# Bicistronic mRNAs to enhance membrane protein overexpression

Marino, Jacopo ; Hohl, Michael ; Seeger, Markus A ; Zerbe, Oliver ; Geertsma, Eric

DOI: https://doi.org/10.1016/j.jmb.2014.11.002

# Bicistronic mRNAs to enhance membrane protein overexpression

Jacopo Marino[a*], Michael Hohl[b], Markus A. Seeger[b],

Oliver Zerbe[a] and Eric R. Geertsma[c#]

[a] Department of Chemistry, University of Zurich, Zurich, Switzerland.

[b] Institute of Medical Microbiology, University of Zurich, Zurich, Switzerland.

[c] Institute of Biochemistry, Goethe-University Frankfurt, Frankfurt am Main, Germany

* present address: Gene Center, Department of Chemistry and Biochemistry, University of Munich, Germany

[#] corresponding author. E-mail: geertsma@em.uni-frankfurt.de. Institute of Biochemistry, Biocenter N200/1.08, Goethe-University Frankfurt, Max-von-Laue-Str. 9, D-60438 Frankfurt/M., Germany; phone: +49 (0)69-798-29255; fax: +49 (0)69-798-29244.

**Keywords**

transcriptional fusion, ybeL, protein folding, structural biology, mRNA secondary structure

**Abbreviations**

| | |
|---|---|
| ORF | Open Reading Frame |
| SD | Shine-Dalgarno |
| GFP | Green Fluorescent Protein |
| *E. coli* | *Escherichia coli* |
| DDM | n-Dodecyl β-D-Maltopyranoside |
| IMAC | Immobilized Metal Affinity Chromatography |
| FSEC | Fluorescent Size Exclusion Chromatography |

**Abstract**

Functional overexpression of membrane proteins is essential for their structural and functional characterization. However, functional overexpression is often difficult to achieve, and frequently either no expression or expression as misfolded aggregates is observed. We present an approach for improving the functional overexpression of membrane proteins in *E. coli* using transcriptional fusions. The method involves the use of a small additional RNA sequence upstream to the RNA sequence of the target membrane protein and results in the production of a bicistronic mRNA. In contrast to the common approach of translational fusions to enhance protein expression, transcriptional fusions do not require protease treatment and subsequent removal of the fusion protein. Using this strategy we observed improvements in the quantity and/or the quality of the produced material for several membrane proteins to levels compatible with structural studies. Our analysis revealed that translation of the upstream RNA sequence was not essential for increased expression. Rather, the sequence itself had a large impact on protein yields, suggesting that alternative folding of the transcript was responsible for the observed effect.

## Introduction

The heterologous overexpression of well-folded protein remains one of the main bottlenecks for structural biology of membrane proteins. Despite our expanding knowledge on the factors involved, establishing and optimizing membrane protein overexpression is currently still a largely empirical process. General strategies used for *E. coli* involve the screening of several parameters such as expression hosts, promoters, growth conditions, coding sequences, tags and fusion proteins [1].

The use of fusion proteins for expression in *E. coli* was extensively explored for soluble proteins and later successfully used for membrane proteins as well. In general, fusion proteins represent independent folding domains that do not interact with their fused protein target. Fusion proteins are mostly hydrophilic and vary in size from 10 to 60 kDa [2–4]. They can be fused either to the N- or C-terminus of the target protein. The latter is often employed to monitor the folding state of the target proteins using a fusion partner that allows facile 'read-out', such as GFP [5, 6] or antibiotic resistance markers [7, 8]. Fusions to the N-terminus of the target protein are generally used for their favourable properties in enhancing expression levels [3] and decreasing proteolysis [4]. Both types of fusions can be used to detect and even purify the target protein. Commonly used fusion proteins are MBP, the maltoside binding protein from *E. coli* [9,10], Trx (*E. coli* thioredoxin [11]), SUMO, (small ubiquitin-like modifier from *S. cerevisiae*, [12]), and Mistic (membrane integrating sequence for translation of inner membrane protein constructs from *B. subtilis*) [13]. In addition, novel fusion partners are constantly introduced, such as YnaI and YbeL, two small hydrophilic proteins from *E. coli* that have been recently reported to enhance the production of functional membrane proteins in *E.*

*coli* [14]. Although N-terminal fusions can be important for targeting the nascent membrane protein to the membrane for insertion or for decreasing its susceptibility to proteolysis, the often-observed ability of N-terminal sequences to enhance expression levels is often attributed to their high initiation rate of translation due to favourable mRNA secondary structure [15, 16].

Nevertheless, the application of fusion proteins should be done with some caution. The fusion protein can affect the folding state of the target protein, e.g., if too short linkers are used to connect both proteins [10]. Furthermore, the activity of the target protein can be compromised by the fusion [17–19], and the presence of a flexible fusion protein can interfere with crystallization [20]. A solution to the latter issues is to perform proteolytic cleavage and removal of the fusion protein, but efficient cleavage cannot always be achieved [21].

A potential alternative to translational fusions could be transcriptional fusions. Such fusions are common in prokaryotes and result in transcripts containing multiple open reading frames (ORFs) that are translated separately. The genomic sequence of *E. coli* shows that more than 25% of all operons are polycistronic and approximately 9% of the ORFs have a start codon overlapping with the stop codon of the preceding ORF [22]. This system allows simultaneous and stoichiometry-regulated expression of proteins that are functionally related. Examples of such bacterial operons are found in multiprotein complexes as the ribosome [23], the ATP synthase [24], the anthranilate and tryptophan synthases [25], and photosynthetic complexes [26].

Translation of ORFs that are preceded by another ORF can result from either re-initiation or *de novo* initiation [27]. In re-initiation, the ribosome completing translation of

the upstream ORF remains bound to the transcript. Translation of the downstream ORF is initiated when a start codon nearby or overlapping with the stop codon of the preceding ORF is encountered. The presence of a Shine-Dalgarno (SD) sequence upstream to the start codon enhances the re-initiation efficiency, but is not strictly necessary [27–29]. Translation of the upstream ORF is suggested to be necessary as the helicase activity of the ribosome opens up secondary structures and makes elements as the SD sequence or the start codon accessible [27–30]. For *de novo* initiation, a novel 30S ribosomal subunit binds directly to the SD preceding a secondary ORF on the transcript. This mode of initiation requires an exposed SD not involved in mRNA secondary structure [27].

As the beneficial effects of translational fusions on protein expression are mostly attributed to more favourable secondary structures in the transcript, we investigated whether similar beneficial effects could also be achieved using transcriptional fusions, which would avoid enzymatic cleavage to liberate the target protein. Here we compare the effect on membrane protein overexpression for translational and transcriptional fusions of genes coding for the fusion proteins Mistic, SUMO and YbeL, and demonstrate an application for transcriptional fusions.

## Results

### Overexpression of membrane proteins using transcriptional fusions

To compare translational and transcriptional fusions four evolutionary unrelated polytopic membrane proteins (MP) from *E. coli* were selected that could be functionally expressed without N-terminal fusions and which possessed a cytoplasmic C-terminus [31]. The latter allowed the use of C-terminal GFP fusions to rapidly assess the ratio between well-folded and misfolded protein by gel electrophoresis [6]. Though the fluorescence of C-terminal GFP fusion proteins is an indirect indication for correct folding, a very high correlation between GFP fluorescence and functional and stably folded membrane proteins is generally observed [5, 6, 32]. ORFs coding for the glycine betaine/proline ABC transporter permease ProW [5, 33], the mechanosensitive channel of large conductance MscL [33, 34], the lactose permease LacY [6, 35] and the glutamate transporter GltP [6] were placed under the control of the arabinose promoter [36] and fused to the sequence coding for a C-terminal GFP in a diverse set of universal high-throughput cloning vectors (Fig.1) [37]. These vectors allowed direct expression of the target ORF-GFP alone, with translational fusions to Mistic, SUMO or YbeL, or with transcriptional fusions to *mstX*, *sumo*, or *ybeL* (Fig. 1). Translation using the first two vector sets was initiated using the SD of the vector. Vectors for transcriptional fusions had overlapping stop and start codons, and were either preceded by a SD (for *mstX* and *ybeL*) or a sequence devoid of SD (for *sumo*). The *mstX* gene is naturally part of a two-gene operon [38] and contains a SD sequence that is necessary for efficient translation of the downstream ORF *yugO* [39]. Instead, *ybeL* is not part of an operon, but possesses a GA-rich sequence in the vicinity of its stop codon that is predicted to

function as a SD. Predicted translation efficiencies (RBS-Designer software [40]) were in the same range for all vectors except for the plasmid for transcriptional fusion to *sumo* that did not contain an additional SD preceding the target ORF (Fig. 1). The importance of the SD present on *mstX* and *ybeL* for achieving translation of the downstream ORF was further investigated by creating vectors devoid of the SD sequences, and as well by introducing a SD sequence into *sumo* (Fig. 1, Suppl. 1).

**Improved membrane protein folding quality using transcriptional fusions**

All target proteins in the context of the different vectors were expressed in *E. coli* following the same growth and induction protocol. Expression levels were determined by the combination of *in gel* GFP fluorescence and immunoblotting (Fig. 2). In addition, the GFP fluorescence in whole cells was determined (Suppl. Fig. 1A). All four constructs yielded well-folded membrane proteins without a translation or transcriptional fusion, as judged from the observed GFP fluorescence bands corresponding to the expected sizes of the respective MP-GFP fusions (Fig. 2; left panels). However, large fractions of the produced proteins were misfolded, as indicated by the strong non-fluorescent band migrating 10-15 kDa higher than the fluorescent band (Fig.2; right panels). The presence of N-terminal translational fusion partners overall decreased the functional expression levels for ProW, LacY and GltP, while for MscL increased levels were observed.

Transcriptional fusion with *sumo* did not lead to detectable expression levels for all tested proteins, in contrast to fusions with *mstX* and *ybeL*. The insertion of a SD into the *sumo* DNA sequence was able to alleviate the failing expression, though expression levels remained very low (Suppl. Fig. 1A). Transcriptional fusions to *mstX* or *ybeL* led to a modest increase in functionally expressed protein for ProW and LacY as compared to

expression without fusion (Suppl. Table 4A). Remarkably, the increase in the amount of well-folded material was not paralleled by a similar increase in the amount of misfolded protein. This is most clearly observed for the *ybeL* transcriptional fusions, which led to a significant reduction in the fraction of misfolded membrane protein.

**Translational and transcriptional fusions do not alter expression profiles**

It is well established that very high membrane protein expression rates can exceed the capacity of the cell to properly target, insert and/or fold membrane proteins [6, 41]. This leads to an increased fraction of misfolded membrane proteins at high expression rates. To exclude that the decreased levels of misfolded protein observed for transcriptional fusions to *mstX* and *ybeL* resulted from a decreased transcript concentration and consequently lower expression rates, we tuned the promoter strength by varying the arabinose concentration during induction [6, 36]. To identify the optimal inducer concentration for the four membrane proteins under study, the expression levels of all constructs were determined by measuring GFP fluorescence in whole cells at six different arabinose concentrations (Suppl. Fig. 2). Curves indicate that the relative expressions levels for all constructs at different arabinose concentrations are very similar, with an optimum around $2x10^{-2}$ % (w/v) arabinose. Therefore, the improved ratios between well-folded and misfolded membrane proteins in the fusion constructs cannot be attributed to inappropriately chosen inducer concentrations, but rather represent an intrinsic feature of the constructs.

**Additional assessment of the protein folding quality**

Though GFP fluorescence is a strong indicator of correct folding of the preceding membrane protein, we additionally verified the quality of the GFP-fusion proteins by

determining their *n*-dodecyl β-D-maltoside (DDM) extraction efficiency (Suppl. Table 2). We have chosen DDM as this is a mild detergent that preferentially solubilizes correctly folded MPs [6]. Protein produced using either direct expression or transcriptional fusions could be extracted efficiently (between 75-90%). We noticed that poor extractability was achieved when the four MPs were expressed as translational fusions with Mistic (between 25-50%), while translational fusions with SUMO or YbeL could be extracted with intermediate efficiencies (between 50 and 80%). These reduced solubilization efficiencies suggest that the stability of the proteins produced by translational fusions is decreased.

All samples solubilized by DDM were subsequently analyzed by fluorescence size-exclusion chromatography (FSEC) as an additional means to verify their folding state (Suppl. Fig. 3). All proteins that had intermediate to high solubilization efficiencies were monodisperse (ProW and LacY) or oligodisperse (GltP and MscL). Except for MscL, all proteins showed similar migration profiles as the proteins produced using direct expression. For MscL, direct expression resulted in little detergent-solubilized fluorescent protein that displayed non-uniform migration during FSEC. Increased quantities of more uniform MscL could be obtained only using translational or transcriptional fusions.

**Application of transcriptional fusions to difficult-to-express membrane proteins**

The modest increase in well-folded material using transcriptional fusions with *mstX* and *ybeL* for membrane proteins that had previously been demonstrated to express well [31] made us ask whether this methodology could be applied to membrane proteins that either showed overall low expression levels or large proportions of misfolded material.

We selected *E. coli* AraH and NarK, *E. faecalis* EF583/584, and the human Y4 GPCR based on their reported low expression levels and cytoplasmic carboxy-terminus. AraH is a component of the *E. coli* L-arabinose ABC transporter permease ($AraFG_2H$) and NarK is a nitrite/nitrate antiporter. Both were shown to express poorly [42], though overexpression of the latter proved possible later upon small modifications in the expression conditions [43]. The poorly expressing ABC transporter EF583/584, encoded by two overlapping genes (*EF0583/0584*), was selected based on a high-throughput expression screening which lead to the crystallization of the heterodimeric ABC transporter TM287/288 [44]. As an example of a difficult-to-express eukaryotic membrane protein, we have used the human Y4 receptor, a G-protein coupled receptor (GPCR) that could only be produced with modest expression levels in *E. coli* using N-terminal fusions [45].

Production levels of these four targets, expressed directly or via transcriptional fusion, were determined by the combination of *in gel* GFP fluorescence and immunoblotting (Fig. 3) and whole cell GFP fluorescence (Suppl. Fig. 1B). The induction time for expression of AraH, NarK, EF583/584 and Y4 GPCR was extended by approximately 12 h compared to the well-expressed proteins ProW, MscL, LacY and GltP in order to obtain significant GFP signals.

In contrast to previous observations [42], we could achieve expression of AraH when directly expressed (Fig. 3). Furthermore, the expression conditions applied here also allowed direct expression of NarK. These differences are likely to results from the use of a milder promoter system. Nevertheless, expression levels of all four target proteins were low or resulted in large amounts of misfolded material.

The use of transcriptional fusions to *mstX* or *ybeL* improved the expression levels for all targets, though expression levels of the Y4 GPCR remained very poor (Fig. 3, Suppl. Table 4B). Noteworthy, the low but significant fluorescence observed for the Y4 GPCR in whole cells (Suppl. Fig. 1B) did not translate to a similar in gel GFP signal, neither for full-length Y4-GFP nor for breakdown products or higher molecular weight aggregates (Fig. 3). The cause for this discrepancy is presently unclear. As the in gel GFP fluorescence allows specific assignment of a signal to the full-length GFP fusion protein, in contrast to measurements in whole cells that can only detect bulk fluorescence, we interpret the relatively high GFP fluorescence in whole cells as an overestimation of the Y4 GPCR levels.

Overall, transcriptional fusions to *ybeL* proved best: not only were increased levels of the fluorescent species observed, but these were accompanied by decreased levels of misfolded protein as well, as observed previously. Moreover, the fraction of the folded material could be quantitatively extracted by the mild detergent DDM (Suppl. Fig. 4, Suppl. Table 3).

**Translation of the first ORF is not required for improved expression**

To determine whether translation of the first ORF was required for the beneficial effects of transcriptional fusions, we disrupted the start codons in *mstX* and *ybeL*. Disruption was done by the insertion of a single base in the respective start codons making translation highly unlikely. No major changes in the expression levels of the target proteins were observed upon abolishment of the MstX and YbeL expression, except for an increased production of misfolded Y4 GPCR upon removal of the MstX start codons (Fig. 3). These experiments suggest that the increased expression of the respective

membrane proteins cannot be attributed to re-initiation of ribosomes, which arrive at the 3' end of the *mstX* and *ybeL* genes on the expression constructs. Instead, the cause for the enhanced expression levels of transcriptional fusions must be due to more favorable folding of the messenger RNA.

**3' fragments of *mstX* and *ybeL* do not suffice for improved expression**

We subsequently probed the region of the fusion partner most relevant for the observed effects by making truncations from the 5' side of *mstX*Δ*(*$ATG_{1/79}$*)* and *ybeL*Δ*(*$ATG_1$) and assessing its effect on expression of AraH and NarK. The truncated constructs contained the last 180, 90 and 30 nucleotides of *mstX* and *ybeL* corresponding to approximately 50, 25 and 8% of the respective genes (Fig. 4A). A progressive shortening of *mstX* from the 5' side resulted in decreased expression levels of the fluorescent species for both AraH and NarK (Fig. 4B). Moreover, these decreased functional expression levels were accompanied by increased levels of misfolded protein for the -180 and -90 constructs. The shortest construct, -30, did not show any misfolded protein. Similar observations were made for expression of AraH using a transcriptional fusion to 5' truncations of *ybeL*, though here no misfolded protein was detected at all.

The decreased expression levels of the fluorescent species upon truncating the fusion partners from the 5' side indicate that the beneficial effects of these transcriptional fusions cannot be explained by more favorable secondary structures in the mRNA involving exclusively nucleotide fragments nearby the SD preceding *araH* or *narK*. Rather, this suggests an important role of nucleotides in the 5' half of *mstX* and *ybeL* and potentially a long-range interaction in the mRNA, in establishing conditions for improved protein expression.

**Large scale expression and purification of the ABC transporter EF583/584 from *E. faecalis***

As EF583/584 expressed to one of the lowest levels of the four test targets, we further investigated on this protein to determine the practical use of the transcriptional fusion approach. Cells expressing EF583/584 as such ("direct expression") or as transcriptional fusions to *mstX* or *ybeL*, were grown in identical preparative volumes and the transporter was purified by IMAC. Analyses of the eluted fractions by SDS-PAGE (Fig. 5A) are in good agreement with the expected yields based on *in gel* GFP fluorescence (Fig. 3). Direct expression resulted in barely detectable bands of the dimer, confirming our previous observation that expression yields were insufficient to obtain pure protein in a single IMAC purification step [44]. In contrast, yields were improved using transcriptional coupling to *mstX* and best, in term of purity and quantity, using transcriptional coupling to *ybeL*. Although only the EF584 subunit was tagged with a C-terminal GFP-$His_{10}$, the untagged EF583 chain was co-purified in stoichiometric amounts (Fig. 5A and B). SEC analysis of IMAC-purified EF583/584 revealed a main peak at a retention volume of 12 ml with a $A_{254}/A_{280}$ ratio of 0.55 and a peak width of around 3.5 ml, indicative of a pure and monodisperse protein sample (Fig. 5C, Suppl. Fig. 5). The protein yield of this main peak amounts to 0.1 mg of pure and monodisperse EF583/584 per liter of culture. This suffices to perform crystallization experiments. Moreover, using richer and buffered media such as TB [46] this value is expected to increase further. The minor peak at 9 ml retention volume exhibited an $A_{254}/A_{280}$ ratio of around 1.3 indicating that it consists mostly of co-purified DNA (Suppl. Fig. 5). The elution profile of EF583/584 is very similar to the profile obtained for TM287/288, a heterodimeric ABC transporter from *Thermotoga maritima* for which we obtained well-diffracting crystals [44] (Suppl. Fig. 5).

Due to the additional presence of a C-terminal GFP on EF584, EF583/584 elutes slightly earlier than TM287/288.

The stoichiometric co-purification of the untagged subunit of the heterodimeric ABC transporter EF583/584 as well as a SEC profile similar to a related and crystallized heterodimeric ABC transporter are strong indications that EF583/584 was expressed in a properly folded state. Taken together, these results demonstrate that transcriptional coupling can improve the expression of a membrane protein to such a degree that it becomes amenable to protein purification at decent yields and purity, which is a prerequisite for structural and functional characterization.

## Discussion

The use of N-terminal translational fusion proteins is a well-established method to circumvent low expression levels for both soluble and membrane proteins. In part, the beneficial effects of these fusion proteins can be explained by improved characteristics of the resulting fusion protein compared to the unfused protein. Examples are an increased solubility using N-terminal MBP [9, 10], a correct topology of membrane proteins using signal peptides [47], or a decreased susceptibility towards proteases [48]. Next to this, fusions can also be effective at the pre-translation phase, e.g., by decreasing the stability of secondary structure elements in the mRNA and improving translation initiation [15]. Here we explored the use of fusions at the mRNA level to enhance expression of membrane proteins.

The importance of favourable mRNA secondary structure at the initiation region has been illustrated by several studies. Kudla *et al*. [15] concluded that decreased stability of the secondary structure around the start site correlated best with the increased expression levels in their GFP library. Furthermore, small modifications in the 5' region of the ORF have been shown to improve expression both *in vivo* [42, 49] and *in vitro* [16]. In addition, it was observed that modification of the sequence near the SD [50] or preceding the start codon [51, 52] affect expression levels. Finally, a genome-wide analysis on ~400 bacterial genomes demonstrated that evolution has selected sequences at the initiation region that weaken mRNA secondary structure, thus facilitating translation initiation [53].

We have created bicistronic mRNA cassettes where the stop codon of the preceding gene (*mstX, sumo,* or *ybeL*) overlaps with the start codon of the protein of interest,

thereby mimicking a common genetic organization observed for bacterial operons [22,27] . As opposed to translational fusions, transcriptional fusions offer the advantage that the ribosome produces two distinct protein products during translation, thereby eliminating the need to enzymatically remove the fusion protein during purification. Moreover, this approach eliminates potential interference of the fusion partner in proper folding and functioning of the target protein [17, 19]. To the best of our knowledge, this is the first report on using transcriptional coupling for achieving overexpression of membrane proteins.

Compared to translational fusions, the quality and quantity of the well-characterized target membrane proteins ProW, MscL, LacY and GltP was improved using transcriptional fusions to *mstX* and *ybeL*. The negative results obtained with *sumo* were expected to result from the absence of a SD at its 3' end. However, insertion of a strong SD sequence into *sumo* at an optimal distance from the start codon of the downstream ORF only slightly increased expression levels. As translational fusions to SUMO often lead to overexpression of target proteins, we hypothesize that here the mRNA secondary structure at the 3' end of *sumo* decreases the accessibility of the SD sequence thereby reducing translation initiation of the target proteins.

The modest improvement in expression of well-folded material for the initial target proteins ProW, MscL, LacY and GltP compared to direct expression of these proteins, suggests that for these proteins the translation initiation is not limiting their expression. Rather their downstream processing (e.g., membrane targeting, insertion and folding) represents the bottleneck, as evident from the large amounts of misfolded material. To test the use of transcriptional fusions we thus subsequently used more challenging

targets, selected based on published negative results [42, 44, 45]. We could confirm low expression levels for AraH, EF583/584 and Y4 GPCR. Importantly, for these proteins transcriptional fusions to *mstX* or *ybeL* did lead to a significant increase in the amount of well-folded material. As a final demonstration of the beneficial effects of transcriptional fusions to overexpression of a membrane protein, we produced the heterodimeric ABC transporter EF583/584 using direct expression or transcriptional fusion to either *mstX* or *ybeL,* and purified the resulting protein. Clearly, more well-folded material was obtained using transcriptional fusion to *ybeL* allowing to acquire sufficient quantities for structural and functional studies.

We speculate that increased expression of membrane proteins achieved by transcriptional coupling can be explained by the fact that the translation initiation region in some genes coding for membrane proteins are occluded by secondary mRNA structures, and that the presence of an upstream gene such as *mstX* or *ybeI* would render that region more accessible to ribosomes. Importantly, the observed decreased expression levels of the fluorescent species upon truncating the fusion partners from their 5' side suggest that long-range mRNA interactions are particularly relevant here (Fig. 4B). Although occluded initiation regions will not be the only reason why membrane proteins are difficult to express, an improved understanding of the role of secondary structures in the 5'-UTR could lead to the design of better expressing constructs.

While comparing the different transcriptional fusion partners, we made the interesting observation that the use of *ybeL* significantly reduced the fraction of misfolded membrane proteins, while the amount of folded material remained high (Fig. 2 and 3). We could exclude that this resulted from a decreased protein production rate that did not

exceed the capacity of the downstream targeting, insertion and folding machinery: the expression levels of well-folded target protein using *ybeL* are not lower, but similar to transcriptional fusion to *mstX*; furthermore, decreased production rates would result in maximal expression levels at a higher inducer concentration. Such a shift was not observed, instead, all constructs had a very similar optimal inducer concentration (Suppl. Fig. 2). Next to this, as YbeL is a soluble protein [14] it is also unlikely that it facilitates the migration of the polyribosome-mRNA to the membrane, thereby overcoming the potentially limiting step of targeting the ribosome-nascent chain complex to the membrane. In addition, translation of YbeL is not required for this effect (Fig. 3). Whereas it is clear that membrane protein insertion and folding is a co-translational process, the influence of the transcript on this process is poorly understood.

Taken together, we have shown that transcriptional fusions to *mstX*, but most successfully using *ybeL*, can enhance the expression of membrane proteins. The cause of this effect is most likely the enhanced translation initiation by more favourable secondary structure in the transcript. Transcriptional fusions enhancing expression can avoid the use of translational fusions. Thereby problems associated with cleavage of the folding partner and interference with proper folding and functioning can be avoided. Though fusion of the target protein to the N-terminus of GFP facilitates the interpretation of expression trials, this is not absolutely required for this approach, and target proteins can be produced with minimal tags if needed. We anticipate that transcriptional fusions may also be advantageous for the production of soluble proteins.

## Materials and methods

### Vector construction

To enable the insertion of various sequences preceding the target ORFs, vector pBXC3GH [37] was modified to contain a unique XhoI site between the start codon and the serine codon that immediately preceded the target ORF. The XhoI site was introduced by PCR using phosphorylated primers. Following PCR and DpnI treatment, the PCR product was column-purified, ligated using T4 DNA ligase and transformed to *E. coli* DB3.1 cells. Relevant regions were verified by sequencing. The final vector was designated pBX-XhoI-C3GH.

Genes coding for Mistic (*mstX*), SUMO (*sumo*) and a shorter version of YbeL (*ybeL 1-360 bp*) were PCR-amplified using primers containing XhoI sites at their 5' termini using pBXNHmistic, pBXNHsumo and pBXNHybeI as template, respectively. GeI-purified PCR products were digested with XhoI, re-purified and ligated into XhoI-digested, dephosphorylated and gel-purified pBX-XhoI-C3GH. Correct orientation of the inserts was verified by sequencing. The final vectors for expression of target ORFs as fusion proteins to Mistic, SUMO, or YbeL were designated pBXMstXC3GH, pSUMOC3GH and pYbeLC3GH, respectively.

Vectors for transcriptional fusions were constructed similarly as described for translational fusions, except that during PCR a TGA stop codon was introduced to *mstX*, *sumo* and *ybeL*. Furthermore, the penultimate codon of *ybeL* was changed from GTC to GTA. Thus, the last four bases of every sequence read ATGA and comprised the TGA stop codon of the upstream ORF and the ATG start codon of the downstream ORF.

Vectors were verified by sequencing and designated pBXMstXTGAC3GH, pBXSUMOTGA3CGH and pBXYbeLTGA3CGH.

Ribosome binding sites in *mstX*, *sumo* and *ybeL* in pBXMstXTGAC3GH, pBXSUMOTGA3CGH and pBXYbeLTGA3CGH were removed (*mstX* and *ybeL*) or introduced (*sumo*) by PCR. Sequence and strength of the RBS's were estimated using RBS Designer [40]. RBS's in *mstX* and *ybeL* were replaced by a polyadenine sequence predicted to not serve as a RBS. Vectors were verified by sequencing and designated pBXMstXTGAC3GHΔSD, pBXSUMOTGAC3GH+SD and pBXybeLTGAC3GHΔSD.

Translation of *mstX* and *ybeL* was avoided by creating a frame shift in relevant start codons (ATG to ATaG). Next to the ATG at position 1, for *mstX* this involved the alternative start codon at position 79 as well. The additional nucleotide was introduced by PCR using pBXMstXTGAC3GH and pBXYbeLTGAC3GH as templates, yielding vectors pBXMstXTGAC3GHΔATG$_1$ and pBXYbeLTGAC3GHΔATG$_1$. For introduction of the additional nucleotide at position 79 in *mstX*, plasmid pBXMstXTGAC3GHΔATG$_1$ was used as template resulting in pBXMstXTGAC3GHΔATG$_{1\text{-}79}$.

Genes coding for *proW, mscL, gltP, araH,* and *narK* were amplified from genomic DNA of *E. coli* MC1061, *lacY* was amplified from pBADcLIC-GFP(*lacY*) [6], ABC V583 was amplified from genomic DNA of *E. faecalis* V583, and the sequence coding for the human Y4 GPCR was amplified from a plasmid obtained as previously described [45]. Gel-purified PCR products were cloned into pINITIAL using FX cloning as described [37]. The sequence-verified ORFs were subsequently used for subcloning into the expression vectors detailed above.

pBXMstXTGAC3GHΔATG$_{1\text{-}79}$ and pBXYbeLTGAC3GHΔATG$_1$ were used as templates to produce FX compatible plasmids where *mstX* and *ybeL* were reduced at sequences corresponding to their last 30, 90 or 180 nucleotides, starting from the codons corresponding to amino acids S102, Y82, I52 (MstX) and G112, D92, F62 (YbeL), respectively. Plasmids were amplified using phosphorylated primers in order to include the SD$_1$ present on the vector, using the same purification strategy described above.

**Whole cell fluorescence determination**

Single colonies of *E. coli* MC1061 transformed with relevant expression vectors were used to start overnight cultures in 700 µl of LB supplemented with 100 µg/ml ampicillin in a 96 deep-well plate. Cells were grown overnight at 37°C under vigorous shaking. A 1% inoculum was used to start cultivation in 700 µl LB-Amp and cells were grown for two hours at 37°C after which the temperature of the shaker was set to 25°C. One hour later, expression was induced by the addition of 0.2 % (w/v) L-arabinose. Cells expressing ProW, MscL, LacY, and GltP were harvested after 5 hours, while expression of AraH, NarK, ABCV583 and Y4 GPCR was carried out overnight. Cells were harvested by centrifugation (3000 x *g* for 10 min) at 4°C, washed with 500 µl ice-cold PBS buffer and finally resuspended in 300 µl. A volume of 150 µl was used to determine whole cell GFP fluorescence in a black 96 well plate; the $OD_{600}$ was determined using 50 µl in a transparent 96-well plate. Fluorescence measurements were carried out in an Infinite M1000 plate reader (Tecan, Switzerland) with excitation at 485 nm and emission at 535 nm. Fluorescence values were normalized by optical densities, and mean values were obtained by performing experiments in triplicates. Experiments were performed three times independently at a minimum.

**Dual electrophoretic mobility analysis**

In gel fluorescence and western blotting analysis were performed as described previously [6], with the exception that identical amounts of cells corresponding to ~ 2 mg of total cellular protein were used that were derived from a pooled sample of three independent cultures. Furthermore, cell pellets were resuspended in 350 µl of 50 mM KPi buffer pH 7.5, supplemented with protease inhibitors tablets (Complete mini, Roche), 10% glycerol, lysozyme (1 mg/ml), and DNase (20 µg/ml).

**Solubilization in DDM and FSEC analysis**

Small scale solubilization tests on AraH, NarK, EF583/584 and Y4 GPCR (Suppl. Fig. 4, Suppl. Table 3) were carried out as described before [6]. Samples corresponding to the solubilized fraction after ultracentrifugation were subsequently used to perform FSEC analysis. Undiluted and diluted (1:10) samples were placed in a 96 well plate. Dilution was done to avoid saturation of the fluorescence detector for well expressing proteins. Samples were run on an analytical size exclusion column (Superdex 200 5/150 GL, GE) at 4°C in 50 mM Tris.HCl, pH 7.5, 150 mM NaCl, and 2 mM DDM, using an Agilent Technologies 1100 Series HPLC coupled to an Agilent 1260 Infinity fluorescence detector.

**Large scale expression and purification of the ABC transporter EF583/584**

The ABC transporter EF583/584 from *E. faecalis* was expressed in *E. coli* MC1061 without any N-terminal fusion using the pBXC3GH vector, or via transcriptional fusion to *mstX* or *ybeL* (pBXMstXTGAC3GH and pBXybeLTGAC3GH). Single colonies were used to inoculate 10 ml of LB-AMP medium and grown overnight at 37°C. The precultures

were diluted 1:100 in 600 ml of LB-AMP per culture flask and three flasks were used for each of the three constructs. Expression of EF583/584 cloned into pBXybeLC3GH was further scaled up to 6 l of LB-AMP. Cultivation was performed as described for whole cell fluorescence determination and induction was done overnight. Membrane vesicle preparation and IMAC were carried out exactly as described previously [44, 45], using 1 % of n-dodecyl-β-D-maltoside (DDM) for solubilization. Size exclusion chromatography was carried out using a Superdex 200 10/300 GL column (GE Healthcare).

**Acknowledgements**

We are indebted to Prof. Raimund Dutzler for helpful discussions and providing generous access to instruments and equipment. This work has been primarily supported by the Swiss National Science Foundation (Grant No. 141074 to O.Z); in addition, support is acknowledged from the German Research Foundation Cluster of Excellence Frankfurt "Macromolecular Complexes" (E.R.G.), Collaborative Research Center SFB807 (E.R.G.) and the Swiss National Science Foundation (SNF Professorship to M.A.S.).

**Author contributions**

J.M. conceived the project and performed most experiments. J.M. and E.R.G. designed experiments. M.H. performed the protein purifications. M.A.S. supervised the purification. O.Z. coordinated the entire project. J.M. and E.R.G. drafted the manuscript. All authors discussed results and contributed to the final manuscript. E.R.G. supervised the project.

**Figure legends**

**Fig. 1. Schematic representation of the three different approaches used to express ProW, MscL, LacY and GltP**. Sequences coding for fusion partners, membrane proteins and GFP are indicated by fp, MP-ORF, and gfp, respectively. The topology of the resulting protein(s) is indicated on the right. The location of the SD sequences in the fusion partners and the translation efficiency (defined as the probability of a given mRNA being bound to a ribosome) were predicted using the RBS-designer software [40, 54]. SD sequences and the overlapping Start/Stop codon are indicated in bold. The translation efficiency of the (first) SD sequence encoded by the pBAD vector was predicted to be $3.0*10^{-1}$.

**Fig. 2 Differential electrophoretic mobility analysis of ProW, MscL, LacY and GltP expressed with a C-terminal GFP**. Proteins were expressed directly, as translational fusions with N-terminal Mistic, SUMO or YbeL, or via transcriptional fusions with *mstX*, *sumo* and *ybeL*. In the left panels the protein is detected by *in gel* GFP fluorescence, in the right panels immunoblots of the same gels are shown where the protein is detected using an anti-His antibody. The lower, fluorescent band, indicative of well-folded protein, and the upper non-fluorescent band, indicative of misfolded material, are indicated by white and black arrows, respectively. Identical amounts of cells were used for each sample.

**Fig. 3. Differential electrophoretic mobility analysis of AraH, NarK, EF583/584 and Y4 GPCR expressed with a C-terminal GFP.** Proteins were expressed directly or as transcriptional fusions to *mstX* or *ybeL*. The two most right lanes of each panel represent expression using transcriptional fusions for which the fusion partner is not expressed

due to disruption of (potential) start codons. In the left panels the protein is detected by *in gel* GFP fluorescence, in the right panels immunoblots of the same gel are shown where the protein is detected using an anti-His antibody. The lower, fluorescent band, indicative of well-folded protein, and the upper non-fluorescent band, indicative of misfolded material, are indicated by white and black arrows, respectively. Identical amounts of cells were used for each sample.

**Fig. 4. Differential electrophoretic mobility analysis of AraH and NarK transcriptionally fused to a 5' truncated *mstX* or *ybeL derivative*.** A) Schematic representation of the nucleotide starting positions of *mstX*Δ($ATG_{1/79}$) and *ybeL*Δ($ATG_1$) 5' truncations. Sequences coding for fusion partners, membrane proteins and GFP are indicated by fp, MP-ORF, and gfp, respectively. B) Proteins with a C-terminal GFP were expressed directly or as transcriptional fusions to *mstX* or *ybeL* derivatives. In the left panels the protein is detected by *in gel* GFP fluorescence, in the right panels immunoblots of the same gel are shown where the protein is detected using an anti-His antibody. The lower, fluorescent band, indicative of well-folded protein, and the upper non-fluorescent band, indicative of misfolded material, are indicated by white and black arrows, respectively. Identical amounts of cells were used for each sample.

**Fig. 5**. **Purification of the ABC transporter EF583/584.** A) Expression was performed directly or via transcriptional coupling to *mstX* or *ybeL.* Proteins were purified under identical conditions using IMAC and three consecutive elution fractions were analyzed by SDS-PAGE and stained with Coomassie. Expression using transcriptional coupling to *ybeL* was further scaled-up and analyzed by SEC. B) Peak fractions were analyzed by

SDS-PAGE and stained with Coomassie. C) Analysis of the IMAC-purified protein by SEC using a Superdex 200 10/300 GL column.

## References


[1] Wang DN, Safferling M, Lemieux MJ, Griffith H, Chen Y, Li XD. Practical aspects of overexpressing bacterial secondary membrane transporters for structural studies. Biochim Biophys Acta - Biomembr 2003;1610:23–36.

[2] Waugh DS. Making the most of affinity tags. Trends Biotechnol 2005;23:316–20.

[3] Young CL, Britton ZT, Robinson AS. Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications. Biotechnol J 2012;7:620–34.

[4] Terpe K. Overview of bacterial expression systems for heterologous protein production: From molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotechnol 2006;72:211–22.

[5] Drew DE, von Heijne G, Nordlund P, de Gier JW. Green fluorescent protein as an indicator to monitor membrane protein overexpression in Escherichia coli. FEBS Lett 2001;507:220–4.

[6] Geertsma ER, Groeneveld M, Slotboom D-J, Poolman B. Quality control of overexpressed membrane proteins. Proc Natl Acad Sci U S A 2008;105:5722–7.

[7] Linares DM, Geertsma ER, Poolman B. Evolved Lactococcus lactis Strains for Enhanced Expression of Recombinant Membrane Proteins. J Mol Biol 2010;401:45–55.

[8] Mansell TJ, Linderman SW, Fisher AC, DeLisa MP. A rapid protein folding assay for the bacterial periplasm. Protein Sci 2010;19:1079–90.

[9] Kapust RB, Waugh DS. Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci 1999;8:1668–74.

[10] Raran-Kurussi S, Waugh DS. The Ability to Enhance the Solubility of Its Fusion Partners Is an Intrinsic Property of Maltose-Binding Protein but Their Folding Is Either Spontaneous or Chaperone-Mediated. PLoS One 2012;7.

[11] LaVallie ER, Lu Z, Diblasio-Smith EA, Collins-Racie LA, McCoy JM. Thioredoxin as a fusion partner for production of soluble recombinant proteins in Escherichia coli. Methods Enzymol 2000;326:322–40.

[12] Peroutka Iii RJ, Orcutt SJ, Strickler JE, Butt TR. SUMO fusion technology for enhanced protein expression and purification in prokaryotes and eukaryotes. Methods Mol Biol 2011;705:15–30.

[13] Roosild TP, Greenwald J, Vega M, Castronovo S, Riek R, Choe S. NMR structure of Mistic, a membrane-integrating protein for membrane protein expression. Science 2005;307:1317–21.

[14] Leviatan S, Sawada K, Moriyama Y, Nelson N. Combinatorial method for overexpression of membrane proteins in Escherichia coli. J Biol Chem 2010;285:23548–56.

[15] Kudla G, Murray AW, Tollervey D, Plotkin JB. Coding-sequence determinants of gene expression in Escherichia coli. Science 2009;324:255–8.

[16] Haberstock S, Roos C, Hoevels Y, Dötsch V, Schnapp G, Pautsch A, et al. A systematic approach to increase the efficiency of membrane protein production in cell-free expression systems. Prot Exp Purif 2012;82:308–16.

[17] Sabaty M, Grosse S, Adryanczyk G, Boiry S, Biaso F, Arnoux P, et al. Detrimental effect of the 6 His C-terminal tag on YedY enzymatic activity and influence of the TAT signal sequence on YedY synthesis. BMC Biochem 2013;14:28.

[18] Singh P, Sharma L, Kulothungan SR, Adkar B V., Prajapati RS, Ali PSS, et al. Effect of Signal Peptide on Stability and Folding of Escherichia coli Thioredoxin. PLoS One 2013;8.

[19] Majtan T, Kraus JP. Folding and activity of mutant cystathionine beta-synthase depends on the position and nature of the purification tag: Characterization of the R266K CBS mutant. Protein Expr Purif 2012;82:317–24.

[20] Smyth DR, Mrozkiewicz MK, McGrath WJ, Listwan P, Kobe B. Crystal structures of fusion proteins with large-affinity tags. Protein Sci 2003;12:1313–22.

[21] Vergis JM, Wiener MC. The variable detergent sensitivity of proteases that are utilized for recombinant protein affinity tag removal. Protein Expr Purif 2011;78:139–42.

[22] Blattner FR. The Complete Genome Sequence of Escherichia coli K-12. Science (80- ) 1997;277:1453–62.

[23] Nomura M. Regulation of ribosome biosynthesis in Escherichia coli and Saccharomyces cerevisiae: diversity and common principles. J Bacteriol 1999;181:6857–64.

[24] Rex G, Surin B, Besse G, Schneppe B, McCarthy JEG. The mechanism of translational coupling in Escherichia coli. Higher order structure in the atpHA mRNA acts as a conformational switch regulating the access of de novo initiating ribosomes. J Biol Chem 1994;269:18118–27.

[25] Zurawski G, Elseviers D, Stauffer G V, Yanofsky C. Translational control of transcription termination at the attenuator of the Escherichia coli tryptophan operon. Proc Natl Acad Sci U S A 1978;75:5988–92.

[26] Choudhary M, Kaplan S. DNA sequence analysis of the photosynthesis region of Rhodobacter sphaeroides 2.4.1. Nucleic Acids Res 2000;28:862–7.

[27] Kozak M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene 2005;361:13–37.

[28] Andrè A, Puca A, Sansone F, Brandi A, Antico G, Calogero RA. Reinitiation of protein synthesis in Escherichia coli can be induced by mRNA cis-elements unrelated to canonical translation initiation signals. FEBS Lett 2000;468:73–8.

[29] Das A, Yanofsky C. A ribosome binding site sequence is necessary for efficient expression of the distal gene of a translationally-coupled gene pair. Nucleic Acids Res 1984;12:4757–68.

[30] Takyar S, Hickerson RP, Noller HF. mRNA helicase activity of the ribosome. Cell 2005;120:49–58.

[31] Daley DO, Rapp M, Granseth E, Melén K, Drew D, von Heijne G. Global topology analysis of the Escherichia coli inner membrane proteome. Science 2005;308:1321–3.

[32] Kawate T, Gouaux E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. Structure 2006;14:673–81.

[33] Romantsov T, Battle AR, Hendel JL, Martinac B, Wood JM. Protein localization in Escherichia coli cells: comparison of the cytoplasmic membrane proteins ProP, LacY, ProW, AqpZ, MscS, and MscL. J Bacteriol 2010;192:912–24.

[34] Chang G, Spencer RH, Lee AT, Barclay MT, Rees DC. Structure of the MscL homolog from Mycobacterium tuberculosis: a gated mechanosensitive ion channel. Science 1998;282:2220–6.

[35] Kumar H, Kasho V, Smirnova I, Finer-Moore JS, Kaback HR, Stroud RM. Structure of sugar-bound LacY. Proc Natl Acad Sci U S A 2014;111:1784–8.

[36] Guzman LM, Belin D, Carson MJ, Beckwith J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol 1995;177:4121–30.

[37] Geertsma ER, Dutzler R. A versatile and efficient high-throughput cloning tool for structural biology. Biochemistry 2011;50:3272–8.

[38] Lundberg ME, Becker EC, Choe S. MstX and a putative potassium channel facilitate biofilm formation in Bacillus subtilis. PLoS One 2013;8:e60993.

[39] Marino J, Bordag N, Keller S, Zerbe O. Mistic's membrane association and its assistance in overexpression of a human GPCR are independent processes. Protein Sci 2014.

[40] Na D, Lee D. RBSDesigner: software for designing synthetic ribosome binding sites that yields a desired level of protein expression. Bioinformatics 2010;26:2633–4.

[41] Wagner S, Klepsch MM, Schlegel S, Appel A, Draheim R, Tarry M, et al. Tuning Escherichia coli for membrane protein overexpression. Proc Natl Acad Sci U S A 2008;105:14371–6.

[42] Nørholm MHH, Toddo S, Virkki MTI, Light S, von Heijne G, Daley DO. Improved production of membrane proteins in Escherichia coli by selective codon substitutions. FEBS Lett 2013;587:2352–

[43] Zheng H, Wisedchaisri G, Gonen T. Crystal structure of a nitrate/nitrite exchanger. Nature 2013;497:647–51.

[44] Hohl M, Briand C, Grütter MG, Seeger MA. Crystal structure of a heterodimeric ABC transporter in its inward-facing conformation. Nat Struct Mol Biol 2012;19:395–402.

[45] Marino J, Geertsma ER, Zerbe O. Topogenesis of heterologously expressed fragments of the human Y4 GPCR. Biochim Biophys Acta 2012;1818:3055–63.

[46] Lessard JC. Growth media for E. coli. Methods Enzymol 2013;533:181–9.

[47] Grisshammer R. Purification of recombinant G-protein-coupled receptors. Methods Enzymol 2009;463:631–45.

[48] Murby M, Uhlén M, Ståhl S. Upstream strategies to minimize proteolytic degradation upon recombinant production in Escherichia coli. Protein Expr Purif 1996;7:129–36.

[49] De Boer HA, Hui AS. Sequences within ribosome binding site affecting messenger RNA translatability and method to direct ribosomes to single messenger RNA species. Methods Enzymol 1990;185:103–14.

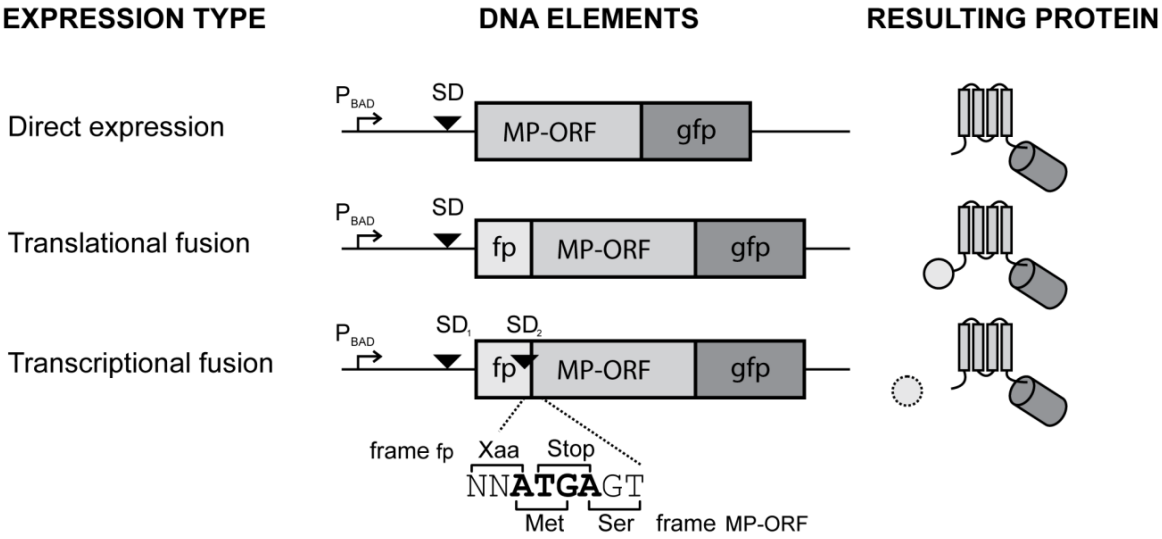
[50] De Boer HA, Hui A, Comstock LJ, Wong E, Vasser M. Portable Shine-Dalgarno regions: a system for a systematic study of defined alterations of nucleotide sequences within E. coli ribosome binding sites. DNA 1983;2:231–5.

[51] Hui A, Hayflick J, Dinkelspiel K, de Boer HA. Mutagenesis of the three bases preceding the start codon of the beta-galactosidase mRNA and its effect on translation in Escherichia coli. EMBO J 1984;3:623–9.

[52] Gheysen D, Iserentant D, Derom C, Fiers W. Systematic alteration of the nucleotide sequence preceding the translation initiation codon and the effects on bacterial expression of the cloned SV40 small-t antigen gene. Gene 1982;17:55–63.

[53] Bentele K, Saffert P, Rauscher R, Ignatova Z, Blüthgen N. Efficient translation initiation dictates codon usage at gene start. Mol Syst Biol 2013;9:675.

[54] Na D, Lee S, Lee D. Mathematical modeling of translation initiation for the estimation of its efficiency to computationally design mRNA sequences with desired expression levels in prokaryotes. BMC Syst Biol 2010;4:71.




| 3' DNA sequence fusion partner | Predicted translation efficiency (a.u.) |
|---|---|
| *mstX* CTCTCAGTATCTG**AAGAAG**GCGAAAAAGA**ATGA** | $2.4*10^{-1}$ |
| *sumo* ATTGACGTGTTCCAGCAACACACCGGTGG**ATGA** | $1.4*10^{-2}$ |
| *ybeL* CATGGGGTTTATCACAGC**GGAGAA**GTGGT**ATGA** | $1.0*10^{-1}$ |
| *mstX* Δ(SD) CTCTCAGTATCTGAAAAAAGCGAAAAAGA**ATGA** | $1.6*10^{-2}$ |
| *sumo* +(SD) ATTGACGTGTTCCAG**CAGGAG**GCCGGTGG**ATGA** | $3.3*10^{-1}$ |
| *ybeL* Δ(SD) CATGGGGTTTATCACAGCCCAAACGTGGT**ATGA** | $2.0*10^{-3}$ |

Figure 1

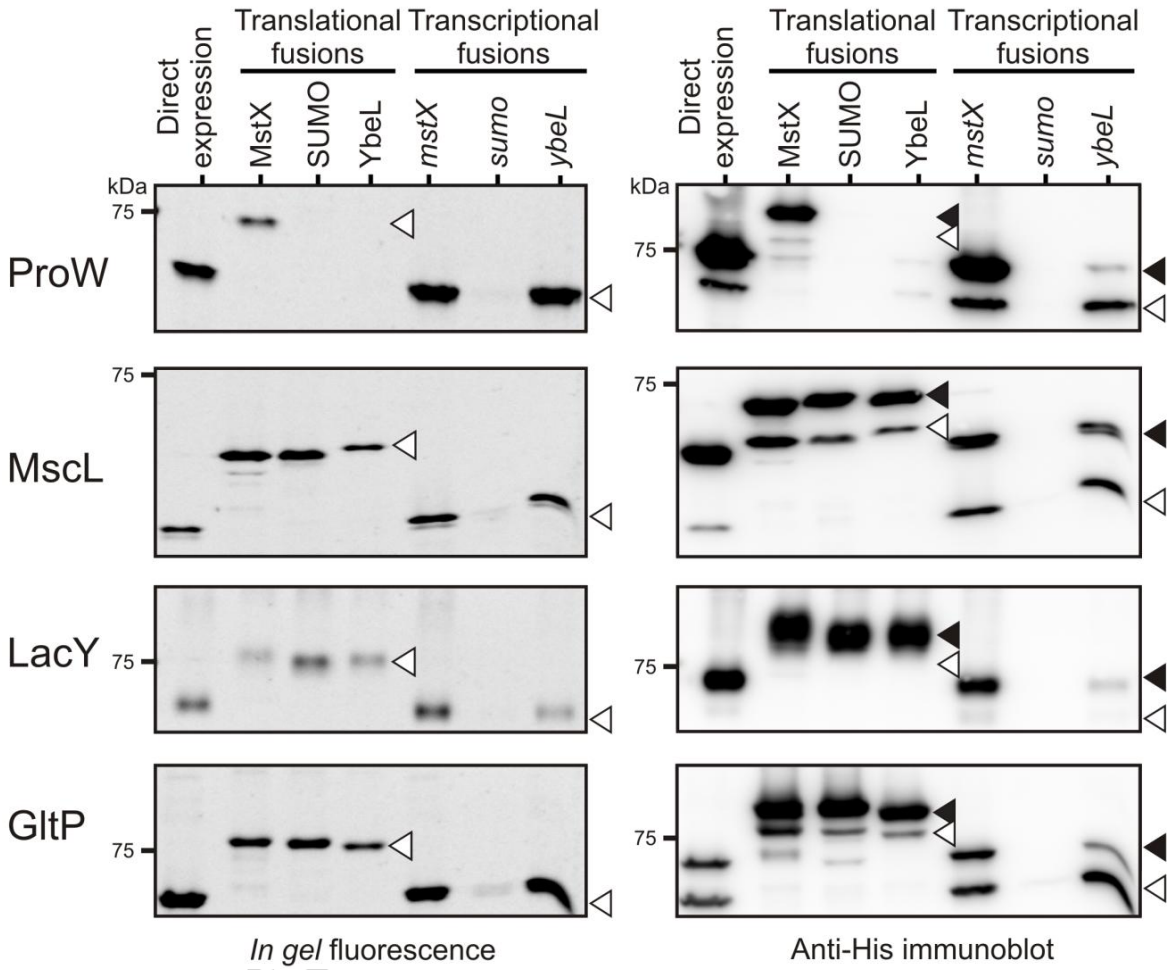



Figure 2

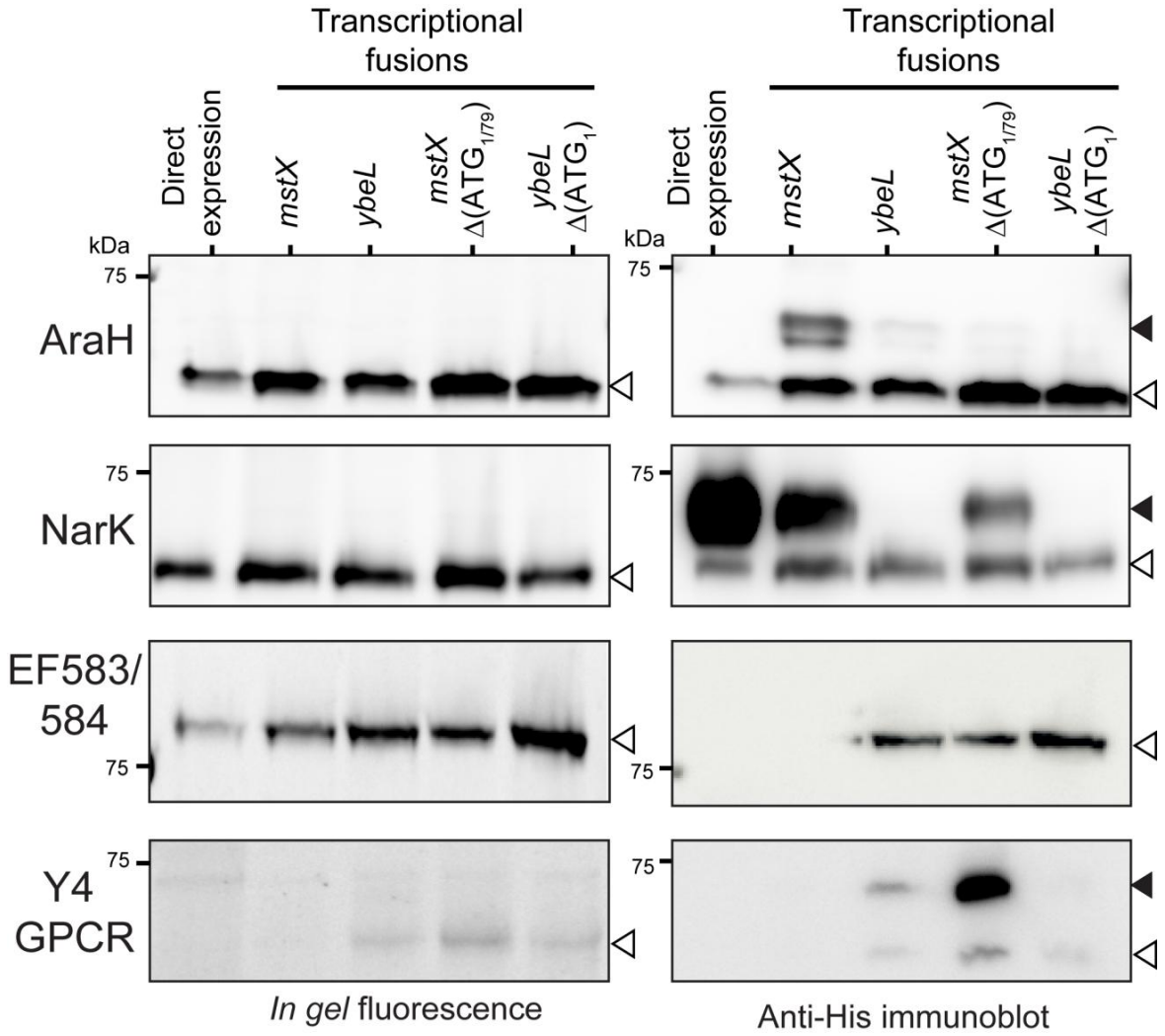



Figure 3

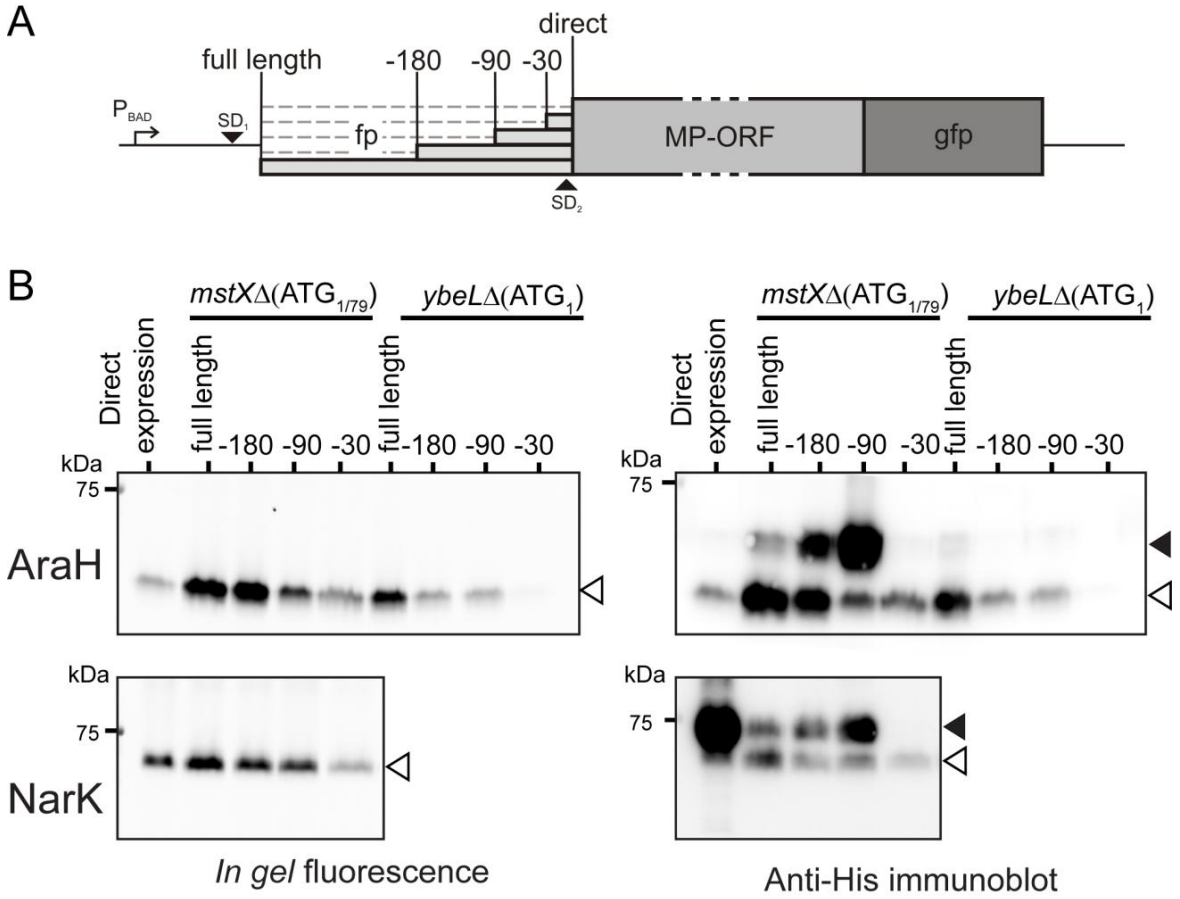



Figure 4

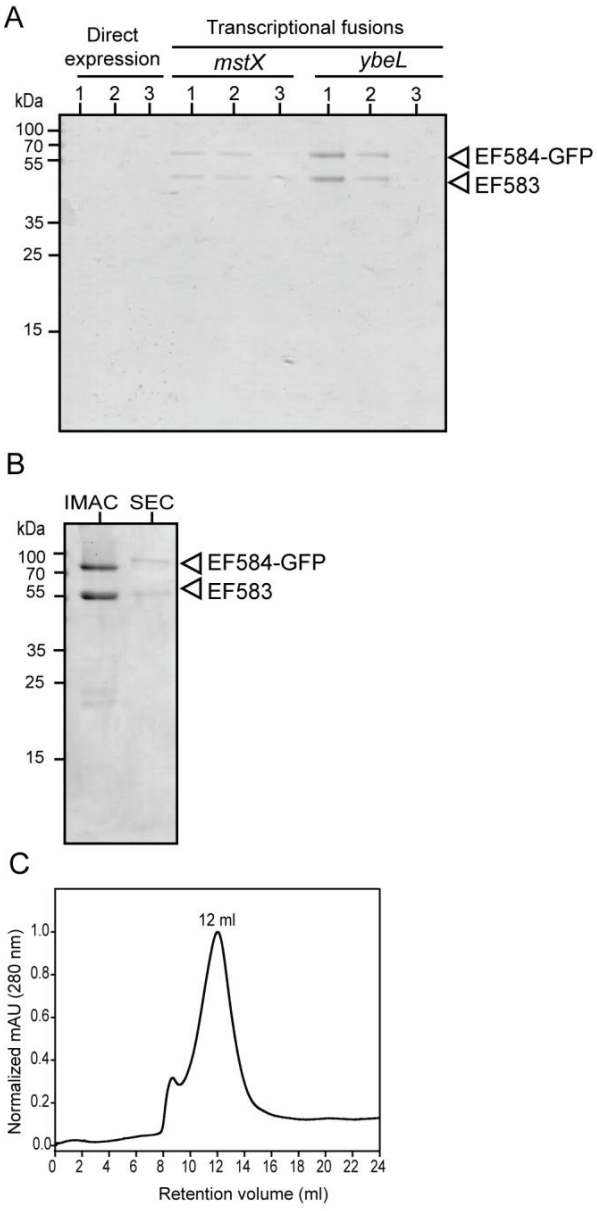



Figure 5