

RESEARCH

Open Access



An efficient expression tag library based on self-assembling amphipathic peptides

Weixin Zhao^{1,2}, Song Liu^{1,2*} , Guocheng Du^{2,3} and Jingwen Zhou^{1,2,4}

Abstract

Background: Self-assembling amphipathic peptides (SAPs) may improve protein production or induce the formation of inclusion bodies by fusing them to the N-terminus of proteins. However, they do not function uniformly well with all target enzymes and systematic research on how the composition of SAPs influence the production of fusion protein is still limited.

Results: To improve the efficiency of SAPs, we studied factors that might be involved in SAP-mediated protein production using S1 (AEAEAKAK)₂ as the original SAP and green fluorescent protein (GFP) as the reporter. The results indicate that hydrophobicity and net charges of SAPs play a key role in protein expression. As hydrophobicity regulation tend to cause the formation of insoluble inclusion bodies of protein, an expression tag library composed of SAPs, which varied in net charge (from +1 to +20), was constructed based on the random amplification of S1nV1 (ANANA-RAR)₁₀. The efficiency of the library was validated by polygalacturonate lyase (PGL), lipoxygenase (LOX), L-asparaginase (ASN) and transglutaminase (MTG). To accelerate preliminary screening, each enzyme was fused at the C-terminus with GFP. Among the four enzyme fusions, the SAPs with +2 – +6 net charges were optimal for protein expression. Finally, application of the library improved the expression of PGL, LOX, ASN, and MTG by 8.3, 3.5, 2.64, and 3.68-fold relative to that of the corresponding wild-type enzyme, respectively.

Conclusions: This is the first report to study key factors of SAPs as an expression tag to enhance recombinant enzyme production. The SAP library could be used as a novel plug-and-play protein-engineering method to screen for enzymes or proteins with enhanced production.

Keywords: Self-assembling amphipathic peptides, Expression tags, Hydrophobicity, Positive charge, High-throughput screening

Background

Escherichia coli is preferred for heterologous protein production because of its short growth period, simple transformation process, uncomplicated nutritional and sterility requirements, and extensive research foundation [1]. To improve the protein production in this expression system, efforts have been directed toward promoter screening [2], N-terminal codon optimization [3], fusing with an expression tag at the terminus [4], and culture condition optimization [5].

Due to the high efficiency of protein production, a series of tags have been developed, including glutathione S-transferase (GST) [6], maltose binding protein (MBP) [7], small ubiquitin modifier (SUMO) [8], and N-utilization substance (NusA) [9]. These proteins or tags could regulate the process of protein transcription and translation [10] or help to fold the protein properly [4], thus leading to enhanced expression of the target protein. However, a specific fusion tag does not work efficiently in all cases [11], and the biological activity of functional proteins could even be inhibited by the fused tags [12]. Thus, it is desirable to improve the universality of expression tags without sacrificing the biological properties of the target proteins.

Self-assembling amphipathic peptides (SAPs) are short peptides constituted by alternating hydrophobic

*Correspondence: liusong@jiangnan.edu.cn

¹ National Engineering Laboratory for Cereal Fermentation Technology, Jiangnan University, Wuxi 214122, China

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



and hydrophilic residues [13]. We previously showed that S1 (AEAEAKAK)₂, a SAP originating from the Zuotin protein sequence, improved the production and thermal stability of the LOX protein when fused to its N-terminus in *E. coli* [14]. The positive effects of SAPs on enzyme activity or stability were also observed in the SAP fusions of amylase [15] and nitrile hydratase [16]. We then generated an S1 variant (AEAEAHAH)₂ with the ability to benefit the production, thermal stability, and purification yield of recombinant proteins in *E. coli* [17]. These findings of positive effects on both protein production and properties suggested that the desired expression tags could be derived from SAPs.

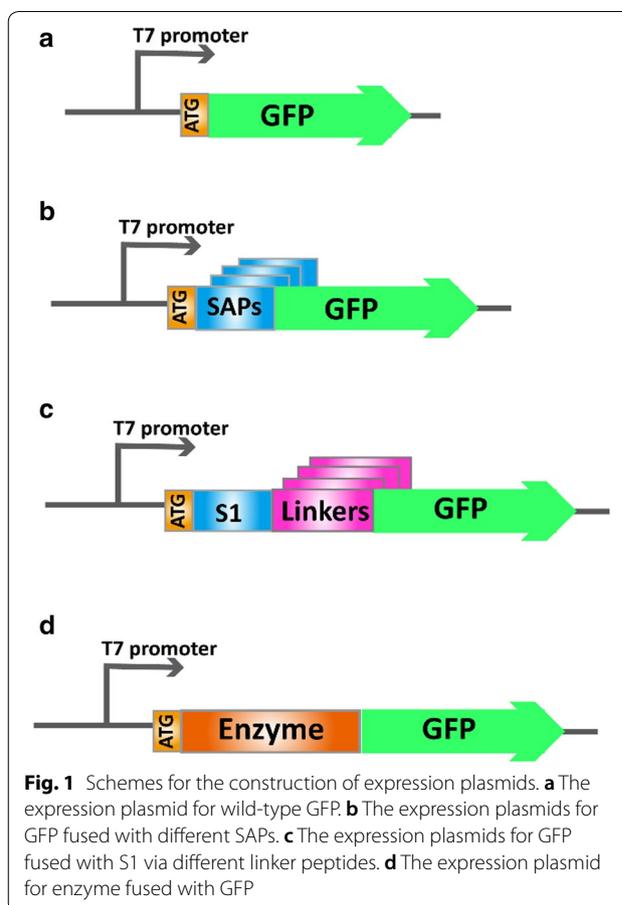
Although there is a certain degree of universality for enhancing protein production, fusion with SAPs could not improve the production of all proteins to an acceptable extent in *E. coli* [17]. The mechanism by which the production enhancement function of SAPs can be maximized remains unexplored. It was reported that a SAP (LELELKLK)₂ with high hydrophobicity induced in vivo assembly of active protein aggregates after fusing to the terminus of proteins in *E. coli* [18]. These findings implied that the hydrophobicity or charges of the SAPs play an important role in the production of protein fusions. In addition, as the bridge of the fusion protein, linker regions are also important for the construction of bioactive fusion proteins, with the length and rigidity of linker peptides directly affecting the orientation of the linked proteins [19]. Based on these reports, SAP fusion may be optimized for enhancing the production of a specific protein by adjusting SAP and linker compositions.

Here we explored the factors that could increase the efficiency of SAPs as an expression fusion tag in *E. coli*. First, we studied the key factors (the composition of SAPs and linker peptides) that might be involved in SAP fusion production using GFP as a reporter. Then, an expression tag library composed of SAPs which varied in net charge was constructed based on the random amplification of S1nv1 (ANANARAR)₁₀, an S1 variant. Finally, the efficiency of the library was validated using four different enzymes.

Materials and methods

Strains and plasmids

Escherichia coli JM109 and *E. coli* BL21 (DE3) cells (Novagen, Madison, WI, USA) were used for gene cloning and protein expression, respectively. Plasmid pET-22b(+)/gfp (Fig. 1a) encoding wild-type GFP (gfp) from *Aequorea victoria* [20] was constructed as described previously [17].



Plasmid construction

The gene fragments of S1 and its derivatives, which varied in hydrophobic residues (Table 1, S1hv1–S1hv6), hydrophilic residues (Table 1, S1cv1–S1cv5), and net charge (Table 1, S1nv1 and S1nv2), were chemically synthesized and cloned into the *Nde*I and *Nco*I sites of pET-22b(+)/gfp by Sangon Biotech (Shanghai, China), yielding the plasmids expressing GFP fusions with different SAPs at the N-terminus (Fig. 1b). To avoid the effect of codons on protein expression, all genes encoding SAPs were synthesized according to the codon usage preferred by *E. coli* [21].

DNA manipulations in our study were based on standard protocols and the related primer pairs are shown in Additional file 1: Table S1. The plasmids expressing the GFP fusions containing different S1 units (Table 1, S1lv1–S1lv9) (Fig. 1b) were constructed by whole plasmid polymerase chain reaction (PCR) as shown in Additional file 1: Figure S1. The plasmid encoding GFP fused with S1 (Table 1, Additional file 1: Table S1) was used as the template and S1lv1-F/S1-R were the primer pairs. PCR was conducted using the followed amplification program:

Table 1 The amino acid sequence of SAPs designed in this study

Schemes	SAPs	Amino acid sequence	GRAVY ^a	Net charge
Hydrophobic residues	S1	(AEAEAKAK) ₂	-0.95	0
	S1hv1	(IEIEIKIK) ₂	0.4	0
	S1hv2	(LELELKLK) ₂	0.05	0
	S1hv3	(VEVEVKVK) ₂	0.25	0
	S1hv4	(FEFEFKFK) ₂	-0.45	0
	S1hv5	(GEGEGK GK) ₂	-2.05	0
Hydrophilic residues	S1hv6	(PEPEPKPK) ₂	-2.65	0
	S1cv1	(ADADAH AH) ₂	-0.78	0
	S1cv2	(AEAEAH AH) ₂	-0.78	0
	S1cv3	(ADADAK AK) ₂	-0.95	0
	S1cv4	(AEAEAR AR) ₂	-1.1	0
Length	S1cv5	(ADADAR AR) ₂	-1.1	0
	S1lv1	AEAEAKAK	-0.95	0
	S1lv2	(AEAEAKAK) ₃	-0.95	0
	S1lv3	(AEAEAKAK) ₄	-0.95	0
	S1lv4	(AEAEAKAK) ₅	-0.95	0
	S1lv5	(AEAEAKAK) ₆	-0.95	0
	S1lv6	(AEAEAKAK) ₇	-0.95	0
	S1lv7	(AEAEAKAK) ₈	-0.95	0
	S1lv8	(AEAEAKAK) ₉	-0.95	0
Net charge	S1lv9	(AEAEAKAK) ₁₀	-0.95	0
	S1nv1	(ANANAR AR) ₁₀	-1.1	+20
	S1nv2	(ANANAD AD) ₁₀	-1.1	-20
	S1nv3	(ANANAD AD) ₈	-1.1	-16
	S1nv4	(ANANAD AD) ₇	-1.1	-14
	S1nv5	(ANANAD AD) ₆	-1.1	-12
	S1nv6	(ANANAD AD) ₄	-1.1	-8
	S1nv7	(ANANAD AD) ₃	-1.1	-6
	S1nv8	(ANANAD AD) ₂	-1.1	-4
	S1nv9	ANANAD AD	-1.1	-2
	S1nv10	(ANANAR AR) ₂	-1.1	+4
	S1nv11	(ANANAR AR) ₃	-1.1	+6
	S1nv12	(ANANAR AR) ₄	-1.1	+8
	S1nv13	(ANANAR AR) ₅	-1.1	+10
	S1nv14	(ANANAR AR) ₆	-1.1	+12
	S1nv15	(ANANAR AR) ₈	-1.1	+16
	S1nv16	(ANANAR AR) ₉	-1.1	+18
S1nv17	ANANAR ARANAR	-1.064	+3	

^a The total average of hydrophobicity (GRAVY, <https://web.expasy.org/>) was used to indicate the hydrophobicity of SAPs

an initial denaturation of 3 min at 98 °C, followed by 34 cycles of 10 s at 98 °C, 10 s at 55 °C, and 6 min at 72 °C. The length of SAP may be altered by changing the PCR annealing temperature, yielding several GFP fusions containing different number of S1 units. Ten fusions with SAPs ranging from a half to five S1 units (Table 1, S1lv1–S1lv9) were selected.

Plasmids expressing GFP fused with S1nv1 were used as the templates to construct the fusions containing

SAPs carrying different positive net charges (Table 1, S1nv10–S1nv17) using the primer pairs S1nv1-F/S1-R. Plasmids expressing GFP fused with S1nv2 were used as templates to construct fusions containing SAPs carrying different negative net charges (Table 1, S1nv3–S1nv9) using the primer pairs S1nv2-F/S1-R. The mutation of the net charge variation (Fig. 1b) were introduced through whole plasmid PCR using the corresponding templates and primers, and PCR was conducted using the followed

amplification program: initial denaturation of 3 min at 98 °C, followed by 34 cycles of 10 s at 98 °C, 10 s at 55 °C, and 6 min at 72 °C. To remove the primary template, *DpnI* was used to digest the PCR products, followed by purification and ligation by T4 ligase (Shanghai, China) to obtain circular plasmids. Then the circular plasmids were transformed into *E. coli* BL21 (DE3) for GFP fusion expression.

The gene fragments of linker peptides (Table 2) were separately synthesized and inserted into the *NcoI* site of the plasmids of GFP fused with S1 by Sangon Biotech (Shanghai, China), yielding the plasmids expressing GFP fused with S1 via different linkers (Fig. 1c). To avoid the effect of codons on the protein expression, all of the genes encoding linkers were synthesized according to the codon usage preferred by *E. coli* [21].

Library construction for screening fusions with enhanced production

The PGL gene from the genome of *Bacillus subtilis* WSHB04-02 (CCTCCM 204082), LOX from *Pseudomonas aeruginosa* BBE (CCTCC M2011185), ASN from *E. coli* (*E. coli* strain BL21, TaKaRa), and MTG from *Streptomyces mobaraense* (CICC 11018) were amplified using the corresponding primer pairs *pgl-F/pgl-R*, *lox-F/lox-R*, *asn-F/asn-R*, and *mtg-F/mtg-R*, respectively.

The target enzyme gene was cloned into the *NcoI* site of pET-22b(+)/*gfp*, yielding the plasmid expressing enzyme-GFP (wild-type enzyme fused with GFP) (Fig. 1d). Then the gene fragments of the enzyme were inserted into the *NcoI* site of the plasmid expressing GFP fused with *S1nv1* (Fig. 1b), yielding the plasmid pET-22b(+)/*S1nv1-enzyme-gfp* expressing *S1nv1-enzyme-GFP* (enzyme fused with *S1nv1* at the N-terminus and GFP at the C-terminus) (Fig. 2).

The general scheme for library construction is shown in Fig. 2. First, the charge mutation was introduced into the *S1nv1* coding region of pET-22b(+) derivatives by a whole plasmid PCR, using a constant forward primer (*nSAP-F*) and the specific reverse primer *nSAP-R*. Second, the linearized fragments were ligated by T4 ligase (Shanghai, China), yielding the mixed plasmid libraries of pET-22b(+)/*nS1nv1-enzyme-gfp* expressing the *nS1nv1-enzyme-GFP* (enzyme fused with different units of *S1nv1* at the N-terminus and GFP at the C-terminus). Third, a library of SAP fusions was constructed after protein production and detection based on RFU/OD₆₀₀. Finally, after obtaining fusions with enhanced fluorescence intensity, the plasmids pET-22b(+)/*nS1nv1-enzyme-gfp* of the positive mutants were used as templates with specific forward primers *Dgfp-up* and a constant reverse primer *Dgfp-down* to remove the GFP gene, yielding the plasmids expressing recombinant enzymes fused with specific

Table 2 Amino acid sequences of the linker peptides designed in this study

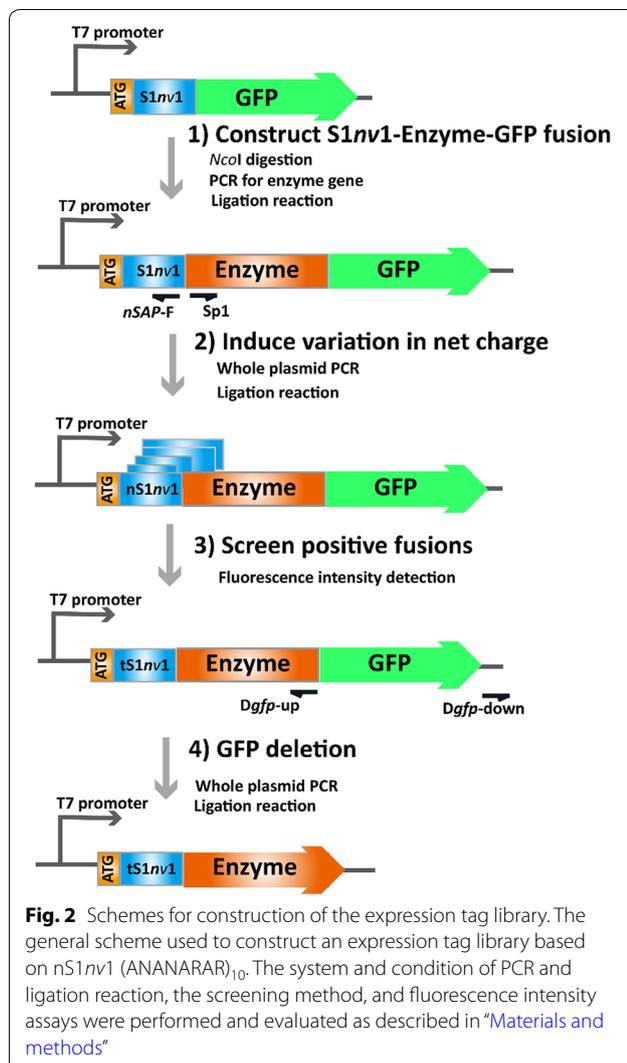
Linkers	Linker sequences
1	EAAAK
2	EAAAKEAAAK
3	EAAAKEAAAKEAAAK
4	EAAAKEAAAKEAAAKEAAAK
5	EAAAKEAAAKEAAAKEAAAKEAAAK
6	EAAAKEAAAKEAAAKEAAAKGGGGS
7	EAAAKEAAAKEAAAKGGGSGGGGS
8	EAAAKEAAAKGGGSGGGSGGGGS
9	EAAAKGGGSGGGSGGGSGGGGS
10	EAAAKGGGSGGGSGGGSEAAAK
11	EAAAKGGGSGGGSEAAAKGGGGS
12	EAAAKEAAAKGGGSGGGSEAAAK
13	EAAAKGGGSGGGSEAAAKEAAAK
14	EAAAKGGGSEAAAKEAAAKEAAAK
15	EAAAKEAAAKGGGSEAAAKEAAAK
16	EAAAKEAAAKEAAAKGGGSEAAAK
17	EAAAKGGGSEAAAKGGGSEAAAK
18	EAAAKEAAAKGGGSEAAAKGGGGS
19	EAAAKGGGSEAAAKGGGSGGGGS
20	EAAAKEAAAKGGGSGGGSEAAAK
21	GGGGS
22	GGGSGGGGS
23	GGGSGGGSGGGGS
24	GGGSGGGSGGGSGGGSGGGGS
25	GGGSGGGSGGGSGGGSGGGSGGGGS
26	GGGSGGGSGGGSGGGSGGGSEAAAK
27	GGGSGGGSGGGSGGGSEAAAKEAAAK
28	GGGSGGGSEAAAKEAAAKEAAAK
29	GGGSEAAAKEAAAKEAAAKEAAAK
30	GGGSEAAAKEAAAKEAAAKGGGGS
31	GGGSEAAAKEAAAKGGGSEAAAK
32	GGGSGGGSEAAAKEAAAKGGGGS
33	GGGSEAAAKEAAAKGGGSGGGGS
34	GGGSEAAAKGGGSGGGSGGGGS
35	GGGSGGGSEAAAKGGGSGGGGS
36	GGGSGGGSGGGSEAAAKGGGGS
37	GGGSEAAAKGGGSEAAAKGGGGS
38	GGGSGGGSEAAAKGGGSEAAAK
39	GGGSEAAAKGGGSEAAAKEAAAK
40	GGGSGGGSEAAAKEAAAKGGGGS

The rigid unit was (EAAAK) while the flexible unit was (GGGGS)

units of *S1nv1*. Specifically, the *S1nv1* peptide for MTG was fused with a pro-peptide in its C-terminus [22].

Culture conditions

The transformations containing the plasmids were transferred to Luria–Bertani (LB) media containing



100 µg/mL ampicillin. After overnight cultivation at 37 °C, 3% (v/v), *E. coli* seed cultures were inoculated into Terrific Broth (TB) medium containing 100 µg/mL ampicillin and cultivated at 37 °C. Upon reaching an optical density at 600 nm (OD₆₀₀) of 0.6–0.8 in TB medium, protein production was induced by adding isopropyl β-D-1-thiogalactopyranoside (IPTG). Induction conditions in 96-well plates were as follows: PGL was induced for 4 h at 30 °C with 0.04 mM IPTG; LOX was induced for 8 h at 25 °C with 1 mM IPTG; ASN was induced for 8 h at 25 °C with 1 mM IPTG; and MTG was induced for 8 h at 25 °C with 0.05 mM IPTG. Induction conditions in shake flasks were as follows: PGL was induced for 24 h at 30 °C with 0.04 mM IPTG; LOX was induced for 24 h at 20 °C with 1 mM IPTG; ASN was induced for 12 h at 25 °C with 1 mM IPTG; and MTG was induced for 24 h at 25 °C with 0.05 mM IPTG.

Fluorescent spectral analysis

The recombinant strains for GFP production or the fusion libraries cultured in the 96-wells or shake flasks under the corresponding culture conditions were harvested and washed twice by phosphate buffer solution (PBS, 50 mM, pH 7.5). Whole cell fluorescence and cell density (OD₆₀₀) were measured on a Cytation 3 imaging reader system (BioTek, Winooski, VT, USA). The corresponding wild-type strain *E. coli* BL21 (DE3) was used as the negative control, and its fluorescence intensity was subtracted as the background. The emission and excitation wavelength of GFP were 520 and 488 nm, respectively.

In Silico Analysis

The grand average of hydrophobicity (GRAVY, <https://web.expasy.org/>) was introduced to measure the hydrophobicity of SAPs. The increase in GRAVY value is an indication of strong hydrophobicity.

Protein production and enzymatic properties determination

After shaker cultivation, the enzymatic activities of PGL, LOX, ASN, and MTG fusions were measured and purified as previously reported [23–26]. The production of the recombinant protein was determined from the specific and crude enzymatic activity of each protein.

The dynamic thermal stabilities of the pure enzymes were determined by measuring residual activity after incubating the enzyme solutions at the corresponding incubation temperature. The half-life ($t_{1/2}$) was calculated using an exponential fitting of the data points [27]. The data were analyzed by fitting to first-order plots and the first-order rate constants (k_d) were determined by linear regression of \ln (residual activity) versus the incubation time (t). The time required for the residual activity to be reduced by half was calculated using the following equation: $t_{1/2} = \ln 2 / k_d$.

Gel electrophoresis and protein concentration assay

The fermentation broth was centrifuged at 8000 rpm for 10 min, and the cell pellets were washed once by PBS (50 mM, pH 7.5) and re-suspended in 50 mM PBS (pH 7.5) containing 0.1 M NaCl. The re-suspended cell solution was prepared to have an absorbency of 8.0 at 600 nm. Then cells were lysed by ultra-sonication and the operational process was conducted as reported previously [17]. Aliquots of *E. coli* lysate, supernatant, and pellets (re-suspended fraction) were mixed with SDS loading buffer (NuPAGE1 LDS Sample Buffer 4×, Fisher Scientific) at a ratio of 3:1 and sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) was performed as described previously [17].

Results

Effect of hydrophobic residues of SAP on GFP production

For expressing wild-type GFP from *A. Victoria* [20], the GFP gene was cloned into pET-22b(+) as shown in Fig. 1a. To consider the effects of SAP hydrophobicity on GFP production, the hydrophobic alanine (Ala) residues of S1 were replaced with isoleucine (Ile), leucine (Leu), valine (Val), phenylalanine (Phe), glycine (Gly), and proline (Pro) residues, yielding S1hv1, S1hv2, S1hv3, S1hv4, S1hv5, and S1hv6 (Table 1). The six S1 mutants showed GRAVY values ranges from 0.4 to -2.65 , suggesting a progressive variation in their hydrophobicity. Based on the constructs described in Fig. 1b, S1 and its mutants with different hydrophobicity were fused to the N-terminus of the GFP individually. Cells expressing the target proteins were harvested at the early exponential phase, and the GFP production presented as relative fluorescence intensity (fluorescence intensity divided by the corresponding OD_{600} , RFU/ OD_{600}). The relative fluorescence intensity of each GFP fusion was divided by that of GFP without SAPs, yielding the normalized fluorescence intensity value. As shown in Table 1, in contrast to the wild-type GFP, the yield of the GFP fusions was enhanced by the SAPs with the hydrophobicity close to that of the S1 peptide. SDS-PAGE analysis indicated that two S1 variants with strong hydrophobicity (S1hv1 and S1hv3) induced the formation of insoluble inclusion bodies. The other fusions with S1 variants (S1hv2, S1hv4, S1hv5, and S1hv6) showed a remarkable decrease in soluble fraction in contrast to that of S1 (Fig. 3).

Effect of hydrophilic residues of SAP on GFP production

To understand the role of hydrophilic residues in SAPs, S1 variants with different hydrophilic residues were synthesized, including S1cv1 (glutamic acid and lysine residues were replaced by aspartic acid and histidine

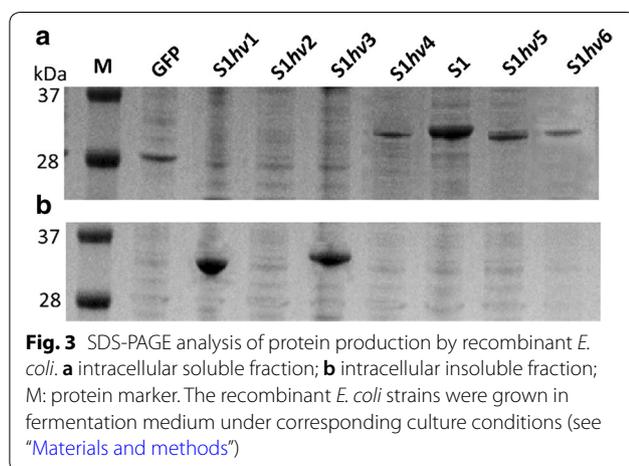
residues, respectively), S1cv2 (lysine residues were replaced by histidine residues), S1cv3 (glutamic acid residues were replaced by aspartic acid residues), S1cv4 (lysine residues were replaced by arginine residues), and S1cv5 (glutamic acid and lysine residues were replaced by aspartic acid and arginine residues, respectively) (Table 1). These SAPs were then fused separately to the N-terminus of the GFP (Fig. 1b). As described in Fig. 4b, all the SAP fusions showed remarkable increases in fluorescence intensity compared with that of GFP, and the S1 variants with the same hydrophobicity displayed different fluorescence intensity. In contrast to S1 variants with changed hydrophilic residues (Fig. 4a), the fluorescence intensity of each fusion here exhibited relative lower fluctuation. These results suggested that the type of hydrophilic residues have little effect on the efficiency of the SAPs.

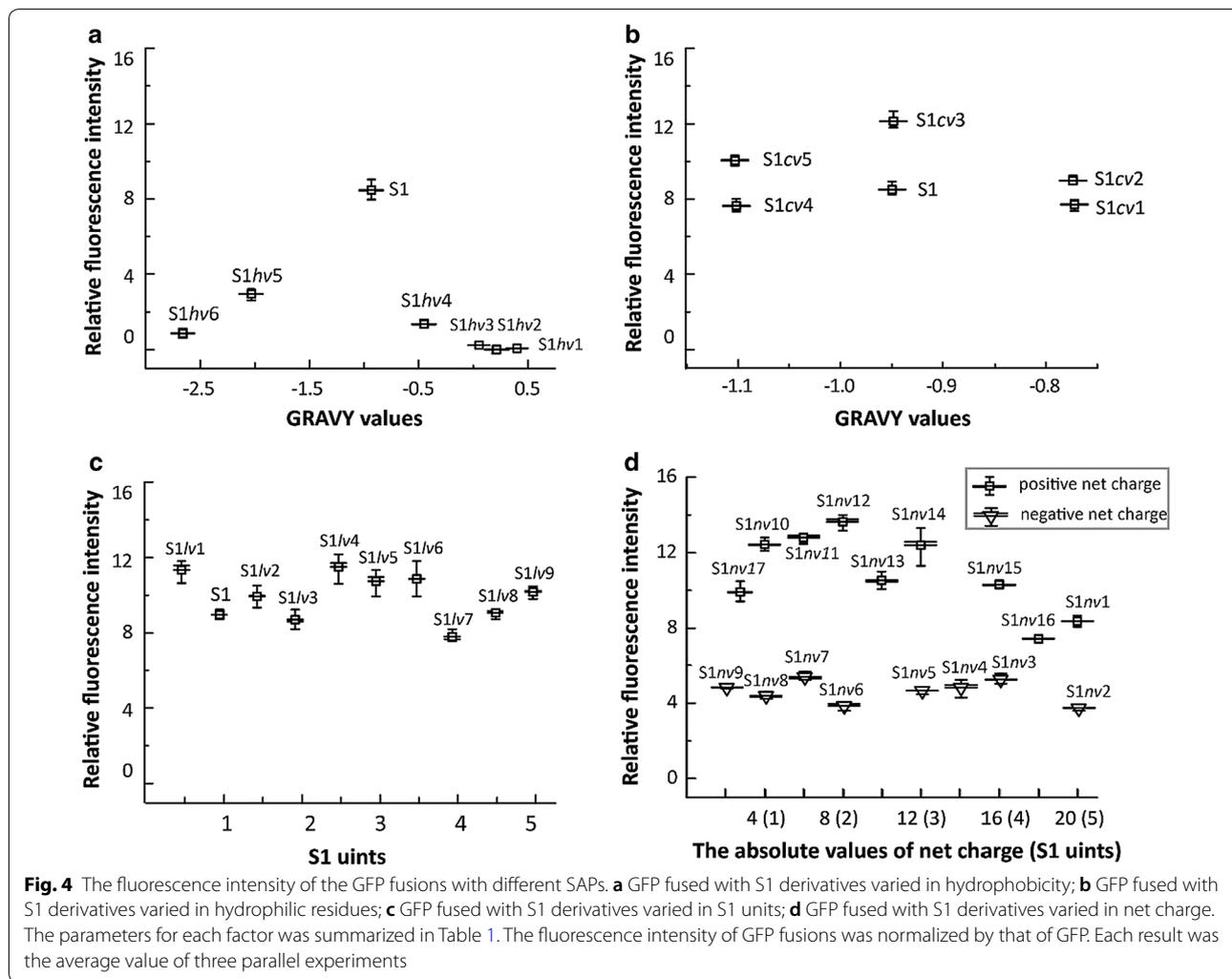
Effect of SAP length on GFP production

To investigate how the length of SAPs influences protein production, the S1 variants with 5, 1/2, 3/2, 2, 5/2, 3, 7/2, 4, and 9/2 S1 units were generated through PCR (Additional file 1: Fig. S1), and named S1lv1, S1lv2, S1lv3, S1lv4, S1lv5, S1lv6, S1lv7, S1lv8, and S1lv9, respectively. As shown in Fig. 1b, each S1 variant was fused to the N-terminus of GFP. It was found that the fluorescence intensities of GFP fusions were not changed greatly by the extension of S1 units, indicating that SAP length only had a slight effect on protein production (Fig. 4).

Effect of the SAP net charge on GFP production

To achieve a variation in the net charge of SAPs, a PCR procedure (Fig. S1) was conducted using the SAP units (ANANARAR)₁₀ and (ANANADAD)₁₀ as templates, yielding SAPs carrying net charge values ranging from +4 to +20 and from -2 to -20 (Table 1). For unknown reasons, we could not obtain SAPs with net charges of -18 , -10 , $+2$, and $+14$ using the PCR procedure (Table 1). Each SAP with a varied net charge was fused to the N-terminus of GFP (Fig. 1b). As shown in Fig. 4, all of the SAPs with different net charges increased GFP production. Under the same SAP length, SAPs with positive net charges produced the fluorescence intensities of GFP fusions 0.95–2.52-times higher than those of SAPs with negative net charges. In particular, the fluorescence intensities of the GFP fused with those SAPs carrying net charges ranging from +4 to +16 were over tenfold higher than that of GFP. The SAP Hence, regulating the net charge of the SAPs could effectively improve the production of SAP fusions.





Effect of linker on GFP production

Previous studies [28] have reported that the length and flexibility of linker peptides exert an important influence on the structure or production of the fusion proteins. Generally, a combination of the flexible (GGGGS) and rigid (EAAAK) linker units could generate linker peptides which vary in length and flexibility [29]. To examine the effect of linker length on GFP fusion production, S1 was fused to the N-terminus of the GFP via the linker peptides composed of 1 to 5 units of rigid or flexible linker units, respectively (Table 2, Fig. 1c). It was shown that the production of the GFP fusions was not changed greatly with variation in the length of flexible and rigid linker peptides (Additional file 1: Fig. S2A). To investigate the flexibility effects, the GFP was fused with S1 via linker peptides that were totally composed of 5 linker units and varied in the ratio of flexible (GGGGS) and rigid (EAAAK) units (Table 2, Fig. 1c). Similarly, changes

in flexibility have little effect on the production of GFP fusions (Additional file 1: Fig. S2B).

SAP library construction for enhanced protein production

Based on key factor analysis, the variables of the SAP library for protein expression were downsized to the number of net charges. Because SAPs with positive net charges were more effective than those with negative net charges, when related to protein expression, the net charges of the SAPs were restricted to the values ranging from +1 to +20.

The gene of the target enzyme was inserted into the gene of S1nv1-GFP, resulting in plasmids expressing S1nv1-enzyme-GFP, then fusions with SAP carrying different positive charges were generated by a PCR procedure using plasmids expressing S1nv1-enzyme-GFP as a template. The scheme for constructing the library is illustrated in Fig. 2.

Evaluation of the SAP library using four enzymes

Four enzymes, which included PGL from *B. Subtilis* WSHB04-02 [30], LOX from *P. Aeruginosa* BBE [23], MTG from *S. Mobaraense* [25], and ASN from *E. coli* strain BL21 (DE3) [31], were used to verify the efficiency of the library. Each enzyme fused with GFP was constructed (Fig. 1d) and used as the control sample. As shown in Fig. 5, all of the cells expressing enzymes fused with SAP and GFP exhibited enhanced fluorescence intensity in contrast to cells carrying the corresponding enzyme fused with GFP. Moreover, the SAPs that were screened covered the net charges ranging from +1 to +20, and those SAPs with net charge +4 (S1nv10), +6 (S1nv11), +3 (S1nv17), and +3 (S1nv17) (Table 1) achieved the highest fluorescence intensity with PGL, LOX, ASN, and MTG, respectively. To confirm the positive relationship between the fluorescence intensity and protein production, the GFP tag was removed from fusions with SAP and GFP for each enzyme, yielding the enzyme solely fused with SAP (Fig. 2). As shown in Additional file 1: Figure S3, the yield of enzymes fused with SAPs displayed a substantial positive correlation

with fluorescence intensity of the corresponding enzyme fusion with SAP and GFP in the case of the four enzymes (PGL, LOX, ASN, and MTG). Moreover, as shown in Fig. 6, in contrast to the corresponding wild-type enzyme, the PGL fused with S1nv10, LOX fused with S1nv11, ASN fused with S1nv17, and MTG fused with S1nv17 exhibited 8.3, 3.5, 3.68, and 2.64-fold increases in protein yield, respectively. SDS-PAGE analysis showed the same trend in the protein expression of wild-type and SAP-fusion of each enzyme (Fig. 6).

Enzyme characterization of the SAP fusions

The influence of SAPs on the catalytic efficiencies and thermal stabilities of the fusions were also assayed. As shown in Table 3, the specific activities changed slightly whereas the thermal stabilities of the SAP fusions improved significantly. The specific activities of S1nv10-PGL, S1nv11-LOX, and S1nv17-ASN presented 0.79, 0.26, and 0.23-fold increase in specific activity, and 2.1, 3.82, and 0.98-fold increase in $t_{1/2}$ relative to the corresponding wild-type, respectively. For MTG, because the active enzyme was produced by the

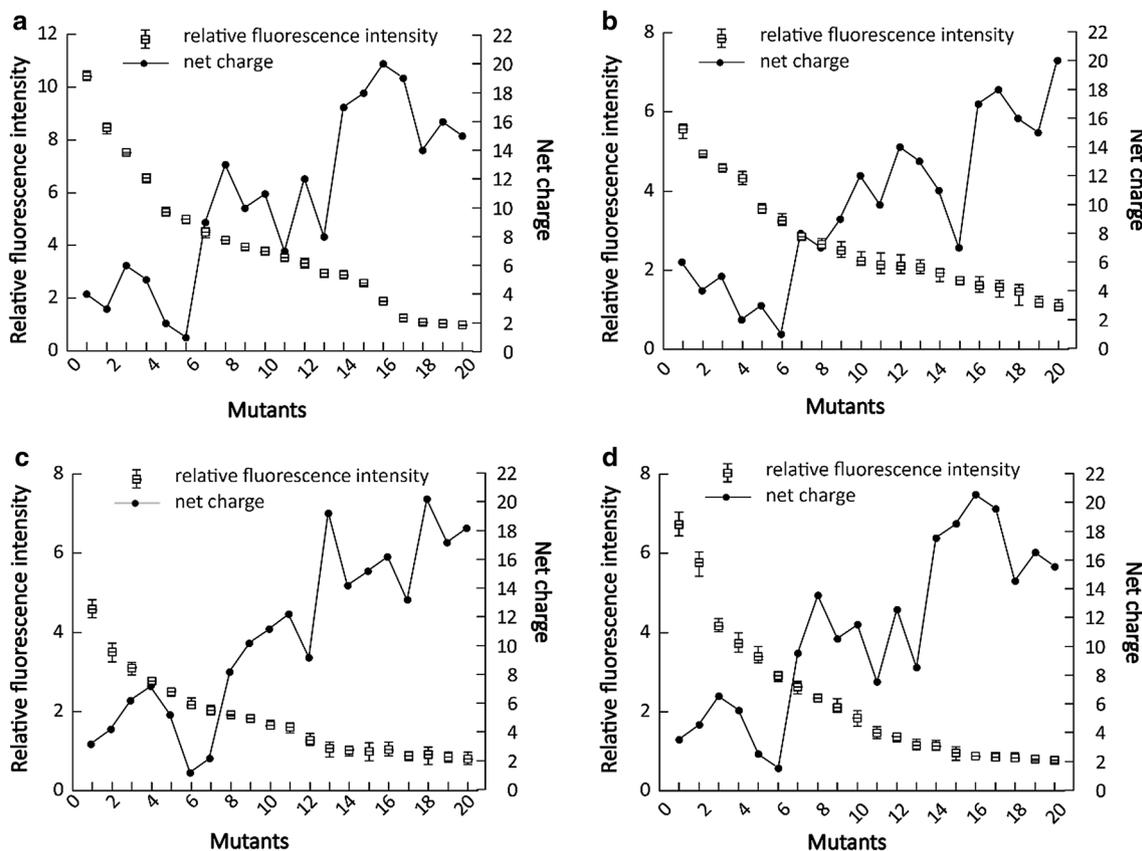


Fig. 5 Fluorescence intensity and SAP net charge of each enzyme fused with SAP and GFP. **a** PGL fusions; **b** LOX fusions; **c** ASN fusions; **d** MTG fusions. The relative fluorescence intensity of each fusion was divided by that of the corresponding wild-type enzyme, yielding the normalized relative production yield. Each result was the average value of three parallel experiments

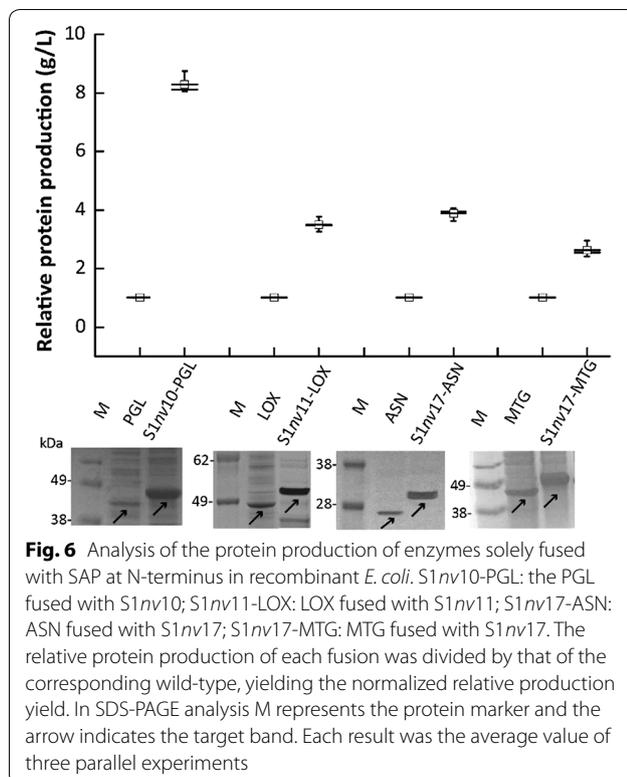


Table 3 Enzymatic properties of wild-type enzymes and fusion proteins

Enzyme	Value \pm SD			
	Km (g L ⁻¹ #, mM*, μ mol L ⁻¹ "#)	kcat (min ⁻¹ #, S ⁻¹ *, S ⁻¹ "#)	Specific activity (U mg ⁻¹)	t1/2 (min)
PGL	0.27 \pm 0.011	12.69 \pm 0.4	279.14 \pm 3.5	5.3 \pm 0.32
S1nv10-PGL	0.47 \pm 0.07	65.6 \pm 6.8	499.6 \pm 8.4	16.4 \pm 1.3
LOX	0.057 \pm 0.004	24.7 \pm 2.97	30.2 \pm 1.22	9.7 \pm 0.43
S1nv11-LOX	0.052 \pm 0.006	81.2 \pm 3.6	38.23 \pm 1.87	46.8 \pm 1.5
ASN	14.2 \pm 0.31	49.2 \pm 1.9	17.34 \pm 0.16	12.3 \pm 0.42
S1nv17-ASN	21.4 \pm 0.34	68.6 \pm 2.5	21.34 \pm 0.6	24.4 \pm 0.65

Each result was the average value of three parallel experiments. SD is the standard deviation of three experiments

#, * and " indicated the units for PGL, LOX, and ASN, respectively

removal of its N-terminal pro-peptide, the enzymatic properties of the fusions were nearly not changed compared with the wild-type (data were not shown).

Discussion

Traditional optimization strategies, based on promoters, ribosome-binding sites, or untranslated region modifications, have been used to improve the production of recombinant enzymes [32, 33]. It should

be noted, that some proteins remain poorly expressed under these strategies probably due to differences in amino acid sequence [34]. To solve this problem, sequence modification [10] of the target protein has received much attention in recent years, including the addition of expression fusion tags to the N-terminus of target proteins [35]. However, the fusion of proteins with expression tags could not improve the production of all proteins [36]. SAPs, an emerging fusion tag, have been successfully used for improving protein production [17], purification [37], and thermal stability or catalytic activity [14]. In this study, an expression tag library composed of SAPs, which varied in net charge, was constructed in *E. coli*. The usefulness of this library was validated by expressing PGL, LOX, ASN, and MTG, with their optimized SAP fusions exhibiting 8.3, 3.5, 2.64, and 3.68-fold increases in production yield, respectively, relative to the corresponding wild-type enzyme. Thus, the results indicated the high efficiency of this SAP-based library in *E. coli*.

Key factor analysis indicated that hydrophobicity and net charges of SAPs played a key role in SAP fusion expression. It was shown that changes in the hydrophobicity of S1 tend to cause the formation of insoluble inclusion bodies of protein. Moreover, the SAP positive net charge was more efficient for protein expression than those with a negative net charge. Thus, the SAP tag library only included 20 types of SAPs with net charges ranging from +1 to +20. Because the SAP length had little effect on the expression of SAP fusions, the variation in SAP net charge was achieved by altering the length of the positive charged S1nv1 (ANANARAR)₁₀ through a PCR procedure, thereby, avoiding the tedious gene cloning of each SAP fusion one by one. Although only 96 colonies of each enzyme fusion were screened, the SAP library still achieved a considerable rate of those fusions with enhanced protein expression, which was much higher than that of irrational-screening strategies [38]. Moreover, in contrast to N-terminal modification [39] or site-directed mutagenesis strategies [40], application of the SAP library could also improve the thermal stabilities or activities of the enzymes. Due to the application of single-factor experiment, the synergistic effect among different factors were not considered in the current study, such as the effect of the SAP hydrophobicity on the protein expression under the optimized net charges in SAPs. Further investigation will be performed on the effects of different combinations of the four factors.

In this study, the S1 variants with a hydrophobicity that was lower or higher than S1 showed reduced fusion protein expression. To date, the mechanism for the effect of hydrophobicity on protein expression is still unclear. Among the six S1 hydrophobicity variants, SAPs with

Ile and Leu induced the GFP inclusion body formation. This result was consistent with a previous study in which active inclusion bodies were obtained after fusions with a SAP containing several Leu residues [41]. Thus, the formation of inclusion bodies may partly account for reduced expression. Considering the high sensitivity of the fusion expression to SAP hydrophobicity, further investigation will be performed on the precise regulation of SAP hydrophobicity to improve the diversity of the SAP library.

As indicated by the key factor analysis, the positive net charge of SAPs could enhance the protein production more effectively in contrast to those with no or negative net charges. It has been reported that the positive charge on the N-terminus of newly synthesized peptides could interact with the negatively charged ribosomal exit tunnel to slow down the initiation rate of protein expression [42]. The relatively low initiation rate may benefit correct protein folding through facilitating interactions between the emerging peptide and the chaperone proteins, and these interactions may increase with the number of positively charged residues [43]. As SAP was fused at the N-terminus of the target proteins, it was possible that SAPs with positive net charges enhanced the protein expression through similar electronic interactions. However, the optimal number of the net charge in SAPs for protein expression ranged from +2 to +6 among the four enzymes tested, and the additive effects on the expression were not observed. Notably, the N-terminal sequence with the ability to affect the protein expression was approximately limited to the first 18 amino acid residues (around) [42]. Because the variation in positive net charge was obtained by changing the length of the positive charged $S1_{nv1}$ (ANANARAR)₁₀, the $S1_{nv1}$ variants with over six positive net charges were composed of more than 24 residues. The positively charged residues after the first 18 amino acid residues of the SAP may reduce the expression efficiency due to the additional interactions with the charged ribosomal exit tunnel. Increasing positively charged residues without extending the SAP sequence may further enhance fusion expression.

Conclusion

In summary, we proposed an efficient expression tag library based on SAPs in *E. coli*, where a positive net charge was selected as the major variable. As a proof-of-concept, four enzymes showed substantial increases in protein production without sacrificing specific activities and thermal stabilities, suggesting that the strategy was efficient. This study provided a simple and convenient strategy to enhance the production of recombinant proteins and established a basis for the application of SAP fusion.

Additional file

Additional file 1: Table S1. Primers used in this study. Figure S1.

Schemes for the construction of GFP fused with SAP varied in S1 unit. In the first round of PCR, the forward primers S1/v1-F separately bind to four sites of the S1 encoding region (1, 2, 3, and 4) under low annealing temperature, yielding I, II, III, and IV linearized plasmids. Then in the second round of PCR, these linearized plasmids are used as the templates to amplify the linearized plasmids encoding GFP fused with different lengths of SAPs. After the ligation reaction, the circular plasmids encoding GFP fused with nS1 are obtained. The PCR and ligation reaction system and conditions are performed as described in [Materials and Methods](#).

Figure S2. The fluorescence intensity of GFP fused with S1 via different linker composition in *E. coli*. (A) GFP fused with S1 via different linker units; (B) GFP fused with S1 via 5 linker units containing a different number of flexible linker units. All linker sequences are listed in [Table 2](#). The rigid and flexible linker units referred to EAAAK and GGGGS, respectively. The fluorescence intensity of GFP fusions was normalized by that of GFP. Each result was the average value of three parallel experiments. **Figure S3.**

The correlation analysis of protein production and fluorescence intensity. (A): PGL and its GFP fusions; (B): LOX and its GFP fusions; (C): ASN and its GFP fusions; (D): MTG and its GFP fusions. The relative protein production and relative fluorescence intensity were measured as described in [Materials and Methods](#). Each result was the average value of three parallel experiments.

Abbreviations

SAPs: self-assembling amphipathic peptides; PGL: poly galacturonate lyase; LOX: lipoxigenase; ASN: L-asparaginase; MTG: transglutaminase; *E. coli*: *Escherichia coli*.

Acknowledgements

Not applicable.

Authors' contributions

SL supervised the project and revised the manuscript, WXZ carried out the experiments and drafted the manuscript. GCD and JWZ revised the manuscript. All authors read and approved the final manuscript.

Funding

This work is financially supported by the National Natural Science Foundation of China (31771913 and 31401638), Key Research and Development Program of Jiangsu Province (BE2016629) and the National First-class Discipline Program of Light Industry Technology and Engineering (LITE2018-08).

Availability of data and materials

All data generated or analyzed during this study are included in this published article and additional file.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ National Engineering Laboratory for Cereal Fermentation Technology, Jiangnan University, Wuxi 214122, China. ² School of Biotechnology, Jiangnan University, Wuxi 214122, China. ³ The Key Laboratory of Carbohydrate Chemistry and Biotechnology, Ministry of Education, Jiangnan University, Wuxi 214122, China. ⁴ Jiangsu Provisional Research Center for Bioactive Product Processing Technology, Jiangnan University, Wuxi 214122, China.

Received: 16 December 2018 Accepted: 17 May 2019

Published online: 27 May 2019

References

- Kolaj O, Spada S, Robin S, Wall JG. Use of folding modulators to improve heterologous protein production in *Escherichia coli*. *Microb Cell Fact*. 2009;8:17.
- Terpe K. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol*. 2006;72:211–22.
- Goodman DB, Church GM, Kosuri S. Causes and effects of N-Terminal codon bias in bacterial genes. *Science*. 2013;342:475–9.
- Esposito D, Chatterjee DK. Enhancement of soluble protein expression through the use of fusion tags. *Curr Opin Biotech*. 2006;17:353–8.
- Garcia-Fraga B, da Silva AF, Lopez-Seijas J, Sieiro C. Optimized expression conditions for enhancing production of two recombinant chitinolytic enzymes from different prokaryote domains. *Bioproc Biosyst Eng*. 2015;38:2477–86.
- Nygren PA, Stahl S, Uhlen M. Engineering proteins to facilitate bio-processing. *Trends Biotechnol*. 1994;12:184–8.
- Fox JD, Waugh DS. Maltose-binding protein as a solubility enhancer. *Methods Mol Biol*. 1996;205: p. 99–117.
- Marblestone JG, Edavettal SC, Lim Y, Lim P, Zuo X, Butt TR. Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. *Protein Sci*. 2006;15:182–9.
- Davis GD, Elisee C, Newham DM, Harrison RG. New fusion protein systems designed to give soluble expression in *Escherichia coli*. *Biotechnol Bioeng*. 1999;65:382–8.
- Kudla G, Murray AW, Tollervey D, Plotkin JB. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science*. 2009;324:255–8.
- Hammarstrom M, Helligren N, Van den Berg S, Berglund H, Hard T. Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*. *Protein Sci*. 2002;11:313–21.
- Arnau J, Lauritzen C, Petersen GE, Pedersen J. Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. *Protein Expr Purif*. 2006;48:1–13.
- Zhang S, Lockshin C, Herbert A, Winter E, Rich A. Zootin, a putative Z-DNA binding protein in *Saccharomyces cerevisiae*. *Embo J*. 1992;11:3787–96.
- Lu XY, Liu S, Zhang DX, Zhou XM, Wang M, Liu Y, Wu J, Du GC, Chen J. Enhanced thermal stability and specific activity of *Pseudomonas aeruginosa* lipoxigenase by fusing with self-assembling amphipathic peptides. *Appl Microbiol Biotechnol*. 2013;97:9419–27.
- Yang H, Lu X, Liu L, Li J, Shin HD, Chen RR, Du G, Chen J. Fusion of an oligopeptide to the N terminus of an alkaline alpha-amylase from *Alkalimonas amylolytica* simultaneously improves the enzyme's catalytic efficiency, thermal stability, and resistance to oxidation. *Appl Environ Microbiol*. 2006;79:3049–58. <https://doi.org/10.3189/172756506781828700>.
- Liu Y, Cui W, Liu Z, Cui Y, Xia Y, Kobayashi M, Zhou Z. Enhancement of thermo-stability and product tolerance of *Pseudomonas putida* nitrile hydratase by fusing with self-assembling peptide. *J Biosci Bioeng*. 2014;118:249–52.
- Zhao W, Liu L, Du G, Liu S. A multifunctional tag with the ability to benefit the expression, purification, thermostability and activity of recombinant proteins. *J Biotechnol*. 2018;283:1–10.
- Wu W, Xing L, Zhou BH, Lin ZL. Active protein aggregates induced by terminally attached self-assembling peptide ELK16 in *Escherichia coli*. *Microb Cell Fact*. 2011;10:1.
- Gokhale RS, Khosla C. Role of linkers in communication between protein modules. *Curr Opin Biotech*. 2000;4:22–7.
- Yang F, Moss LG, Phillips GN. The molecular structure of green fluorescent protein. *Nat Biotechnol*. 1996;14:1246–51.
- Athey J, Alexaki A, Osipova E, Rostovtsev A, Santana-Quintero LV, Katneni U, Simonyan V, Kimchi-Sarfaty C. A new and updated resource for codon usage tables. *BMC Bioinform*. 2017;18:10.
- Liu S, Zhang DX, Wang M, Cui WJ, Chen KK, Du GC, Chen J, Zhou ZM. The order of expression is a key factor in the production of active transglutaminase in *Escherichia coli* by co-expression with its pro-peptide. *Microb Cell Fact*. 2011;10:7.
- Lu XY, Zhang J, Liu S, Zhang DX, Xu Z, Wu J, Li JH, Du GC, Chen J. Overproduction, purification, and characterization of extracellular lipoxigenase of *Pseudomonas aeruginosa* in *Escherichia coli*. *Appl Microbiol Biotechnol*. 2013;97:5793–800.
- Wang Y, Wang ZH, Du GC, Hua ZZ, Liu LM, Li JH, Chen J. Enhancement of alkaline polygalacturonate lyase production in recombinant *Pichia pastoris* according to the ratio of methanol to cell concentration. *Bioresour Technol*. 2009;100:1343–9.
- Zhang DX, Wang M, Wu J, Cui L, Du GC, Chen J. Two different proteases from *Streptomyces hygroscopicus* are involved in transglutaminase activation. *J Agr Food Chem*. 2008;56:10261–4.
- Feng Y, Liu S, Jiao Y, Gao H, Wang M, Du G, Chen J. Enhanced extracellular production of L-asparaginase from *Bacillus subtilis* 168 by *B. subtilis* WB600 through a combined strategy. *Appl Microbiol Biotechnol*. 2017;101:1509–20.
- O'Fagain C. Enzyme stabilization-recent experimental progress. *Enzyme Microb Tech*. 2003;33:137–49.
- Chen XY, Zaro JL, Shen WC. Fusion protein linkers: property, design and functionality. *Adv Drug Deliver Rev*. 2013;65:1357–69.
- Li G, Huang Z, Zhang C, Dong B-J, Guo R-H, Yue H-W, Yan L-T, Xing X-H. Construction of a linker library with widely controllable flexibility for fusion protein design. *Appl Microbiol Biotechnol*. 2016;100:215–25.
- Wang Z, Wang Y, Zhang D, Li J, Hua Z, Du G, Chen J. Enhancement of cell viability and alkaline polygalacturonate lyase production by sorbitol co-feeding with methanol in *Pichia pastoris* fermentation. *Bioresour Technol*. 2010;101:1318–23.
- Campbell HA, Mashburn LT, Boyse EA, Old LJ. Two L-asparaginases from *Escherichia coli* B. Their separation, purification, and antitumor activity. *Biochemistry*. 1967;6:721–30.
- Mutalik VK, Guimaraes JC, Cambray G, Lam C, Christoffersen MJ, Mai QA, Tran AB, Paull M, Keasling JD, Arkin AP, Endy D. Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat Methods*. 2013;10:354.
- Zhou SH, Ding RP, Chen J, Du GC, Li HZ, Zhou JW. Obtaining a panel of cascade promoter-5'-UTR complexes in *Escherichia coli*. *ACS Synth Biol*. 2017;6:1065–75.
- Kataeva I, Chang J, Xu H, Luan CH, Zhou JZ, Uversky VN, Lin DW, Horanyi P, Liu ZJ, Ljungdahl LG, et al. Improving solubility of *Shewanella oneidensis* MR-1 and *Clostridium thermocellum* JW-20 proteins expressed into *Escherichia coli*. *J Proteome Res*. 2005;4:1942–51.
- Yadav DK, Yadav N, Yadav S, Haque S, Tuteja N. An insight into fusion technology aiding efficient recombinant protein production for functional proteomics. *Arch Biochem Biophys*. 2016;612:57–77.
- Kosobokova EN, Skrypnik KA, Kosorukov VS. Overview of fusion tags for recombinant proteins. *Biochemistry-Moscow*. 2016;81:187–200.
- Lin ZL, Zhao Q, Xing L, Zhou BH, Wang X. Aggregating tags for column-free protein purification. *Biotech J*. 2015;10:1877–86.
- Copley SD. Toward a systems biology perspective on enzyme evolution. *J Biol Chem*. 2012;287:3–10.
- Borujeni AE, Cetnar D, Farasat I, Smith A, Lundgren N, Salis HM. Precise quantification of translation inhibition by mRNA structures that overlap with the ribosomal footprint in N-terminal coding sequences. *Nucleic Acids Res*. 2017;45:5437–48.
- Karshikoff A, Nilsson L, Ladenstein R. Rigidity versus flexibility: the dilemma of understanding protein thermal stability. *Febs J*. 2015;282:3899–917.
- Lin Z, Zhou B, Wu W, Xing L, Zhao Q. Self-assembling amphipathic alpha-helical peptides induce the formation of active protein aggregates in vivo. *Faraday Discuss*. 2013;166:243.
- Tuller T, Zur H. Multiple roles of the coding sequence 5' end in gene expression regulation. *Nucleic Acids Res*. 2015;43:13–28.
- Charneski CA, Hurst LD. Positively charged residues are the major determinants of ribosomal velocity. *PLoS Biol*. 2013;11:20.

Publisher's Note

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