Heliyon 10 (2024) e26996

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

SAS: Split antibiotic selection for identifying chaperones that improve protein solubility

Emily McNutt, Na Ke, Alexandre Thurman, James B. Eaglesham, Mehmet Berkmen *

New England Biolabs, 240 County Road, Ipswich, MA 01938, USA

ARTICLE INFO

Keywords:
E. coli
Recombinant protein expression
Chaperones
Genetic selection

ABSTRACT

Background: Heterologous expression of active, native-folded protein in *Escherichia coli* is critical in both academic research and biotechnology settings. When expressing non-native recombinant proteins in *E. coli*, obtaining soluble and active protein can be challenging. Numerous techniques can be used to enhance a proteins solubility, and largely focus on either altering the expression strain, plasmid vector features, growth conditions, or the protein coding sequence itself. However, there is no one-size-fits-all approach for addressing issues with protein solubility, and it can be both time and labor intensive to find a solution. An alternative approach is to use the co-expression of chaperones to assist with increasing protein solubility. By designing a genetic system where protein solubility is linked to viability, the appropriate protein folding factor can be selected for any given protein of interest. To this end, we developed a Split Antibiotic Selection (SAS) whereby an insoluble protein is inserted in-frame within the coding sequence of the hygromycin B resistance protein, aminoglycoside 7″-phosphotransferase-Ia (APH(7″)), to generate a tripartite fusion. By creating this tripartite fusion with APH(7″), the solubility of the inserted protein can be assessed by measuring the level of hygromycin B resistance of the cells.

Results: We demonstrate the functionality of this system using a known protein and co-chaperone pair, the human mitochondrial Hsp70 ATPase domain ($ATPase^{70}$) and its co-chaperone human escort protein (Hep). Insertion of the insoluble $ATPase^{70}$ within APH(7") renders the tripartite fusion insoluble and results in sensitivity to hygromycin B. Antibiotic resistance can be rescued by expression of the co-chaperone Hep which assists in the folding of the APH(7")-$ATPase^{70}$-APH(7") tripartite fusion and find that cellular hygromycin B resistance correlates with the total soluble fusion protein. Finally, using a diverse chaperone library, we find that SAS can be used in a pooled genetic selection to identify chaperones capable of improving client protein solubility.

Conclusions: The tripartite APH(7") fusion links the *in vivo* solubility of the inserted protein of interest to hygromycin B resistance. This construct can be used in conjunction with a chaperone library to select for chaperones that increase the solubility of the inserted protein. This selection system can be applied to a variety of client proteins and eliminates the need to individually test chaperone-protein pairs to identify those that increase solubility.

* Corresponding author.
E-mail address: berkmen@neb.com (M. Berkmen).

https://doi.org/10.1016/j.heliyon.2024.e26996
Received 12 October 2023; Received in revised form 15 February 2024; Accepted 22 February 2024
Available online 1 March 2024

1. Introduction

Escherichia coli is one of the most widely used organisms for protein production. Despite decades of progress in heterologous protein expression, the folding of many proteins, especially large multi-domain eukaryotic proteins, remains a significant challenge [1]. Heterologous protein production can account for up to 50% of *E. coli's* total cellular protein content and thus puts strain on the native transcriptional and translational machinery [2]. Additionally, the speed of coupled transcription and translation in bacteria can be problematic for producing large eukaryotic proteins. Key protein folding factors from the native host are also absent, causing further protein folding problems. Combined, these factors can lead to misfolding and/or aggregation of the recombinant protein [3]. Over the years many strategies have been developed to improve heterologous protein production [1,4,5].

For each recombinant protein being expressed, numerous variables influence the amount of soluble protein being produced by the cell including the protein sequence, expression vector, cell line, and expression conditions. Many of these factors may need to be optimized to achieve expression of active, soluble protein in *E. coli*. Upon selecting the protein of interest (POI) to be expressed, several avenues to improve protein solubility can be explored. A protein sequence can be altered by eliminating rare codons, reducing stretches of hydrophobic residues, or simply by selecting alternative homologs of the desired protein to increase chances of successful expression.

[6–8]. Any changes to the protein sequence itself come with the risk of altering the activity of the protein. Once a protein sequence has been selected, the appropriate expression system must be designed. The promoter, origin of replication, and the inclusion of solubility tags (MBP, DsbC, SUMO, NusA, etc.) can all be altered to enhance protein expression [6,9–12]. After designing and building the vector harboring the POI, an *E. coli* expression strain must be chosen. There are many strains that have been engineered for the purpose of heterologous protein expression which are reviewed here [13]. The choice of strain can have a profound impact on protein expression, and it is often difficult to predict which strain of *E. coli* will give the best results. Finally, the growth conditions of the strain expressing the recombinant protein need to be optimized. The growth media, temperature, concentration of inducer, and time of induction can all be altered to maximize protein production.

Optimizing the multiple parameters involved in protein expression can be time consuming and costly. The previously described methods modify either the expression conditions, the protein, the expression vector, or the strain in which the proteins are expressed. An additional approach is to use the co-expression of molecular chaperones to improve recombinant protein expression. Chaperones are proteins which interact with and assist client proteins during the transition from unfolded to natively folded protein, without being present in the final complex. Chaperones are frequently used to enhance the expression of heterologous proteins [14,15]. Two approaches are often used: 1) overexpressing native *E. coli* chaperones which may be at capacity during recombinant expression of a client protein [16,17] and 2) expressing chaperones not native to *E. coli* [18–21]. However, there is no single chaperone that will universally enhance heterologous protein expression, thus chaperones assisting in expression of a POI are usually identified empirically, which is a time-consuming process.

Here, we developed a genetic selection capable of identifying chaperones that enhance solubility of a client protein from a diverse chaperone library by linking *in vivo*, the solubility of POI to cell viability. This genetic linkage can be achieved in several ways. For example, a POI can be directly fused to the N-terminus of an antibiotic resistance protein (e.g. kanamycin [22] or chloramphenicol [23]). While applicable to a broad range of proteins in high-throughput genetic selections, there are drawbacks to using these single-fusion systems. Chiefly, that proteolytic cleavage events can separate the reporter from the protein, uncoupling antibiotic resistance from protein solubility [24,25]. An alternative approach involves creation of a tripartite fusion in which a POI is fused to multiple partners which must function in concert to produce a selectable or screenable *in vivo* phenotype. The first tripartite selection system used a fusion of the twin-arginine translocation (Tat) sequence, a POI, and the β-lactamase gene [26]. Proteins must be fully folded to be transported to the periplasm by the Tat machinery. Directing this construct through the Tat pathway serves as an internal quality control step that directly indicates the folding state of the POI. The β-lactamase protein is only active in the periplasm. Therefore, tripartite fusion will only confer ampicillin resistance once it has been correctly folded and transported to the periplasm via the Tat pathway. Using this system, DeLisa et al., assessed a variant library of the Alzheimer's Aβ42 peptide and demonstrated that they could select for proteins with enhanced solubility by selecting for ampicillin resistant colonies [26]. However, β-lactamase activity could still be separated from protein solubility if proteolytic cleavage occurs in the periplasm.

A more generally applicable tripartite system involves embedding a POI in-frame within an antibiotic resistance gene [27]. β-lactamase tolerates insertion of a foreign protein in a surface exposed loop. Neither half of the split protein is capable of providing resistance to the β-lactam antibiotic penicillin V in isolation [28], but insertion of a soluble POI at this split site results in a functional β-lactamase protein capable of conferring antibiotic resistance. Conversely, if the inserted protein is insoluble the complex will be misfolded and potentially degraded, resulting in the cells exhibiting sensitivity to antibiotics. Bardwell et al. demonstrated that the relative solubility of the inserted protein correlated with penicillin V resistance allowing for selection of proteins with increased solubility from a mutant library. Other iterations of the periplasmic tripartite fusion systems have been used for the selection of increased protein solubility by linking cadmium resistance, ampicillin resistance, or maltose metabolism to protein solubility [29,30] Cytoplasmic tripartite fusions have also been developed and used split constructs of aminoglycoside-3′-phosphotransferase IIa (APH (3′)), adenylyltransferase (ANT(3″)), and nourseothricin acetyltransferase (NAT), which confer resistance to kanamycin, spectinomycin and nourseothricin, respectively [31].

Thus far these tripartite selection systems have been used in combination with a mutagenic approach for increasing the POI solubility, either by mutagenizing POI itself or mutagenizing the *E. coli* chromosome. The objective of this study was to design a cytoplasmic selection system that couples protein solubility to viability and use this system in combination with a curated library of chaperones. Thus, by selecting for antibiotic resistance, one can rapidly and accurately identify chaperones that increase POI

solubility. To accomplish this, we created a unique tripartite construct, where the human Hsp70 ATPase domain (ATPase^{70}), a model eukaryotic insoluble POI, was inserted into the hygromycin B resistance gene aminoglycoside 7″-phosphotransferase-Ia (APH(7″)). Using a library of constitutively expressed chaperones, we demonstrate selection of a protein-folding factor capable of enhancing ATPase^{70} solubility – its native co-chaperone Hep. We further show that the hygromycin resistance conferred by the ATPase^{70}–APH (7") tripartite fusion correlates with the solubility of the fusion construct. We anticipate our selection system and chaperone library may be broadly applicable to different POI and potentially enable identification of protein-folding factors capable of enhancing expression of insoluble protein constructs.

2. Methods

2.1. Culture growth conditions

For liquid cultures, cells were grown in 5 ml rich media (10 g/L soy peptone, 5 g/L yeast extract, 5 g/L NaCl, NaOH to pH 7.2) in test tubes at 37 °C with 30 μg/ml chloramphenicol, 40 μg/ml kanamycin, or a combination of the two added to the media when appropriate. For solid cultures, cells were grown in 25 ml rich agar plates (rich media with 1.5% bacteriological agar) in the presence of appropriate antibiotics. When cells were grown under selective conditions hygromycin B (50 μg/ml) was added to the media.

2.2. Strain and plasmid construction

Bacterial strains, plasmids and primers used in this work are described in Supplementary Tables S1 and S2, and were constructed using standard molecular and genetic techniques [32]. All cloned plasmids were constructed using NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs, cat# E2621S). NEB T7 Express competent *E. coli* (New England Biolabs, cat# C2566) was used for selection and protein expression studies. NEB 10-beta competent *E. coli* (New England Biolabs, cat# C3019) was used for plasmid cloning procedures and transformed by heat shock, following manufacturer's instructions. Transformed cells were then selected on rich agar plates containing the appropriate antibiotic.

The tripartite fusion was constructed using the pAL vector described by Studier et al. [33], referred to here as pST7a. The hygromycin B resistance gene, aminoglycoside 7″-phosphotransferase-Ia (GenBank ID: WP_063842185.1), was ordered from IDT and cloned into the pST7a vector. pST7a was cut with restriction enzymes *Nco*I (New England Biolabs, cat# R0193S) and *Xba*I (New England Biolabs, cat# R0145S), APH(7") was amplified using primers NK1036 and NK1041 and cloned into pST7a, creating pST7a-HygR. The ATPase domain of human mitochondrial Hsp70 (GenBank ID: AAA67526.1) was ordered from IDT and cloned into the pST7a-HygR at position G108 of APH(7″) using primers NK1109 and NK1110. The resulting plasmid was called pST7a-hygRN-mtHsp70ATPase-hygRC.

The pACYC derivative pAL0 was used to express the library of protein folding factors. pAL0 was made by cloning the *rrnB* promoter sequence (ordered from IDT) into pBAD34. Restriction enzymes *Nsi*I (New England Biolabs, cat#R0127) and *Nco*I (New England Biolabs, cat#R0193S) were used to linearize pBAD34. The resulting pAL0 vector carries a strong constitutive promoter and chloramphenicol-resistant marker.

Protein folding factors were cloned into pAL0 that was digested with *Nco*I and *Sal*I (New England Biolabs, cat#R0138S). Proteins native to *E. coli* were cloned from genomic DNA extracted from *E. coli* K-12 using the Monarch® Genomic DNA Purification Kit (New England Biolabs, cat# T3010S). Protein folding factors from other organisms were codon-optimized for *E.coli* expression and ordered from IDT. All primers used for cloning the pAL library are listed in Table S2.

2.3. Spot titer assays

Cultures were grown overnight at 37 °C in rich media with the appropriate antibiotic. The cells were normalized to $OD_{600} = 1.0$ with rich media and subsequently serial diluted in 10-fold steps from 10° down to 10^{-6}. 2 μL of cells were spotted onto rich agar (growth controls) and rich agar supplemented with hygromycin B. Cells were grown at 37 °C for 24 or 48 h, respectively.

2.4. Chaperone selection

NEB T7 Express competent *E. coli* (New England Biolabs, cat# C2566) cells expressing the tripartite APH(7″)-ATPase^{70}-APH(7″) construct (strain MB6827 in Table S1) were transformed with either 100 ng of the pAL-chaperone library or 50 ng of empty pAL0 vector as a negative control via electroporation. Cells were recovered in 1 mL of SOC and grown for 2 h at 30 °C. The transformants were serially diluted in 10-fold increments from 10° to 10^{-6}. 100 μl of each dilution was replica plated onto non-selective and selective media. The transformants were incubated at 37 °C for 24 h (growth controls) or 48 h under hygromycin B selection. Colonies were counted on the non-selective plates to obtain the viable count. Colonies were counted on selective plates to determine the signal to noise ratio and percentage of colonies that were hygromycin B resistant.

Hygromycin B resistant colonies were patched on selective media and grown overnight at 37 °C to confirm hygromycin B resistance. Hygromycin B resistant patches were inoculated into rich media with appropriate antibiotics and grown overnight at 37 °C. Plasmids were extracted from the overnight cultures using the Monarch® Plasmid Miniprep Kit (New England Biolabs, cat# T1010S) and sequenced using primers EM1 and EM2 to identify the chaperone suppressing hygromycin B sensitivity.

2.5. Western blotting

To assay the solubility of the APH(7″) tripartite fusion construct, overnight cultures of 5 mL rich media with the appropriate antibiotics were grown and 250 μL was used as the inoculum for 25 ml rich media shake flask cultures and grown at 37 °C until OD_{600} reached 0.8, at which time expression of the tripartite fusion construct was induced with 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG). The cultures were then grown for an additional 2 h at 37 °C.

From each 25 mL culture, 1 mL of cells were harvested by centrifugation and cell densities were then standardized to the lowest OD_{600} value using lysis buffer (1X PBS supplemented with1X Halt™ Protease Inhibitor Thermo cat#87786) and lysed by sonication (QSONICA Q500 sonicator 4-Tip Horn with 1/8″ tips: 30% amplitude, 2 min total sonication time, pulses of 2 s on – 4 s off). The lysates were then clarified by centrifugation for 60 min at 20,000 g at 4 °C. Total protein in the soluble fraction was quantified by Nanodrop (1 mg = 1 A280) and standardized to the lowest protein concentration. Samples were mixed with protein loading dye (New England Biolabs, cat#B7703S), heated for 5 min at 95 °C, and 10 μl of each sample was loaded into a Novex™ WedgeWell™ 10–20% Tris-Glycine gel (Invitrogen cat#XP10205BOX). Proteins were transferred using Trans-Blot®Turbo™ Midi 0.2 μM PVDF Transfer Packs (BioRad cat#1704157). A semi-dry transfer was done using the Trans-Blot®Turbo™ Transfer System (BioRad#1704150) at 25 V for 10 min. Following transfer, the blocking, washes, and antibody treatment steps were performed with Intercept® (PBS) Blocking Buffer and Diluent kit (Li-Cor cat#927–76003) according to the manufacturer's protocol. Mouse anti-FLAG primary antibody (Cell Signaling Technology® cat#14793) was used in a 1:5000 dilution to detect FLAG-tagged constructs and incubated with gentle rocking for 1 h at room temperature. IRDye® 680RD Goat anti-Rabbit (Li-Cor cat#926–68071) was used as secondary antibody and diluted to 1:10,000. The secondary antibody was incubated for 1 h at room temperature with gentle rocking. The blot was scanned with a Licor Odyssey Infrared Imaging System.

3. Results

3.1. Designing a split enzyme construct to link protein solubility to antibiotic resistance

Typically, recombinant proteins in *E. coli* are expressed in the cytoplasm. We sought to create a selection system capable of identifying chaperones that solubilize recombinant POI *in situ*, avoiding the need for laborious optimization steps which may often fail to improve protein solubility [5]. Thus, we designed a tripartite fusion in which a client protein is embedded within a the hygromycin B resistance gene, APH(7"), an antibiotic resistance gene which functions in the cytoplasm (Fig. 1A) [34]. We have termed this approach SAS for Split Antibiotic Selection. APH(7") was an ideal candidate for a cytoplasmic selection system because it is not a common antibiotic resistance used for plasmid maintenance, the antibiotic Hygromycin B is common and affordable making our system accessible, and split protein constructs of APH(7") homologs have successfully been created by other groups [35].

Structurally, APH(7″) consists of two "lobes". The N-terminal lobe contains the nucleotide binding domain where the C-terminal lobe binds to hygromycin B and contains the catalytic residues (Fig. 1B) [36]. To link protein solubility to hygromycin B resistance, a site had to be selected within APH(7″) where a protein of interest would be inserted. A folded, soluble protein inserted into the hygromycin B resistance gene should allow the N- and C-terminal halves of APH(7″) to come together and exhibit phosphotransferase activity, therefore conferring hygromycin B resistance. However, if the inserted protein is insoluble, it should lead to misfolding and eventual aggregation or degradation of the complex, thereby rendering the cells sensitive to hygromycin B. The insertion site within APH(7″) should separate the two domains and demonstrate a range of antibiotic sensitivity and resistance when fused with insoluble and soluble proteins. It has been demonstrated that loops in the kanamycin resistance protein, aminoglycoside-3′-phosphotransferase

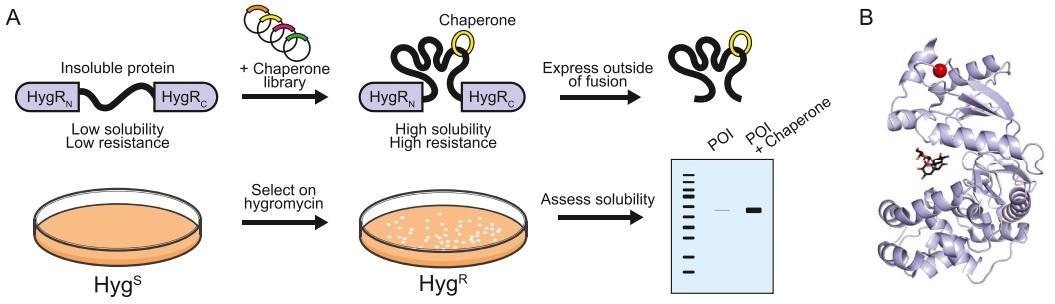


Fig. 1. A split antibiotic selection (SAS) to identify chaperones that increase the solubility of a heterologously expressed POI A) A workflow for using SAS to identify chaperones that solubilize a given protein. An insoluble POI (black curved line), is inserted into the hygromycin B resistance gene APH(7″) (HygR, purple). When the inserted protein is insoluble, this construct is unstable leading to hygromycin B sensitivity. When the inserted protein is solubilized, APH(7″) is functional and provides hygromycin B resistance. By transforming a library of chaperones into the *E. coli* strains expressing this fusion and selecting for hygromycin B resistant colonies, one can quickly identify chaperons (gold ring) that improve solubility of a given protein. The protein of interest can then be co-expressed with the identified solubilizing chaperone outside of the fusion, to confirm that the chaperone increases POI solubility independent of APH(7″) B) The crystal structure of *Streptomyces hygroscopicus* APH(7″) (PDB ID:6IY9) represented as a cartoon. The split site at residue G108 is shown as a red sphere. Hygromycin B (center) is shown in black sticks. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

IIa, were the most capable of tolerating an inserted protein without disrupting its phosphotransferase activity [31]. Based on these criteria we selected residue G108 to be the site to split APH(7″) (Fig. 1B, red sphere).

3.2. The tripartite APH(7″) construct couples hygromycin B resistance to POI solubility

To assess the efficacy of our tripartite system, we measured the hygromycin B resistance of cells expressing APH(7″) with a known insoluble protein inserted within APH(7") at position G108. The ATPase domain of human mitochondrial Hsp70 (ATPase^{70}) is completely insoluble when expressed cytoplasmically in *E. coli* unless expressed with its co-chaperone, human escort protein (Hep) [19]. We inserted ATPase^{70} proceeding residue G108 in APH(7″), creating the APH(7″)-ATPase^{70}-APH(7″) tripartite fusion construct. Cells expressing this split construct containing an insoluble cargo were sensitive to hygromycin B as assessed by spot titer assay (Fig. 2). Introduction of the co-chaperone Hep on a constitutive (*rrnB* promoter) helper plasmid, pAL, resulted in a strong increase in hygromycin B resistance, similar to a cell line expressing a wild-type APH(7″) control (Fig. 2). Based on these data, we concluded that a POI inserted into APH(7″) at position G108 will only disrupt the enzyme activity, resulting in hygromycin B sensitivity, if the inserted protein is insoluble. Therefore, the hygromycin B resistance of the cells expressing the tripartite fusion can be corelated to the relative solubility of the inserted POI. These results indicate that G108 is an appropriate split site for APH(7″).

3.3. The tripartite APH(7″) construct can be used to select for chaperones that improve POI solubility

We next utilized the ATPase^{70} SAS cassette in a model genetic selection to identify chaperones capable of improving client protein expression from a pooled library. For this approach, we designed a cytoplasmically-expressed chaperone library with 40 different protein folding factors called the pAL library (Table 1). The pAL library consists of native *E. coli* chaperones, chaperones from other organisms, redox enzymes, and transcription factors involved envelope stress, heat shock, and the general stress response. Importantly, this library includes the ATPase^{70} co-chaperone Hep as an internal positive control.

We performed a pooled SAS genetic selection by transforming cells expressing APH(7″)-ATPase^{70}-APH(7″) with either the pAL library, or empty pAL0 vector, and selected for hygromycin B resistant colonies (Fig. 3A). We found that 0.005% of the cells transformed with pAL0 were hygromycin B resistant, whereas ~2% of the cells transformed with the pAL library were hygromycin B resistant (Fig. 3B). Therefore, the signal to noise ratio of the selection is 400:1. This high signal to noise ratio indicates that our genetic selection is robust. Colonies of pAL library transformants were subsequently patched on hygromycin B media to confirm their resistance. pAL plasmids were extracted from the hygromycin B resistant cells and sequenced to identify chaperones responsible for solubilizing the APH(7″)-ATPase^{70}-APH(7″) construct. All the colonies surveyed in this manner harbored the pAL-Hep plasmid. This result demonstrates that the selection system is functional and specific, as the split APH(7″) construct enables one to select chaperones that solubilize an insoluble protein. Additionally, this indicates that the overexpression of native *E. coli* chaperones is not sufficient to solubilize the APH(7″)-ATPase^{70}-APH(7″) construct in the context of these experimental conditions.

3.4. Hygromycin B resistance correlates to the solubility of the APH(7″)-ATPase^{70}-APH(7″) tripartite fusion

To confirm that the observed increase in hygromycin B resistance of APH(7″)-ATPase^{70}-APH(7″) was due to Hep increasing the solubility of the tripartite fusion, we assessed the level of soluble fusion protein via western blots. We created C-terminal FLAG-tagged constructs of both APH(7″) and APH(7″)-ATPase^{70}-APH(7″) and assessed the levels of soluble tripartite fusion protein when expressed with and without Hep. As expected, the wild-type APH(7″) was soluble in *E. coli* (Fig. 4, lane 1). Conversely, the split construct, APH(7″)-ATPase^{70}-APH(7″), was not observed in the soluble fraction when expressed on its own, or alongside empty pAL vector (Fig. 4, lane 2 and 3). This solubility defect could be rescued by the introduction of the pAL-Hep plasmid (Fig. 4, lane 4). These results mirror those of the ATPase^{70} domain solubility when expressed with or without Hep, independent of the split hygromycin B construct [19]. Taken

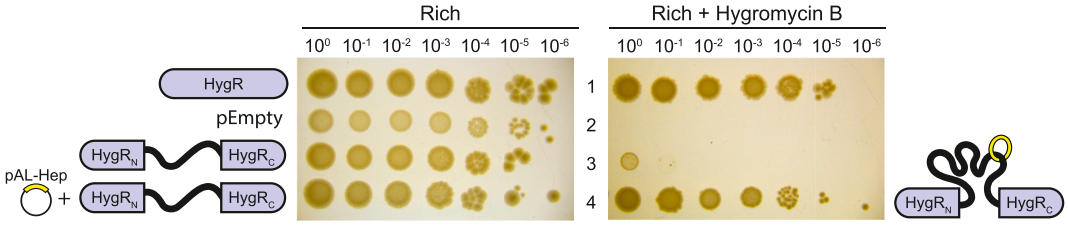


Fig. 2. Hygromycin B resistance conferred by the tripartite APH(7″) fusion is linked to POI solubility
Hygromycin resistance of the tripartite fusion correlates to the solubility of the inserted cargo protein. Spot titer assays comparing the growth of strains on nonselective (rich) and selective (rich + hygromycin B) media. Lane 1: Cells expressing wild-type APH(7″), shown as HygR in purple, are resistant to hygromycin B (MB6826). Lane 2: *E. coli* harboring an empty vector is sensitive hygromycin B (MB4853). Lane 3: Cells expressing the split construct, APH(7″)-ATPase^{70}-APH(7″) are shown as HygR$_N$ and HygR$_C$ in purple, sandwiching the insoluble ATPase^{70} protein which is represented as a black line (MB6827). This strain is also sensitive to hygromycin B. (4) Expression of the co-chaperone Hep alongside the APH(7″)-ATPase^{70}-APH (7″) construct results in an increase in hygromycin B resistance. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 1
The pAL chaperone library.

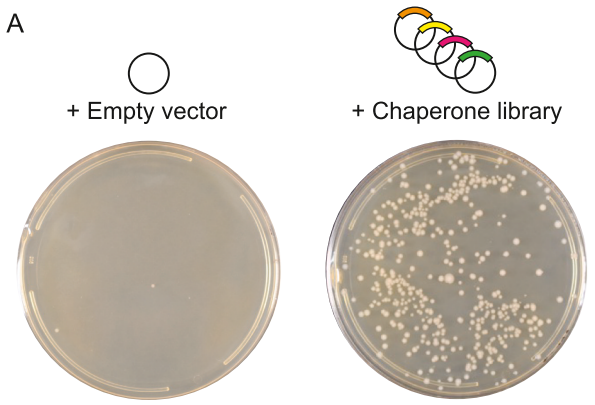
Plasmid	Gene(s)	Function
pAL0	Empty vector	Negative control
pAL1	GroEL	Chaperonin
pAL2	GroEL-ES	Chaperonin
pAL3	Hsp33	Holdase
pAL4	DsbC	Isomerase
pAL5	DnaK-DnaJ-GrpE	Foldase
pAL6	FkpA	Holdase
pAL7	Tig-FkpA	Holdase
pAL8	SurA	Holdase
pAL9	SlyD	PPIase
pAL10	Human PPIB	PPIase
pAL11	Spy	Holdase
pAL12	ClpB	Disaggregase
pAL13	CspA	RNA chaperone
pAL14	Skp	Holdase
pAL15	KatG	Reductase
pAL16	DegP_S210A	Protease
pAL17	rpoH	Sigma factor
pAL18	rpoH I54 N	Sigma factor
pAL19	lon	Protease
pAL20	rpoS	Sigma factor
pAL21	rpoE	Sigma factor
pAL22	IbpAB	Disaggregase
pAL24	Citrobacter FevR	Transcription factor
pAL25	Geobacillus GrpE	Nucleotide exchange factor
pAL26	DsbM	Oxidase
pAL27	Human Hep	Holdase
pAL28	Tkodakarensis MsrAB	Reductase
pAL29	Mtuberculosis MsrA	Reductase
pAL30	Mtuberculosis MsrB	Reductase
pAL31	AgsA	Disaggregase
pAL32	Hsp90	Foldase
pAL102	Saccharomyces MPD1	Isomerase/Foldase
pAL103	Saccharomyces MPD2	Isomerase/Foldase
pAL104	Human PDI	Isomerase/Foldase
pAL105	Erv1	Sulfhydryl oxidase
pAL106	DsbG_KQ	Reductase
pAL107	Human BIP	PPIase
pAL1001	T4 ligase	DNA Ligase
pAL1002	*E. coli* DNA ligase	DNA Ligase
pAL1003	Chlorella ligase	DNA Ligase

together, our data demonstrate that hygromycin B resistance correlates with the solubility of the APH(7″) construct and support that the SAS tripartite fusion construct can serve as a reporter for the solubility of an inserted cargo protein.

4. Discussion

The purpose of this study was to design a cytoplasmic tripartite selection system that could be used in combination with a library of protein folding factors to select for the chaperones that solubilize a given insoluble POI (Fig. 1A). To this end, we used the hygromycin B resistance protein, APH(7″) as the basis for our tripartite construct (Fig. 1B). We inserted a known insoluble protein, the ATPase domain of human mitochondrial Hsp70, into a surface exposed loop at residue G108 of APH(7″). Cells expressing this fusion were sensitive to hygromycin B, demonstrating that the functionality of APH(7″) is disrupted by insertion of an insoluble protein (Fig. 2). However, cells expressing APH(7″)-ATPase^{70}-APH(7″) and a plasmid borne copy of the co-chaperone Hep were resistant to hygromycin B. These data demonstrate that insertion of a soluble protein will not disrupt the function of APH(7″) and further shows that the hygromycin B sensitivity of cells expressing APH(7″)-ATPase^{70}-APH(7″) tripartite fusion is due to insolubility of the inserted protein, and not a general disruption of function due to the presence of a foreign protein within APH(7″). We demonstrate that using a tripartite construct in combination with a library of chaperones enables the selection of chaperones improving solubility of a heterologous protein (Fig. 3). In support of these *in vivo* findings, Western blot analysis of these strains confirmed that the hygromycin B resistance conferred by the fusion corelates to the solubility of the APH(7″) construct (Fig. 4).

Following the success of the split APH(7″) construct, several other insoluble proteins were cloned into the tripartite fusion. The insoluble proteins all exhibited hygromycin B sensitivity relative to the unsplit construct. Unfortunately, none of these constructs showed improved hygromycin B resistance when transformed with the chaperone library. This result could be for several reasons. First, it is possible that none of these proteins would ever be solubilized by any of the chaperones in our library. This may be the case for APH (7″) which could only be solubilized by its escort protein Hep. The pAL library mostly encompasses native *E. coli* chaperones and



B

	pAL0	pAL library
CFUs/mL	$1.7x10^8$	$1.4x10^8$
HygR CFUs/mL	$9.0x10^3$	$4.0x10^6$
% HygR colonies	0.005%	2.8%

Fig. 3. SAS can select for chaperones that increase POI solubility from the pAL library
The split APH(7″) construct can be used in a genetic selection to identify chaperones that improve POI expression. (A) Cells expressing the APH(7″)-ATPase^{70}-APH(7″) fusion were transformed with either empty pAL vector, or the pAL chaperone library and plated on B hygromycin containing media. Selection plates of a 10^{-3} dilution of the transformants are shown. (B) Table depicting the total CFUs/mL, CFUs/mL that are hygromycin B resistant, and the percentage of colonies that are hygromycin B resistant.

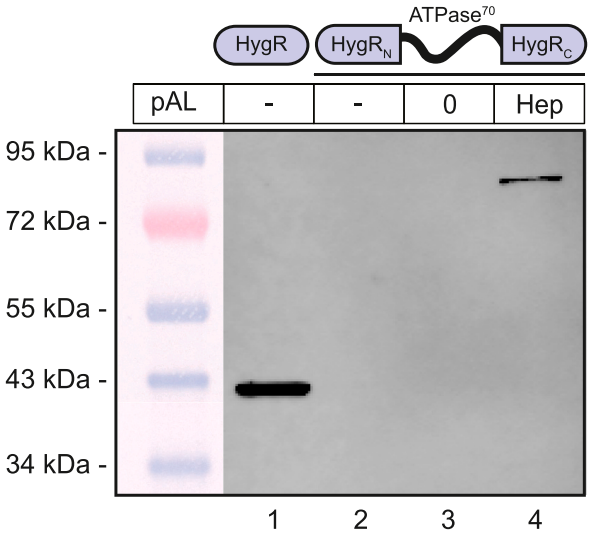


Fig. 4. Hygromycin B resistance correlates to the amount of soluble APH(7″)-ATPase^{70}-APH(7″) protein in the cell
Western blot analysis of soluble lysate demonstrates that hygromycin B resistance, shown in Fig. 2, correlates to the solubility of the APH(7″) tripartite fusion. Lane 1: The unsplit wild-type APH(7″) protein, denoted HygR, in purple, (38 kD) is soluble when expressed in *E. coli* (MB8355). Lane 2: The tripartite fusion, APH(7″)-ATPase^{70}-APH(7″), depicted as $HygR_N$ and $HygR_C$ in purple, sandwiching the insoluble ATPase^{70} protein which is represented as a black line (80 kD), is not detected in the soluble fraction when expressed on its own (MB8356) or (Lane 3) in combination with the empty vector, pAL0 (MB8359). Lane 4: When the co-chaperone, pAL-Hep, is expressed APH(7″)-ATPase^{70}-APH(7″) is detected in the soluble fraction (MB8360). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

protein folding factors, thus there could be a key chaperone native to the protein's host organism that is required for soluble expression, as with Hep and ATPase^{70}. Another possibility is that higher levels of expression of the chaperone library may be required when a chaperone has numerous interacting partners, as the native proteins in *E. coli* that are folded by these chaperones could dilute out the population of unoccupied chaperones in the cell. Similarly, it could be that the split APH(7″) construct is better at sensing highly specific protein-protein interactions and is not sensitive enough to account for more general increases in solubility from a non-specific

protein-chaperone interaction. Notably, the size and or structure of the inserted POI could influence the ability of the two halves of the APH(7″) protein to come together. In particular, large POIs could physically prevent the reconstitution of the functional APH(7″) protein. With this tripartite approach there is no guarantee that a given POI, even when solubilized, would necessarily result in a Hyg^R phenotype.

In the future, additional insoluble proteins could be cloned into the split APH(7″) construct. Although the insoluble proteins we assessed did not work, this does not preclude the possibility that a chaperone that increases soluble expression of other proteins could be selected from our library. In addition to testing new cargo proteins, increasing the diversity of our chaperone library could improve the chances of success. Specifically, the inclusion of more eukaryotic folding factors such as Hep, could assist with the expression of eukaryotic proteins, which are notoriously difficult to express in *E. coli*.

Improvements to the APH(7″) construct itself could also be explored, such as including linkers between the split site and inserted POI. The inclusion of linkers has been used for other tripartite fusion approaches [27,31,35,37]. To test this possibility, we added a 17-mer glycine serine linker on either side of the split site at G108. Unfortunately, the inclusion of these linkers resulted in a decreased signal to noise ratio when comparing the hygromycin resistance of APH(7″)-ATPase^{70}-APH(7″) expressed with and without Hep. In the future, additional linker lengths and could be tested.

Despite these drawbacks, SAS has great potential not only to select for chaperones that increase protein solubility, but also as a platform for chaperone discovery. Genomic DNA libraries from the host organism of the insoluble protein or a curated library of putative chaperones could be used in place of the strictly defined pAL library. With this approach, any uncharacterized protein that increases the solubility of any cargo protein in this system would provide evidence that the selected folding factor exhibits chaperone activity *in vivo*. Following the selection, these chaperones could then be further characterized *in vitro*. Additionally, SAS could be used to study known chaperone-protein interactions. Alterations to either the chaperone or cargo could be made, and the ability of the chaperone to fold the cargo protein could be assessed using this system. This opens up the opportunity for structure function analysis of chaperone protein interactions that do not naturally exhibit a phenotype when disrupted. Ultimately, there is no one size fits all approach for addressing difficulty with soluble expression of heterologous proteins in *E. coli*. We hope that SAS will serve as a genetic tool both for solubilizing important client proteins, as a selection system to discovering new protein folding factors, and for other applications which require monitoring solubility of client proteins *in vivo*.

CRediT authorship contribution statement

Emily McNutt: Writing – review & editing, Investigation, Data curation. **Na Ke:** Writing – review & editing, Investigation, Data curation, Conceptualization. **Alexandre Thurman:** Data curation. **James B. Eaglesham:** Writing – review & editing, Methodology, Investigation. **Mehmet Berkmen:** Writing – review & editing, Writing – original draft, Project administration, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e26996.

References

[1] S. Sahdev, S.K. Khattar, K.S. Saini, Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies, Mol. Cell. Biochem. 307 (2008) 249–264.
[2] G.L. Rosano, E.A. Ceccarelli, Recombinant protein expression in *Escherichia coli*: advances and challenges, Front. Microbiol. 5 (2014) 172.
[3] F. Baneyx, M. Mujacic, Recombinant protein folding and misfolding in Escherichia coli, Nat. Biotechnol. 22 (2004) 1399–1408.
[4] D.M. Francis, R. Page, Strategies to optimize protein expression in *E. coli*, Curr Protoc Protein Sci Chapter 5 (2010). Unit 5 24 21-29.
[5] A.A. Tungekar, A. Castillo-Corujo, L.W. Ruddock, So you want to express your protein in *Escherichia coli*? Essays Biochem. 65 (2021) 247–260.
[6] G.E. Dale, C. Broger, H. Langen, A.D. Arcy, D. Stüber, Improving protein solubility through rationally designed amino acid replacements: solubilization of the trimethoprim-resistant type S1 dihydrofolate reductase, Protein Eng. Des. Sel. 7 (1994) 933–939.
[7] M. Murby, E. Samuelsson, T.N. Nguyen, L. Mignard, U. Power, H. Binz, et al., Hydrophobicity engineering to increase solubility and stability of a recombinant protein from respiratory syncytial virus, Eur. J. Biochem. 230 (1995) 38–44.
[8] Al-Hawash A. B., Z.X., Ma F. (2017) Strategies of codon optimization for high-level heterologous protein expression in microbial expression systems, Gene Reports 9: 46-53..
[9] Z. Zhang, Z.-H. Li, F. Wang, M. Fang, C.-C. Yin, Z.-Y. Zhou, et al., Overexpression of DsbC and DsbG markedly improves soluble and functional expression of single-chain Fv antibodies in Escherichia coli, Protein Expr. Purif. 26 (2002) 218–228.
[10] V. De Marco, G. Stier, S. Blandin, A. de Marco, The solubility and stability of recombinant proteins are increased by their fusion to NusA, Biochem. Biophys. Res. Commun. 322 (2004) 766–771.
[11] T.R. Butt, S.C. Edavettal, J.P. Hall, M.R. Mattern, SUMO fusion technology for difficult-to-express proteins, Protein Expr. Purif. 43 (2005) 1–9.
[12] K. Guild, Y. Zhang, R. Stacy, E. Mundt, S. Benbow, A. Green, P.J. Myler, Wheat germ cell-free expression system as a pathway to improve protein yield and solubility for the SSGCID pipeline, Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun. 67 (2011) 1027–1031.
[13] D.M. Francis, R. Page, Strategies to optimize protein expression in *E. coli*, Curr Protoc Protein Sci Chapter 5 (2010), 5 24 21-25 24 29.

[14] O. Kolaj, S. Spada, S. Robin, J.G. Wall, Use of folding modulators to improve heterologous protein production in *Escherichia coli*, Microb. Cell Factories 8 (2009) 9.
[15] M. Mamipour, M. Yousefi, M. Hasanzadeh, An overview on molecular chaperones enhancing solubility of expressed recombinant proteins with correct folding, Int. J. Biol. Macromol. 102 (2017) 367–375.
[16] K. Nishihara, M. Kanemori, H. Yanagi, T. Yura, Overexpression of trigger factor prevents aggregation of recombinant proteins in *Escherichia coli*, Appl. Environ. Microbiol. 66 (2000) 884–889.
[17] A. Haacke, G. Fendrich, P. Ramage, M. Geiser, Chaperone over-expression in *Escherichia coli*: apparent increased yields of soluble recombinant protein kinases are due mainly to soluble aggregates, Protein Expr. Purif. 64 (2009) 185–193.
[18] M. Ferrer, T.N. Chernikova, K.N. Timmis, P.N. Golyshin, Expression of a temperature-sensitive esterase in a novel chaperone-based *Escherichia coli* strain, Appl. Environ. Microbiol. 70 (2004) 4499–4504.
[19] P. Zhai, C. Stanworth, S. Liu, J.J. Silberg, The human escort protein Hep binds to the ATPase domain of mitochondrial hsp70 and regulates ATP hydrolysis, J. Biol. Chem. 283 (2008) 26098–26106.
[20] A.H. Baek, E.Y. Jeon, S.M. Lee, J.B. Park, Expression levels of chaperones influence biotransformation activity of recombinant *Escherichia coli* expressing *Micrococcus luteus* alcohol dehydrogenase and *Pseudomonas putida* Baeyer-Villiger monooxygenase, Biotechnol. Bioeng. 112 (2015) 889–895.
[21] S. Peng, Z. Chu, J. Lu, D. Li, Y. Wang, S. Yang, Y. Zhang, Co-expression of chaperones from P. furiosus enhanced the soluble expression of the recombinant hyperthermophilic alpha-amylase in *E. coli*, Cell Stress Chaperones 21 (2016) 477–484.
[22] H. Chautard, E. Blas-Galindo, T. Menguy, L. Grand'Moursel, F. Cava, J. Berenguer, M. Delcourt, An activity-independent selection system of thermostable protein variants, Nat. Methods 4 (2007) 919–921.
[23] K.L. Maxwell, A.K. Mittermaier, J.D. Forman-Kay, A.R. Davidson, A simple *in vivo* assay for increased protein solubility, Protein Sci. 8 (1999) 1908–1911.
[24] K. Tsumoto, M. Umetsu, I. Kumagai, D. Ejima, T. Arakawa, Solubilization of active green fluorescent protein from insoluble particles by guanidine and arginine, Biochem. Biophys. Res. Commun. 312 (2003) 1383–1386.
[25] P. Philibert, P. Martineau, Directed evolution of single-chain Fv for cytoplasmic expression using the beta-galactosidase complementation assay results in proteins highly susceptible to protease degradation and aggregation, Microb. Cell Factories 3 (2004) 16.
[26] A.C. Fisher, W. Kim, M.P. DeLisa, Genetic selection for protein solubility enabled by the folding quality control feature of the twin-arginine translocation pathway, Protein Sci. 15 (2006) 449–458.
[27] L. Foit, G.J. Morgan, M.J. Kern, L.R. Steimer, A.A. von Hacht, J. Titchmarsh, et al., Optimizing protein stability *in vivo*, Mol Cell 36 (2009) 861–871.
[28] A. Galarneau, M. Primeau, L.E. Trudeau, S.W. Michnick, Beta-lactamase protein fragment complementation assays as *in vivo* and *in vitro* sensors of protein protein interactions, Nat. Biotechnol. 20 (2002) 619–622.
[29] S. Quan, P. Koldewey, T. Tapley, N. Kirsch, K.M. Ruane, J. Pfizenmaier, et al., Genetic selection designed to stabilize proteins uncovers a chaperone called Spy, Nat. Struct. Mol. Biol. 18 (2011) 262–269.
[30] C.W. Lennon, M. Thamsen, E.T. Friman, A. Cacciaglia, V. Sachsenhauser, F.A. Sorgenfrei, et al., Folding optimization *in vivo* uncovers new chaperones, J. Mol. Biol. 427 (2015) 2983–2994.
[31] A. Malik, A. Mueller-Schickert, J.C. Bardwell, Cytosolic selection systems to study protein stability, J. Bacteriol. 196 (2014) 4333–4343.
[32] J. Sambrook, E.F. Fritsch, Maniatis, *Molecular Cloning: a Laboratory Manual.* Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory, 1989.
[33] F.W. Studier, T7 expression systems for inducible production of proteins from cloned genes in E. coli, Curr. Protoc. Mol. Biol. 124 (2018) e63.
[34] R.C. Pittenger, R.N. Wolfe, M.M. Hoehn, P.N. Marks, W.A. Daily, G.J. Mc, Hygromycin. I. Preliminary studies on the production and biologic activity of a new antibiotic, Antibiot. Chemother. 3 (1953) 1268–1278.
[35] S.W. Michnick, I. Remy, F.X. Campbell-Valois, A. Vallee-Belisle, J.N. Pelletier, Detection of protein-protein interactions by protein fragment complementation strategies, Methods Enzymol. 328 (2000) 208–230.
[36] M. Takenoya, T. Shimamura, R. Yamanaka, Y. Adachi, S. Ito, Y. Sasaki, et al., Structural basis for the substrate recognition of aminoglycoside 7''-phosphotransferase-Ia from *Streptomyces hygroscopicus*, Acta Crystallogr F Struct Biol Commun 75 (2019) 599–607.
[37] P. Secco, E. D'Agostini, R. Marzari, M. Licciulli, R. Di Niro, S. D'Angelo, et al., Antibody library selection by the beta-lactamase protein fragment complementation assay, Protein Eng. Des. Sel. 22 (2009) 149–158.