

RESEARCH

Open Access



# Increased production of periplasmic proteins in *Escherichia coli* by directed evolution of the translation initiation region

Kiavash Mirzadeh<sup>1,2†</sup>, Patrick J. Shilling<sup>1†</sup>, Rageia Elfageih<sup>1</sup>, Alister J. Cumming<sup>1</sup>, Huanhuan L. Cui<sup>3</sup>, Maja Rennig<sup>4</sup>, Morten H. H. Nørholm<sup>2,4</sup> and Daniel O. Daley<sup>1,2\*</sup> 

## Abstract

**Background:** Recombinant proteins are often engineered with an N-terminal signal peptide, which facilitates their secretion to the oxidising environment of the periplasm (gram-negative bacteria) or the culture supernatant (gram-positive bacteria). A commonly encountered problem is that the signal peptide influences the synthesis and secretion of the recombinant protein in an unpredictable manner. A molecular understanding of this phenomenon is highly sought after, as it could lead to improved methods for producing recombinant proteins in bacterial cell factories.

**Results:** Herein we demonstrate that signal peptides contribute to an unpredictable translation initiation region. A directed evolution approach that selects a new translation initiation region, whilst leaving the amino acid sequence of the signal peptide unchanged, can increase production levels of secreted recombinant proteins. The approach can increase production of single chain antibody fragments, hormones and other recombinant proteins in the periplasm of *E. coli*.

**Conclusions:** The study demonstrates that signal peptide performance is coupled to the efficiency of the translation initiation region.

**Keywords:** Protein secretion, Signal peptide, Recombinant protein expression, Bacteria, Translation initiation, Directed evolution

## Background

Bacterial cell factories are widely used in the biotech and pharmaceutical industries for the production of high-value recombinant proteins. Classic examples include industrial enzymes, hormones and antibody fragments, which generate billions of dollars in revenue annually [1, 2]. These recombinant proteins are typically engineered with an N-terminal signal peptide so that they will be secreted out of the bacterial cytoplasm [3, 4]. For

industrial enzymes, which are usually produced in gram-positive bacteria such as *Bacillus subtilis*, secretion from the cytoplasm to the culture supernatant simplifies purification and downstream processing. For hormones and antibody fragments, which are usually produced in gram-negative bacteria like *Escherichia coli*, secretion from the cytoplasm to the oxidising environment of the periplasm is necessary for the formation of disulfide bonds that are essential for protein folding and activity [5, 6].

Secretion out of the bacterial cytoplasm is usually mediated by the general secretion pore (Sec) [7, 8]. Sec is a major hub for protein trafficking as it inserts proteins into the cytoplasmic membrane, and secretes proteins to the envelope and beyond. Secreted proteins are typically targeted to Sec by a cleavable N-terminal signal peptide.

\*Correspondence: ddaley@dbb.su.se

<sup>†</sup>Kiavash Mirzadeh, Patrick J. Shilling have contributed equally to this work

<sup>1</sup> Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden

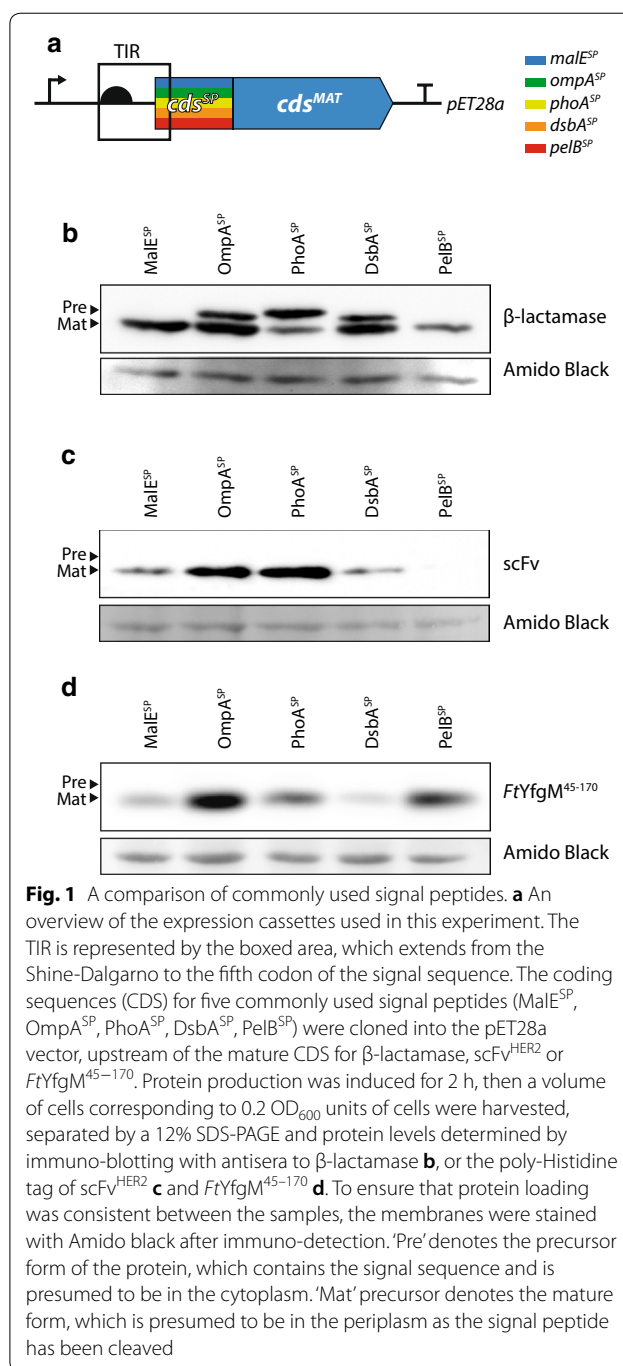
Full list of author information is available at the end of the article



Signal peptides vary in length and amino acid sequence, but have a distinctive tripartite structure that includes a positively-charged N-terminal region, a hydrophobic core, and a polar C-terminal cleavage site that contains the signal peptidase recognition site (Ala-X-Ala) [9, 10]. They also have a distinctive codon usage, which includes a biased use of the AAA (Lys) codon at the second position, and a high frequency of non-optimal codons [11–15]. It has been suggested that the signal peptide slows folding of the protein in the cytoplasm and targets it to Sec in a predominantly unfolded confirmation [16]. Upon arrival at Sec the signal peptide also promotes binding to the SecA chaperone, thereby allosterically activating Sec for protein secretion [17]. Given these multiple roles it is likely that signal peptides have co-evolved with the protein that they translocate, as well as with the secretion machinery.

Signal peptides have unpredictable effects on the production yields of recombinant proteins. For example, a signal peptide that supports a high-level of protein synthesis and secretion for one recombinant protein often supports a low-level of protein synthesis and secretion for another [3, 4, 18]. Herein we refer to this phenomenon as signal peptide performance. Since it is not possible to predict how well a signal peptide will perform with a given recombinant protein, it is common practice to screen signal peptide-libraries for one that supports a high-level of protein synthesis and secretion [3, 4, 18]. This approach is both time-consuming and expensive. A molecular understanding of signal peptide performance is therefore highly sought after, as it could lead to new methods for (1) identifying suitable signal peptides, and (2) rationally engineering signal peptides that increase production yields in bacterial cell factories.

In this study we have noted that signal peptides have a significant impact on translation initiation; the rate-limiting step in protein synthesis [19–23]. The efficiency of translation initiation is dependent on the nucleotide sequence of the translation initiation region (TIR) [24–27], a stretch of approximately thirty nucleotides that extends from the Shine-Dalgarno region to the fifth codon of the coding sequence (i.e. the first ribosomal footprint) [28]. When signal peptides are fused to recombinant proteins, the TIR is formed ad hoc by genetically fusing the 5' UTR encoded in the plasmid to the coding sequence of the signal peptide (Fig. 1a). Thus each time a different signal peptide is fused, a different TIR is generated. Herein we demonstrate that TIRs formed by genetic fusion are not optimal for protein production, and that protein production yields can be increased using a directed evolution approach that selects a new translation initiation region, whilst leaving the amino acid sequence of the signal peptide unchanged. Taken together the data



**Fig. 1** A comparison of commonly used signal peptides. **a** An overview of the expression cassettes used in this experiment. The TIR is represented by the boxed area, which extends from the Shine-Dalgarno to the fifth codon of the signal sequence. The coding sequences (CDS) for five commonly used signal peptides (MalE<sup>SP</sup>, OmpA<sup>SP</sup>, PhoA<sup>SP</sup>, DsbA<sup>SP</sup>, PelB<sup>SP</sup>) were cloned into the pET28a vector, upstream of the mature CDS for  $\beta$ -lactamase, scFv<sup>HER2</sup> or FtYfgM<sup>45-170</sup>. Protein production was induced for 2 h, then a volume of cells corresponding to 0.2 OD<sub>600</sub> units of cells were harvested, separated by a 12% SDS-PAGE and protein levels determined by immuno-blotting with antisera to  $\beta$ -lactamase **b**, or the poly-Histidine tag of scFv<sup>HER2</sup> **c** and FtYfgM<sup>45-170</sup> **d**. To ensure that protein loading was consistent between the samples, the membranes were stained with Amido black after immuno-detection. 'Pre' denotes the precursor form of the protein, which contains the signal sequence and is presumed to be in the cytoplasm. 'Mat' precursor denotes the mature form, which is presumed to be in the periplasm as the signal peptide has been cleaved

indicate that signal sequence performance is tightly coupled to the nucleotide sequence of the TIR.

## Results

### Production of periplasmic proteins with commonly used signal peptides

Five signal peptides that are commonly used for the production of recombinant proteins in the periplasm of

*E. coli* were selected (MalE<sup>SP</sup>, OmpA<sup>SP</sup>, PhoA<sup>SP</sup>, DsbA<sup>SP</sup> and PelB<sup>SP</sup>; Table 1). The coding sequences for these signal peptides were cloned into the commonly used pET28a expression plasmid, upstream of the coding sequence for  $\beta$ -lactamase (Fig. 1a). To determine how efficiently the signal peptides supported the synthesis and secretion of  $\beta$ -lactamase the expression plasmids were transformed into the *E. coli* strain BL21(DE3) *pLysS* and a mild induction protocol was used to initiate transcription (0.05 mM IPTG for 2 h at 30 °C). Following the induction period, whole cells were collected, and proteins were separated by SDS-PAGE and immuno-blotted, so that the secreted (Mature) and non-secreted (Precursor)  $\beta$ -lactamase could be distinguished. The experiment indicated that MalE<sup>SP</sup>, OmpA<sup>SP</sup>, PhoA<sup>SP</sup>, DsbA<sup>SP</sup> supported a comparatively high-level of  $\beta$ -lactamase synthesis, whereas PelB<sup>SP</sup> did not (Fig. 1b). Moreover, MalE<sup>SP</sup> and PelB<sup>SP</sup> were effective in supporting the secretion of  $\beta$ -lactamase to the periplasm, as only the Mature sized protein was detected. In contrast OmpA<sup>SP</sup>, PhoA<sup>SP</sup> and DsbA<sup>SP</sup> were less effective in supporting the secretion of  $\beta$ -lactamase as there was a prominent Precursor band, indicating that the signal peptide was not cleaved. The latter observation most likely means that the  $\beta$ -lactamase was not secreted to the periplasm.

To evaluate the performance of the signal peptides with other recombinant proteins, we fused them to a single chain variable fragment that recognises the human epidermal growth factor receptor protein 2 protein (scFv<sup>HER2</sup>) and a soluble fragment of the periplasmic chaperone YfgM from *Francisella tularensis* (FtYfgM<sup>45–170</sup>). Again there were differences in synthesis levels of scFv<sup>HER2</sup> and FtYfgM<sup>45–170</sup> when using different signal peptides (Fig. 1c, d). Taken together, these observations demonstrate that signal peptide performance is varied and unpredictable during the synthesis and secretion of recombinant periplasmic proteins. This conclusion is supported by a large body of published work, however a molecular explanation for the phenomenon remains elusive (reviewed in [3, 4]).

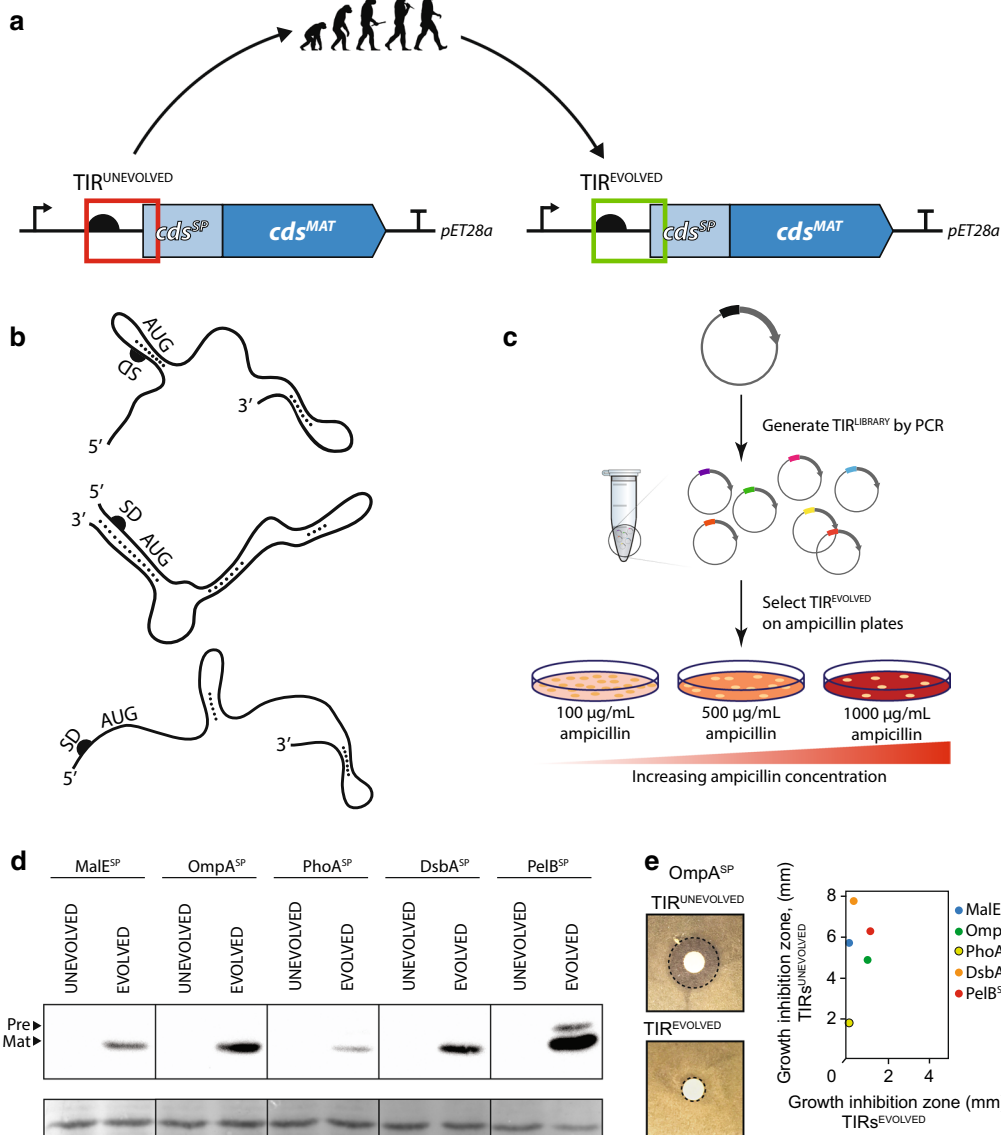
**Table 1** Signal peptides used in this study

Signal peptide	Sequence	Origin
MalE <sup>SP</sup>	MKIKTGARILALSALTTMMFSASALA	<i>Escherichia coli</i>
OmpA <sup>SP</sup>	MKKTAIAlAVALAGFATVAQA	<i>Escherichia coli</i>
PhoA <sup>SP</sup>	MKQSTIALALLPLLFPTVTKA	<i>Escherichia coli</i>
DsbA <sup>SP</sup>	MKKIWLALAGLVLAFSASA	<i>Escherichia coli</i>
PelB <sup>SP</sup>	MKYLLPTAAAGLLLLAAQPAMA	<i>Erwinia carotovora</i>

### Signal peptide performance is coupled to translation initiation

In our recombinant expression plasmids, the TIR was formed by fusing the 5'UTR from the expression plasmid with the coding sequence of the signal peptide (Fig. 2a, b). Thus each time a different signal peptide was used a different TIR was generated. We refer to each of these TIRs as unevolved (i.e. TIR<sup>UNEVOLVED</sup>) as they were formed by ad hoc genetic fusion rather than co-evolution with the host cell ribosomes. We hypothesised that a TIR<sup>UNEVOLVED</sup> would not be optimal for protein production.

We employed a directed evolution approach to select a TIR that was more suitable for protein production. In the directed evolution experiment, TIR<sup>LIBRARIES</sup> were created from the expression plasmids encoding the MalE<sup>SP</sup>, OmpA<sup>SP</sup>, PhoA<sup>SP</sup>, DsbA<sup>SP</sup> and PelB<sup>SP</sup> fused to  $\beta$ -lactamase. In the design of the TIR<sup>LIBRARIES</sup>, the six nucleotides immediately upstream from the AUG start codon were completely randomised, and the six nucleotides immediately downstream were randomised with synonymous codon changes only (Fig. 2c) [29, 30]. Each TIR<sup>LIBRARY</sup> theoretically contained >18,000 expression plasmids with a different TIR. The TIR<sup>LIBRARIES</sup> were transformed into BL21(DE3) *pLysS* and plated onto LB agar containing 0.05 mM IPTG and different concentrations of ampicillin (Fig. 2c). Twenty colonies that were resistant to a high concentration of ampicillin were initially selected and sequenced (Additional file 1: Fig. 1). The expression of  $\beta$ -lactamase from each unique TIR was compared by immuno-blotting (Additional file 1: Fig. 2). Those expression plasmids that supported a high level of  $\beta$ -lactamase were selected. The evolved TIRs (TIR<sup>EVOLVED</sup>) are presented in Table 2. Each TIR<sup>EVOLVED</sup> was then re-constructed in a fresh expression vector, and the production levels of  $\beta$ -lactamase again compared by immuno-blotting. After a 2-h induction period we observed that more  $\beta$ -lactamase was synthesised when a TIR<sup>EVOLVED</sup> was used (Fig. 2d). Note that production of  $\beta$ -lactamase from each TIR<sup>UNEVOLVED</sup> was undetectable on these blots because the difference from the TIR<sup>EVOLVED</sup> was too large to capture at this time point (see below). Consistent with this observation, disc diffusion assays confirmed that the TIR<sup>EVOLVED</sup> supported a higher level of resistance to ampicillin than the TIR<sup>UNEVOLVED</sup> (Fig. 2e). We could not explain the difference in production levels from the TIR<sup>UNEVOLVED</sup>/TIR<sup>EVOLVED</sup> pairs using mRNA structure predictions. Nevertheless, the data underscores the fact that the synthesis of secreted recombinant proteins in bacterial cell factories is tightly coupled to the nucleotide sequence of the TIR.



**Fig. 2** Improved signal peptide performance following directed evolution of the TIR **a** A directed evolution approach was used to convert a TIR<sup>UNEVOLVED</sup> to a TIR<sup>EVOLVED</sup>. The TIR is defined as the region from the Shine-Dalgarno (half-dome) to codon 5 of the signal peptide. **b** mRNA has a high propensity to form structures, thus a TIR<sup>UNEVOLVED</sup> can be sequestered into short- (top) or long-range structures (middle). Directed evolution should select a TIR<sup>EVOLVED</sup> that is less structured and more accessible to the ribosome during translation initiation (bottom). **c** An overview of the directed evolution process. A TIR<sup>LIBRARY</sup> was constructed by completely randomising the six nucleotides immediately upstream of the AUG start codon, and partially randomising the six nucleotides immediately downstream of the AUG start codon (allowing synonymous codons changes only). The TIR<sup>LIBRARY</sup> was transformed into *E. coli* BL21 (DE3) pLysS and plated on increasing concentrations of ampicillin. A TIR<sup>EVOLVED</sup> was identified on the plate containing the highest concentration of ampicillin relative to the TIR<sup>UNEVOLVED</sup> variant. **d**  $\beta$ -lactamase production levels from TIR<sup>UNEVOLVED</sup>/TIR<sup>EVOLVED</sup> pairs was assessed by immuno-blotting. In this experiment,  $\beta$ -lactamase production was induced for 2 h, then a volume of cells corresponding to 0.2 OD<sub>600</sub> units of cells were harvested, separated by a 12% SDS-PAGE and protein levels were determined by immuno-blotting with antisera to  $\beta$ -lactamase. To ensure that protein loading was consistent between the samples, the membrane was stained with Amido black after immuno-detection. 'Pre' denotes the precursor form of the protein, which contains the signal peptide fused version of  $\beta$ -lactamase, which we presume to be in the cytoplasm as the signal peptide is still present. 'Mat' precursor denotes the mature version of  $\beta$ -lactamase, which presumably is in the periplasm as the signal peptide has been cleaved. **e**  $\beta$ -lactamase activity from TIR<sup>UNEVOLVED</sup>/TIR<sup>EVOLVED</sup> pairs was assessed using the disc diffusion assay. Here a filter disc containing 2 mg of ampicillin was placed on top of an LB-agar plate containing a lawn of bacteria expressing  $\beta$ -lactamase from either a TIR<sup>UNEVOLVED</sup> or a TIR<sup>EVOLVED</sup>. The diameter of the growth-inhibition zone was measured for each experiment. In all cases, a TIR<sup>EVOLVED</sup> conferred more resistant to ampicillin than TIR<sup>UNEVOLVED</sup>

**Table 2** Nucleotide sequences of the TIR<sup>UNEVOLVED</sup> and corresponding TIR<sup>EVOLVED</sup> used in this study. The TIR is defined as the region from the Shine-Dalgarno to codon five of the signal peptide

Signal peptide	TIR	Sequence
MalE <sup>SP</sup>	UNEVOLVED	GTTTAACTTTAAGAAGGAGATATACCGATGAAAATAAAACAGGTGCACGC
	EVOLVED	GTTTAACTTTAAGAAGGAGAT <u>CGTATTATGAAAATT</u> AAAACAGGTGCACGC
OmpA <sup>SP</sup>	UNEVOLVED	GTTTAACTTTAAGAAGGAGATATACCGATGAAAAAGACAGCTATCGCGATT
	EVOLVED	GTTTAACTTTAAGAAGGAGAT <u>GTTTCGTATGAAGAAG</u> ACAGCTATCGCGATT
PhoA <sup>SP</sup>	UNEVOLVED	GTTTAACTTTAAGAAGGAGATATACCGATGAAACAAAGCACTATTGCACTG
	EVOLVED	GTTTAACTTTAAGAAGGAGAT <u>TAAACGTATGAAGCAA</u> AGCACTATTGCACTG
DsbA <sup>SP</sup>	UNEVOLVED	GTTTAACTTTAAGAAGGAGATATACCGATGAAAAAGATTTGGCTGGCGCTG
	EVOLVED	GTTTAACTTTAAGAAGGAGAT <u>CGTAGGATGAAAAAG</u> ATTTGGCTGGCGCTG
PelB <sup>SP</sup>	UNEVOLVED	GTTTAACTTTAAGAAGGAGATATACCGATGAAATACCTGCTGCCGACCGCT
	EVOLVED	GTTTAACTTTAAGAAGGAGATAT <u>CAGGATGAAGTAT</u> CTGCTGCCGACCGCT

<sup>1</sup> Underlined region was randomised during the directed evolution process

<sup>2</sup> Nucleotides marked in bold text were changed in TIR<sup>SYN\_EVOLVED</sup>

<sup>3</sup> Shine-Dalgarno sequence marked in blue <sup>4</sup>Start methionine (ATG) marked in grey

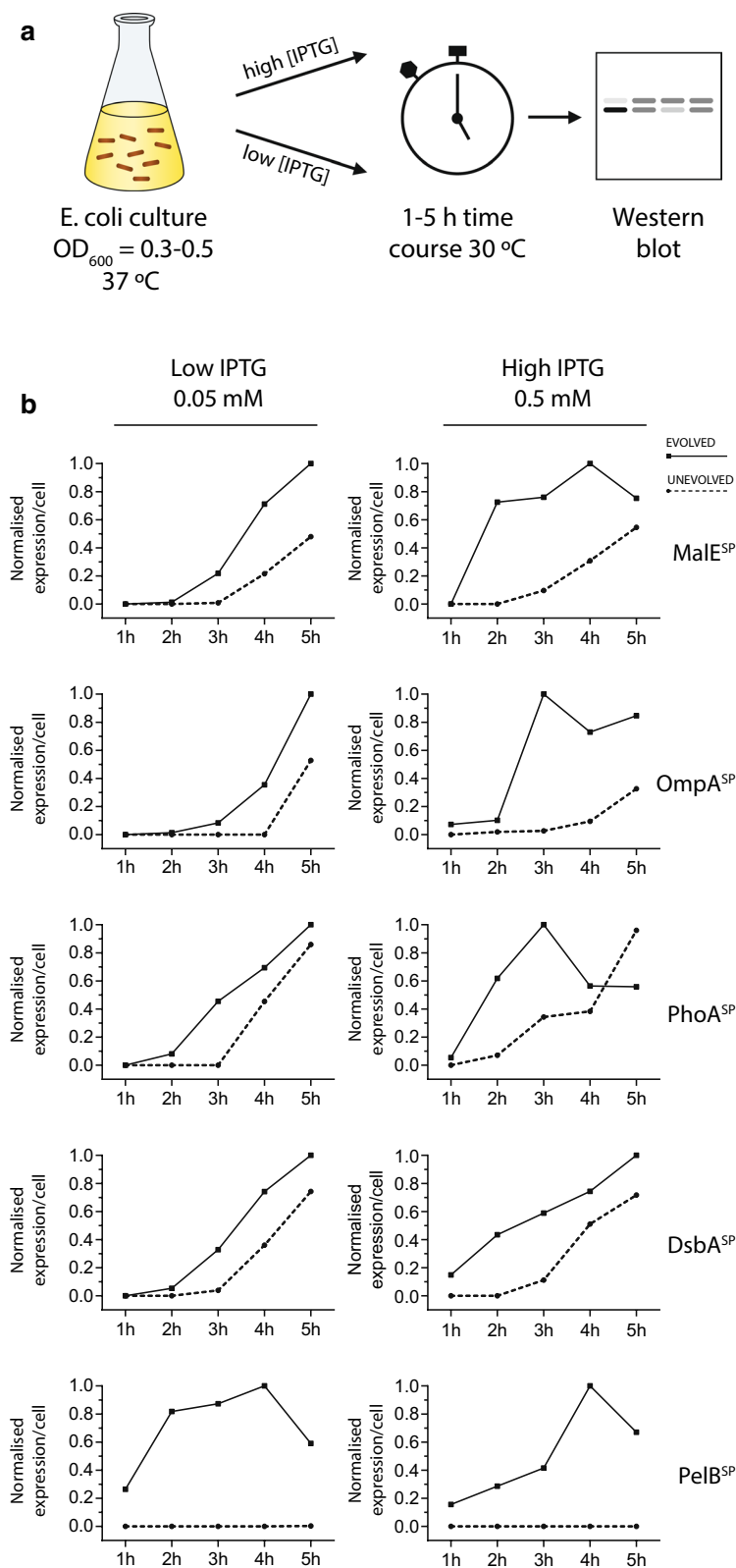
### Secretion is not a bottle-neck when synthesis levels are increased

In the previous series of experiments a mild induction protocol was used (0.05 mM IPTG for 2 h at 30 °C), so that differences in protein synthesis could be assessed in the absence of a metabolic load on the cell. The concern about metabolic load largely related to the Sec translocon, which is believed to be a bottleneck in the production of periplasmic proteins [31, 32]. When production levels of periplasmic proteins are too high, the translocon can become saturated and the recombinant protein may be retained in the cytoplasm. To determine if expression plasmids with a TIR<sup>EVOLVED</sup> would saturate the Sec translocon we monitored production over a 5-h period, as well as with a higher (0.5 mM) IPTG concentration (Fig. 3a). We observed higher production levels of  $\beta$ -lactamase at a higher IPTG concentration (data not shown). And at all but one late time-point, a TIR<sup>EVOLVED</sup> secreted more periplasmic  $\beta$ -lactamase than the corresponding TIR<sup>UNEVOLVED</sup> (Fig. 3b). This observation was made at

both low and high concentrations of IPTG. These time-course experiments therefore indicated that the Sec translocon was able to cope with the increased synthesis levels that were reached using a TIR<sup>EVOLVED</sup>.

### A TIR<sup>EVOLVED</sup> can be used as a generic solution

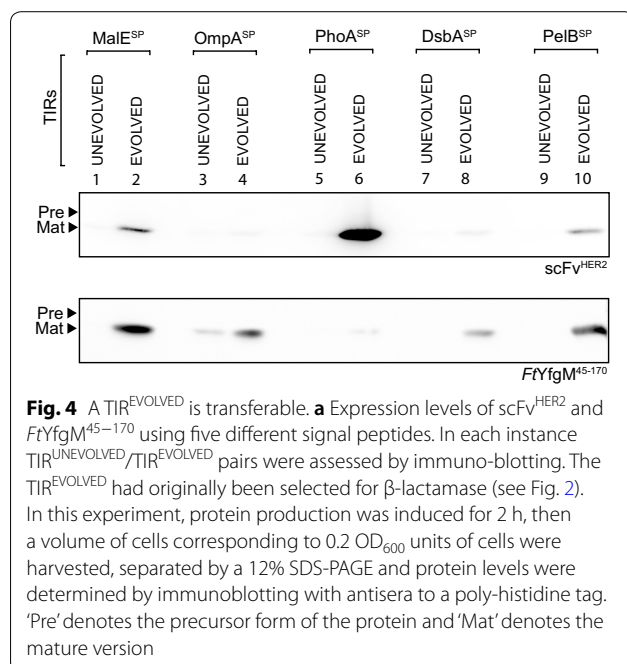
The five TIRs<sup>EVOLVED</sup> presented in the previous section were easily selected because  $\beta$ -lactamase is compatible with high-throughput screening. High-throughput screening can be more challenging with other recombinant periplasmic proteins, so it might not be feasible to screen a large library for a TIR<sup>EVOLVED</sup>. We therefore asked whether a TIR<sup>EVOLVED</sup> that we had selected using  $\beta$ -lactamase could be used to increase the production of other recombinant proteins in the periplasm of *E. coli*. In this set of experiments the coding sequences of scFv<sup>HER2</sup> and FtYfgM<sup>45–170</sup> were expressed as fusions to the original five signal peptides, using both the TIR<sup>UNEVOLVED</sup> and TIR<sup>EVOLVED</sup> pairs. The expression plasmids were again transformed into BL21(DE3) *pLysS* and production was



**Fig. 3** Time-course analysis of  $\beta$ -lactamase production. **a** An illustration of the experimental workflow used. **b** At each time point, a volume of cells was extracted, then separated by SDS-PAGE and immuno-blotted with antisera to  $\beta$ -lactamase. Band intensities were obtained from immuno-blots by densitometric analysis and normalised to the highest-value

monitored using a mild induction protocol (0.05 mM IPTG for 2 h at 30 °C). As we had observed for  $\beta$ -lactamase, the TIR<sup>EVOLVED</sup> always produced more protein than the corresponding TIR<sup>UNEVOLVED</sup> (Fig. 4). Note that production from each TIR<sup>UNEVOLVED</sup> was undetectable on these blots because the difference from the TIR<sup>EVOLVED</sup> was too large to capture. We still noted that signal peptide performance was varied; the most effective signal peptide for production of scFv<sup>HER2</sup> was PhoA<sup>SP</sup>, whilst the most effective for FtYfgM<sup>45–170</sup> was MalE<sup>SP</sup>. Thus, signal peptide performance might partly be explained by compatibility of the signal peptide.

A similar approach was taken to produce the human growth hormone (hGH). Here we observed that the most effective TIR<sup>EVOLVED</sup> for production of hGH was the one coupled to the PelB<sup>SP</sup> (Fig. 5a). To assess how much more protein was produced the N-terminally His-tagged hGH was purified by Immobilised Metal Affinity Chromatography (IMAC), the His-tag removed by proteolytic processing, and the sample polished by Size Exclusion Chromatography (SEC) (Fig. 5b). The yield of purified hGH was more than threefold higher using the TIR<sup>EVOLVED</sup> compared to the TIR<sup>UNEVOLVED</sup> (2.56 mg/L vs 0.79 mg/L). Importantly, we could not detect any difference in the quality of the purified hGH, as judged by monodispersity of the sample following SEC (Fig. 5c), the proportion of protein that had formed disulphide bonds (Fig. 5d), or the activity of the protein when tested by the MTS cell proliferation assay (Fig. 5e).



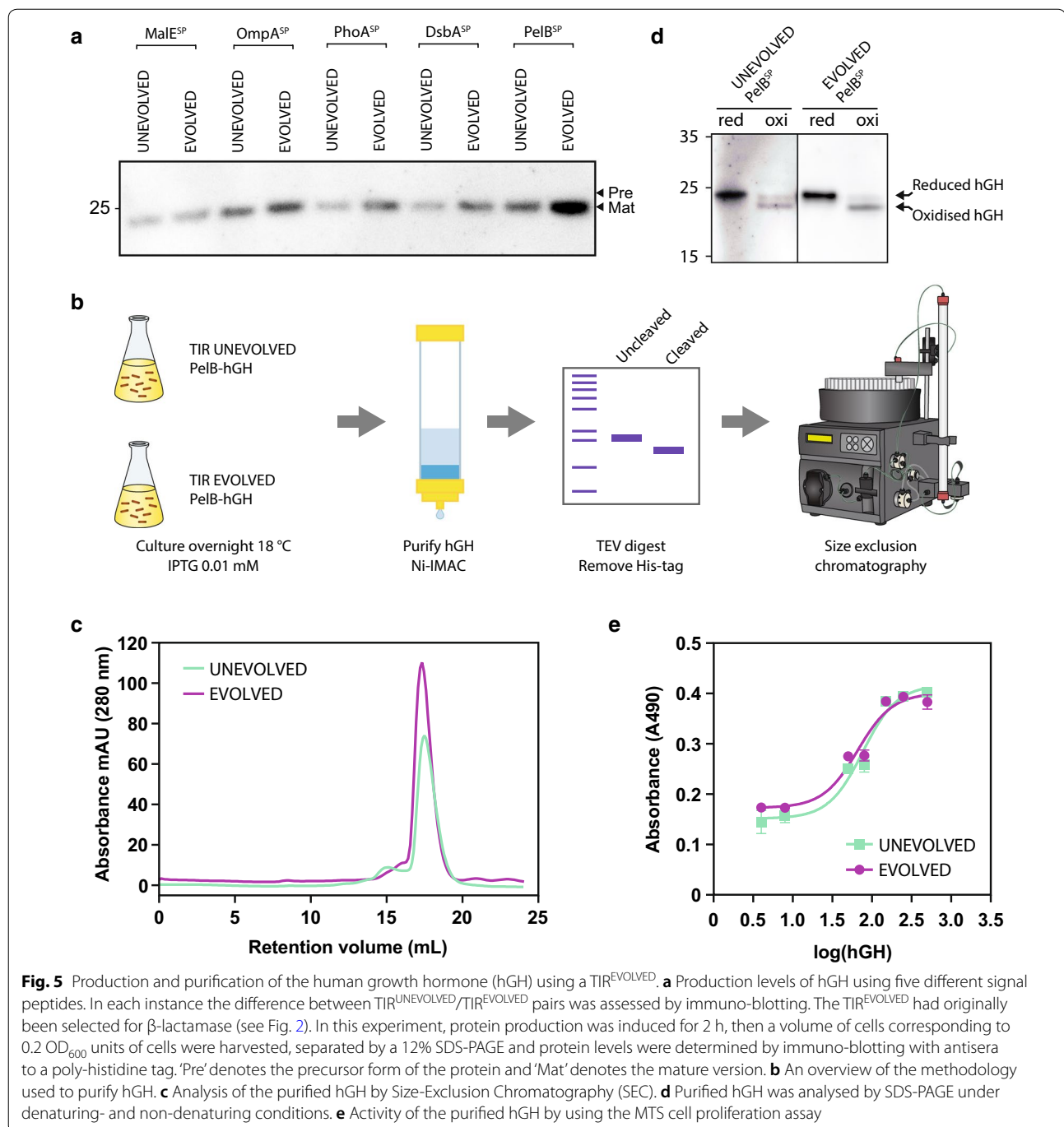
Taken together, this series of experiments indicate that the pET28a-based vectors containing signal peptides with an embedded TIR<sup>EVOLVED</sup> can be used as a generic solution to increase production of single chain antibody fragments, hormones and other recombinant proteins in the periplasm of *E. coli* without compromising protein quality.

## Discussion

Why does a given signal peptide support a high-level of synthesis and secretion for one recombinant protein but not another? In this study we compared production levels of recombinant periplasmic proteins using different signal peptides. Initially the TIR was generated by ad hoc cloning (TIR<sup>UNEVOLVED</sup>), but we also utilised a directed evolution to select the evolved TIRs (TIR<sup>EVOLVED</sup>). Although the different TIRs did not influence the amino acid sequence of the recombinant protein being produced, they did affect the production levels; in all cases increased production was observed with a TIR<sup>EVOLVED</sup>. The data demonstrate that the nucleotide sequence of the TIR contributes to signal peptide performance in bacterial cell factories. This conclusion is best exemplified by our experiments with the PelB<sup>SP</sup>, which gave the lowest levels of  $\beta$ -lactamase production from a TIR<sup>UNEVOLVED</sup> but the highest when expressed from a TIR<sup>EVOLVED</sup>. PelB<sup>SP</sup> had therefore been converted from a 'poor-performing' signal peptide to a 'top-performing' signal peptide without changing a single amino acid.

Two previous studies have noted that nucleotide changes in the signal sequence can influence signal peptide performance. Ng and Sarkar noted that synonymous changes to the Usp45sp signal peptide in *Lactococcus lactis* increased production levels of a nuclease and an amylase by approximately 15% [33]. And Punginelli and co-workers noted that non-synonymous nucleotide changes in the signal peptide of the Tat-dependent formate dehydrogenase increased production levels by up to 60-fold in *E. coli* [34]. Both of these studies suggested that increased production yields correlated with reduced mRNA structures in the TIR. We were unable to correlate production yields to predicted mRNA structures in our study (see also [29, 35]). Further work is required to understand how structures within the mRNA and interactions with the ribosome allow certain TIRs to support higher levels of recombinant protein production.

Is there more to understand about signal peptide performance? Although the performance of MalE<sup>SP</sup>, OmpA<sup>SP</sup>, PhoA<sup>SP</sup>, DsbA<sup>SP</sup> and PelB<sup>SP</sup> was improved by generating a TIR<sup>EVOLVED</sup>, we still noted differences in the production levels of the recombinant proteins. These differences could potentially be explained by (1) differences in the efficiency of translational elongation of the signal



peptides, or (2) compatibility between the signal peptide and the recombinant protein. The latter point is raised because the signal peptide acts to slow folding of the protein in the cytoplasm [16] so it could be expected that the physico-chemical characteristics of the signal peptide could impact on their performance. We therefore believe there is more to understand about 'fitting' of signal peptides with recombinant proteins. Until that point in time,

there is still merit in testing a small selection of different signal peptides.

Improving signal peptide performance by optimisation of the TIR is a simple and inexpensive way to increase the production yields of secreted proteins in bacterial cell factories. It will be compatible with other published methods; such as those that use titratable promoters to tune transcription rates of secreted proteins [36].



At present there are no general rules for identifying an optimal TIR. It is therefore common practice to use bioinformatic prediction programs, such as RBS calculator [37], UTR designer [38] and RBS designer [39]. Herein we used a directed evolution approach to screen large TIR<sup>LIBRARIES</sup>. We were able to screen these TIR<sup>LIBRARIES</sup> as we fused signal peptides to the  $\beta$ -lactamase protein, which confers resistance to  $\beta$ -lactam antibiotics. For proteins where no simple screening assay is available it is possible to directly couple [18], or translationally-couple  $\beta$ -lactamase to the recombinant protein [40–42]. It is also possible to use the pET28a-based TIRs<sup>EVOLVED</sup>, which we identified in this study and which improved production yields of a single chain antibody fragment, a hormone and another recombinant protein in *Escherichia coli*. However there may be limitations if these TIRs are transferred to other expression vectors, where the 5'UTR differs. Thus if other expression vectors are being used, we believe it would be better to create and screen a bespoke TIR<sup>LIBRARY</sup> as demonstrated here.

## Methods

### Molecular cloning

The sequences encoding MalE<sup>SP</sup>, OmpA<sup>SP</sup>, PhoA<sup>SP</sup>, DsbA<sup>SP</sup>, PelB<sup>SP</sup>,  $\beta$ -lactamase, hGH and FtYfgM<sup>45–170</sup> were chemically synthesised (Genscript, USA). The sequence encoding scFv<sup>HER2</sup> was obtained from the pHP2-15 plasmid [43]. All sequences are listed in Additional file 1: Table 1. To generate expression clones, the coding sequences and the pET28a vector were amplified by PCR using the Q5 polymerase (New England Biolabs, UK). The coding sequences were then cloned between the *NcoI* and *NdeI* restriction enzyme sites using the Gibson cloning method. Enzymes used for Gibson cloning were obtained from New England Biolabs, UK.

### Synthetic evolution of the TIR

TIR<sup>LIBRARIES</sup> were generated by amplifying the expression plasmids by PCR, using overlapping primers as previously described [29, 30]. The forward primer was approximately 45 nucleotides in length and was partly degenerate. The design enabled complete randomization of the six nucleotides upstream of the AUG start codon, and partial randomization of the six nucleotides downstream stream of the AUG start codon (synonymous codons only). The reverse primer was always the same sequence (5'-CTCCTTCTTAAAGTTAAACAA AATTATTTCTAGAGGGGAATTGTTATC-3'). It overlapped with the forward primer by 13 nucleotides thus allowing circularization of the PCR product by homologous recombination in *E. coli* MC1061. The PCR was carried out using the Q5 polymerase (New England Biolabs, UK) in a program that consisted of 94 °C for 5 min and

then 30 cycles of 95 °C for 45 s, 48–68 °C for 45 s (using a gradient thermocycler), 72 °C for 6 min and a final elongation step of 72 °C for 5 min. Specific PCR products that were amplified at the lowest annealing temperature were treated with *DpnI*, then transformed into chemically competent *E. coli* MC1061. The transformation was seeded into 100 mL of Luria–Bertani containing 50  $\mu$ g/mL kanamycin and incubated overnight at 37 °C. Isolation of the TIR<sup>LIBRARIES</sup> was carried out using ten E.N.Z.A DNA mini kit purification columns (Omega Biotek, USA) and pooling of the eluates.

TIR<sup>LIBRARIES</sup> were screened by transforming chemically competent BL21(*DE3*) *pLysS* and identifying clones that survived on the highest concentration of ampicillin. Here 0.5  $\mu$ g of the TIR<sup>LIBRARY</sup> was transformed into 50  $\mu$ L of chemically competent BL21(*DE3*) *pLysS* using standard protocols. The entire transformation was then seeded into 3 mL of LB containing 50  $\mu$ g/mL kanamycin and 34  $\mu$ g/mL chloramphenicol. Cultures were grown at 37 °C with shaking for 16 h. Cultures were then back-diluted (1:50) into 5 mL of LB containing 50  $\mu$ g/mL kanamycin and 34  $\mu$ g/mL chloramphenicol and incubated as before until an OD<sub>600</sub> of ~0.3 was reached. Expression of the coding sequence was induced by streaking a volume of cells corresponding to 0.002 OD<sub>600</sub> units on LB agar containing 0.05 mM isopropyl- $\beta$ -D thiogalactopyranoside (IPTG) and increasing concentrations of ampicillin (100–5000  $\mu$ g/mL). Note that kanamycin and chloramphenicol were omitted from the plates. The plates were then incubated for 16 h at 37 °C. Colonies formed at higher ampicillin concentrations were selected for further analysis and sequencing (Eurofins MWG operon, Germany).

### Immuno-blotting

Cultures were grown at 37 °C with shaking for 16 h, then back-diluted (1:50) into 5 mL of LB containing 50  $\mu$ g/mL kanamycin and 34  $\mu$ g/mL chloramphenicol and incubated as before until an OD<sub>600</sub> of ~0.3 to 0.5 was reached. Expression of the coding sequence was induced with 0.05 mM IPTG for 2 h at 30 °C. A volume of cells corresponding to an OD<sub>600</sub> of either 0.02 or 0.2 was harvested by centrifugation then resuspended in 2  $\times$  Laemmli loading buffer [125 mM Tris–HCl pH 6.8, 4% SDS, 3% Glycerol, 0.02% bromphenol blue, 20%  $\beta$ -mercaptoethanol]. Proteins were separated by 12% SDS-PAGE then transferred to a nitrocellulose membrane using a semi-dry transfer apparatus (Bio-Rad, USA). The nitrocellulose membranes were probed with an antibody against either  $\beta$ -lactamase (Thermo Scientific, USA) or the poly-histidine tag (His-Probe, ThermoFisher Scientific, USA). Binding was detected using anti-mouse IgG linked to horseradish peroxidase (GE healthcare, USA) and a SuperSignal West femto luminol/enhancer solution

(ThermoFisher Scientific, USA). Luminescence emitting from the nitrocellulose membrane was detected using an Azure Biosystems c600 device.

#### Disc diffusion assays

Cells were grown in LB containing 50 µg/mL kanamycin and 34 µg/mL chlormphenicol until an OD<sub>600</sub> of ~0.3. A volume of cells corresponding to an OD<sub>600</sub> of 0.002 was then plated onto LB agar (lacking all antibiotics). A sterile filter disc containing 2 mg ampicillin was then placed on top of the cells and the plates were incubated at 37 °C for 16 h. Zones of growth inhibition were measured using a standard ruler.

#### Purification of hGH

Expression plasmids harboring pET28a *pelB-hGH* were transformed into the expression host BL21(DE3) pLysS and grown on LB agar plates containing 50 µg/mL kanamycin and 34 µg/mL chloramphenicol. Single colonies were used to inoculate 100 mL of LB plus antibiotics medium which was grown overnight at 37 °C with shaking at 180 RPM. Overnight pre-cultures were used to inoculate 2 L flasks containing 1 L of LB media plus antibiotics, to a starting OD<sub>600</sub> of 0.05. Cultures were grown to an OD<sub>600</sub> of 0.7, at which point, flasks were incubated on ice for 10 min. Induction proceeded with the addition of 0.01 mM IPTG and incubation for 16 h at 18 °C with shaking at 180 RPM. Cells were harvested for 20 min at 4000g. Cell pellets were resuspended in 50 mL suspension buffer [50 mM Tris pH 8.0, 500 mM NaCl, 20 mM imidazole pH 8.0 and 1x protease inhibitor cocktail (cOmplete, Roche, USA)]. Cell suspensions were homogenized with a glass dounce homogenizer followed by cell disruption using an Avestin emulsiflex C3 high-pressure homogenizer (Avestin, Canada). Cell debris was removed by centrifugation at 20,000g for 30 min. Samples were applied to 2.5 mL Ni-Sepharose (GE Healthcare) and batch incubated at 4 °C for 1 h on a benchtop roller. The column was washed with 20 column volumes (50 mL) of wash buffer (50 mM Tris pH 8.0, 500 mM NaCl and 50 mM imidazole pH 8.0), followed by elution with 30 mL of elution buffer (50 mM Tris pH 8.0, 500 mM NaCl and 500 mM imidazole pH 8.0). The elution fraction was concentrated and buffer exchanged (50 mM Tris pH 8.0, 150 mM NaCl and 20 mM imidazole) using a centrifugal filter with a nominal MWCO of 10 kDa (Amicon, Merck Millipore). The N-terminal his-tag was proteolytically removed with TEV protease (purified in-house) at a 1:10 weight ratio and allowed to incubate overnight at 4 °C. Samples were reverse Ni purified, concentrated and applied to size exclusion chromatography using a Superdex 200 10/300 GL column (GE Healthcare, Sweden) in 50 mM

Tris pH 8.0 and 100 mM NaCl. Relevant fractions were pooled, and concentrated. Sample concentration was measured by the BCA protein assay kit (Pierce, ThermoFisher Scientific, USA) and protein quality assessed by SDS-PAGE. Calculation of final yield per liter was determined by accounting of final volume, final OD at the conclusion of expression, and final concentration of purified hGH.

#### MTS cell proliferation assay

The breast cancer MCF7 cell line (ATCC) was maintained in RPMI-1640 medium containing 10% FBS, 2 mM glutamine and 1% penicillin streptomycin (Gibco/Thermo Fisher Scientific) at 37 °C in a humidified atmosphere at 5% CO<sub>2</sub>. Cell proliferation following titration of purified hGH was determined according to the CellTiter 96 AQueous Non-Radioactive Cell Proliferation assay (MTS) protocol (Promega). Briefly, 1 × 10<sup>4</sup> MCF7 cells were seeded in triplicate, in 100 µL aliquots into 96 well plates, followed by serum starvation for 24 h, prior to commencing the proliferation assay. Serially diluted hGH was added to the medium at a final concentration ranging from 0 to 400 ng/mL. Cell proliferation was assessed after 48 h of incubation, by addition of MTS and the electron coupling reagent PMS. The conversion of MTS to formazan was measured by absorbance at 490 nm using a SpectraMax plate reader. Background absorbance was corrected by subtraction of wells containing RPMI. hGH EC50 was calculated using GraphPad Prism 8.1.0.

#### Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12934-020-01339-8>.

**Additional file 1: Fig. 1.** Selection of clones from TIR<sup>LIBRARIES</sup>. **Figure 2.** Expression levels of β-lactamase using TIRs selected from TIR<sup>LIBRARIES</sup>. **Table 1.** Coding sequences used in this study

#### Acknowledgements

Open access funding provided by Stockholm University. We are grateful to Johan Rockberg (KTH, Sweden) for the pHP2-15 plasmid that encoded the scFv<sup>HER2</sup>.

#### Authors' contributions

KM, PJS and DOD conceived the study and designed the experiments. KM, PJS, AJC, RE and HLC acquired the data. All authors interpreted the data. KM, PJS and DOD wrote the manuscript. All authors read and approved the final manuscript.

#### Funding

DOD was supported by a grant from the Swedish Research Council (2017-00704) and Carl Trygger stiftelse (CTS 18:78). MR and MHHN were supported by a grant from the Novo Nordisk Foundation.

#### Availability of data and materials

All data generated or analysed during this study are included in this published article [and its additional files].

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The directed evolution process used in the study is patent protected (European Patent no. 3234146; US application US15/535,686). Specific TIRs identified in the study are patent pending.

**Author details**

<sup>1</sup> Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden. <sup>2</sup> CloneOpt AB, Stockholm, Sweden. <sup>3</sup> Department of Medicine (Solna), Division of Microbial Pathogenesis, Bioclinicum, Karolinska Institutet, Stockholm, Sweden. <sup>4</sup> Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kgs. Lyngby, Denmark.

Received: 11 February 2020 Accepted: 23 March 2020

Published online: 07 April 2020

**References**

- Walsh G. Biopharmaceutical benchmarks 2014. *Nat Biotechnol*. 2014;32(10):992–1000.
- Singh R, Kumar M, Mittal A, Mehta PK. Microbial enzymes: industrial progress in 21st century. *3 Biotech*. 2016;6(2):174.
- Freudl R. Signal peptides for recombinant protein secretion in bacterial expression systems. *Microb Cell Factories*. 2018;17(1):52.
- Low KO, Muhammad Mahadi N, Illias R. Optimisation of signal peptide for recombinant protein secretion in bacterial hosts. *Appl Microbiol Biotechnol*. 2013;97(9):3811–26.
- Berkmen M. Production of disulfide-bonded proteins in *Escherichia coli*. *Protein Expr Purif*. 2012;82(1):240–51.
- Manta B, Boyd D, Berkmen M. Disulfide Bond Formation in the Periplasm of *Escherichia coli*. *EcoSal Plus*. 2019;8(2). <http://www.asmscience.org/content/journal/ecosalplus/10.1128/ecosalplus.ESP-0012-2018>. Accessed 24 Apr 2019.
- Crane JM, Randall LL. The sec system: protein export in *Escherichia coli*. *EcoSal Plus*. 2017; 7(2). <http://www.asmscience.org/content/journal/ecosalplus/10.1128/ecosalplus.ESP-0002-2017>. Accessed 24 Apr 2019.
- Tsirigotaki A, De Geyter J, Šoštaric' N, Economou A, Karamanou S. Protein export through the bacterial sec pathway. *Nat Rev Microbiol*. 2017;15(1):21–36.
- von Heijne G. Signal sequences. The limits of variation. *J Mol Biol*. 1985;184(1):99–105.
- von Heijne G. The signal peptide. *J Membr Biol*. 1990;115(3):195–201.
- Power PM, Jones RA, Beacham IR, Bucholtz C, Jennings MP. Whole genome analysis reveals a high incidence of non-optimal codons in secretory signal sequences of *Escherichia coli*. *Biochem Biophys Res Commun*. 2004;322(3):1038–44.
- Zalucki YM, Beacham IR, Jennings MP. Biased codon usage in signal peptides: a role in protein export. *Trends Microbiol*. 2009;17(4):146–50.
- Zalucki YM, Power PM, Jennings MP. Selection for efficient translation initiation biases codon usage at second amino acid position in secretory proteins. *Nucleic Acids Res*. 2007;35(17):5748–54.
- Zalucki YM, Gittins KL, Jennings MP. Secretory signal sequence non-optimal codons are required for expression and export of beta-lactamase. *Biochem Biophys Res Commun*. 2008;366(1):135–41.
- Zalucki YM, Beacham IR, Jennings MP. Coupling between codon usage, translation and protein export in *Escherichia coli*. *Biotechnol J*. 2011;6(6):660–7.
- Park S, Liu G, Topping TB, Cover WH, Randall LL. Modulation of folding pathways of exported proteins by the leader sequence. *Science*. 1988;239(4843):1033–5.
- Gouridis G, Karamanou S, Gelis I, Kalodimos CG, Economou A. Signal peptides are allosteric activators of the protein translocase. *Nature*. 2009;462(7271):363–7.
- Selas Castiñeiras T, Williams SG, Hitchcock A, Cole JA, Smith DC, Overton TW. Development of a generic  $\beta$ -lactamase screening system for improved signal peptides for periplasmic targeting of recombinant proteins in *Escherichia coli*. *Sci Rep*. 2018;8(1):6986.
- Laursen BS, Sørensen HP, Mortensen KK, Sperling-Petersen HU. Initiation of protein synthesis in bacteria. *Microbiol Mol Biol Rev*. 2005;69(1):101–23.
- Milón P, Rodnina MV. Kinetic control of translation initiation in bacteria. *Crit Rev Biochem Mol Biol*. 2012;47(4):334–48.
- Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*. 2009;324(5924):218–23.
- McCarthy JE, Gualerzi C. Translational control of prokaryotic gene expression. *Trends Genet TIG*. 1990;6(3):78–85.
- Kozak M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene*. 2005;21(361):13–37.
- Goodman DB, Church GM, Kosuri S. Causes and effects of N-terminal codon bias in bacterial genes. *Science*. 2013;342(6157):475–9.
- Bentele K, Saffert P, Rauscher R, Ignatova Z, Blüthgen N. Efficient translation initiation dictates codon usage at gene start. *Mol Syst Biol*. 2013;18(9):675.
- Scharff LB, Childs L, Walther D, Bock R. Local absence of secondary structure permits translation of mRNAs that lack ribosome-binding sites. *PLoS Genet*. 2011;7(6):e1002155.
- Kudla G, Murray AW, Tollervey D, Plotkin JB. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science*. 2009;324(5924):255–8.
- Reeve B, Hargest T, Gilbert C, Ellis T. Predicting translation initiation rates for designing synthetic biology. *Front Bioeng Biotechnol*. 2014;2:1.
- Mirzadeh K, Martínez V, Toddo S, Guntur S, Herrgård MJ, Elofsson A, et al. Enhanced protein production in *Escherichia coli* by optimization of cloning scars at the vector-coding sequence junction. *ACS Synth Biol*. 2015;4(9):959–65.
- Daley D, Mirzadeh K, Toddo S, Guntur S, Ab C. Selective optimisation of a ribosome binding site for protein production. 2015. <https://www.google.com/patents/WO2016099388A1?cl=da>. Accessed 16 Mar 2018.
- Hjelm A, Karyolimos A, Zhang Z, Rujas E, Vikström D, Slotboom DJ, et al. Tailoring *Escherichia coli* for the l-Rhamnose PBAD promoter-based production of membrane and secretory proteins. *ACS Synth Biol*. 2017;6(6):985–94.
- 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 Factories*. 2013;12:24.
- Ng DTW, Sarkar CA. Engineering signal peptides for enhanced protein secretion from lactococcus lactis. *Appl Environ Microbiol*. 2013;79(1):347–56.
- Punginelli C, Ize B, Stanley NR, Stewart V, Sawers G, Berks BC, et al. mRNA secondary structure modulates translation of tat-dependent formate dehydrogenase N. *J Bacteriol*. 2004;186(18):6311–5.
- Bonde MT, Pedersen M, Klausen MS, Jensen SI, Wulff T, Harrison S, et al. Predictable tuning of protein expression in bacteria. *Nat Methods*. 2016;13(3):233–6.
- Karyolimos A, Ampah-Korsah H, Hillenaar T, Mestre Borrás A, Dolata KM, Sievers S, Riedel K, Daniels R, de Gier JW. Enhancing recombinant protein yields in the *E. coli* periplasm by combining signal peptide and production rate screening. *Front Microbiol*. 2019;10:1511.
- Salis HM, Mirsky EA, Voigt CA. Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol*. 2009;27(10):946–50.
- Seo SW, Yang J-S, Kim I, Yang J, Min BE, Kim S, et al. Predictive design of mRNA translation initiation region to control prokaryotic translation efficiency. *Metab Eng*. 2013;15:67–74.
- Na D, Lee D. RBSDesigner: software for designing synthetic ribosome binding sites that yields a desired level of protein expression. *Bioinforma Oxf Engl*. 2010;26(20):2633–4.
- Rennig M, Martínez V, Mirzadeh K, Dunas F, Röjsäter B, Daley DO, et al. TARSyn: tunable antibiotic resistance devices enabling bacterial synthetic evolution and protein production. *ACS Synth Biol*. 2018;7(2):432–42.
- Rennig M, Daley DO, Nørholm MHH. Selection of highly expressed gene variants in *Escherichia coli* using translationally coupled antibiotic selection markers. *Methods Mol Biol Clifton NJ*. 2018;1671:259–68.
- Ferro R, Rennig M, Hernández-Rollán C, Daley DO, Nørholm MHH. A synbio approach for selection of highly expressed gene variants in gram-positive bacteria. *Microb Cell Factories*. 2018;17(1):37.

43. Hu FJ, Volk A-L, Persson H, Säll A, Borrebaeck C, Uhlen M, et al. Combination of phage and Gram-positive bacterial display of human antibody repertoires enables isolation of functional high affinity binders. *New Biotechnol.* 2018;25(45):80–8.

### **Publisher's Note**

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

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

