

1
2
3 **Enhancing the translational capacity of *E. coli* by resolving the codon bias**
4
5
6

7 Zoltan Lipinszki^{a,b}, Viktor Vernyik^a, Nora Farago^c, Tobias Sari^a, Laszlo G. Puskas^c, Frederick R.
8 Blattner^d, Gyorgy Posfai^{a,#} and Zsuzsanna Gyorfy^{a,#}
9
10
11
12

13
14 ^aInstitute of Biochemistry, Biological Research Centre of the Hungarian Academy of Sciences,
15 Szeged, Hungary
16
17

18
19 ^bMTA SZBK Lendület Laboratory of Cell Cycle Regulation, Biological Research Centre of the
20 Hungarian Academy of Sciences, Szeged, Hungary
21
22

23
24 ^cInstitute of Genetics, Biological Research Centre of the Hungarian Academy of Sciences,
25 Szeged, Hungary
26
27

28 ^dScarab Genomics LLC, Madison, WI, USA
29
30
31

32
33 #corresponding authors: Zsuzsanna Gyorfy, gyorfyzs@brc.hu; Gyorgy Posfai,
34 posfai.gyorgy@brc.mta.hu.
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Abstract

Escherichia coli is a well-established, and popular host for heterologous expression of proteins. The preference in the choice of synonymous codons (codon bias), however, might differ for the host and the original source of the recombinant protein, constituting a potential bottleneck in production. Codon choice affects the efficiency of translation by a complex and poorly understood mechanism. The availability of certain tRNA species is one of the factors that may curtail the capacity of translation. Here we provide a tRNA-overexpressing strategy that allows the resolution of the codon bias, and boosts the translational capacity of the popular host BL21(DE3) when rare codons are encountered. In BL21(DE3)-derived strain, called SixPack, copies of the genes corresponding to the six least abundant tRNA species have been assembled in a synthetic fragment and inserted into a ribosomal RNA operon. This arrangement, while not interfering with the growth properties of the new strain, allows dynamic control of the transcription of the extra tRNA genes, providing significantly elevated levels of the rare tRNAs in exponential growth phase. Results from expression assays of a panel of recombinant proteins of diverse origin and codon composition showed that the performance of SixPack surpassed that of the parental BL21(DE3) or a related strain equipped with a rare tRNA-expressing plasmid.

Keywords: rare tRNA genes, genome editing, efficient translation, recombinant protein production, *E. coli*

E. coli is by far the most widely used host organism for biopharmaceutical heterologous production of recombinant proteins. This expression platform is favored for its simplicity, speed, and low cost. The codon bias discrepancy, however, can seriously hinder protein expression in *E.*

1
2
3 *coli*¹⁻⁵. Choice in the usage of synonymous codons can be different in various organisms, and this
4
5 bias has been shown to correlate with the relative and absolute quantities of individual tRNAs^{6,7}.
6
7 Heterologous expression of a protein with a high ratio of codons occurring infrequently in *E. coli*
8
9 might deplete the corresponding tRNA species, leading to translational frameshifting, codon
10
11 skipping, misincorporations, and protein truncations⁸. Ultimately, the codon bias seriously limits
12
13 the use of *E. coli* as an expression platform.
14
15
16
17
18

19 To overcome this problem, codon optimization or rare tRNA overexpression strategies have been
20
21 applied, with limited success. On one hand, synthesis of the recombinant protein encoding gene
22
23 with an *E. coli* codon preference is labor-intensive and expensive. Moreover, replacing rare
24
25 codons with frequent ones does not necessarily lead to increased yield in protein synthesis. Codon
26
27 choice might affect expression, solubility, and folding of a protein⁹⁻¹⁵, and rare codons can,
28
29 paradoxically, enhance the translation of a gene via reducing the mRNA secondary structure
30
31 ^{7,16,17}. Additionally, codon optimization by gene synthesis is not feasible when testing gene
32
33 libraries. On the other hand, attempts to overexpress rare tRNA species by cloning extra copies of
34
35 the corresponding tRNA encoding genes into a plasmid have their drawbacks as well. In such
36
37 commercially available hosts, maintenance of the tRNA-expressing plasmid requires the addition
38
39 of extra antibiotics (usually the protein synthesis inhibitor chloramphenicol); moreover, the
40
41 recombinant protein-encoding expression plasmid must belong to a different complementation
42
43 group. Using two antibiotics and/or permanently altering the balance of the various tRNA species
44
45 can have a fitness cost on the host, eventually resulting in a low success rate in applications^{10,18}.
46
47
48
49
50
51
52

53 We sought to resolve the codon bias problem by expressing rare tRNAs in a more dynamic
54
55 fashion. We hypothesized that inserting extra copies of the relevant tRNA genes into a ribosomal
56
57
58
59
60

1
2
3 RNA (*rrn*) operon would harmonize their expression with the translational activity and would
4
5 provide enhanced levels of rare tRNAs according to the actual needs. *rrn* operons have key roles
6
7 in bacterial physiology and economy. Synthesis of rRNA quickly reacts to environmental
8
9 conditions, determining ribosome availability via regulating the expression of ribosomal proteins
10
11 by a translational feedback mechanism¹⁹. In *E. coli*, there are seven nearly identical copies of
12
13 rRNA operons. Synthesis of rRNA is driven by the strongest promoters found in the genome²⁰,
14
15 and the rate of transcription can change more than an order of magnitude between poor and rich
16
17 nutrient conditions²¹. While most of the tRNA genes are scattered around the chromosome, all
18
19 rRNA operons carry certain (but not any of the rare) tRNA genes as well, co-transcribed with the
20
21 rRNA genes.
22
23
24
25
26
27

28 We inserted extra copies of the six tRNA genes (*argX*, *glyT*, *leuW*, *proL*, *argU*, and *ileX*)
29
30 corresponding to the minor codons of *E. coli* (CGG, GGA, CUA, CCC, AGA/AGG, and AUA,
31
32 respectively) into one of the ribosomal RNA operons (*rrnD*) of BL21(DE3), the widely used *E.*
33
34 *coli* expression host. We demonstrated that expression of these tRNA genes shows a marked
35
36 increase in exponential growth phase, compared to the unmodified host. By testing the expression
37
38 of a panel of recombinant proteins with a high ratio of rare codons in their genes, we showed that,
39
40 in most cases, the modified strain that we named SixPack performs better and shows significantly
41
42 enhanced expression of the heterologous proteins in classical IPTG-induction as well as in an
43
44 auto-induction system. Moreover, compared to the commercially available strain
45
46 Rosetta2(DE3)pLysS (Merck), which carries a similar array of extra tRNA genes on a plasmid
47
48 (pLysSRARE2; Merck), SixPack proves to be a superior expression platform.
49
50
51
52
53
54
55

56 Results

57
58
59
60

Design and genomic insertion of extra copies of tRNA genes

Hypothetically, cloning extra copies of the genes corresponding to rare tRNAs into a *rrn* operon allows their controlled expression, as the activity of the operon is tightly regulated by nutrient availability and other physiological conditions²². This might ensure that expression of the rare tRNAs is coupled to the actual need for translational capacity.

The genes encoding for the six least abundant tRNA species were combined in a single, synthetic DNA fragment. The original copies of these genes are either single genes whose expression is controlled by their own promoter (*proL*, *argU*, or *ileX*), or are parts of polycistronic operons (*argX*, *glyT*, or *leuW*) (Figure S1). Since the process of maturation of the primary transcripts of tRNA genes is not thoroughly understood, the nucleotide sequences surrounding the genes in the synthetic fragment were designed to mimic the natural tRNA operons. The intergenic and flanking regions of the *argX-glyT-leuW-proL* segment were essentially identical to those of the natural *argX-hisR-leuT-proM* polycistronic operon. For *argU* and *ileX*, exact copies of the genes and their flanking sequences were fused to the 3' end of the segment (Figure S2).

For the genomic insertion site *rrnD*, one of the seven *rrn* operons was selected. Although the structure, sequence, and activity of the *rrn* operons of *E. coli* are very similar, the 3' end of *rrnD* (containing *trhV* and *rrfF*) is unique, allowing specific targeting of the locus²⁰.

The 1207-base synthetic operon carrying the six tRNA genes was engineered into the 3' region of *rrnD* (Figure 1) by a multistep process using λ RED recombineering, followed by elimination of

1
2
3 the markers via CRISPR-Cas9-stimulated homologous recombination (Figure S3), resulting in
4
5 strain SixPack.
6
7
8
9

10 **Growth properties of SixPack and control strains**

11
12 Altering the expression level of rare tRNAs might affect the speed of translation, co-translational
13
14 folding of proteins and, consequently, the growth rate of *E. coli*²³. Apparently, insertion of the
15
16 extra tRNA genes into *rrnD* did not disturb the homeostasis of the cell. The growth parameters of
17
18 SixPack were identical to those of the parental BL21(DE3) in both LB and AIM (auto-induction
19
20 medium). Signaling normal cellular operation, expression levels of two chaperons (DnaK and
21
22 GroEL, the increase of which might be indicative of abnormal folding of native proteins) were
23
24 not significantly altered (Figure S4). Noticeably, the commercially available
25
26 Rosetta2(DE3)pLysS strain had a longer lag phase, lower growth rate, and lower maximum OD
27
28 in both media, presumably owing to the burden of maintaining the plasmid and the presence of
29
30 the antibiotic chloramphenicol in the media (Figure 2 and Table S1).
31
32
33
34
35
36
37

38 **Elevated expression of rare tRNAs**

39
40 Expression of the six rare tRNA genes was measured by qRT-PCR on first-strand cDNA
41
42 generated from total RNA isolated from cultures of BL21(DE3) and SixPack in the exponential
43
44 phase. Primers were designed to amplify both the unprocessed and the matured forms of the
45
46 various tRNA species. In the case of SixPack, the measured rare tRNAs represented transcripts
47
48 originating both from their original genomic copies and from the newly inserted fragment in
49
50 *rrnD*. The rare tRNA ratios were normalized with the ratios of three control (abundant) tRNA
51
52 species.
53
54
55
56
57
58
59
60

1
2
3 Compared to parental BL21(DE3), SixPack showed elevated expression (1.8- to 3.8-fold) of the
4
5 six rare tRNAs (Figure 3).
6
7
8
9

10 Despite being on the same cistron and transcribed together, increases in the expression levels of
11
12 the extra tRNA genes were diverse (*argU* showing the highest and *glyT* the lowest increase). This
13
14 might be due to several factors, including tRNA processing, stability, regulation, and diverse
15
16 expression levels of the original genomic copies.
17
18
19
20

21 **Individual impact of elevated levels of the rare tRNAs on protein expression**

22
23 The functional effect of the increased levels of rare tRNAs was first tested separately for each
24
25 tRNA species. To magnify the effect of rare tRNAs, we constructed IPTG-inducible, pETDuet-1-
26
27 based (Novagen) test plasmids expressing modified versions of the GFP gene under the
28
29 regulation of the T7 promoter. GFP genes carrying runs of rare codons corresponded to each of
30
31 the six extra tRNA genes (one GFP version corresponding to *proL*, *ileX*, *argX*, *glyT*, and *leuW*
32
33 each, and two GFP versions for *argU*, as it relates to both AGA and AGG codons). The tandem
34
35 rare codons (3 or 5 copies) were inserted into the GFP gene right after the ATG codon.
36
37
38
39

40 Expressions of the modified GFPs were detected by fluorescence measurements and visualized
41
42 by denaturing polyacrylamide gel electrophoresis (SDS-PAGE).
43
44
45

46
47 IPTG- (or lactose-) induced expression of the GFP and its variants with rare codons had a
48
49 negative but diverse impact on the growth of the strains. In LB medium, SixPack and BL21(DE3)
50
51 usually showed similar growth patterns, and in most cases displayed higher growth parameters
52
53 than Rosetta2(DE3)pLysS (Figure S5). In AIM, peculiarly, Rosetta2(DE3)pLysS reached the
54
55 highest maximal optical density when GFP variants with rare codons were introduced (Figure
56
57
58
59
60

1
2
3 S6). The reason might lie in the low metabolic burden, attributable to the inefficient expression of
4 the particular protein. (We note that using standard AIM, the induction of the T7 polymerase and,
5 consequently, T7 promoter-driven genes might be below full capacity in Rosetta2(DE3)pLysS,
6 and this might contribute to the low expression level.)
7
8
9
10
11
12

13
14 Total fluorescence measurements of GFP production, however, showed a clear advantage of
15 SixPack in both LB (Figure 4) and AIM (Figure S7). In both media, SixPack showed higher
16 expression of the modified GFPs than the parental BL21(DE3) did. In one exception, there was
17 no difference (3xGGA-GFP; GGA codon relates to *glyT*) (Figure S7H). Compared to
18 Rosetta2(DE3)pLysS, SixPack showed higher expression in all cases. As an additional control,
19 expression of GFP and 5xAGA-GFP were also tested in the other commercial strain
20 Rosetta2(DE3). This strain, presumably due to the lack of T7 lysozyme, performed better than
21 Rosetta2(DE3)pLysS (where introduction of LysS is intended to decrease background expression
22 of the cloned gene), however, the expression levels were significantly lower than in SixPack
23 (Figure S8).
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39

40 Results of the fluorescence measurements were further supported by protein gel assays. Equal
41 amounts of proteins obtained from crude extracts of cultures collected after a 10 h incubation
42 were analyzed on Coomassie Brilliant Blue-stained SDS-PAGE gels. Differences in the amounts
43 of modified GFPs in the three strains were in clear correlation with the fluorescence data (Figure
44 S9).
45
46
47
48
49
50
51
52

53 **Heterologous protein expression performance of SixPack and control strains**

54
55
56
57
58
59
60

1
2
3 The usefulness of SixPack was further tested by expressing a panel of eight protein-coding genes
4 of diverse length and rare codon content (Table S2), selected from various species, including
5 *Drosophila melanogaster*, *Pyrococcus furiosus*, *Saccharomyces cerevisiae*, *Streptococcus*
6 *pyogenes*, and human papillomavirus (HPV). The genes were cloned into pETDuet-1 and
7 expressed in SixPack and control strains under two different conditions (LB and AIM).
8
9

10
11
12
13
14
15
16
17 The growth curves of IPTG- (or lactose-) induced cultures seemed to be quite diverse in LB
18 (Figure S10), but rather uniform in AIM (Figure S11). Production of the specific proteins was
19 visualized by using Coomassie-stained gels and Western blotting (Figure 5). In most cases,
20 SixPack proved to be the best producer. This was especially evident when the ratio of rare codons
21 was higher than 8% (and the ratio of tandems was also high), as in the cases of HumdCas9
22 (dCas9 optimized to human codon preference), HumdCas9-GFP fusion, and *Pfu* DNA
23 polymerase. In the case of EFT1, the ratio was only 4.3% (35 AGA and 2 CTA codons); but
24 SixPack still produced the highest amount of this protein. It is conceivable that the translation
25 efficiency is influenced not only by the total amount of rare codons, but also by the occurrence of
26 tandem situations, the location of rare codons in the gene (e.g., close or far from the start site),
27 and the ratios of the different rare codon species³. In other cases (Asl, dCas9, Flfl, and HPV16
28 L2), when the rare codon ratio was below 8%, production by SixPack was similar to that of
29 BL21(DE3) in LB and similar to or better than that in AIM. In all cases, the performance of
30 Rosetta2(DE3)pLysS lagged behind that of SixPack (especially evidently in AIM) or was similar
31 at most (Figure 6).
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52

53 Usefulness of SixPack was further demonstrated by comparing the production of two versions of
54 the same protein (dCas9, 5.4% rare codons vs. HumdCas9, 8.6% rare codons), which can be
55
56
57
58
59
60

1
2
3 viewed as rare codon-containing and codon-optimized versions of the same protein. In this
4
5 particular case, higher expression of HumdCas9 by SixPack versus dCas9 by BL21(DE3) clearly
6
7 demonstrated the advantage of SixPack (Figure S12).
8
9

11 **Discussion**

12
13
14
15
16
17 Depletion of rare tRNA species evoked by the expression of heterologous proteins can
18
19 compromise product yield and translational fidelity. Several attempts, both genetic
20
21 (overexpression of rare tRNA genes from a plasmid, or codon optimization of the heterologous
22
23 gene by synthesis) and non-genetic ones (optimizing the growth conditions, or using *in vitro*
24
25 protein synthesis systems) have been applied with varying degrees of success^{2, 24-28}. We offer
26
27 here a solution that resolves codon bias discrepancies with a high success rate, does not require
28
29 additional antibiotics, and performs better or as well as the widely used BL21(DE3) *E. coli* strain.
30
31
32
33
34

35 The novelty of the approach lies in placing the extra copies of rare tRNA genes into a ribosomal
36
37 operon of the host chromosome. This arrangement offers the following benefits: (i) lack of
38
39 physiological burden associated with plasmid replication and maintenance, (ii) genetic stability
40
41 without the use of an antibiotic, and (iii) dynamic changes in the expression of rare tRNA genes,
42
43 paralleling the actual translational activity.
44
45
46
47
48

49 A ribosomal operon is a natural choice as a location for inserting tRNA genes. While most tRNA
50
51 genes are scattered elsewhere in the genome, each of the *rrn* operons carries at least one tRNA
52
53 gene (typically coding for abundant tRNAs). As anticipated, increased amounts of the rare
54
55 tRNAs, corresponding to the inserted genes, were detected. However, when compared to the
56
57
58
59
60

1
2
3 levels in the parental host BL21(DE3), the rate of increase in the exponential phase was only
4
5 modest. An interplay of several factors, including co-factors of tRNA maturation, tRNA stability,
6
7 and feedback regulation of the genes at the original locations, might mitigate the increase. This
8
9 limited elevation in rare tRNA abundance might even be a fortuitous feature. Apparently, the
10
11 increase is not high enough (the rare tRNAs remain minor components in the tRNA pool) to
12
13 disturb normal physiological operation.
14
15
16
17
18

19 The modest overexpression of the rare tRNAs had, however, a significant, positive effect on
20
21 heterologous protein expression, mitigating the effect of tRNA depletion. Individually, with the
22
23 exception of *glyT* in AIM, all the genes caused increased expression (up to 12-fold) of specific
24
25 GFP variants carrying tandem copies of the corresponding rare codon. The combined effect of the
26
27 six extra genes was then clearly demonstrated by expressing a panel of proteins of various origin
28
29 and codon bias. Expression in SixPack (amount of target protein per cell mass or per culture
30
31 volume) was at least as good as in any of the control strains and in fact, was better in most cases
32
33 (in some cases even 20-fold). In general, with the increasing number of rare codons and their
34
35 tandem copies, the advantage of SixPack became increasingly evident. SixPack compared
36
37 especially favorably to control strains when applying AIM, which is a popular medium for
38
39 biotechnological production of proteins in large volumes (i.e., in bioreactors)²⁹.
40
41
42
43
44
45
46

47 The aim of this study was to improve the performance of BL21(DE3) for the heterologous
48
49 expression of proteins. We do not claim that other strategies, like codon optimization, cannot
50
51 work better in certain cases, but the simplicity of the use of SixPack is definitely an attractive
52
53 feature.
54
55
56
57
58
59
60

1
2
3 In summary, insertion of the six rare tRNA genes into the genome of BL21(DE3) did not result in
4 changes in its basic characteristics like morphology, growth properties, or transformation
5 efficiency. The beneficial features (no need for extra plasmid and antibiotics, no need for codon
6 optimization, simplicity when expressing gene libraries, high success rate when expressing
7 proteins with different codon bias) render SixPack a useful, and likely superior, alternative host
8 for expression of heterologous proteins.
9
10
11
12
13
14
15
16
17
18

19 **Materials and Methods**

20 **Medium**

21
22
23
24
25
26 In all experiments involving bacterial culture growth, standard LB or AIM (auto-induction
27 medium; LB broth base including trace elements; Formedium LTD, England) were used.

28
29 Antibiotics were used in the following concentrations: 100 µg/ml ampicillin (Ap) and 24 µg/ml
30 chloramphenicol (Cam).
31
32
33
34
35
36
37

38 ***E. coli* strains**

39
40 BL21(DE3) was used as the parental strain. It expresses T7 RNA polymerase from the λDE3
41 lysogen inserted into the genome under the regulation of the *lacUV5* promoter. This polymerase
42 can drive the expression of the genes of interest via the T7 promoter³⁰.
43
44
45
46

47 Rosetta2(DE3)pLysS is a derivative of BL21(DE3). In this strain, pLysSRARE2 plasmid encodes
48 seven genes of rare tRNAs (*argU*, *argX*, *argW*, *glyT*, *leuW*, *ileX*, and *proL*, recognizing the
49 AGA/AGG, CGG, AGG, GGA, CUA, AUA, and CCC codons, respectively) under the control of
50
51
52
53
54
55
56
57
58
59
60

1
2
3 their own promoter. pLysSRARE2 also contains chloramphenicol resistance- and lysosyme-
4
5 encoding genes ³¹.
6

7 To create the SixPack strain, BL21(DE3) was modified by inserting extra copies of six rare tRNA
8
9 genes (*argU*, *argX*, *glyT*, *leuW*, *ileX*, and *proL*) into its *rrnD* operon (Figure 1). These tRNAs
10
11 recognize the same rare codons as the tRNAs expressed from the pLysSRARE2 plasmid (AGG is
12
13 recognized by *argX* and *argW* in Rosetta2(DE3)pLysS, but only by *argX* in SixPack).
14
15

16 Rosetta2(DE3) strain (Merck) supplies the same tRNA genes on pRARE2 plasmid, without
17
18 lysozyme expression.
19
20
21
22

23 **Genomic construct**

24
25 The rare tRNA genes were cloned into pSG76A (replicating by R6K ori) in three pieces. The first
26
27 part (synthesized by Thermo Fisher Scientific GENEART GmbH; Germany) contained the *argX*,
28
29 *glyT*, *leuW*, and *proL* tRNA genes. The structure of this DNA fragment was based on the
30
31 polycistronic operon of *E. coli* coding for *argX*, *hisR*, *leuT*, and *proM* tRNA genes; we replaced
32
33 the *hisR*, *leuT*, and *proM* genes with *glyT*, *leuW*, and *proL* rare tRNA genes, but kept *argX* and
34
35 the original intergenic regions. The second and third pieces coding for *argU* and *ileX*,
36
37 respectively, were directly amplified from the *E. coli* genome with their own intergenic regions
38
39 (Figures S1 and S2). The plasmid carrying the six tRNA genes was linearized by PCR using
40
41 overhanging primers containing homologous regions (50- or 100-nt) with the target site (unique
42
43 sequence of the 3' end of *rrnD*), and transformed into BL21(DE3) in the presence of λ RED
44
45 recombinase ³². After successful recombination into the genome, the CRISPR-Cas9 system was
46
47 used to introduce a double-strand break in the ampicillin resistance gene on the insert ³³. The
48
49 cells' own RecA system repaired the cleavage by homologous recombination using the 47-nt box
50
51
52
53
54
55
56
57
58
59
60

1
2
3 on the plasmid homologous to the adjacent *rrnD* sequence downstream from the site of insertion
4
5 (Figures S3 and S13)³⁴.
6
7
8
9

10 **Expression plasmids**

11
12 The coding sequences of different test proteins were cloned into pETDuet-1 (Novagen). A list of
13 these genes and the cloning sites is given in Table S3. The GFP-encoding gene was amplified
14 from pCA24N (the cloning plasmid of the ASKA collection)³⁵, and subcloned into pETDuet-1.
15
16 GFP gene variants containing runs of rare codons were created by PCR using pETDUET/GFP as
17 a template. The coding DNA sequence of asterless (*Asl*) was subcloned from the GH02902
18
19 *Drosophila* Gold cDNA clone into pETDuet-1 following standard procedures. Cloning of Flf1
20 into pETDuet-1 has been described elsewhere³⁶. dCas9 was cloned from the pdCas9 plasmid #
21 46569 from Addgene³⁷. The HumdCas9 gene was amplified from pMLM3705 plasmid # 47754
22 from Addgene³⁸. pETDUET/HumdCas9-GFP expresses HumdCas9 and GFP as a fusion protein.
23
24 EFT1, the gene of elongation factor 2, was amplified from genomic DNA of *Saccharomyces*
25 *cerevisiae* strain EMY 74.7 and was a kind gift from Tamas Feher³⁹.
26
27 The HPV16 L2 gene (subcloned in Addgene plasmid # 72473) was a kind gift from Vilmos
28 Tubak. The gene of *Pfu* DNA polymerase was amplified from *Pyrococcus furiosus* genomic
29 DNA (DSM 3638) which was a kind gift from Vilmos Tubak. BL21(DE3) harbouring an empty
30 pETDuet-1 was used as a control.
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48

49 **Quantitative real-time PCR (qRT-PCR)**

50
51 RNA was isolated from exponential phase cultures grown in LB, using the E.Z.N.A. Bacterial
52 RNA Kit (VWR, USA). To obtain cDNA, 1 µg of RNA was then reverse transcribed using the
53 High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific) in a final volume of
54
55
56
57
58
59
60

1
2
3 10 μ l by using 1 pmole of the reverse primers (Table S4). After dilution with 20 μ l of water, 1 μ l
4
5 of the diluted reaction mixture was used as a template in qRT-PCR with 10 μ l of qPCRBIO
6
7 SyGreen Mix (PCR Biosystems) and 1 μ l of gene-specific primer mix, according to the following
8
9 protocol: 10 min at 95°C followed by 40 cycles of 95°C for 25 sec, 60°C for 25 sec, and 72°C for
10
11 15 sec. qRT-PCR was performed in a LightCycler® Nano Real-Time PCR System (Roche
12
13 Diagnostics GmbH; Mannheim, Germany). Transcripts of *proK*, *alaU*, and *gltW* genes were used
14
15 as house-keeping tRNA probes.
16
17
18
19
20

21 **Measuring and calculating growth parameters**

22
23 To measure growth parameters, a Synergy 2 automated microplate reader machine (BioTek,
24
25 USA) was used. Aliquots of 1 μ l each of BL21(DE3), SixPack, and Rosetta2(DE3)pLysS
26
27 overnight starter cultures were transferred into 100 μ l fresh LB or AIM medium in dedicated 96-
28
29 well plates. Rosetta2(DE3)pLysS cultures were supplemented with Cam. Absorbance at 600 nm
30
31 was measured every 5 min for 24 h at 37°C with continuous shaking. Growth parameters (length
32
33 of the lag phase, doubling time and OD increment) were calculated by using previously described
34
35 methods ⁴⁰.
36
37
38
39
40
41

42 **Fluorescence measurements**

43
44 Starter cultures of BL21(DE3), SixPack, Rosetta2(DE3)pLysS and Rosetta2(DE3) strains
45
46 expressing GFP or GFP variants were grown overnight in LB in a Synergy 2 automated
47
48 microplate reader machine (BioTek, USA). Starter cultures (1 μ l each) were re-inoculated into
49
50 100 μ l fresh LB or AIM media supplemented with Ap and isopropyl β -D-1-thiogalactopyranoside
51
52 (IPTG) (plus Cam, in the case of Rosetta2(DE3)pLysS and Rosetta2(DE3)). IPTG was used in
53
54
55
56
57
58
59
60

1
2
3 concentrations optimized for highest induction (0.05 mM for BL21(DE3) and SixPack, 0.1 mM
4
5 for Rosetta2(DE3) and 0.5 mM for Rosetta2(DE3)pLysS). Measurements were made every 5
6
7 min.
8
9

12 **Protein expression and whole cell lysate preparation**

14 Protein expressions in BL21(DE3) and SixPack were induced with 0.05 mM IPTG, and those in
15
16 Rosetta2(DE3)pLysS were induced with 0.5 mM IPTG in 3ml of LB or AIM (supplemented with
17
18 antibiotics) for 10 h. Bacteria were harvested by centrifugation, resuspended in 150 μ l/absorption
19
20 unit (at 600 nm) of phosphate-buffered saline, pH 7.4, supplemented with 1 U/ml
21
22 BenzonaseNuclease (Merck) and 2 mM $MgCl_2$, kept on ice for 10 min, mixed with 4x Laemmli
23
24 sample buffer and boiled for 5 min.
25
26
27
28
29

30 **SDS-PAGE and Western blotting**

32 Equal amounts of protein samples were run on 7% or 10% SDS-PAGE gels. Gels were fixed for
33
34 10 min in 10% acetic acid, stained for 15 min with Coomassie Brilliant Blue (0.1% CBB in 50%
35
36 methanol and 10% acetic acid), and differentiated overnight in 7% acetic acid and 10% methanol.
37
38 For immunoblotting, proteins were blotted onto a nitrocellulose membrane (GE Healthcare) and
39
40 probed with anti-His mouse monoclonal (Thermo Fisher Scientific, #MA1-21315; dilution 1:3-
41
42 5000), anti-Flfl rat polyclonal ³⁶ (dilution 1: 15,000), anti-Asl rabbit polyclonal ⁴¹ (dilution 1:
43
44 20,000), anti-GroEL rabbit polyclonal ⁴² (dilution 1:5,000) or anti-DnaK1 mouse monoclonal
45
46 (dilution 1:5,000, Stressgen SPA-880) antibodies following standard procedures. Stained gels and
47
48 X-ray films were scanned at 600 dpi resolution for image processing.
49
50
51
52
53
54
55

56 **Funding**

1
2
3
4
5 This work was supported by The National Research, Development and Innovation Office (OTKA
6 K116455 to GP and OTKA-PD115404 to ZL), the Ministry for National Economy of Hungary
7 (GINOP-2.3.2-15-2016-00001), the Hungarian Academy of Sciences (Bolyai Fellowship
8 (bo_329_15), and Lendület Program Grant (LP2017-7/2017) to ZL).
9
10
11
12
13
14
15
16

17 **Acknowledgements**

18
19
20
21 The authors would like to thank Vilmos Tubak (BRC) for the *Pirococcus furiosus* genomic DNA
22 and the HPV16 L2-expressing plasmid, Tamas Feher (BRC) for the EFT1-coding plasmid, and
23 Laszlo Vigh (BRC) for DnaK1 and GroEL antibodies, David Glover (University of Cambridge)
24 for Asl antibody, and the *Drosophila* Genomics Resource Center (NIH grant 2P40OD010949) for
25 asterless cDNA.
26
27
28
29
30
31
32
33
34

35 **References**

- 36
37
38
39 1. Deana, A., Ehrlich, R., and Reiss, C. (1998) Silent mutations in the *Escherichia coli ompA*
40 leader peptide region strongly affect transcription and translation *in vivo*. *Nucleic Acids Res.* 26,
41 4778-4782.
42
43
44
45 2. Novy, R., Drott, D., Yaeger, K., and Mierendorf, R. (2001) Overcoming the codon bias of *E.*
46 *coli* for enhanced protein expression. In *Innovations, Newsletter of Novagen, Inc.* 12, 1–4.
47
48
49 3. Kim, S., and Lee, S.-B. (2006) Rare codon clusters at 5'-end influence heterologous expression
50 of archaeal gene in *Escherichia coli*. *Protein Expr. Purif.* 50, 49–57.
51
52
53
54
55
56
57
58
59
60

- 1
2
3 4. Harris, R.-P., and Kilby, P.-M. (2014) Amino acid misincorporation in recombinant
4
5 biopharmaceutical products. *Curr. Opin. Biotechnol.* 30, 45–50.
6
- 7
8 5. Boël, G., Letso, R., Neely, H., Price, W.-N., Wong, K.-H., Su, M., Luff, J., Valecha, M.,
9
10 Everett, J.-K., Acton, T.-B., Xiao R., Montelione, G.-T., Aalberts, D.-P., and Hunt, J.-F. (2016)
11
12 Codon influence on protein expression in *E. coli* correlates with mRNA levels. *Nature* 529,
13
14 358-363.
15
- 16
17 6. Ikemura, T. (1981) Correlation between the abundance of *Escherichia coli* transfer RNAs and
18
19 the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon
20
21 choice that is optimal for the *E. coli* translational system. *J. Mol. Biol.* 151, 389-409.
22
- 23
24 7. Novoa, E.-M., and Ribas de Pouplana, L. (2012) Speeding with control: codon usage, tRNAs,
25
26 and ribosomes. *Trends Genet.* 28, 574-581.
27
- 28
29 8. McNulty, D.-E., Claffee, B.-A., Huddleston, M.-J., and Kane, J.-F. (2003) Mistranslational
30
31 errors associated with the rare arginine codon CGG in *Escherichia coli*. *Protein Expr. Purif.* 27,
32
33 365–374.
34
- 35
36 9. Thangadurai, C., Suthakaran, P., Barfal, P., Anandaraj, B., Pradhan, S.-N., Boneya, H.-K.,
37
38 Ramalingam, S., and Murugan, V. (2008) Rare codon priority and its position specificity at the
39
40 5' of the gene modulates heterologous protein expression in *Escherichia coli*. *Biochem.*
41
42 *Biophys. Res. Commun.* 376, 647–652.
43
- 44
45 10. Kudla, G., Murray, A.-W., Tollervey, D., and Plotkin, J.-B. (2009) Coding-sequence
46
47 determinants of gene expression in *Escherichia coli*. *Science* 324, 255-258.
48
- 49
50 11. Rosano, G.-L., and Ceccarelli, E.-A. (2009) Rare codon content affects the solubility of
51
52 recombinant proteins in a codon bias-adjusted *Escherichia coli* strain. *Microb. Cell Fact.* 8, 41.
53
- 54
55 12. Mauro, V.-P., and Chappell, S.-A. (2014) A critical analysis of codon optimization in human
56
57 therapeutics. *Trends Mol. Med.* 20, 604-613.
58
59
60

13. Borel, F., Härtle, M., and Leberman, R. (1993) In vivo overexpression and purification of *Escherichia coli* tRNA(ser). *FEBS Lett.* 324, 162-166.
14. Pavon-Eternod, M., Gomes, S., Rosner, M.-R., and Pan, T. (2013) Overexpression of initiator methionine tRNA leads to global reprogramming of tRNA expression and increased proliferation in human epithelial cells. *RNA* 19, 461-466.
15. Shaham, G., and Tuller, T. (2018) Genome scale analysis of *Escherichia coli* with a comprehensive prokaryotic sequence-based biophysical model of translation initiation and elongation. *DNA Res.* 25, 195-205.
16. Goodman, D.-B., Church, G.-M., and Kosuri, S. (2013) Causes and effects of N-terminal codon bias in bacterial genes. *Science* 342, 475-479.
17. Karimi, Z., Nezafat, N., Negahdaripour, M., Berenjian, A., Hemmati, S., and Ghasemi, Y. (2015) The effect of rare codons following the ATG start codon on expression of human granulocyte-colony stimulating factor in *Escherichia coli*. *Protein Expr. Purif.* 114, 108-114.
18. Sogaard, K.-M., and Nørholm, M.-H. (2016) Side effects of extra tRNA supplied in a typical bacterial protein production scenario. *Protein Sci.* 25, 2102-2108.
19. Nomura, M., Yates, J.-L., Dean, D., and Post, L.-E. (1980) Feedback regulation of ribosomal protein gene expression in *Escherichia coli*: structural homology of ribosomal RNA and ribosomal protein mRNA. *Proc. Natl. Acad. Sci. U.S.A.* 77, 7084-7088.
20. Jin, D.-J., Cagliero, C., and Zhou, Y.-N. (2012) Growth rate regulation in *Escherichia coli*. *FEMS Microbiol. Rev.* 36, 269-287.
21. Milo, R., and Phillips, R. (2015) How many ribosomes are in the cell? In *Cell biology by the numbers*. Garland Science. ISBN13 9780815345374
22. Paul, B.-J., Ross, W., Gaal, T., and Gourse, R.-L. (2004) rRNA transcription in *Escherichia coli*. *Annu. Rev. Genet.* 38, 749-770.

- 1
2
3 23. Tuller, T., and Zur, H. (2015) Multiple roles of the coding sequence 5' end in gene expression
4 regulation. *Nucleic Acids Res.* 43, 13-28.
5
6
7 24. Mualif, S.-A., Teow, S.-Y., Omar, T.-C., Chew, Y.-W., Yusoff, N.-M., and Ali, S.-A. (2015)
8 Engineering and validation of a vector for concomitant expression of rare transfer RNA (tRNA)
9 and HIV-1 nef Genes in *Escherichia coli*. *PLoS One* 10:e0130446.
10
11
12 25. Burgess-Brown, N.-A., Sharma, S., Sobott, F., Loenarz, C., Oppermann, U., and Gileadi, O.
13 (2008) Codon optimization can improve expression of human genes in *Escherichia coli*: A
14 multi-gene study. *Protein Expr. Purif.* 59, 94–102.
15
16
17 26. Hutterer, K.-M., Zhang, Z., Michaels, M.-L., Belouski, E., Hong, R.-W., Shah, B., Berge, M.,
18 Barkhordarian, H., Le, E., Smith, S., Winters, D., Abrosion, F., Hecht, R., and Liu, J. (2012)
19 Targeted codon optimization improves translational fidelity for an Fc fusion protein.
20
21
22 *Biotechnol. Bioeng.* 109, 2770-2777.
23
24
25 27. Chen, H., Xu, Z., and Cen, P. (2006) High-level expression of human beta-defensin-2 gene
26 with rare codons in *E. coli* cell-free system. *Protein Pept. Lett.* 13, 155-162.
27
28
29 28. Chumpolkulwong, N., Sakamoto, K., Hayashi, A., Iraha, F., Shinya, N., Matsuda, N., Kiga,
30 D., Urushibata, A., Shirouzu, M., Oki, K., Kigawa, T., and Yokoyama, S. (2006) Translation of
31 'rare' codons in a cell-free protein synthesis system from *Escherichia coli*. *J. Struct. Funct.*
32
33
34 *Genomics* 7, 31-36.
35
36
37 29. Studier, F.-W. (2005) Protein production by auto-induction in high-density shaking cultures.
38
39
40 *Protein Expr. Purif.* 41, 207–234.
41
42
43 30. Studier, F.-W., Rosenberg, A.-H., Dunn, J.-J., and Dubendorff, J.-W. (1990) Use of T7 RNA
44 polymerase to direct expression of cloned genes. *Methods Enzymol.* 185, 60-89.
45
46
47 31. Studier, F.-W. (1991) Use of bacteriophage T7 lysozyme to improve an inducible T7
48
49
50 expression system. *J. Mol. Biol.* 219, 37-44.
51
52
53
54
55
56
57
58
59
60

- 1
2
3 32. Muyrers, J.-P., Zhang, Y., Testa, G., and Stewart, A.-F. (1999) Rapid modification of
4
5 bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res.* 27, 1555–1557.
6
7
8 33. Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L.-A. (2013) RNA-guided editing
9
10 of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* 31, 233-9.
11
12
13 34. Kowalczykowski, S.-C. (2000) Initiation of genetic recombination and recombination-
14
15 dependent replication. *Trends Biochem. Sci.* 25, 156-165.
16
17 35. Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., and
18
19 Mori, H. (2005) Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set
20
21 of *E. coli* K-12 ORF archive): unique resources for biological research. *DNA Res.* 12, 291–299.
22
23
24 36. Lipinszki, Z., Lefevre, S., Savoian, M.-S., Singleton, M.-R., Glover, D.-M., and Przewloka,
25
26 M.-R. (2015) Centromeric binding and activity of Protein Phosphatase 4. *Nat. Commun.* 6,
27
28 5894.
29
30
31 37. Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., and Marraffini, L.-A. (2013)
32
33 Programmable repression and activation of bacterial gene expression using an engineered
34
35 CRISPR-Cas system. *Nucleic Acids Res.* 41, 7429-7437.
36
37
38 38. Maeder, M.-L., Linder, S.-J., Cascio, V.-M., Fu, Y., Ho, Q.-H., and Joung, J.-K. (2013)
39
40 CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods* 10, 977-979.
41
42
43 39. Johnson, R.-E., Torres-Ramos, C.-A., Izumi, T., Mitra, S., Prakash, S., and Prakash, L. (1998)
44
45 Identification of *APN2*, the *Saccharomyces cerevisiae* homolog of the major human AP
46
47 endonuclease *HAPI*, and its role in the repair of abasic sites. *Genes Dev.* 12, 3137–3143.
48
49
50 40. Warringer, J., Ericson, E., Fernandez, L., Nerman, O., and Blomberg, A. (2003) High-
51
52 resolution yeast phenomics resolves different physiological features in the saline response.
53
54 *Proc. Natl. Acad. Sci. U.S.A.* 100, 15724–15729.
55
56
57
58
59
60

- 1
2
3 41. Dzhindzhev, N.-S., Yu, Q.-D., Weiskopf, K., Tzolovsky, G., Cunha-Ferreira, I., Riparbelli,
4
5 M., Rodrigues-Martins, A., Bettencourt-Dias, M., Callaini, G., and Glover, D.-M. (2010)
6
7 Asterless is a scaffold for the onset of centriole assembly. *Nature* 467, 714-718.
8
9
10 42. Lehel, C., Wada, H., Kovacs, E., Torok, Z., Gombos, Z., Horvath, I., Murata, N., and Vigh, L.
11
12 (1992) Heat shock protein synthesis of the cyanobacterium *Synechocystis* PCC 6803:
13
14 purification of the GroEL-related chaperonin. *Plant Mol. Biol.* 18, 327-36.
15
16
17
18

19 Figure legends

20
21
22
23

24 Figure 1. Schematic picture of *rrnD* showing the six extra tRNA genes (dark blue) inserted in the
25
26 3' end region between *thrV* and *rrfF*.
27
28
29

30 Figure 2. Growth curves of SixPack and control strains BL21(DE3) (marked BL) and
31
32 Rosetta2(DE3)pLysS (marked R) in LB (A) and in AIM (B). (Averages of three independent
33
34 experiments.) There is no significant difference between the growth parameters of BL21(DE3)
35
36 and SixPack (two sample t-test) (see also Table S2).
37
38
39

40
41
42 Figure 3. Ratios of the different tRNA species between SixPack and BL21(DE3) in the log phase.
43
44 (Averages of five independent experiments, each comprising three technical repetitions.)
45

46
47 Statistical analyses were performed on ddCt values (one-sample t-test, against zero); $p < 0,05$ was
48
49 measured in all cases.
50
51
52

53
54 Figure 4. Expression of GFP (A) and modified GFP versions (B-H) in BL21(DE3) (marked BL),
55
56 SixPack, and Rosetta2(DE3)pLysS (marked R) in LB, monitored by fluorescence measurements.
57
58
59
60

1
2
3 (The curves represent the averages of four independent experiments.) Statistical analysis was
4 performed on 20 h-time point fluorescence values, the differences between SixPack vs.
5 BL21(DE3) were significant in cases **B-F** ($p < 0,05$; paired t-test).
6
7
8
9

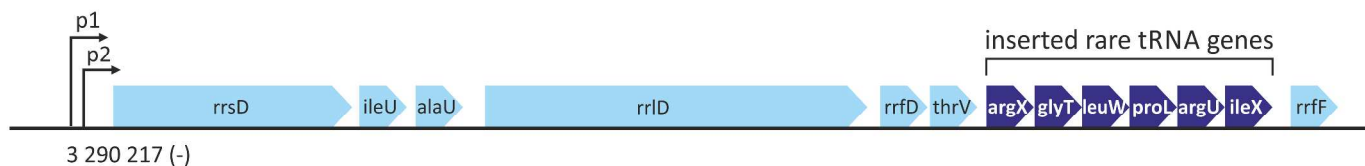
10
11
12 Figure 5. Production of eight heterologous proteins (**A-H** for AsI, dCas9, HumdCas9,
13 HumdCas9-GFP fusion, EFT1, Flfl, HPV16 L2, and *Pfu*, respectively) detected on protein gels
14 by Coomassie Brilliant Blue staining and Western blotting. Proteins were extracted from
15 BL21(DE3) (marked BL), SixPack and Rosetta2(DE3)pLysS (marked R) after 10 h growth in LB
16 or AIM. Asterisks indicate cases where SixPack produced significantly higher amount of test
17 proteins compared to BL21(DE3), based on densitometry data of Western blots of three
18 independent experiments ($p < 0,05$; paired t-test). “Empty” indicates BL21(DE3) harboring
19 pETDuet-1.
20
21
22
23
24
25
26
27
28
29
30
31
32

33 Figure 6. Summary of heterologous protein production of BL21(DE3) (marked BL), SixPack, and
34 Rosetta2(DE3)pLysS (marked R) measured in LB and AIM. The “+” and “-“ marks indicate the
35 protein production of the strains relative to SixPack: 100-68% (+++), 67-35% (++), 34-2% (+),
36 1% or lower (-). Table **A** is based on the fluorescence values of GFP and GFP variants detected
37 20 h after induction. Table **B** is based on results of Western blots, validated by densitometry.
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Circles label the cases when SixPack performed better than the parental strain BL21(DE3) in a statistically significant manner.

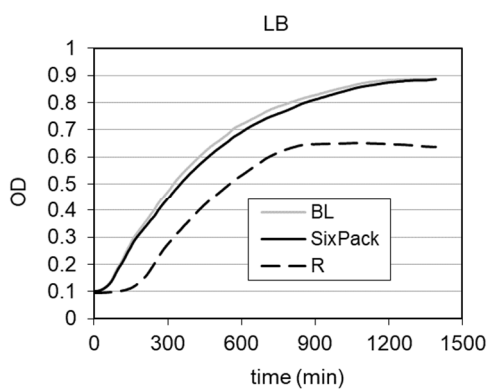
1
2
3 **Figures**
4
5
6
7

8 **Figure 1.**
9
10
11



27 **Figure 2.**
28
29
30
31

32 **A**



B

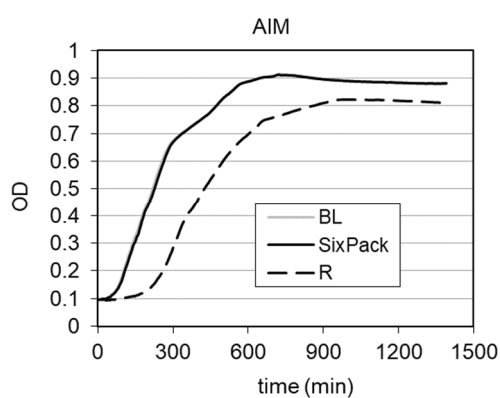


Figure 3.

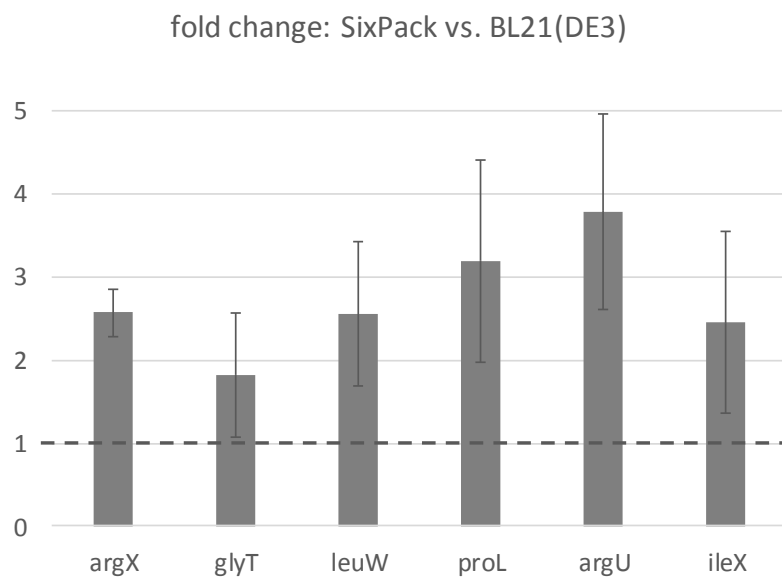
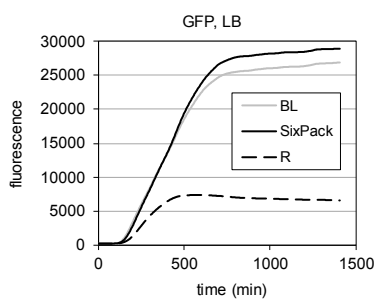
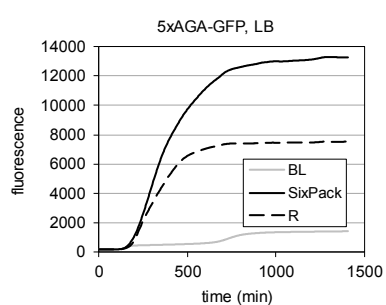


Figure 4.

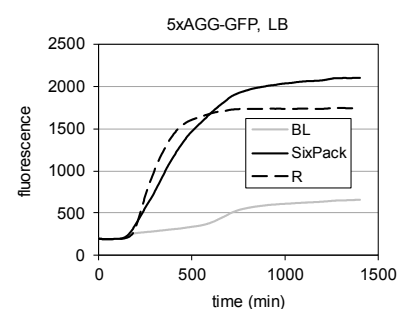
A



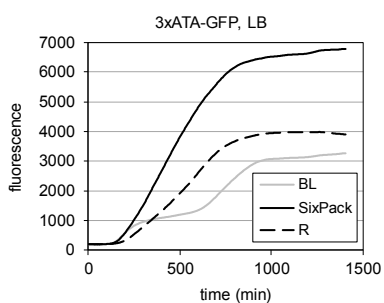
B



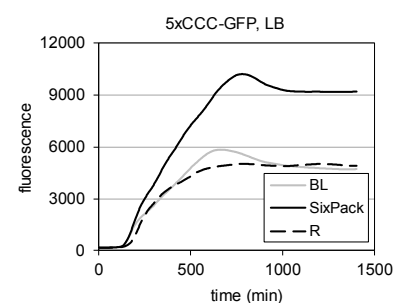
C



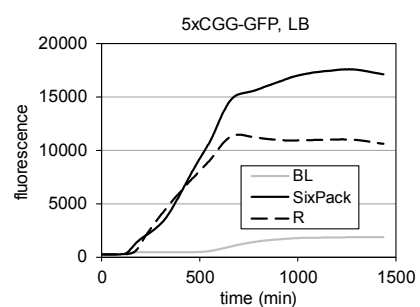
D



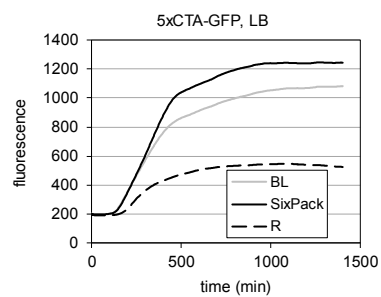
E



F



G



H

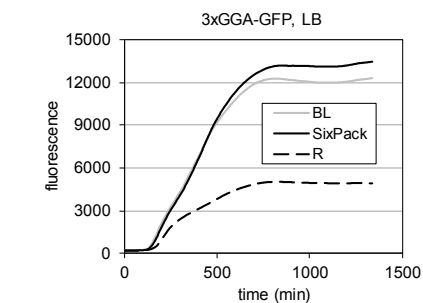
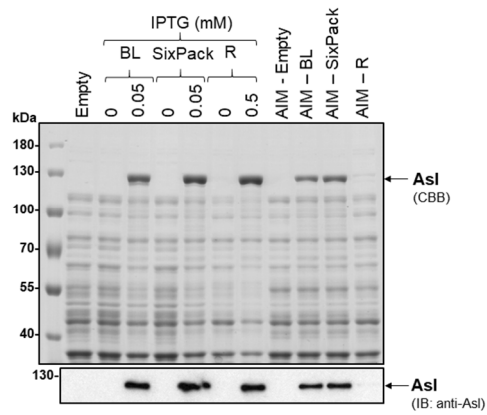
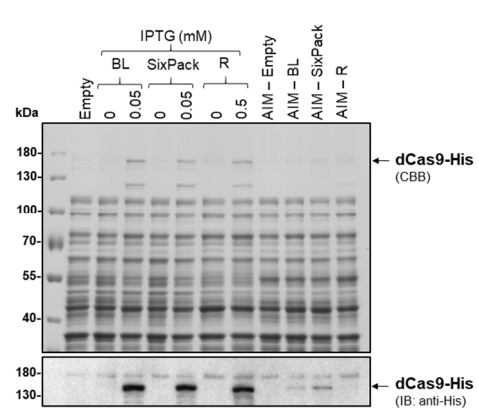


Figure 5.

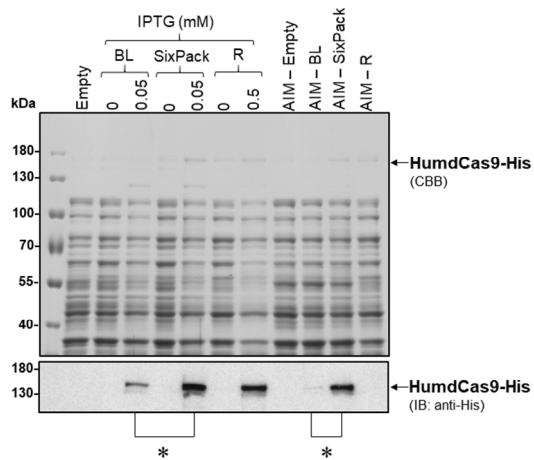
A



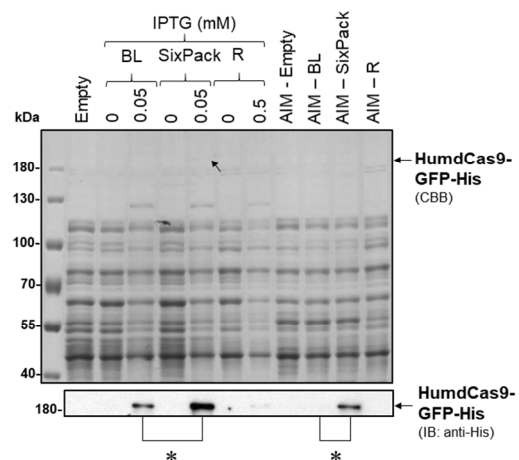
B



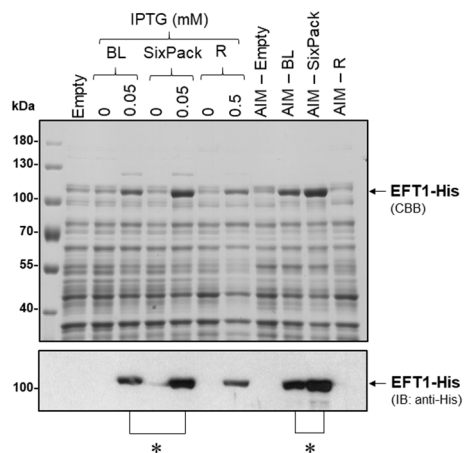
C



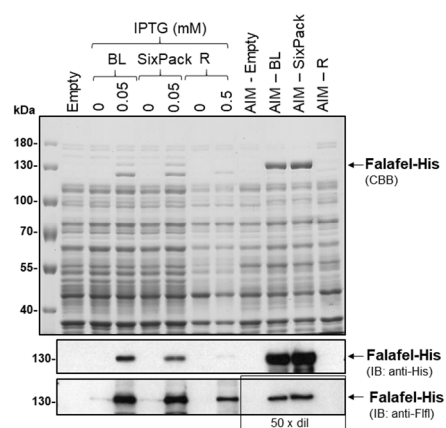
D



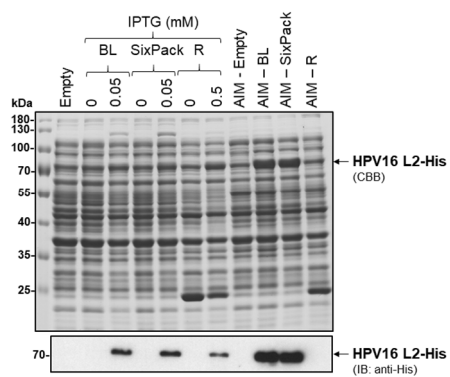
E



F



G



H

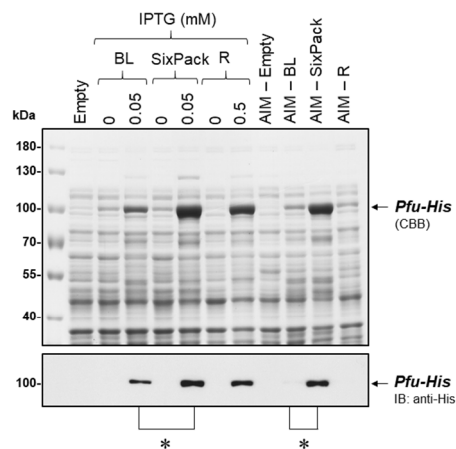


Figure 6.

A

	LB, IPTG			AIM		
	BL	SixPack	R	BL	SixPack	R
GFP	+++	+++	+	+++	+++	+
5xAGA-GFP	+	+++	++	+	+++	+
5xAGG-GFP	+	+++	+++	++	+++	+++
3xATA-GFP	++	+++	++	++	+++	++
5xCCC-GFP	++	+++	++	++	+++	+
5xCGG-GFP	+	+++	++	+	+++	++
5xCTA-GFP	+++	+++	++	+++	+++	++
3xGGA-GFP	+++	+++	++	+++	+++	+

B

	LB, IPTG			AIM		
	BL	SixPack	R	BL	SixPack	R
AsI	+++	+++	+++	++	++	-
dCas9	++	++	++	+	++	-
HumdCas9	+	+++	++	+	+++	-
HumdCas9-GFP	+	+++	-	-	++	-
EFT1	++	+++	+	++	+++	-
FifI	++	++	+	+++	+++	-
HPV16 L2	++	++	+	+++	+++	-
<i>Pfu</i>	+	+++	++	-	+++	-

SixPack

ACS Synthetic Biology Page 30 of 30

BL21(DE3)
chromosome

with
extra copies of
rare tRNA genes



rare tRNAs

ACS Paragon Plus Environment

heterologous
protein production

SixPack

BL21(DE3)

time

mRNA

rare codons

1

2

3

4

5

6

