

REVIEW

Open Access

Recombinant polypeptide production in *E. coli*: towards a rational approach to improve the yields of functional proteins

Ario de Marco^{1,2}

Abstract

The development of complementary technologies enabled the successful production of recombinant polypeptides in bacteria and opened to biology researchers new avenues as obtaining suitable amounts of proteins necessary for their experimental work became easy, fast, and inexpensive. Nevertheless, the recombinant approach remained somehow unpredictable, since many constructs resisted to apparent production or accumulated as aggregates. Several factors and physical/chemical conditions that could improve the accumulation of native-like protein were identified. At the same time, it was acknowledged that the outcome of most of them was erratic and that almost any protein required its own specific optimized set of conditions to achieve its correct folding. The attempt to understand the critical points specific for recombinant protein production missed the goal of setting universally useful protocols, but contributed to the increase of the rate of success by proposing always new empiric combinations. Nevertheless, the results published in the recent literature allow for a better comprehension of some key mechanisms controlling protein production in *E. coli* and could enable the elaboration of rational methodologies for improving the quantitative and qualitative features of the produced polypeptides. This result will be achieved when the identification of the limiting step that impairs the accomplishment of the native folding for any single construct will become straightforward. This minireview will discuss how factors such as the expression rate, the folding machinery, and the secretion efficiency may impact the final protein yields.

Keywords: Carrier protein, Chaperones, Expression rate, Fusion proteins, Inclusion bodies, Isomerases, Promoter, Regulative sequences, Secretion mechanisms

Introduction

The intuition that recombinant expression could represent a metabolic burden for the host cell arose already in the infancy of this technology [1]. Consequently, it was proposed to abandon the initial approaches that pushed the expression rates by exploiting high-copy number plasmids and strong (as well as leaky) promoters in favor of tighter control and slower expression rate obtained with suitable plasmids and reduced growth temperatures [2,3]. For instance, in the *lac* promoter-regulated system both the addition of glucose and the overexpression of the *lacI* repressor and of the T7 lysozyme repress gene expression

leakage effectively, impairing negative effects on plasmid segregation during cell replication [4]. Recombinant expression can be regulated also by controlling the plasmid number per cell. This factor depends on the origin of replication sequences: whereas the pUC-based plasmids can accumulate in hundreds of copies per single cell, the pBR322-based plasmids produce tens of copies and the pSC101 and p15A systems just few.

It was also observed that the induction of stress conditions by ethanol, benzyl alcohol or salt addition as well as by transient heat shock could improve the folding capacity of the treated bacteria [5]. All these perturbations stimulate the cell machinery specialized in protecting the macromolecule structure and the observed positive effect in terms of final recombinant protein yields was attributed to the accumulation of chaperones. This assumption was experimentally proven by demonstrating that chaperone

Correspondence: ario.de-marco@curie.fr

¹Department of Biomedical Sciences and Engineering, University of Nova Gorica (UNG), Glavni Trg 9 - SI-5261, Vipava, Slovenia

²Therapeutic Antibody Platform, Institut Curie, 3-5 Impasse Reille, Paris 75014, France

overexpression could prevent the aggregation of recombinant proteins and it represented a significant biotechnological improvement [6]. Such approach has been successively optimized but still remains cumbersome due to the necessity of comparing several different chaperone combinations [7] (Figure 1). As an alternative, strains that could improve the production of a single target protein or a group of structurally similar proteins were generated by random mutagenesis [8-10]. Both deletion and overexpression of *E. coli* proteins resulted in the increased accumulation of recombinant proteins. The mechanisms involved vary from improved metabolism to the optimization of the protein quality control machinery [11]. The “Walker strains” [12] are spontaneous chromosomal mutants isolated by empirical selection of bacteria expressing toxic proteins and that were able to grow in the presence of both the expression inducer IPTG and ampicillin. They allowed for the recovery of several functional recombinant membrane proteins and created a wave of hope before the

realization dawned that they were not miraculous, but constituted just another tool that could work in specific conditions. More interesting for the design of a more general methodology was probably the discovery of the molecular mechanism by which these strains managed to improve the recombinant protein yields.

Expressing less to fold better

The Walker strains were introduced as “mutant hosts that allow the synthesis of SOME membrane and globular proteins” [12] and for more than ten years the reasons of this (partial) success remained unknown. It was surprising to discover that the mutations present in the different strains were indeed convergent and involved a single mechanism, namely they contributed to the reduced efficiency of the *lac* UV5 promoter that controls the expression of T7 RNA polymerase [13,14]. It was the first evidence that reducing the T7 RNA polymerase level in strains depending on this enzyme for recombinant expression could be a key element

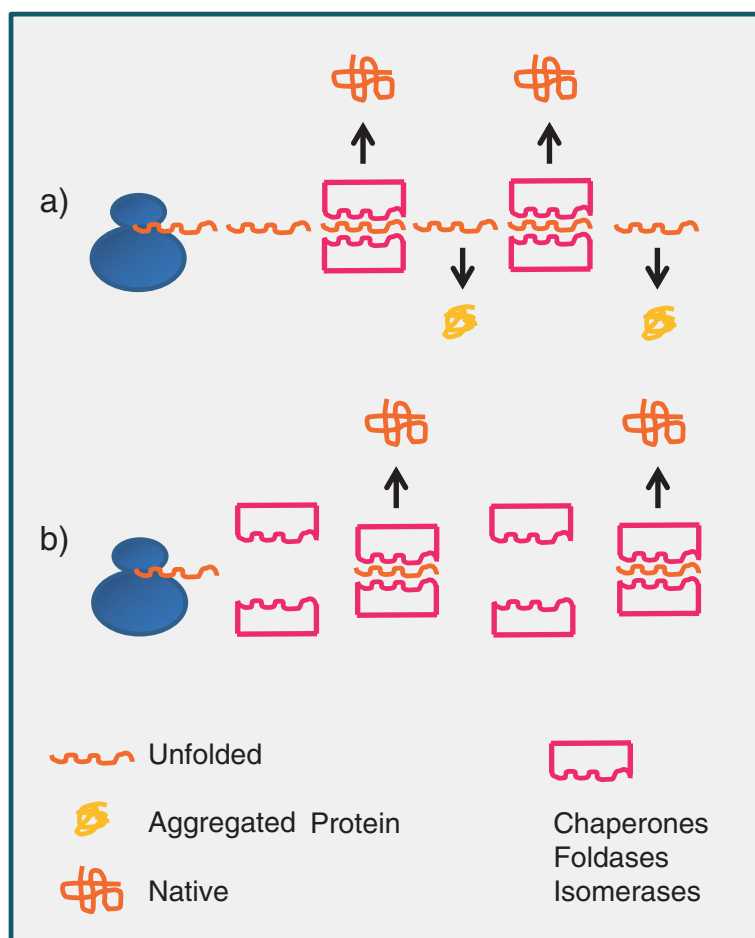


Figure 1 Expression rate controls the protein native folding in the cytoplasm. The recombinant polypeptide synthesized by the ribosomal machinery is mostly dependent on one/more chaperones and foldases to reach its native conformation. Disequilibrium between expression rate and folding capacity (a) will result in the accumulation of aggregates. This fate can be prevented by both reducing the expression rate of the recombinant protein and favoring the accumulation of chaperones and foldases (b).

towards the control of the quality of the expressed protein and the increase of its accumulation. To assess such hypothesis, the control of the polymerase accumulation was obtained by means of the tunable expression of its inhibitor T7 lysozyme. The results confirmed that, in a specific range, there was an inverse correlation between polymerase availability and target protein accumulation [13]. This output is in agreement with the very early supposition that reducing the expression rate could limit the cell metabolic burden and promote protein folding quality [15]. The same effect of improved protein yield under conditions of promoter down-regulation was proved more recently for the arabinose-inducible expression system P(BAD) [16]. A spontaneous and less efficient mutant allowed for lower cell toxicity and higher productivity of secreted membrane proteins.

From a practical point of view, the synthesis of fewer polypeptides per cell would result in more correctly folded proteins because the cell machinery would work more effectively, i.e. within its limits. This condition avoids the accumulation of unfolded polypeptides that tend to aggregate and to produce a toxic effect. Clearly, not all of the proteins have the same requirements. For instance, it has been estimated that only 85 of all the *E. coli* proteins have strict GroEL-dependence for completing their folding [17], whereas 180 strongly depend on DnaK and a minority need both chaperones [18]. Therefore, the production of recombinant proteins that are either strictly GroEL-dependent or can use any/no chaperone to reach their native structure will exert a totally different pressure on the host cell in terms of competition for limiting components of the folding machinery. Both repressor leakiness and high expression rates can amplify the risk of accumulating misfolded proteins. The tight regulation of the recombinant expression is consequently crucial to limit the metabolic burden and the toxicity. From this perspective, it is interesting to learn that leakiness can be reduced 10-fold by a single mutation in the core domain of the *lac* repressor and that such mutation does not prevent IPTG binding and recombinant protein production [19]. A complementary approach should investigate what chaperones are necessary for any particular protein to fold correctly in its original eukaryotic cell with the aim of overexpressing the prokaryotic homologues in the recombinant system. For instance, a systematic work has identified the substrates of human Hsp90 [20] and such proteins could probably benefit from the co-expression of the *E. coli* orthologue chaperone ClpB when produced in bacteria.

Minimal metabolic perturbation

The production of transcripts originating from a single gene copy of the target DNA integrated in the chromosome represents a solution for obtaining sufficient yields of heterologous protein and minimizing the metabolic burden

for the host cells. In such conditions, the cell physiology is not perturbed by the presence of high-copy number instable plasmids and by the massive production of antibiotic resistance enzymes [21]. Sufficient protein yields are assured by enabling highly effective transcription and translation. Such result was obtained for the *XylS/Pm* expression system by combining several stimulatory elements in a unique expression cassette comprising the promoter, the 5'-untranslated mRNA region, and the *XylS* regulator coding sequence [22]. The synergic effect of single point mutations in the sequences which regulate the transcription and translation steps allowed improving the final protein yields of 75 times. Both increased transcript production and its improved translation contributed to the final protein accumulation [23]. Interestingly, the optimal level of the stimulation was protein-specific and higher transcriptional levels did not automatically correspond to higher yields of functional protein. These results were explained as the effect of the post-translational folding machinery saturation, a bottleneck that caused protein misfolding and aggregation. Such example underlines again how the final yields always depend on the limiting step in the long chain of events that leads from transcription to correctly folded proteins.

Stoichiometric limitations

The question as to elevated expression rates can result in low yields finds its answer in the observation that proteins, in the majority of cases, depend on molecular factors to reach their functional structure. These factors can be as different as foldases, chaperones, interacting partners in the case of larger complexes, or membrane proteins involved in cell trafficking. The efficiency of the overall process leading to the production of a fully active protein relies on the availability of all the essential partners. Therefore, it is necessary to identify the critical molecular element involved in the folding process of a specific polypeptide that is the first to become limiting. This allows concentrating on its overproduction without causing the unaffordable energetic expense that the indiscriminate production of any foldase would cause to the cell. For instance, a specific chaperone overexpression is meaningful only when the recombinant protein requires their massive support to fold correctly, otherwise it represents a metabolic competition for the target recombinant protein [7]. In the case of secreted recombinant proteins, the Sec-translocon appears a major bottleneck because its carrier capacity can be easily saturated, as illustrated by Schlegel et al. [24]. As a consequence, unfolded protein intermediates accumulate in the cytoplasm and precipitate into aggregates (Figure 2). When toxic, these impair correct cell growth and division, leading to diminished biomass formation. The reduction of the recombinant translational activity can restore the physiological conditions and promote the periplasmic accumulation of processed and correctly

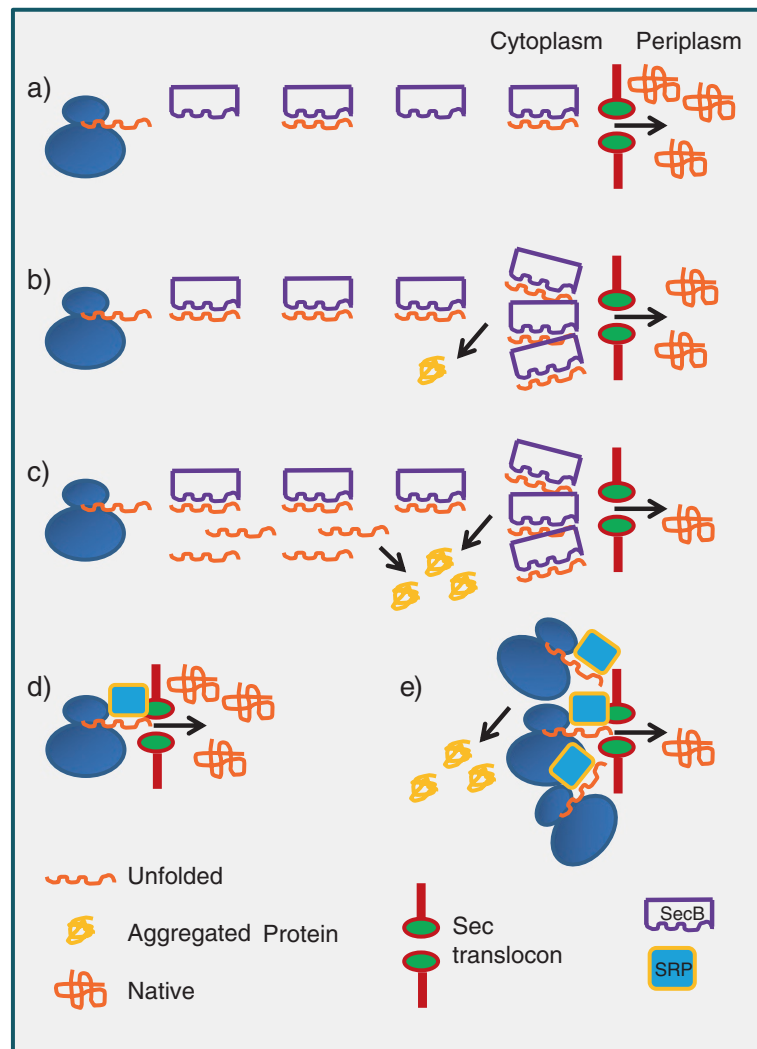


Figure 2 Expression rate determines the folding rate of secreted proteins. In optimal conditions (a), the newly synthesized polypeptides can be successfully delivered to the translocon by the SecB co-chaperone and transported for folding into the periplasm. At higher expression rates, the polypeptides will progressively aggregate because either the Sec translocon (b) or/and SecB (c) will become limiting. The co-translational SRP-based secretion route could overcome the SecB paucity (d), but shares the same export machinery and, therefore, any construct in excess to the translocon capacity will remain trapped in the cytoplasm (e).

folded heterologous proteins. This accomplishment recalls the results obtained with a library of vectors used to produce secreted proteins the components of which possess translational strengths covering a 10-fold range [25]. It was obtained by random alteration of the translation initiation regions of the signal sequence and the resulting vectors were tested in combination with a set of heterologous proteins. The results showed that the translation rate optimum is narrow and protein-specific. There is an upper limit above which the production drops due to the saturation of the system and a lower limit below which the protein synthesis machinery is not completely exploited.

In the past it was also demonstrated that secreting thermodynamically favored constructs by the co-translational SRP route could facilitate the production of functional proteins because this option avoided the SecB-dependent accumulation in the cytoplasm of partially folded, inactive or toxic intermediates [26,27]. Also the deletion of the gene responsible for the expression of the Trigger Factor (TF) chaperone contributed to limit the cell toxicity and to increase the accumulation of recombinant SRP-dependent proteins [16,28]. Since TF binds 1:1 the polypeptides emerging from the ribosomes, it was argued that TF competes with the signal recognition particle for the nascent proteins. Therefore, in the

absence of TF the cytoplasmic accumulation of non-productive protein intermediates would be drastically reduced. Nevertheless, too elevated expression rates would finally result in secretion impairment independently of the deletion strategy and of the chosen export route, since SecB and SRP share the same translocon unit [24] (Figure 2).

Summarizing, the improvement of the recombinant disulfide bond-dependent protein production relies on the capacity of calibrating the secretion and the translocon rates. The illustrated data emphasize also the limits of production strategies depending on the periplasmic accumulation of heterologous proteins. Even after successful export through the Sec-translocon, the proteins will undergo molecular crowding in the reduced periplasmic space and will be prone to aggregation [29].

The consequence is that the yields of periplasmically accumulated recombinant proteins remain low. A theoretical possibility exists of folding the (disulfide-independent) proteins in the bacterial cytoplasm, where the folding machinery is more effective, and then of exporting them into the periplasm by means of the Twin-arginine translocation pathway. The factors regulating this export route have been recently investigated [30] and it can have some biotechnological applications, such as the possibility of evaluating the correct folding of GFP-fused proteins [31]. Disulfide-bond dependent proteins can be forced to fold in the cytoplasm of mutated bacteria (oxidizing strains such as AD494 or Origami) in which the reducing pathways have been partially or totally impaired [32]. Others have proposed the secretion of such proteins into the medium [33,34]. Nevertheless, controlling the parameters that regulate the protein accumulation in the medium is cumbersome and the conditions can critically change when productions are scaled-up due to different leakage rates [34]. Concerning the “oxidizing strains”, the results are often deceiving in terms of yield [32,35,36], probably because the non-physiological conditions imposed to the cells slow down their growth. The overproduction of isomerases can help the correct folding of disulfide bond-dependent proteins. Nevertheless, DsbC isomerase used as a fusion tag is not sufficient to obtain reliable yields of functional target protein [32,36], even though it is statistically more effective than other carriers [36]. So far, the most innovative and complete approach for producing functional disulfide-rich proteins remains the model proposed by the Ruddock's group that enables oxidation and assisted folding by co-expressing both sulfhydryl oxidase and disulfide isomerase (DsbC or PDI) in the cytoplasm of wild type bacteria [37]. In the absence of sulfhydryl oxidase, most of the disulfide-dependent fusion proteins fused to DsbC that accumulated in the cytoplasm of wild type bacteria formed soluble aggregates [36] in which probably the (folded) carrier keeps in solution the aggregate target protein. According to the model initially proposed by

Nominé et al. [38] (Figure 3), such soluble aggregates of fusion proteins are micelles with a core composed by tightly aggregated target proteins and an external hydrophilic layer formed of folded carriers. The amphiphilic nature of these structures has been recently harnessed for producing nanopills with bacterial binding capacity [39]. In the proposed model, the micelles are formed by conjugated proteins in which the external layer has the capacity to bind specifically the biological target. The cargo, composed by aggregated but partially active polypeptides, is then released by means of controlled enzymatic cleavage.

Tuning protein aggregates

Aggregate formation is a constant of recombinant protein production. In contrast to previous hypotheses, we know now that aggregated proteins can conserve a consistent biological activity. Consequently, it is meaningful to study the aggregation process *in vivo* and to understand how chemical-physical factors can tune their characteristics [40] to improve their biotechnological value. Two opposite approaches have been proposed to cope with recombinant protein aggregation: try to limit it or to force the process and accumulate useful (functional and non-toxic) aggregates. The reversible dynamic of the aggregation-disaggregation process has been forced to improve the yields of soluble recombinant proteins by stimulating the bacterial cells to dissolve their inclusion bodies *in vivo* [7,41]. As an alternative, aggregation has been exploited as a valuable means for accumulating functional recombinant protein in the inclusion bodies and simplifying its purification for biotechnological applications [42] (Figure 3). From this perspective, the identification of genetic backgrounds favorable for the production of particularly dense inclusion bodies is justified by the interest in improving the protein processing, obtaining higher titers in fermentation and simplified purification steps due to the reduced volumes and faster sedimentation. Pandey et al. [43] induced the production of recombinant model proteins in a bacterial library of randomly obtained knock-out mutants and succeeded in isolating clones producing more/denser inclusion bodies by means of percoll density gradient centrifugation. Unluckily, the authors did not evaluate whether the protein packed into denser inclusion bodies had structural and functional characteristics comparable to the control material or have lost them, as it has been observed when lipase A was fused to the self-assembling peptide ELK16 to improve its precipitation [44].

The biological value of proteins trapped in the inclusion bodies can be related to their physical structure rather than to their functional features. For instance, controlling the fermentation parameters allowed for the obtaining of inclusion bodies with variable geometries. When used to

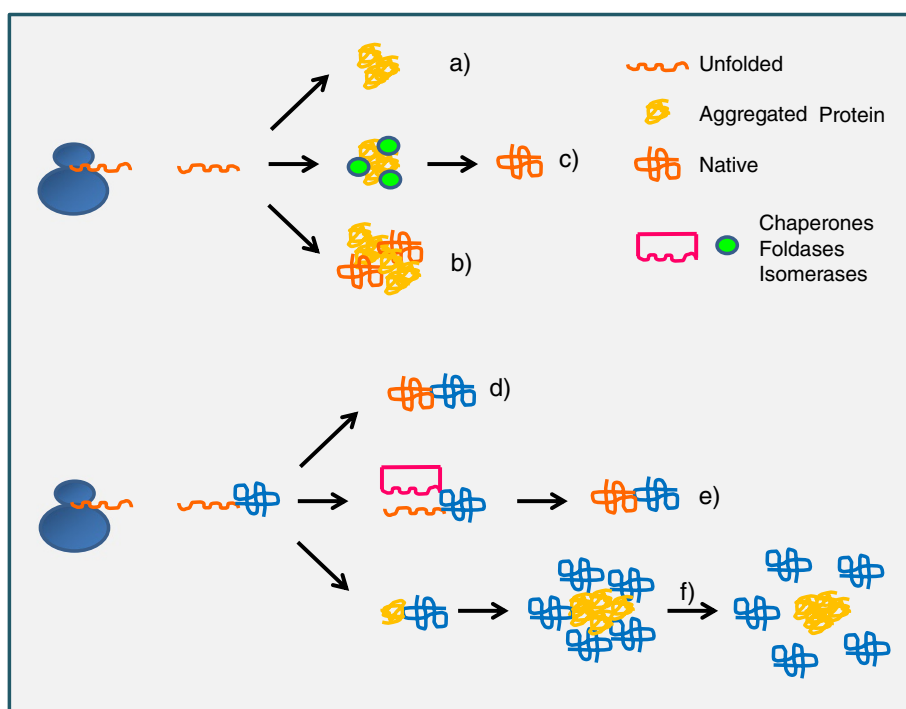


Figure 3 Characteristics of the recombinant protein aggregates. Recombinant polypeptides can aggregate in the inclusion bodies that are either completely inactive (a) or trap different amounts of active protein (b). Furthermore, the polypeptides can be released from the inclusion bodies and a network of chaperones and foldases/isomerases can convert misfolded constructs into native proteins (c). Highly soluble carriers fused to the target proteins can simplify the purification step (d) or keep in solution thermodynamically unfavorable partners for the time necessary to the folding machinery to complete its work (e). Nevertheless, the partner can finally fail to fold, remaining (partially) misfolded. Such misfolded domains exposing hydrophobic residues tend to form micelles that are kept in solution by the soluble carriers that constitute an external hydrophilic layer (f). Such soluble aggregates can be used to deliver the cargo in vivo and the carriers can be cleaved off when the biologically relevant target is reached.

treat surfaces, they facilitate the adhesion of eukaryotic cells and promote tissue culture and differentiation [45].

Conclusions

Successful recombinant protein production is often strictly necessary to accumulate enough material for basic research studies as well as for biotechnological and clinical applications. A stimulating article [46] indicates that research stops where no reagents are available. Specifically, the authors remark that proteins that have a clear genetic link with human diseases, but for which reliable antibodies and chemical inhibitors are not accessible, are systematically less studied than similar proteins for which reagents are available. It would be interesting to analyze whether the reagent availability is related to the ease in obtaining functional and natively structured recombinant protein, since it can be expected that missing this step would represent a critical bottleneck. The knowledge accrued in the last 20 years conceives recombinant protein production as a long chain of successive events. Any of them can be limiting or can become the limiting after another obstacle has been removed. Therefore, the process optimization can be seen as a progressive removal of impairments: for a specific

construct, too many transcripts blocks the translational machinery because there are not enough tRNA molecules for a specific codon, but in a strain overexpressing that tRNA the accumulated polypeptides can fail to fold due to DnaK shortage, and so on. Now we are probably approaching the comprehension of each single step involved in the whole process. For instance, several proteins that can improve protein secretion have been identified, the mechanisms by which low expression rate can lead to higher protein yields have been described in details, bacterial strains and vectors calibrated for working at physiological conditions are now available. The future efforts should be dedicated to a rational method to identify quickly and effectively the process weakness for any given protein to select the suitable counteraction. Combinatorial platforms exist for comparing experimentally protein production conditions by varying factors as different as bacterial strains, growth temperature, osmolyte concentration, fusion to different tags, promoter and transcription factor strength, mRNA stability and translation efficiency, chaperone co-expression, or signal peptide sequence [5,7,9,47-52]. By contrast, predictive models are still missing. The lack of searchable data might be

one of the reasons and, consequently, it would be beneficial for the community to make the effort of collecting and annotating experimental data in a standardized manner to facilitate their metadata analysis. Certainly, this effort requires the input of extra energy and can be perceived as a limitation of the research freedom, both reasons for the scarce success of previous initiatives [53]. Nevertheless, such information is necessary to confirm some intuitions and to support statistically the cause-effect correlations evidenced by the already available results. The final accomplishment would be the development of holistic models able to consider all the elements that contribute to the protein production process.

Competing interests

The author declares that he has no competing interests.

Acknowledgments

This work was supported in part by the Creative Core AHA-MOMENT grant from Slovene Ministry of Science and European Fund for Regional Development and by Cross-Border Cooperation Programme Italy-Slovenia 2007–2013, (Project PROTEO, Code N. CB166). The author wishes to thank Alicja Gruszka for her comments.

Received: 25 September 2013 Accepted: 30 October 2013

Published: 1 November 2013

References

- Bentley WE, Mirjalili N, Andersen DC, Davis RH, Kompala DS: Plasmid-encoded protein: the principal factor in the "metabolic burden" associated with recombinant bacteria. *Biotechnol Bioeng* 1990, **35**:668–681.
- Stueber D, Bujard H: Transcription from efficient promoters can interfere with plasmid replication and diminish expression of plasmid specified genes. *EMBO J* 1982, **1**:1399–1404.
- Battistoni A, Carri MT, Mazzetti AP, Rotilio G: Temperature-dependent protein folding in vivo – lower growth temperature increases yield of two genetic variants of *Xenopus laevis* Cu, Zn superoxide dismutase in *Escherichia coli*. *Biochem Biophys Res Commun* 1992, **186**:1339–1344.
- Popov M, Petrov S, Nacheva G, Ivanov I, Reichl U: Effects of a recombinant gene expression on ColE1-like plasmid segregation in *Escherichia coli*. *BMC Biotechnol* 2011, **11**:18.
- de Marco A, Vigh L, Diamant S, Goloubinoff P: Native folding of aggregation-prone recombinant proteins in *Escherichia coli* by osmolytes, plasmid- or benzyl alcohol-overexpressed molecular chaperones. *Cell Stress Chaperones* 2005, **10**:329–339.
- Goloubinoff P, Gatenby AA, Lorimer GH: GroE heat-shock proteins promote assembling of foreign prokaryotic ribulose biphosphate carboxylase oligomers in *Escherichia coli*. *Nature* 1989, **337**:44–47.
- de Marco A, Deuerling E, Mogk A, Tomoyasu T, Bukau B: Chaperone-based procedure to increase yields of soluble recombinant proteins produced in *E. coli*. *BMC Biotechnol* 2007, **7**:32.
- Nakashima N, Tamura T: Conditional gene silencing of multiple genes with antisense RNAs and generation of a mutator strain of *Escherichia coli*. *Nucleic Acids Res* 2009, **37**:e103.
- Makino T, Skretas G, Kang T-H, Georgiou G: Comprehensive engineering of *Escherichia coli* for enhanced expression of IgG antibodies. *Metab Eng* 2011, **13**:241–251.
- Makino T, Skretas G, Georgiou G: Strain engineering for improved expression of recombinant proteins in bacteria. *Microb Cell Fact* 2011, **10**:32.
- Skretas G, Makino T, Varadarajan N, Pogson M, Georgiou G: Multi-copy genes that enhance the yield of mammalian G protein-coupled receptors in *Escherichia coli*. *Metab Eng* 2012, **14**:591–602.
- 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 level. *J Mol Biol* 1996, **260**:289–298.
- Wagner S, Klepsch MM, Schlegel S, Appel A, Draheim R, Tarry M, Högbom M, van Wijk KJ, Slotboom DJ, Persson JO, de Gier J-W: Tuning *Escherichia coli* for membrane protein overexpression. *Proc Natl Acad Sci U S A* 2008, **105**:14371–14376.
- 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**:648–659.
- Hoffmann F, Rinas U: Stress induced by recombinant protein production in *Escherichia coli*. *Adv Biochem Eng Biotechnol* 2004, **89**:73–92.
- Nannenga BL, Baneyx F: Enhanced expression of membrane proteins in *E. coli* with a P(BAD) promoter mutant: synergies with chaperone pathway engineering strategies. *Microb Cell Fact* 2011, **10**:105.
- Kerner MJ, Naylor DJ, Ishihama Y, Maier T, Chang HC, Stines AP, Georgopoulos C, Frishman D, Hayer-Hartl M, Mann M, Hartl FU: Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell* 2005, **122**:209–220.
- Calloni G, Chen T, Schermann SM, Chang H, Genevaux P, Agostini F, Tartaglia GG, Hayer-Hartl M, Hartl FU: DnaK functions as a central hub in the *E. coli* chaperone network. *Cell Reprogram* 2012, **1**:251–264.
- Gatti-Lafranconi P, Dijkman WP, Devenish SRA, Hollfelder F: A single mutation in the core domain of the lac repressor reduces leakiness. *Microb Cell Fact* 2013, **12**:67.
- Taipale M, Jarosz DF, Lindquist S: Hsp90 at the hub of protein homeostasis: emerging mechanistic insights. *Nat Rev Mol Cell Biol* 2010, **11**:515–528.
- Vethanayagam JG, Flower AM: Decreased gene expression from T7 promoters may be due to impaired production of active T7 RNA polymerase. *Microb Cell Fact* 2005, **4**:3.
- Zwick F, Lale R, Valla S: Strong stimulation of recombinant protein production in *Escherichia coli* by combining stimulatory control elements in an expression cassette. *Microb Cell Fact* 2012, **11**:133.
- Balzer S, Kucharova V, Megerle J, Lale R, Brautaset T, Valla S: A comparative analysis of the proprieties of regulated promoter systems commonly used for recombinant gene expression in *Escherichia coli*. *Microb Cell Fact* 2013, **12**:26.
- Schlegel S, Rujas E, Ytterberg AJ, Zubarev RA, Luirink J, de Gier J-W: Optimizing heterologous protein production in the periplasm of *E. coli* by regulating gene expression levels. *Microb Cell Fact* 2013, **12**:24.
- Simmons LC, Yansura DG: Translational level is a critical factor for the secretion of heterologous proteins in *Escherichia coli*. *Nat Biotechnol* 1996, **14**:629–634.
- Steiner D, Forrer P, Stumpp MT, Plückthun A: Signal sequences directing cotranslational translocation expands the range of proteins amenable to phage display. *Nat Biotechnol* 2006, **24**:823–831.
- Monegal A, Olichon A, Bery N, Filleron T, Favre G, de Marco A: Single heavy chain antibodies with VH hallmarks are positively selected during panning of llama (*Lama glama*) naïve libraries. *Develop Comp Immunol* 2012, **36**:150–156.
- Nannenga BL, Baneyx F: Reprogramming chaperone pathways to improve membrane protein expression in *Escherichia coli*. *Protein Sci* 2011, **20**:1411–1420.
- Hunke S, Betton JM: Temperature effect on inclusion body formation and stress response in the periplasm of *Escherichia coli*. *Mol Microbiol* 2003, **50**:1579–1589.
- Lee SJ, Park EH, Kim YO, Nam BH, Kim DG: Control of total GFP expression by alterations of the 3' region nucleotide sequence. *Microb Cell Fact* 2013, **12**:68.
- Fisher AC, Rocco MA, DeLisa MP: Genetic selection of solubility-enhanced proteins using the twin-arginine translocation system. *Meth Mol Biol* 2011, **705**:53–67.
- de Marco A: Strategies for successful recombinant expression of disulfide bond-dependent proteins in *Escherichia coli*. *Microb Cell Fact* 2009, **8**:26.
- van der Wal FJ, Koningstein G, ten Hagen CM, Oudega B, Luirink J: Optimization of bacteriocin release protein (BRP)-mediated protein release by *Escherichia coli*: random mutagenesis of the pCloDF13-derived BRP gene to uncouple lethality and quasi-lysis from protein release. *Appl Environ Microbiol* 1998, **64**:392–398.
- Ukkonen K, Veijola J, Vasala A, Neubauer P: Effect of culture medium. Host strain and oxygen transfer on recombinant Fab antibody fragment yield and leakage to medium in shaken *E. coli* cultures. *Microb Cell Fact* 2013, **12**:73.
- de Marco A: Recent contributions in the field of the recombinant expression of disulfide bonded protein in bacteria. *Microb Cell Fact* 2012, **11**:129.
- Nozach H, Fruchart-Gaillard C, Fenaille F, Beau F, Pereira Ramos OH, Douzi B, Saez NJ, Moutiez M, Servent D, Gondry M, Thai R, Cuniasse P, Vincenzelli R,

- Dive V: High throughput screening identifies disulfide isomerase DsbC as a very efficient partner for recombinant expression of small disulfide-rich proteins in *E. coli*. *Microb Cell Fact* 2013, **12**:37.
37. Nguyen VD, Hatahet F, Salo KE, Enlund E, Zhang C, Ruddock LW: Pre-expression of a sulfhydryl-oxidase significantly increases the yields of eukaryotic disulfide bond containing proteins expressed in the cytoplasm of *E. coli*. *Microb Cell Fact* 2011, **10**:1.
 38. Nominé Y, Ristriani T, Laurent C, Lefèvre JF, Weiss E, Travé G: Formation of soluble inclusion bodies by hpv e6 oncoprotein fused to maltose-binding protein. *Protein Expr Purif* 2001, **23**:22–32.
 39. Talafova K, Hrabarova E, Chorvat D, Nahalka J: Bacterial inclusion bodies as potential synthetic devices for pathogen recognition and a therapeutic substance release. *Microb Cell Fact* 2013, **12**:16.
 40. Ami D, Natalello A, Lotti M, Doglia SM: Why and how protein aggregation has to be studied in vivo. *Microb Cell Fact* 2013, **12**:17.
 41. Carrió MM, Villaverde A: Protein aggregation as bacterial inclusion bodies is reversible. *FEBS Lett* 2001, **489**:29–33.
 42. Villaverde A, García-Fruitos E, Rinas U, Seras-Franzoso J, Kosoy A, Corchero J, Vazquez E: Packaging protein drugs as bacterial inclusion bodies for therapeutic applications. *Microb Cell Fact* 2012, **11**:76.
 43. Pandey N, Sachan A, Chen Q, Ruebling-Jass K, Bhalla R, Panguluri KK, Rouviere PE, Cheng Q: Screening and identification of genetic loci involved in producing more/denser inclusion bodies in *Escherichia coli*. *Microb Cell Fact* 2013, **12**:43.
 44. Wu W, Xing L, Zhou B, Lin Z: Active protein aggregates induced by terminally attached self-assembling peptide ELK16 in *Escherichia coli*. *Microb Cell Fact* 2011, **10**:9.
 45. García-Fruitos E, Seras-Franzoso J, Vazquez E, Villaverde A: Tunable geometry of bacterial inclusion bodies as substrate materials for tissue engineering. *Nanotechnology* 2010, **21**:205101.
 46. Edwards AM, Isserlin R, Bader GD, Frye SV, Willson TM, Yu FH: Too many roads not taken. *Nature* 2011, **470**:163–165.
 47. Aune V, Bakke I, Drabløs F, Lale F, Brautaset T, Valla S: Direct evolution of the transcription factor XylS for the development of improved expression systems. *Microb Biotechnol* 2010, **3**:38–47.
 48. Bandmann N, Nygren P-Å: Combinatorial expression vector engineering for tuning of recombinant protein production in *Escherichia coli*. *Nucleic Acids Res* 2007, **35**:e32.
 49. de Valdivia El G, Isaksson LA: A codon window in mRNA downstream of the initiation codon where NCG codons give strongly reduced gene expression in *Escherichia coli*. *Nucleic Acids Res* 2004, **32**:5198–5205.
 50. Ahn JH, Keum JW, Kim DM: High-throughput, combinatorial engineering of initial codons for tunable expression of recombinant proteins. *J Proteome Res* 2008, **7**:2107–2113.
 51. Heggeset TMB, Kucharova V, Nærdal I, Valla S, Sletta H, Ellingsen TE, Brautaset T: Combinatorial mutagenesis and selection of improved signal sequences and their application for high-level production of translocated heterologous proteins in *Escherichia coli*. *Appl Environ Microbiol* 2013, **79**:559–568.
 52. Vincentelli R, Cimino A, Geerlof A, Kubo A, Satou Y, Cambillau C: High-throughput protein expression screening and purification in *Escherichia coli*. *Methods* 2011, **55**:65–72.
 53. Buckle AM, Bate MA, Androulakis S, Cinquanta M, Basquin J, Bonneau F, Chatterjee DK, Cittaro D, Gräslund S, Gruszka A, Page R, Suppmann S, Wheeler JX, Agostini D, Taussig M, Taylor CF, Bottomley SP, Villaverde A, de Marco A: Recombinant protein quality evaluation: proposal for a minimal information standard. *Stand Genomic Sci* 2011, **5**:195–197.

doi:10.1186/1475-2859-12-101

Cite this article as: de Marco: Recombinant polypeptide production in *E. coli*: towards a rational approach to improve the yields of functional proteins. *Microbial Cell Factories* 2013 **12**:101.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

