptive response consisting of the selection of mutants with decreased, or no, T7 RNA polymerase activity. The final effect will be lower or no exogenous protein production. This could also explain why the use of low IPTG concentrations (<0.1 mM) can reduce the observed toxicity effect. In summary, in a cell

with a limited number of ribosomes, excessive amounts of exogenous mRNA may outcompete endogenous mRNA, impairing the synthesis of endogenous proteins and, ultimately, cell viability. Therefore, transformed cells stressed by high recombinant expression may have two options: die or accumulate mutations that reduce or impair T7 RNA polymerase activity. These conclusions are supported by the presented experimental data but are somewhat in contrast with the daily lab experience, which shows that recombinant protein yields differ across several logs. How to explain it if toxicity would be exclusively due to mRNA amounts and not to clone-specific features, at the level of the same mRNA or of the resulting protein? Should one hypothesize that there is an early mutation that enables the selection of a large cell subpopulation able to produce low levels of the recombinant protein? Should we assume that all the recombinant proteins obtained by inducing T7 RNA polymerase with more than 0.1 mM IPTG are produced by mutant strains? Should we accept the reductive theory that there is only a single limiting element/step in the expression chain (from transcription until posttranslational modifications) or rather consider more potential bottlenecks? What is missing in this article is at least one experiment: if I select an adapted colony (according to the authors' classification, one from the "large green" population, productive because mutations should have kept low the expression rate) and start a new culture from it, would I get a homogeneous and productive population or again a segregation of cells with different characteristics? Are the productive cells genetically homogeneous or a mix of (functionally convergent) mutations? It would also be useful to monitor the relationship between the expression rate and aggregate formation [71] to exclude effects due to (the toxicity of) protein aggregates.

Considering these observations, the optimization of the expression level of (toxic) recombinant proteins has been approached from an interesting perspective by Li et al. [72]. They developed a CRISPR-based strategy to create a large unbiased library of bacterial hosts characterized by variable ribosomal binding site sequences for T7 RNA polymerase that resulted in highly different capacities to express the model enzyme. The rationale of this method derives from repeated observations that lower expression levels of difficult-to-produce proteins often result in higher yields of recombinant proteins [6, 73]. However, whereas the previous options, regardless of the target of the inhibitory mutation, enabled one single downregulated expression level to which any target protein had to adapt, possessing a library of T7 polymerases with uniformly distributed activity represents an evident advantage: statistically, it allows the identification of the most suitable translation rate for any recombinant construct. In summary, the mutagenesis process results in

a reduction in recombinant expression, which, according to James et al. [70], is driven by the stress induced by the accumulation of exogenous mRNA, is somehow anticipated by providing a library of mutants with a differential capacity to express exogenous mRNA. The data are promising and would be interesting for future work to characterize the biophysical/biochemical features of the produced proteins. Alternative approaches aimed at calibrating translation efficiency could also be expected to optimize recombinant protein production. Nevertheless, in contrast to the above-described line of results, the preparation of a mutant library of the T7 promoter did not lead to the identification of significantly more productive candidates, although these mutants should allow variable translation rates [74]. Notably, combining the efforts devoted to both improving and decreasing T7 RNA polymerase activity with the aim of identifying the target protein characteristics that determine why different constructs might benefit from opposite drivers would be exciting. This knowledge might allow the prediction of which strategy would work best for each protein instead of a random trial-and-error optimization approach. Currently, there is at least a consensus on the necessity of finely tunable control of T7 RNA polymerase expression. In this context, there are indeed several systems that should allow the modulation of polymerase activity, but they are not particularly successful. A more accurate design based on arabinose control was proposed by Stargard et al. [75], who conceived a system (based on the expression host BL21-AI<*gp2*>), in which cell growth is decoupled from recombinant protein production and allows simultaneous tuning of recombinant protein expression. This is possible through the expression of a phage-derived inhibitor peptide that blocks *E. coli* RNA polymerase but not T7 RNA polymerase, which controls recombinant expression in host bacteria [76]. The assumption is that recombinant expression might be inhibited by cell metabolite shortages. This solution is the opposite of that proposed by [70], who suggested that inhibition of host physiological metabolism is the cause of bacterial decline and that mutations lead to a greater share of bacteria with reduced or no T7 RNA polymerase activity. It would be interesting to determine the morphological characteristics and degree of T7 RNA polymerase homo/heterogeneity in BL21-AI<*gp2*> cells at the end of the recombinant culture.

There are further indications that somehow connect the mRNA features and metabolic burden issues. Modifying the sequence corresponding to the ribosome binding site of exogenous mRNA can positively affect recombinant protein expression [77]. Improving the efficiency of the translation step of exogenous constructs might allow decent protein yields while decreasing the level of transcription activity and reducing competition

with endogenous mRNA, since recombinant mRNAs will occupy the ribosome for a shorter period. However, ribosome stalling did not affect host cell growth in another case [78], in which it rather seemed that the stress induced by the recombinant expression of Fab fragments was at the level of the membrane and secretion molecular structures. The picture does not become clearer considering the results obtained in a project aimed at improving the expression of recombinant proteins fused to CASPON™. The 5' region of the tag sequence was modified to optimize the ribosome-mRNA interaction and increase the protein yield. Although the highest translation efficiency was obtained with a construct carrying an expression enhancer element, characterized by favorable interaction with the ribosome, the highest recombinant protein accumulation was obtained with a construct with comparatively low transcriptional efficiency and no expression enhancer sequence. A further positive contribution was introduced by silent mutations of the translated N-term that reduced the tendency of the mRNA to assume secondary structures. The authors attributed the better performance of such a construct to the putatively greater accessibility of its ribosome binding site [9]. Remarkably, the significant increase in recombinant protein yield (more than 5 times) induced no stress to the host cells, since their growth rate was not affected, and no accumulation of aggregates/inclusion bodies was observed despite the silent mutations that generated sequences for rare codons. The results might be interpreted otherwise, namely, that the higher yields are due not to improvement but to a worse global efficiency of the selected construct. This condition would result in a slower translation rate, which is more compatible with the metabolic and folding capacities of the host cells. In this sense, the system introduced by Li et al. [79], who engineered a controllable plasmid replication system to regulate the available gene copies and indirectly control gene expression, might be useful.

Impact of recombinant proteins on the metabolism

A surprising discovery of recent years is that the metabolic burden induced by recombinant protein expression in bacteria is not due to energy limitations, as generally accepted, but rather to the accumulation of ATP and precursors of glycolysis. This is true for both the expression system based on the BL21(DE3)/T7 promoter [80] and for the combination of TG1 (an *E. coli* K12 strain) and the tac promoter [81]. This apparently leads to an unbalanced metabolic repertoire that results in the observed growth inhibition. These results provide further support for the use of downregulated expression conditions that can be obtained by either controlling the energy supply or tuning the expression rate more accurately, for example, using promoters that enable more effective tuning

of polymerase activity [5, 82]. From this perspective, the possibility of switching cell resources from biomass formation to recombinant protein synthesis and *viceversa* would make sense [75, 76]. In contrast, it is difficult to anticipate the effect of optimizing the 16 S ribosomal RNA for faster translation [83]: might it result in less stress for the cells because ribosomes can deal with more mRNA (exogenous and endogenous) or more stress because then there are not enough downstream foldases, chaperones or secretion machineries? This report does not provide information about the quality of the produced proteins, preventing any possibility of evaluating this strategy.

Notably, great differences exist even among bacterial strains, and these differences may affect the final outcome. For example, the differences in glucose metabolism between the K12 and B. *coli* strains seem to be the basis of the experimentally observed significantly higher rate of misincorporation of noncanonical amino acids in K12-type HMS174(DE3) bacteria than in the BL21(DE3) strain [84]. The conditions that lead to critical accumulation of such amino acids are generally typical of high-density fermentation, but since any modification that is necessary for adoption in the advanced process phases requires longer setting work, the reported data suggest that beginning the whole process with the most robust BL21(DE3) bacteria is preferable. As a potential alternative, it has been proposed to supplement the growth medium with amino acids, assuming that the metabolic burden (and amino acid misincorporation?) is the consequence of (strain dependent?) reduced capacity to translate endogenous proteins necessary for host metabolism caused by amino acid shortages [85].

Conclusions

It is somehow astonishing that, although *E. coli* has been investigated for decades and has become, at least in research labs, the most commonly used organism for producing recombinant proteins more than 30 years ago, many theoretical and technical advancements still accumulate. Paradoxically, this abundance can become a hindrance, since it is not only difficult to follow the constant updates but also almost impossible to test all the proposed solutions aimed at improving the different steps of the technology, even when rational approaches are applied to guide the choice [86]. Some of these methods can be technically too demanding for a standard lab, for example, providing precisely controlled feeding during fermentation or coordinating more expression steps simultaneously. The consequence is that improving the efficiency of one of them might negatively affect the downstream process because the capacity of involved structures will be undersized with respect to the upstream optimized step. The problem is perhaps

because we do not truly know what "metabolic burden" means, despite several single mechanisms have been thoroughly described [5]. On the one hand, it is clear that the expression of an exogenous protein induces stress in the host cell, but what is, or are, the bottleneck(s)? Is there competition between endogenous metabolism and the supplementary need for biological building blocks, for ribosomes or possibly even for chaperones or secretion structures? Is it always true that "less is more", namely, that decreasing recombinant expression is positive because it avoids pushing the host capacity to its limits, despite evidence that apparently succeeded in overcoming such an expected ceiling, yielding several hundreds of mg/L recombinant proteins [23]?

At the lab level, setting aside more theoretical issues, identifying the rationale suitable for deciding which innovation is worthy of investment in time and resources remains a problem. The choice is usually challenged by the restricted number of examples reported in the literature. Many (recent) proposals base their conclusions on experiments performed with few samples and, in most cases, only one or just a few groups tested the new technologies. Comparative works are rare, and the characterization of the final proteins is often performed differently and is usually not sufficiently accurate and complete, at least with respect to exhaustive standards proposed recently [87]. There is an extremely broad spectrum of experimental accuracy among the published results, highly heterogeneous methods of data reporting and very variable access to lab instrumentation. For example, in recent years, the BOKU unit has made impressive contributions in terms of information, but it uses very sophisticated fermentation conditions that are absent in almost any standard laboratory "simply" interested in producing recombinant proteins as intermediate reagents for biological experiments. Of course, this is not BOKU's fault, nevertheless, the example indicates that data recovered in a single lab may represent a sort of "systematic deviation" with respect to other, common external conditions, and consequently, it might be difficult to repeat some experiments and combine the information generated by independent labs that work differently.

There are several options, among those described in this review, that seem potentially powerful and that will be worth following through the publications that will appear in the next years (Fig. 1). At the same time, it seems that we must accept that there is no one-size-fits-all approach to this topic and, consequently, multiple parameters require simultaneous consideration to carry out recombinant protein production optimization in *E. coli*. Since new information carried further questions rather than providing definitive answers, it would be beneficial to guide the development by means of benchmarking works aimed at validating the recently

proposed methods using larger experimental sets. Filling this knowledge gap would remove odd data and misleading conclusions, allowing concentrating on solid options. Furthermore, the adoption of more uniform data collection and organization would be highly beneficial since it would render acquired information suitable for artificial intelligence-based metadata analysis. This procedure would be highly desirable for evaluating not only the effectiveness of the proposed innovations but also its transferability in terms of required skills and resources.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

References

1. Peleg Y, Unger T. Resolving bottlenecks for recombinant protein expression in *E. Coli*. Methods Mol Biol. 2012;800:173–86. https://doi.org/10.1007/978-1-61779-349-3_12.
2. Gileadi O. Recombinant protein expression in *E. Coli*: a historical perspective. Methods Mol Biol. 2017;1586:3–10. https://doi.org/10.1007/978-1-4939-6887-9_1.
3. Royes J, Talbot P, Le Bon C, Moncoq K, Uzan M, Zito F, Miroux B. Membrane protein production in *Escherichia coli*: protocols and rules. Methods Mol Biol. 2022;2507:19–39. https://doi.org/10.1007/978-1-0716-2368-8_2.
4. Schütz A, Bernhard F, Berrow N, Buyel JF, Ferreira-da-Silva F, Haustraete J, van den Heuvel J, Hoffmann JE, de Marco A, Peleg Y, Suppmann S, Unger T, Vanhoucke M, Witt S, Remans K. A concise guide to choosing suitable gene expression systems for recombinant protein production. STAR Protoc. 2023;4:102572. https://doi.org/10.1016/j.xpro.2023.102572.
5. Snoeck S, Guidi C, De Mey M. Metabolic burden explained: stress symptoms and its related responses induced by (over)expression of (heterologous) proteins in *Escherichia coli*. Microb Cell Fact. 2024;23:96. https://doi.org/10.1186/s12934-024-02370-9.
6. Schlegel S, Löfblom J, Lee C, Hjelm A, Klepsch M, Strous M, Drew D, Slotboom DJ, de Gier JW. Optimizing membrane protein overexpression in the *Escherichia coli* strain Lemo21(DE3). J Mol Biol. 2012;423:648–59. https://doi.org/10.1016/j.jmb.2012.07.019.
7. 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;3:214. https://doi.org/10.1038/s42003-020-0939-8.

8. Peleg Y, Vincentelli R, Collins BM, Chen KE, Livingstone EK, Weeratunga S, et al. Community-wide experimental evaluation of the PROSS Stability-Design Method. J Mol Biol. 2021;433:166964. https://doi.org/10.1016/j.jmb.2021.166964.

9. Köppl C, Buchinger W, Striedner G, Cserjan-Puschmann M. Modifications of the 5′region of the CASPONTM tag's mRNA further enhance soluble recombinant protein production in *Escherichia coli*. Microb Cell Fact. 2024;23:86. https://doi.org/10.1186/s12934-024-02350-z.

10. Saez NJ, Vincentelli R. High-throughput expression screening and purification of recombinant proteins in *E. Coli*. Methods Mol Biol. 2014;1091:33–53. https://doi.org/10.1007/978-1-62703-691-7_3.

11. Luke J, Carnes AE, Hodgson CP, Williams JA. Improved antibiotic-free DNA vaccine vectors utilizing a novel RNA based plasmid selection system. Vaccine. 2009;27:6454–9.

12. Peubez I, Chaudet N, Mignon C, Hild G, Husson S, Courtois V, De Luca K, Speck D, Sodoyer R. Antibiotic-free selection in *E. Coli*: new considerations for optimal design and improved production. Microb Cell Fact. 2010;9:65. https://doi.org/10.1186/1475-2859-9-65.

13. Kroll J, Klinter S, Schneider C, Voss I, Steinbüchel A. Plasmid addiction systems: perspectives and applications in biotechnology. Microb Biotechnol. 2010;3:634–57. https://doi.org/10.1111/j.1751-7915.2010.00170.x.

14. Reschner A, Scohy S, Vandermeulen G, Daukandt M, Jacques C, Michel B, et al. Use of Staby(®) technology for development and production of DNA vaccines free of antibiotic resistance gene. Hum Vaccin Immunother. 2013;9:2203–10. https://doi.org/10.4161/hv.25086.

15. Kang CW, Lim HG, Yang J, Noh MH, Seo SW, Jung GY. Synthetic auxotrophs for stable and tunable maintenance of plasmid copy number. Metab Eng. 2018;48:121–8. https://doi.org/10.1016/j.ymben.2018.05.020.

16. Sathesh-Prabu C, Tiwari R, Lee SK. Substrate-inducible and antibiotic-free high-level 4-hydroxyvaleric acid production in engineered *Escherichia coli*. Front Bioeng Biotechnol. 2022;10:960907. https://doi.org/10.3389/fbioe.2022.960907.

17. Brechun KE, Förschle M, Schmidt M, Kranz H. Method for plasmid-based antibiotic-free fermentation. Microb Cell Fact. 2024;23:18. https://doi.org/10.1186/s12934-023-02291-z.

18. Guillén-Chable F, Arenas-Sosa I, Islas-Flores I, Corzo G, Martinez-Liu C, Estrada G. Antibacterial activity and phospholipid recognition of the recombinant defensin J1–1 from *Capsicum* Genus. Protein Expr Purif. 2017;136:45–51. https://doi.org/10.1016/j.pep.2017.06.007.

19. Gomez-Lugo JJ, Casillas-Vega NG, Gomez-Loredo A, Balderas-Renteria I, Zarate X. High-yield expression and purification of Scygonadin, an antimicrobial peptide, using the small metal-binding protein SmbP. Microorganisms. 2024;12:278. https://doi.org/10.3390/microorganisms12020278.

20. López-Cano A, Martínez-Miguel M, Guasch J, Ratera I, Arís A, Garcia-Fruitós E. Exploring the impact of the recombinant *Escherichia coli* strain on defensins antimicrobial activity: BL21 versus origami strain. Microb Cell Fact. 2022;21:77. https://doi.org/10.1186/s12934-022-01803-7.

21. Djender S, Schneider A, Beugnet A, Crepin R, Desrumeaux KE, Romani C, Moutel S, Perez F, de Marco A. Bacterial cytoplasm as an effective cell compartment for producing functional VHH-based affinity reagents and Camelidae IgG-like recombinant antibodies. Microb Cell Fact. 2014;13:140. https://doi.org/10.1186/s12934-014-0140-1.

22. Bertelsen AB, Hackney CM, Bayer CN, Kjelgaard LD, Rennig M, Christensen B, et al. DisCoTune: versatile auxiliary plasmids for the production of disulphide-containing proteins and peptides in the *E. Coli* T7 system. Microb Biotechnol. 2021;14:2566–80. https://doi.org/10.1111/1751-7915.13895.

23. Hennigan JN, Menacho-Melgar R, Sarkar P, Golovsky M, Lynch MD. Scalable, robust, high-throughput expression & purification of nanobodies enabled by 2-stage dynamic control. Metab Eng. 2024;85:116–30. https://doi.org/10.1016/j.ymben.2024.07.012.

24. Lingg N, Kröß C, Engele P, Öhlknecht C, Köppl C, Fischer A, et al. CASPON platform technology: Ultrafast circularly permuted caspase-2 cleaves tagged fusion proteins before all 20 natural amino acids at the N-terminus. N Biotechnol. 2022;71:37–46. https://doi.org/10.1016/j.nbt.2022.07.002.

25. Köppl C, Lingg N, Fischer A, Kröß C, Loibl J, Buchinger W, et al. Fusion Tag Design influences Soluble recombinant protein production in *Escherichia coli*. Int J Mol Sci. 2022;23:7678. https://doi.org/10.3390/ijms23147678.

26. Gibisch M, Müller M, Tauer C, Albrecht B, Hahn R, Cserjan-Puschmann M, Striedner G. A production platform for disulfide-bonded peptides in the

27. periplasm of *Escherichia coli*. Microb Cell Fact. 2024;23:166. https://doi.org/10.1186/s12934-024-02446-6.

27. Rettenbacher LA, von der Haar T. A quantitative interpretation of oxidative protein folding activity in *Escherichia coli*. Microb Cell Fact. 2022;21:268. https://doi.org/10.1186/s12934-022-01982-3.

28. Natarajan A, Jaroentomeechai T, Cabrera-Sánchez M, Mohammed JC, Cox EC, Young O, et al. Engineering orthogonal human O-linked glycoprotein biosynthesis in bacteria. Nat Chem Biol. 2020;16:1062–70. https://doi.org/10.1038/s41589-020-0595-9.

29. Wardman JF, Sim L, Liu J, Howard TA, Geissner A, Danby PM, Boraston AB, Wakarchuk WW, Withers SG. A high-throughput screening platform for enzymes active on mucin-type O-glycoproteins. Nat Chem Biol. 2023;19:1246–55. https://doi.org/10.1038/s41589-023-01405-3.

30. Pratama F, Linton D, Dixon N. Genetic and process engineering strategies for enhanced recombinant N-glycoprotein production in bacteria. Microb Cell Fact. 2021;20:198. https://doi.org/10.1186/s12934-021-01689-x.

31. Passmore IJ, Faulds-Pain A, Abouelhadid S, Harrison MA, Hall CL, Hitchen P, Dell A, Heap JT, Wren BW. A combinatorial DNA assembly approach to biosynthesis of N-linked glycans in *E. Coli*. Glycobiology. 2023;33:138–49. https://doi.org/10.1093/glycob/cwac082.

32. Bao Z, Gao Y, Song Y, Ding N, Li W, Wu Q, Zhang X, Zheng Y, Li J, Hu X. Construction of an *Escherichia coli* chassis for efficient biosynthesis of human-like N-linked glycoproteins. Front Bioeng Biotechnol. 2024;12:1370685. https://doi.org/10.3389/fbioe.2024.1370685.

33. Dow JM, Mauri M, Scott TA, Wren BW. Improving protein glycan coupling technology (PGCT) for glycoconjugate vaccine production. Expert Rev Vaccines. 2020;19:507–27. https://doi.org/10.1080/14760584.2020.1775077.

34. Samaras JJ, Mauri M, Kay EJ, Wren BW, Micheletti M. Development of an automated platform for the optimal production of glycoconjugate vaccines expressed in *Escherichia coli*. Microb Cell Fact. 2021;20:104. https://doi.org/10.1186/s12934-021-01588-1.

35. Kay EJ, Mauri M, Willcocks SJ, Scott TA, Cuccui J, Wren BW. Engineering a suite of *E. Coli* strains for enhanced expression of bacterial polysaccharides and glycoconjugate vaccines. Microb Cell Fact. 2022;21:66. https://doi.org/10.1186/s12934-022-01792-7.

36. Kay EJ, Dooda MK, Bryant JC, Reid AJ, Wren BW, Troutman JM, Jorgenson MA. Engineering *Escherichia coli* for increased Und-P availability leads to material improvements in glycan expression technology. Microb Cell Fact. 2024;23:72. https://doi.org/10.1186/s12934-024-02339-8.

37. De Marni ML, Monegal A, Venturini S, Vinati S, Carbone R, de Marco A. Antibody purification-independent microarrays (PIM) by direct bacteria spotting on TiO$_2$-treated slides. Methods. 2012;56:317–25. https://doi.org/10.1016/j.ymeth.2011.06.008.

38. Jose J, Maas RM, Teese MG. Autodisplay of enzymes–molecular basis and perspectives. J Biotechnol. 2012;161:92–103. https://doi.org/10.1016/j.jbiotec.2012.04.001.

39. Zhang L, Tan L, Liu M, Chen Y, Yang Y, Zhang Y, Zhao G. Quantitative measurement of cell-surface displayed proteins based on split-GFP assembly. Microb Cell Fact. 2024;23:108. https://doi.org/10.1186/s12934-024-02386-1.

40. Pedelacq J-D, Cabantous S. Development and applications of superfolder and split fluorescent protein detection systems in biology. Int J Mol Sci. 2019;20:3479. https://doi.org/10.3390/ijms20143479.

41. Müller C, Igwe CL, Wiechert W, Oldiges M. Scaling production of GFP1-10 detector protein in *E. Coli* for secretion screening by split GFP assay. Microb Cell Fact. 2021;20:191. https://doi.org/10.1186/s12934-021-01672-6.

42. Kim Y, Kim D, Hieu NM, Byun H, Ahn JH. PySupercharge: a python algorithm for enabling ABC transporter bacterial secretion of all proteins through amino acid mutation. Microb Cell Fact. 2024;23:115. https://doi.org/10.1186/s12934-024-02342-z.

43. Zhu W, Wang Y, Lv L, Wang H, Shi W, Liu Z, Yang W, Zhu J, Lu H. SHTXTHHly, an extracellular secretion platform for the preparation of bioactive peptides and proteins in *Escherichia coli*. Microb Cell Fact. 2022;21:128. https://doi.org/10.1186/s12934-022-01856-8.

44. Ge J, Wang X, Bai Y, Wang Y, Wang Y, Tu T, Qin X, Su X, Luo H, Yao B, Huang H, Zhang J. Engineering *Escherichia coli* for efficient assembly of heme proteins. Microb Cell Fact. 2023;22:59. https://doi.org/10.1186/s12934-023-02067-5.

45. Poborsky M, Crocoll C, Motawie MS, Halkier BA. Systematic engineering pinpoints a versatile strategy for the expression of functional cytochrome P450 enzymes in *Escherichia coli* cell factories. Microb Cell Fact. 2023;22:219. https://doi.org/10.1186/s12934-023-02219-7.

46. García-Fruitós E, González-Montalbán N, Morell M, Vera A, Ferraz RM, Arís A, Ventura S, Villaverde A. Aggregation as bacterial inclusion bodies does not

imply inactivation of enzymes and fluorescent proteins. Microb Cell Fact. 2005;4:27. https://doi.org/10.1186/1475-2859-4-27.

47. Arié JP, Miot M, Sassoon N, Betton JM. Formation of active inclusion bodies in the periplasm of *Escherichia coli*. Mol Microbiol. 2006;62:427–37. https://doi.org/10.1111/j.1365-2958.2006.05394.x.

48. López-Laguna H, Sánchez JM, Carratalá JV, Rojas-Peña M, Sánchez-García L, Parladé E, et al. Biofabrication of functional protein nanoparticles through simple his-tag engineering. ACS Sustain Chem Eng. 2021;9:12341–54. https://doi.org/10.1021/acssuschemeng.1c04256.

49. Helleckes LM, Küsters K, Wagner C, Hamel R, Saborowski R, Marienhagen J, Wiechert W, Oldiges M. High-throughput screening of catalytically active inclusion bodies using laboratory automation and bayesian optimization. Microb Cell Fact. 2024;23:67. https://doi.org/10.1186/s12934-024-02319-y.

50. Komolov AS, Sannikova EP, Gorbunov AA, Gubaidullin II, Plokhikh KS, Konstantinova GE, Bulushova NV, Kuchin SV, Kozlov DG. Synthesis of biologically active proteins as L6KD-SUMO fusions forming inclusion bodies in *Escherichia coli*. Biotechnol Bioeng. 2024;121:535–50. https://doi.org/10.1002/bit.28587.

51. Zhou B, Xing L, Wu W, Zhang XE, Lin Z. Small surfactant-like peptides can drive soluble proteins into active aggregates. Microb Cell Fact. 2012;11:10. https://doi.org/10.1186/1475-2859-11-10.

52. Ma J, Liu P, Cai S, Wu T, Chen D, Zhu C, Li S. Discovery and Identification of a Novel tag of HlyA60 for protein active aggregate formation in *Escherichia coli*. J Agric Food Chem. 2024;72:493–503. https://doi.org/10.1021/acs.jafc.3c05860.

53. Yuan H, Prabhala SV, Coolbaugh MJ, Stimple SD, Wood DW. Improved self-cleaving precipitation tags for efficient column free bioseparations. Protein Expr Purif. 2024;224:106578. https://doi.org/10.1016/j.pep.2024.106578.

54. Yeong V, Wang JW, Horn JM, Obermeyer AC. Intracellular phase separation of globular proteins facilitated by short cationic peptides. Nat Commun. 2022;13:7882. https://doi.org/10.1038/s41467-022-35529-2.

55. Liao J, Yeong V, Obermeyer AC. Charge-patterned disordered peptides Tune Intracellular phase separation in Bacteria. ACS Synth Biol. 2024;13:598–612. https://doi.org/10.1021/acssynbio.3c00564.

56. Dao TP, Rajendran A, Galagedera SKK, Haws W, Castañeda CA. Short disordered termini and proline-rich domain are major regulators of UBQLN1/2/4 phase separation. Biophys J. 2024;123:1449–57. https://doi.org/10.1016/j.bpj.2023.11.3401.

57. Veggiani G, Giabbai B, Semrau MS, Medagli B, Riccio V, Bajc G, Storici P, de Marco A. Comparative analysis of fusion tags used to functionalize recombinant antibodies. Protein Expr Purif. 2020;166:105505. https://doi.org/10.1016/j.pep.2019.105505.

58. Van Zyl WF, Van Staden AD, Dicks LMT, Trindade M. Use of the mCherry fluorescent protein to optimize the expression of class I lanthipeptides in *Escherichia coli*. Microb Cell Fact. 2023;22:149. https://doi.org/10.1186/s12934-023-02162-7.

59. Dümmler A, Lawrence AM, de Marco A. Simplified screening for the detection of soluble fusion constructs expressed in *E. Coli* using a modular set of vectors. Microb Cell Fact. 2005;4:34. https://doi.org/10.1186/1475-2859-4-34.

60. Jo BH. An intrinsically disordered peptide tag that confers an unusual solubility to aggregation-prone proteins. Appl Environ Microbiol. 2022;88:e0009722. https://doi.org/10.1128/aem.00097-22.

61. Jo BH. Improved Solubility and Stability of a Thermostable Carbonic Anhydrase via Fusion with Marine-Derived intrinsically disordered solubility enhancers. Int J Mol Sci. 2024;25:1139. https://doi.org/10.3390/ijms25021139.

62. Tang NC, Su JC, Shmidov Y, Kelly G, Deshpande S, Sirohi P, Peterson N, Chilkoti A. Synthetic intrinsically disordered protein fusion tags that enhance protein solubility. Nat Commun. 2024;15:3727.

63. Bourzac K. Serious errors plague DNA tool that's a workhorse of biology. Nature. 2024;631:487–8. https://doi.org/10.1038/d41586-024-02280-1.

64. Moses A, Bhalla P, Thompson A, Lai L, Coskun FS, Seroogy CM, de la Morena MT, Wysocki CA, van Oers NSC. Comprehensive phenotypic analysis of diverse FOXN1 variants. J Allergy Clin Immunol. 2023;152:1273–e129115. https://doi.org/10.1016/j.jaci.2023.06.019.

65. Nguyen LT, Rananaware SR, Yang LG, Macaluso NC, Ocana-Ortiz JE, Meister KS, et al. Engineering highly thermostable Cas12b via de novo structural analyses for one-pot detection of nucleic acids. Cell Rep Med. 2023;4:101037. https://doi.org/10.1016/j.xcrm.2023.101037.

66. Toyama Y, Shimada I. NMR characterization of RNA binding property of the DEAD-box RNA helicase DDX3X and its implications for helicase activity. Nat Commun. 2024;15:3303. https://doi.org/10.1038/s41467-024-47659-w.

67. Amer-Sarsour F, Falik D, Berdichevsky Y, Kordonsky A, Eid S, Rabinski T, et al. Disease-associated polyalanine expansion mutations impair

UBA6-dependent ubiquitination. EMBO J. 2024;43:250–76. https://doi.org/10.1038/s44318-023-00018-9.

68. Mirzadeh K, Shilling PJ, Elfageih R, Cumming AJ, Cui HL, Rennig M, Nørholm MHH, Daley DO. Increased production of periplasmic proteins in *Escherichia coli* by directed evolution of the translation initiation region. Microb Cell Fact. 2020;19:85. https://doi.org/10.1186/s12934-020-01339-8.

69. Khananisho D, Cumming AJ, Kulakova D, Shilling PJ, Daley DO. Tips for efficiently maintaining pET expression plasmids. Curr Genet. 2023;69:277–87. https://doi.org/10.1007/s00294-023-01276-0.

70. James J, Yarnall B, Koranteng A, Gibson J, Rahman T, Doyle DA. Protein over-expression in *Escherichia coli* triggers adaptation analogous to antimicrobial resistance. Microb Cell Fact. 2021;20:13. https://doi.org/10.1186/s12934-020-01462-6.

71. Schultz T, Martinez L, de Marco A. The evaluation of the factors that cause aggregation during recombinant expression in *E. Coli* is simplified by the employment of an aggregation-sensitive reporter. Microb Cell Fact. 2006;5:28. https://doi.org/10.1186/1475-2859-5-28.

72. 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. https://doi.org/10.1021/acssynbio.1c00463.

73. Sun XM, Zhang ZX, Wang LR, Wang JG, Liang Y, Yang HF, et al. 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. https://doi.org/10.1002/bit.27558.

74. 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. https://doi.org/10.1007/s12010-019-03113-y.

75. 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;20:27. https://doi.org/10.1186/s12934-021-01512-7.

76. 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. https://doi.org/10.1021/acssynbio.0c00028.

77. Zhang Y, Chen H, Zhang Y, Yin H, Zhou C, Wang Y. Direct RBS Engineering of the biosynthesis-related gene cluster for efficient productivity of violaceins in *E. Coli*. Microb Cell Fact. 2021;20:38. https://doi.org/10.1186/s12934-021-01518-1.

78. Vazulka S, Schiavinato M, Tauer C, Wagenknecht M, Cserjan-Puschmann M, Striedner G. RNA-seq reveals multifaceted gene expression response to Fab production in *Escherichia coli* fed-batch processes with particular focus on ribosome stalling. Microb Cell Fact. 2024;23:14. https://doi.org/10.1186/s12934-023-02278-w.

79. 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. https://doi.org/10.1016/j.ymben.2022.01.003.

80. Li Z, Rinas U. Recombinant protein production-associated metabolic burden reflects anabolic constraints and reveals similarities to a carbon overfeeding response. Biotechnol Bioeng. 2021;118:94–105.

81. 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;20:169. https://doi.org/10.1186/s12934-021-01661-9.

82. Du F, Liu YQ, 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;20:189. https://doi.org/10.1186/s12934-021-01680-6.

83. Liu F, Bratulić S, Costello A, Miettinen TP, Badran AH. Directed evolution of rRNA improves translation kinetics and recombinant protein yield. Nat Commun. 2021;12:5638. https://doi.org/10.1038/s41467-021-25852-5.

84. Mayer F, Cserjan-Puschmann M, Haslinger B, Shpylovyi A, Dalik T, Sam C, Hahn R, Striedner G. Strain specific properties of *Escherichia coli* can prevent noncanonical amino acid misincorporation caused by scale-related process heterogeneities. Microb Cell Fact. 2022;21:170. https://doi.org/10.1186/s12934-022-01895-1.

85. 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. https://doi.org/10.1002/bit.27371.

86. 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. https://doi.org/10.1007/s00253-020-10454-w.

87. de Marco A, Berrow N, Lebendiker M, Garcia-Alai M, Knauer SH, Lopez-Mendez B, et al. Quality control of protein reagents for the improvement of research data reproducibility. Nat Commun. 2021;12:2795. https://doi.org/10.1038/s41467-021-23167-z.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.