1 **Enhancing the translational capacity of *E. coli* by resolving the codon bias**

2

3 **Running head: Enhanced translational capacity *E. coli* strain**

4

5 Zoltan Lipinszki[a,b], Viktor Vernyik[a], Nora Farago[c], Tobias Sari[a], Laszlo G. Puskas[c], Frederick R.

6 Blattner[d], Gyorgy Posfai[a, #] and Zsuzsanna Gyorfy[a,#]

7

8 [a]Institute of Biochemistry, Biological Research Centre of the Hungarian Academy of Sciences,

9 Szeged, Hungary

10 [b]MTA SZBK Lendület Laboratory of Cell Cycle Regulation, Biological Research Centre of the

11 Hungarian Academy of Sciences, Szeged, Hungary

12 [c]Institute of Genetics, Biological Research Centre of the Hungarian Academy of Sciences,

13 Szeged, Hungary

14 [d]Scarab Genomics LLC, Madison, WI, USA

15

16 #Address correspondence to Zsuzsanna Gyorfy, gyorfyzs@brc.hu; Gyorgy Posfai,

17 posfai.gyorgy@brc.mta.hu.

18

19 Word count of the abstract: 209

20 Word count of the text: 4034

21

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 heterologous 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.

Importance

Codon composition not fitting the codon bias of the expression host frequently compromises the efficient production of foreign proteins in *E. coli*. Various attempts to remedy the problem (codon optimization by gene synthesis, expression of rare tRNAs from a plasmid) proved to be

2

46   unsatisfying. Our new approach, adjusting the tRNA pool by co-expressing extra copies of rare

47   tRNA genes with ribosomal RNA genes, does not affect normal cell physiology, and seems to be

48   a superior solution in terms of simplicity, cost, and yield.

49

50   **Introduction**

51

52   *E. coli* is by far the most widely used host organism for biopharmaceutical heterologous

53   production of recombinant proteins. This expression platform is favored for its simplicity, speed,

54   and low cost. The codon bias discrepancy, however, can seriously hinder protein expression in *E.*

55   *coli* (1-5). Choice in the usage of synonymous codons can be different in various organisms, and

56   this bias has been shown to correlate with the relative and absolute quantities of individual tRNAs

57   (6,7). Heterologous expression of a protein with a high ratio of codons occurring infrequently in

58   *E. coli* might deplete the corresponding tRNA species, leading to translational frameshifting,

59   codon skipping, misincorporations, and protein truncations (8). Ultimately, the codon bias

60   seriously limits the use of *E. coli* as an expression platform.

61

62   To overcome this problem, codon optimization or rare tRNA overexpression strategies have been

63   applied, with limited success. On one hand, synthesis of the recombinant protein encoding gene

64   with an *E. coli* codon preference is labor-intensive and expensive. Moreover, replacing rare

65   codons with frequent ones does not necessarily lead to increased yield in protein synthesis. Codon

66   choice might affect expression, solubility, and folding of a protein (9-12); and rare codons can,

67   paradoxically, enhance the translation of a gene via reducing the mRNA secondary structure

68   (7,13,14). Additionally, codon optimization by gene synthesis is not feasible when testing gene

69   libraries. On the other hand, attempts to overexpress rare tRNA species by cloning extra copies of

3

70   the corresponding tRNA encoding genes into a plasmid have their drawbacks as well. In such

71   commercially available hosts, maintenance of the tRNA-expressing plasmid requires the addition

72   of extra antibiotics (usually the protein synthesis inhibitor chloramphenicol); moreover, the

73   recombinant protein-encoding expression plasmid must belong to a different complementation

74   group. Using two antibiotics and/or permanently altering the balance of the various tRNA species

75   can have a fitness cost on the host, eventually resulting in a low success rate in applications

76   (10,15).

77

78   We sought to resolve the codon bias problem by expressing rare tRNAs in a more dynamic

79   fashion. We hypothesized that inserting extra copies of the relevant tRNA genes into a ribosomal

80   RNA (*rrn*) operon would harmonize their expression with the translational activity and would

81   provide enhanced levels of rare tRNAs according to the actual needs. *rrn* operons have key roles

82   in bacterial physiology and economy. Synthesis of rRNA quickly reacts to environmental

83   conditions, determining ribosome availability via regulating the expression of ribosomal proteins

84   by a translational feedback mechanism (16). In *E. coli*, there are seven nearly identical copies of

85   rRNA operons. Synthesis of rRNA is driven by the strongest promoters found in the genome

86   (17), and the rate of transcription can change more than an order of magnitude between poor and

87   rich nutrient conditions (18). While most of the tRNA genes are scattered around the

88   chromosome, all rRNA operons carry certain (but not any of the rare) tRNA genes as well, co-

89   transcribed with the rRNA genes.

90

91   We inserted extra copies of the six tRNA genes (*argX*, *glyT*, *leuW*, *proL*, *argU,* and *ileX*)

92   corresponding to the minor codons of *E. coli* (CGG, GGA, CUA, CCC, AGA/AGG, and AUA,

93   respectively) into one of the ribosomal RNA operons (*rrnD*) of BL21(DE3), the widely used *E.*

4

94   *coli* expression host. We demonstrated that expression of these tRNA genes varies with growth

95   rate and shows a marked increase compared to the unmodified host. By testing the expression of a

96   panel of recombinant proteins with a high ratio of rare codons in their genes, we showed that, in

97   most cases, the modified strain named SixPack performs better and shows significantly enhanced

98   expression of the heterologous proteins in classical IPTG-induction as well as in an auto-

99   induction system. Moreover, compared to the commercially available strain Rosetta2(DE3)pLysS

100   (Merck), which carries a similar array of extra tRNA genes on a plasmid (pLysSRARE2; Merck),

101   SixPack proves to be a superior expression platform.

102

103   **Results**

104

105   **Design and genomic insertion of extra copies of tRNA genes**

106   Hypothetically, cloning extra copies of the genes corresponding to rare tRNAs into a *rrn* operon

107   allows their controlled expression, as the activity of the operon is tightly regulated by nutrient

108   availability and other physiological conditions (19). This might ensure that expression of the rare

109   tRNAs does not cause a permanent extra burden for the cell.

110

111   The genes encoding for the six least abundant tRNA species were combined in a single, synthetic

112   DNA fragment. The original copies of these genes are either single genes whose expression is

113   controlled by their own promoter (*proL*, *argU,* or *ileX*), or are parts of polycistronic operons

114   (*argX*, *glyT,* or *leuW*) (Fig. S1). Since the process of maturation of the primary transcripts of

115   tRNA genes is not thoroughly understood, the nucleotide sequences surrounding the genes in the

116   synthetic fragment were designed to mimic the natural tRNA operons. The intergenic and

117   flanking regions of the *argX-glyT-leuW-proL* segment were essentially identical to those of the

118   natural *argX-hisR-leuT-proM* polycistronic operon. For *argU* and *ileX*, exact copies of the genes

119   and their flanking sequences were fused to the 3' end of the segment (Fig. S2).

120

121   For the genomic insertion site *rrnD*, one of the seven *rrn* operons was selected. Although the

122   structure, sequence, and activity of the *rrn* operons of *E. coli* are very similar, the 3' end of *rrnD*

123   (containing *trhV* and *rrfF*) is unique, allowing specific targeting of the locus (17).

124

125   The 1207-base synthetic operon carrying the six tRNA genes was engineered into the 3' region of

126   *rrnD* (Fig. 1) by a multistep process using λRED recombineering, followed by elimination of the

127   markers via CRISPR-Cas9-stimulated homologous recombination (Fig. S3), resulting in strain

128   SixPack.

129

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

131   Apparently, insertion of the extra tRNA genes into *rrnD* did not disturb the homeostasis of the

132   cell. The growth parameters of SixPack were identical to those of the parental BL21(DE3) in both

133   LB and AIM (auto-induction medium). Noticeably, the commercially available

134   Rosetta2(DE3)pLysS strain had a longer lag phase, lower growth rate, and lower maximum OD

135   in both media, presumably owing to the burden of maintaining the plasmid and the presence of

136   the antibiotic chloramphenicol in the media (Fig. 2 and Table S1).

137

138   **Elevated expression of rare tRNAs**

139   Expression of the six rare tRNA genes was measured by qRT-PCR on first-strand cDNA

140   generated from total RNA isolated from cultures of BL21(DE3) and SixPack in the exponential

141   phase ($OD_{600} = 0.45$) and in the early stationary phase ($OD_{600} = 4.5$). Primers were designed to

142   amplify both the unprocessed and the matured forms of the various tRNA species. In the case of

143   SixPack, the measured rare tRNAs represented transcripts originating both from their original

144   genomic copies and from the newly inserted fragment in *rrnD*. The rare tRNA ratios were

145   normalized with the ratios of three control (abundant) tRNA species.

146

147   Compared to parental BL21(DE3), SixPack showed elevated expression of the six rare tRNAs.

148   The difference was more pronounced in the exponential phase (1.4- to 3.8-fold) than in the

149   stationary phase (0.8- to 3.0-fold) (Fig. 3).

150

151   Despite being on the same cistron and transcribed together, increases in the expression levels of

152   the extra tRNA genes were diverse (*proL* showing the highest and *glyT* the lowest increase). This

153   might be due to several factors, including tRNA processing, stability, regulation, and diverse

154   expression levels of the original genomic copies.

155

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

157   The functional effect of the increased levels of rare tRNAs was first tested separately for each

158   tRNA species. To magnify the effect of rare tRNAs, we constructed IPTG-inducible, pETDuet-1-

159   based (Novagen) test plasmids expressing modified versions of the GFP gene under the

160   regulation of the T7 promoter. GFP genes carrying runs of rare codons corresponded to each of

161   the six extra tRNA genes (one GFP version corresponding to *proL*, *ileX*, *argX*, *glyT,* and l*euW*

162   each, and two GFP versions for *argU*, as it relates to both AGA and AGG codons). The tandem

163   rare codons (3 or 5 copies) were inserted into the GFP gene immediately after the beginning ATG

164   codon. Expressions of the modified GFPs were detected by fluorescence measurements and

165   visualized by denaturing polyacrylamide gel electrophoresis (SDS-PAGE).

166

167 IPTG- (or lactose-) induced expression of the GFP and its variants with rare codons had a

168 negative but diverse impact on the growth of the strains. In LB medium, SixPack and BL21(DE3)

169 usually showed similar growth patterns, and in most cases displayed higher growth parameters

170 than Rosetta2(DE3)pLysS. In AIM, peculiarly, Rosetta2(DE3)pLysS reached the highest

171 maximal optical density when GFP variants with rare codons were expressed (Fig. S4 and S5).

172

173 Total fluorescence measurements of GFP production, however, showed a clear advantage of

174 SixPack in both LB (Fig. 4) and AIM (Fig. S6). In both media, SixPack showed higher

175 expression of the modified GFPs than the parental BL21(DE3) did. In one exception, there was

176 no difference (3xGGA-GFP; GGA codon relates to *glyT*) (Fig. S6H). Compared to

177 Rosetta2(DE3)pLysS, SixPack showed higher expression in all cases.

178

179 Results of the fluorescence measurements were further supported by protein gel assays. Equal

180 amounts of proteins obtained from crude extracts of cultures collected after a 10-h incubation

181 were analyzed on Coomassie Brilliant Blue-stained SDS-PAGE gels. Differences in the amounts

182 of modified GFPs in the three strains were in clear correlation with the fluorescence data (Fig.

183 S7).

184

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

186 The usefulness of SixPack was further tested by expressing a panel of eight protein-coding genes

187 of diverse length and rare codon content (Table S2), selected from various species, including

188 *Drosophila melanogaster, Pyrococcus furiosus*, *Saccharomyces cerevisiae*, *Streptococcus*

8

189 *pyogenes,* and human papillomavirus (HPV). The genes were cloned into pETDuet-1 and

190 expressed in SixPack and control strains under two different conditions (LB and AIM).

191

192 The growth curves of IPTG- (or lactose-) induced cultures seemed to be quite diverse in LB (Fig.

193 S8), but rather uniform in AIM (Fig. S9). Production of the specific proteins was visualized by

194 using Coomassie-stained gels and Western blotting (Fig. 5). In most cases, SixPack proved to be

195 the best producer. This was especially evident when the ratio of rare codons was higher than 8%

196 (and the ratio of tandems was also high), as in the cases of HumdCas9 (dCas9 optimized to

197 human codon preference), HumdCas9-GFP fusion, and *Pfu* DNA polymerase. In the case of

198 EFT1, the ratio was only 4.3% (35 AGA and 2 CTA codons); but SixPack still produced the

199 highest amount of this protein. It is conceivable that the translation efficiency is influenced not

200 only by the total amount of rare codons, but also by the occurrence of tandem situations, the

201 location of rare codons in the gene (e.g., close or far from the start site), and the ratios of the

202 different rare codon species (3). In other cases (Asl, dCas9, Flfl, and HPV16 L2), when the rare

203 codon ratio was below 8%, production by SixPack was similar to that of BL21(DE3) in LB and

204 similar to or better than that in AIM. In all cases, the performance of Rosetta2(DE3)pLysS lagged

205 behind that of SixPack (especially evidently in AIM) or was similar at most (Fig. 6).

206

207 **Discussion**

208

209 Depletion of rare tRNA species evoked by the expression of heterologous proteins can

210 compromise product yield and translational fidelity. Several attempts, both genetic

211 (overexpression of rare tRNA genes from a plasmid, or codon optimization of the heterologous

212 gene by synthesis) and non-genetic ones (optimizing the growth conditions, or using *in vitro*

9

213   protein synthesis systems) have been applied with varying degrees of success (20-23). We offer

214   here a solution that resolves codon bias discrepancies with a high success rate, does not require

215   additional antibiotics, and performs better or as well as the widely used BL21(DE3) *E. coli* strain.

216

217   The novelty of the approach lies in placing the extra copies of rare tRNA genes into a ribosomal

218   operon of the host chromosome. This arrangement offers the following benefits related to the rare

219   tRNA expression: (i) lack of physiological burden associated with plasmid replication and

220   maintenance; (ii) genetic stability without the use of an antibiotic; and (iii) dynamic changes in

221   the levels of rare tRNAs, paralleling the actual translational activity. Consequently, elevated

222   levels of rare tRNAs are achieved with minimal interference with normal host physiology.

223

224   A ribosomal operon is a natural choice as a location for inserting tRNA genes. While most tRNA

225   genes are scattered elsewhere in the genome, each of the *rrn* operons carries at least one tRNA

226   gene (typically coding for abundant tRNAs). Among the seven *rrn* operons, *rrnD* is particularly

227   suited for insertion of the extra genes, as it carries unique regions near its 3' end, allowing easy

228   targeting. In this vein, a synthetic segment carrying the genes of the six least abundant tRNAs of

229   *E. coli* (*argX*, *glyT*, *leuW*, *proL*, *argU*, *ileX*) was cloned into *rrnD*, downstream of its own tRNA

230   gene *thrV*.

231

232   As anticipated, increased amounts of the rare tRNAs, corresponding to the inserted genes, were

233   detected by qRT-PCR. Moreover, their levels showed a growth phase-dependent adjustment

234   (higher in exponential phase, lower in stationary phase). However, when compared to the levels

235   detected in the parental host BL21(DE3), the rate of increase in the exponential phase was only

236   modest (1.4- to 3.8-fold), especially when considering the wide dynamic range of *rrn* transcripts

10

237   across the growth cycle. An interplay of several factors, including co-factors of tRNA maturation,

238   tRNA stability, and feedback regulation of the genes at the original locations, might mitigate the

239   increase.

240

241   Although the increase in the abundance of the rare tRNAs did not interfere with cell growth, it

242   had a significant, positive effect on heterologous protein expression. Individually, with the

243   exception of *glyT* in AIM, all the genes caused increased expression (up to 12-fold) of specific

244   GFP variants carrying tandem copies of the corresponding rare codon. The combined effect of the

245   six extra genes was then clearly demonstrated by expressing a panel of proteins of various origin

246   and codon bias. Expression in SixPack (amount of target protein per cell mass or per culture

247   volume) was at least as good as in any of the control strains and in fact, was better in most cases

248   (in some cases even 20-fold). In general, with the increasing number of rare codons and their

249   tandem copies, the advantage of SixPack became increasingly evident. SixPack compared

250   especially favorably to control strains when applying AIM, which is a popular medium for

251   biotechnological production of proteins in large volumes (i.e., in bioreactors) (24).

252

253   Insertion of the six rare tRNA genes into the genome of BL21(DE3) did not result in changes in

254   its basic characteristics like morphology, growth properties, or transformation efficiency. The

255   beneficial features (no need for extra plasmid and antibiotics, no need for codon optimization,

256   simplicity when expressing gene libraries, high success rate when expressing proteins with

257   different codon bias) render SixPack a useful, and likely superior, alternative host for expression

258   of heterologous proteins.

259

260   **Materials and Methods**

11

261

262 **Medium**

263 In all experiments involving bacterial culture growth, standard LB or AIM (auto-induction

264 medium; LB broth base including trace elements; Formedium LTD, England) were used.

265 Antibiotics were used in the following concentrations: 100 μg/ml ampicillin (Ap) and 24 μg/ml

266 chloramphenicol (Cam).

267

268 *E. coli* **strains**

269 BL21(DE3) was used as the parental strain. It expresses T7 RNA polymerase from the λDE3

270 lysogen inserted into the genome under the regulation of the *lacUV5* promoter. This polymerase

271 can drive the expression of the genes of interest via the T7 promoter (25).

272 Rosetta2(DE3)pLysS is a derivative of BL21(DE3). In this strain, pLysSRARE2 plasmid encodes

273 seven genes of rare tRNAs (*argU*, *argX*, *argW*, *glyT*, *leuW*, *ileX*, and *proL*, recognizing the

274 AGA/AGG, CGG, AGG, GGA, CUA, AUA, and CCC codons, respectively) under the control of

275 their own promoter. pLysSRARE2 also contains chloramphenicol resistance- and lysosyme-

276 encoding genes. Due to constitutive lysosyme expression, the T7 polymerase-driven expression

277 of the genes of interest is lower in Rosetta2(DE3)pLysS than in BL21(DE3) or SixPack (26).

278 To create the SixPack strain, BL21(DE3) was modified by inserting extra copies of six rare tRNA

279 genes (*argU*, *argX*, *glyT*, *leuW*, *ileX*, and *proL*) into its *rrnD* operon (Fig. 1). These tRNAs

280 recognize the same rare codons as the tRNAs expressed from the pLysSRARE2 plasmid (AGG is

281 recognized by *argX* and *argW* in Rosetta2(DE3)pLysS, but only by *argX* in SixPack).

282

283 **Genomic construct**

12

284   The rare tRNA genes were cloned into pSG76A (replicating by R6K ori) in three pieces. The first

285   part (synthesized by Thermo Fisher Scientific GENEART GmbH; Germany) contained the *argX*,

286   *glyT*, *leuW,* and *proL* tRNA genes. The structure of this DNA fragment was based on the

287   polycistronic operon of *E. coli* coding for *argX*, *hisR*, *leuT,* and *proM* tRNA genes; we replaced

288   the *hisR*, *leuT,* and *proM* genes with *glyT*, *leuW,* and *proL* rare tRNA genes, but kept *argX* and

289   the original intergenic regions. The second and third pieces coding for *argU* and *ileX*,

290   respectively, were directly amplified from the *E. coli* genome with their own intergenic regions

291   (Fig. S1 and S2).

292   The plasmid carrying the six tRNA genes was linearized by PCR using overhanging primers

293   containing homologous regions (50- or 100-nt) with the target site (unique sequence of the 3' end

294   of *rrnD*), and transformed into BL21(DE3) in the presence of λRED recombinase (27). After

295   successful recombination into the genome, the CRISPR-Cas9 system was used to introduce a

296   double-strand brake in the ampicillin resistance gene on the insert (28). The cells' own RecA

297   system repaired the cleavage by homologous recombination using the 47-nt box on the plasmid

298   homologous to the adjacent *rrnD* sequence downstream from the site of insertion (29) (Fig. S3

299   and S10).

300

301   **Expression plasmids**

302   The coding sequences of different test proteins were cloned into pETDuet-1 (Novagen). A list of

303   these genes and the cloning sites is given in Table S3.

304   The GFP-encoding gene was amplified from pCA24N (the cloning plasmid of the ASKA

305   collection; 30), and subcloned into pETDuet-1. GFP gene variants containing runs of rare codons

306   were created by PCR using pETDUET/GFP as a template.

13

307   The coding DNA sequence of asterless (Asl) was subcloned from the GH02902 Drosophila Gold

308   cDNA clone into pETDuet-1 following standard procedures.

309   Cloning of Flf1 into pETDuet-1 has been described elsewhere (31).

310   dCas9 was cloned from the pdCas9 plasmid # 46569 from Addgene (32).

311   The HumdCas9 gene was amplified from pMLM3705 plasmid # 47754 from Addgene (33).

312   pETDUET/HumdCas9-GFP expresses HumdCas9 and GFP as a fusion protein.

313   EFT1, the gene of elongation factor 2, was amplified from genomic DNA of *Saccharomyces*

314   *cerevisiae* strain EMY 74.7 (34) and was a kind gift from Dr. Tamas Feher.

315   The HPV16 L2 gene (subcloned in Addgene plasmid # 72473) was a kind gift from Dr. Vilmos

316   Tubak.

317   The gene of *Pfu* DNA polymerase was amplified from *Pyrococcus furiosus* genomic DNA (DSM

318   3638) which was a kind gift from Dr. Vilmos Tubak.

319   BL21(DE3) harbouring an empty pETDuet-1 was used as a control.

320

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

322   RNA was isolated from exponential phase ($OD_{600} = 0.45$) and early stationary phase ($OD_{600} =$

323   4.5) cultures grown in LB, using the E.Z.N.A. Bacterial RNA Kit (VWR, USA). To obtain

324   cDNA, 1 μg of RNA was then reverse transcribed using the High-Capacity cDNA Reverse

325   Transcription Kit (Thermo Fisher Scientific) in a final volume of 10 μl by using 1 pmole of the

326   reverse primers (Table S4). After dilution with 20 μl of water, 1 μl of the diluted reaction mixture

327   was used as a template in qRT-PCR with 10 μl of qPCRBIO SyGreen Mix (PCR Biosystems) and

328   1 μl of gene-specific primer mix, according to the following protocol: 10 min at 95°C followed

329   by 40 cycles of 95°C for 25 sec, 60°C for 25 sec, and 72°C for 15 sec. qRT-PCR was performed

14

330   in a LightCycler® Nano Real-Time PCR System (Roche Diagnostics GmbH; Mannheim,

331   Germany). Transcripts of *proK*, *alaU*, and *gltW* genes were used as house-keeping tRNA probes.

332

333   **Measuring and calculating growth parameters**

334   To measure growth parameters, a Synergy 2 automated microplate reader machine (BioTek,

335   USA) was used. Aliquots of 1μl each of BL21(DE3), SixPack, and Rosetta2(DE3)pLysS

336   overnight starter cultures were transferred into 100 μl fresh LB or AIM medium in dedicated 96-

337   well plates. Rosetta2(DE3)pLysS cultures were supplemented with Cam. Absorbance at 600 nm

338   was measured every 5 min for 24 h at 37°C with continuous shaking. Growth parameters (length

339   of the lag phase and the doubling time) were calculated by using previously described methods

340   (35).

341

342   **Fluorescence measurements**

343   Starter cultures of BL21(DE3), SixPack, and Rosetta2(DE3)pLysS strains expressing GFP or

344   GFP variants were grown overnight in LB in a Synergy 2 automated microplate reader machine

345   (BioTek, USA). Starter cultures (1 μl each) were re-inoculated into 100 μl fresh LB or AIM

346   media supplemented with Ap and isopropyl β-D-1-thiogalactopyranoside (IPTG) (plus Cam, in

347   the case of Rosetta2(DE3)pLysS). IPTG was used in concentrations optimized for highest

348   induction (0.05 mM for BL21(DE3) and SixPack, 0.5 mM for Rosetta2(DE3)pLysS).

349   Measurements were made every 5 min.

350

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

15

352   Protein expressions in BL21(DE3) and SixPack were induced with 0.05 mM IPTG, and those in

353   Rosetta2(DE3)pLysS were induced with 0.5 mM IPTG in 3ml of LB or AIM (supplemented with

354   antibiotics) for 10 h. Bacteria were harvested by centrifugation, resuspended in 150 µl/absorption

355   unit (at 600 nm) of phosphate-buffered saline, pH 7.4, supplemented with 1 U/ml

356   BenzonaseNuclease (Merck) and 2 mM $MgCl_2$, kept on ice for 10 min, and then mixed with 4x

357   Laemmli sample buffer and boiled for 5 min.

358

359   **SDS-PAGE and Western blotting**

360   Equal amounts of protein samples were run on 7% or 10% SDS-PAGE gels. Gels were fixed for

361   10 min in 10% acetic acid, stained for 15 min with Coomassie Brilliant Blue (0.1% CBB in 50%

362   methanol and 10% acetic acid), and differentiated overnight in 7% acetic acid and 10% methanol.

363   For immunoblotting, proteins were blotted onto a nitrocellulose membrane (GE Healthcare) and

364   probed with anti-His mouse monoclonal (Thermo Fisher Scientific, #MA1-21315; dilution 1:3-

365   5000), anti-Flfl rat polyclonal (31) (dilution 1: 15,000), or anti-Asl rabbit polyclonal (36)

366   (dilution 1: 20,000) antibodies following standard procedures. Stained gels and X-ray films were

367   scanned at 600 dpi resolution for image processing.

368

369   **Funding**

370

375

16

376   **Acknowledgements**

377

378   The authors would like to thank Dr. Vilmos Tubak for the *Pirococcus furiosus* genomic DNA and

379   the HPV16 L2-expressing plasmid, and Dr. Tamas Feher for the EFT1-coding plasmid.

380

381   **References**

382

383   1. Deana A, Ehrlich R, Reiss C. 1998. Silent mutations in the *Escherichia coli ompA* leader

384   peptide region strongly affect transcription and translation *in vivo*. Nucleic Acids Res 26:4778-

385   4782.

386   2. Novy R, Drott D, Yaeger K, Mierendorf R. 2001. Overcoming the codon bias of *E. coli* for

387   enhanced protein expression. Innovations, Newsletter of Novagen, Inc. 12:1–4.

388   3. Kim S, Lee SB. 2006. Rare codon clusters at 5'-end influence heterologous expression of

389   archaeal gene in *Escherichia coli*. Protein Expr Purif 50:49–57.

390   4. Harris RP Kilby PM. 2014. Amino acid misincorporation in recombinant biopharmaceutical

391   products. Curr Opin Biotechnol 30:45–50.

392   5. Boël G, Letso R, Neely H, Price WN, Wong KH, Su M, Luff J, Valecha M, Everett JK, Acton

393   TB, Xiao R, Montelione GT, Aalberts DP, Hunt JF. 2016. Codon influence on protein

394   expression in *E. coli* correlates with mRNA levels. Nature 529:358-363.

395   6. Ikemura T. 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and

396   the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon

397   choice that is optimal for the *E. coli* translational system. J Mol Biol 151:389-409.

398   7. Novoa EM, Ribas de Pouplana L. 2012. Speeding with control: codon usage, tRNAs, and

399   ribosomes. Trends Genet 28:574-581.

400   8. McNulty DE, Claffee BA, Huddleston MJ, Kane JF. 2003. Mistranslational errors associated

401   with the rare arginine codon CGG in *Escherichia coli*. Protein Expr Purif 27:365–374.

402   9. Thangadurai C, Suthakaran P, Barfal P, Anandaraj B, Pradhan SN, Boneya HK, Ramalingam

403   S, Murugan V. 2008. Rare codon priority and its position specificity at the 5' of the gene

404   modulates heterologous protein expression in *Escherichia coli*. Biochem Biophys Res Commun

405   376:647–652.

406   10. Kudla G, Murray AW, Tollervey D, Plotkin JB. 2009. Coding-sequence determinants of gene

407   expression in *Escherichia coli*. Science 324:255-258.

408   11. Rosano GL, Ceccarelli EA. 2009. Rare codon content affects the solubility of recombinant

409   proteins in a codon bias-adjusted *Escherichia coli* strain. Microb Cell Fact 8:41.

410   12. Mauro VP, Chappell SA. 2014. A critical analysis of codon optimization in human

411   therapeutics. Trends Mol Med 20:604-613.

412   13. Goodman DB, Church GM, Kosuri S. 2013. Causes and effects of N-terminal codon bias in

413   bacterial genes. Science 342:475-479.

414   14. Karimi Z, Nezafat N, Negahdaripour M, Berenjian A, HemmatiS, Ghasemi Y. 2015. The

415   effect of rare codons following the ATG start codon on expression of human granulocyte-

416   colony stimulating factor in *Escherichia coli*. Protein Expr Purif 114:108-114.

417   15. Søgaard KM, Nørholm MH. 2016. Side effects of extra tRNA supplied in a typical bacterial

418   protein production scenario. Protein Sci 25:2102-2108.

419   16. Nomura M, Yates JL, Dean D, Post LE. 1980. Feedback regulation of ribosomal protein gene

420   expression in *Escherichia coli*: structural homology of ribosomal RNA and ribosomal protein

421   mRNA. Proc Natl Acad Sci U S A 77:7084–7088.

422   17. Jin DJ, Cagliero C, Zhou YN. 2012. Growth rate regulation in *Escherichia coli*. FEMS

423   Microbiol Rev 36:269–287.

424   18. Milo R, Phillips R. 2015. Cell biology by the numbers. Garland Science. ISBN13

425   9780815345374

426   19. Paul BJ, RossW, GaalT, Gourse RL. 2004. rRNA transcription in *Escherichia coli*. Annu Rev

427   Genet 38:749–770.

428   20. Burgess-Brown NA, Sharma S, Sobott F, Loenarz C, Oppermann U, Gileadi O. 2008. Codon

429   optimization can improve expression of human genes in *Escherichia coli*: A multi-gene study.

430   Protein Expr Purif 59:94–102.

431   21. Hutterer KM, Zhang Z, Michaels ML, Belouski E, Hong RW, Shah B, Berge M,

432   Barkhordarian H, Le E, Smith S, Winters D, Abroson F, Hecht R, Liu J. 2012. Targeted codon

433   optimization improves translational fidelity for an Fc fusion protein. Biotechnol Bioeng

434   109:2770-2777.

435   22. Chen H, Xu Z, Cen P. 2006. High-level expression of human beta-defensin-2 gene with rare

436   codons in *E. coli* cell-free system. Protein Pept Lett 13:155-162.

437   23. Chumpolkulwong N, Sakamoto K, Hayashi A, Iraha F, Shinya N, Matsuda N, Kiga D,

438   Urushibata A, Shirouzu M, Oki K, Kigawa T, Yokoyama S. 2006. Translation of 'rare' codons

439   in a cell-free protein synthesis system from *Escherichia coli*. J Struct Funct Genomics 7:31-36.

440   24. Studier FW. 2005. Protein production by auto-induction in high-density shaking cultures.

441   Protein Expr Purif 41:207–234.

442   25. Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW. 1990. Use of T7 RNA polymerase to

443   direct expression of cloned genes. Methods Enzymol 185:60-89.

444   26. Studier FW. 1991. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression

445   system. J Mol Biol 219:37-44.

446   27. Muyrers JP, Zhang Y, Testa G, Stewart AF. 1999. Rapid modification of bacterial artificial

447   chromosomes by ET-recombination. Nucleic Acids Res 27:1555–1557.

448   28. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. 2013. RNA-guided editing of bacterial

449   genomes using CRISPR-Cas systems. Nat Biotechnol 31:233-9.

450   29. Kowalczykowski SC. 2000. Initiation of genetic recombination and recombination-dependent

451   replication. Trends Biochem Sci 25:156-165.

452   30. Kitagawa M, Ara T, Arifuzzaman M, Ioka-NakamichiT, Inamoto E, Toyonaga H, Mori H.

453   2005. Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of *E. coli*

454   K-12 ORF archive): unique resources for biological research. DNA Res 12:291–299.

455   31. Lipinszki Z, Lefevre S, Savoian MS, Singleton MR, Glover DM, Przewloka MR. 2015

456   Centromeric binding and activity of Protein Phosphatase 4. Nat Commun 6:5894.

457   32. Bikard D, Jiang W, Samai P, Hochschild A, Zhang F, Marraffini LA. 2013. Programmable

458   repression and activation of bacterial gene expression using an engineered CRISPR-Cas system.

459   Nucleic Acids Res 41:7429-7437.

460   33. Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, Joung JK. 2013. CRISPR RNA-guided

461   activation of endogenous human genes. Nat Methods 10:977-979.

462   34. Johnson RE, Torres-Ramos CA, Izumi T, Mitra S, Prakash S, Prakash L. 1998. Identification

463   of *APN2*, the *Saccharomyces cerevisiae* homolog of the major human AP endonuclease *HAP1*,

464   and its role in the repair of abasic sites. Genes Dev 12:3137–3143.

465   35. Warringer J, Ericson E, Fernandez L, Nerman O, Blomberg A. 2003. High-resolution yeast

466   phenomics resolves different physiological features in the saline response. Proc Natl Acad Sci

467   U S A 100:15724–15729.

468   36. Dzhindzhev NS, Yu QD, Weiskopf K, Tzolovsky G, Cunha-Ferreira I, Riparbelli M,

469   Rodrigues-Martins A, Bettencourt-Dias M, Callaini G, Glover DM. 2010. Asterless is a

470   scaffold for the onset of centriole assembly. Nature 467:714-718.

471

20

472   **Figure legends**

473

474   Figure 1. Schematic picture of *rrnD* showing the six extra tRNA genes (dark blue) inserted in the

475   3' end region between *thrV* and *rrfF*.

476

477   Figure 2. Growth curves of SixPack and control strains BL21(DE3) (marked BL) and

478   Rosetta2(DE3)pLysS (marked R) in LB (**A**) and in AIM (**B**). (Averages of three independent

479   experiments.)

480

481   Figure 3. Ratios of the different tRNA species between SixPack and BL21(DE3) in the

482   exponential growth phase (black columns) and in the early stationary phase (gray columns).

483   (Averages of two independent experiments, each comprising three technical repetitions.)

484

485   Figure 4. Expression of GFP (**A**) and modified GFP versions (**B-H**) in BL21(DE3) (marked BL),

486   SixPack, and Rosetta2(DE3)pLysS (marked R) in LB, monitored by fluorescence measurements.

487   (The curves represent the averages of four independent experiments.)

488

489   Figure 5. Production of eight heterologous proteins (**A-H** for Asl, dCas9, HumdCas9,

490   HumdCas9-GFP fusion, EFT1, Flfl, HPV16 L2, and *Pfu*, respectively) detected on protein gels

491   by Coomassie Brilliant Blue staining and Western blotting. Proteins were extracted from

492   BL21(DE3) (marked BL), SixPack and Rosetta2(DE3)pLysS (marked R) after 10 h growth in LB

493   or AIM.

494

21

495   Figure 6. Summary of heterologous protein production of BL21(DE3) (marked BL), SixPack, and

496   Rosetta2(DE3)pLysS (marked R) measured in LB and AIM. The "+" and "-" marks indicate the

497   relative protein production of the strains. Table **A** is based on the maximal fluorescence values of

498   GFP and GFP variants detected in the first 24 h after induction. Table **B** is based on results of

499   Western blots – validated by densitometry (data not shown). Circles label the cases when SixPack

500   performed better than the parental strain BL21(DE3).

501

502   **Figures**

503

504   Figure 1.

505

506

507

508   Figure 2.

509   A   B

510

511

512   LB   AIM

513

514

515
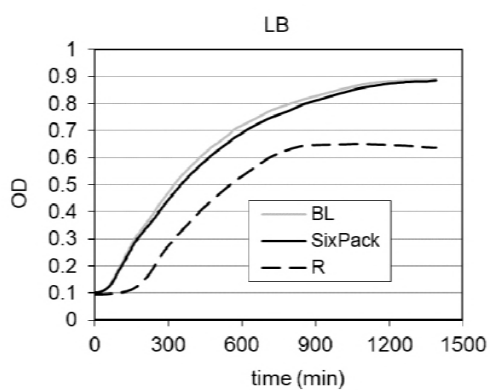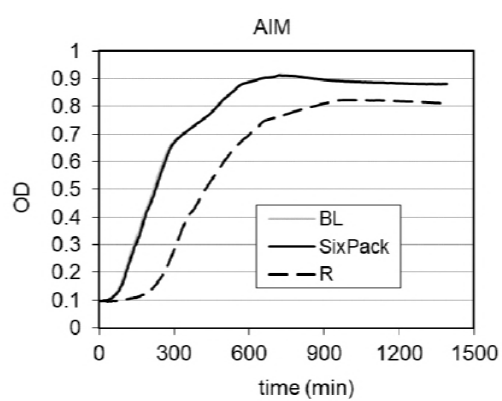
516   Figure 3.

517

518

519

520

521

522

523

524

525

526

527

528
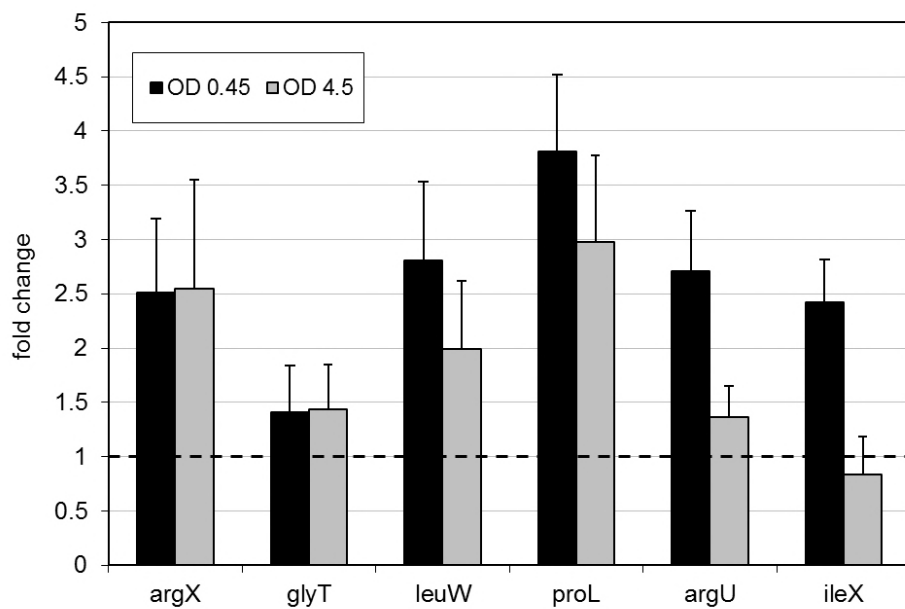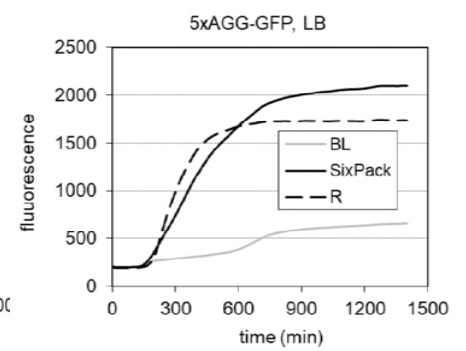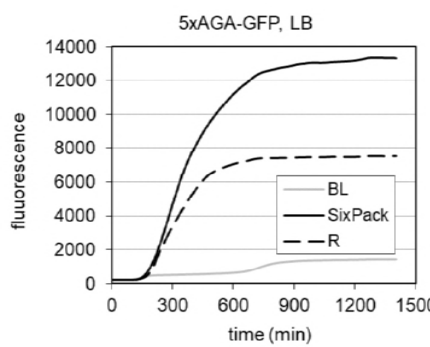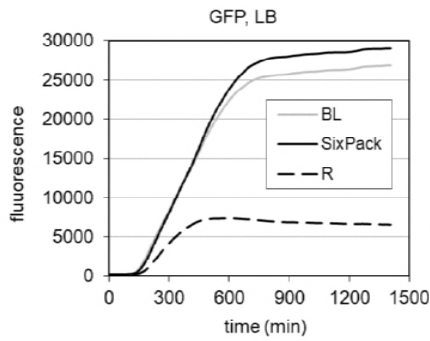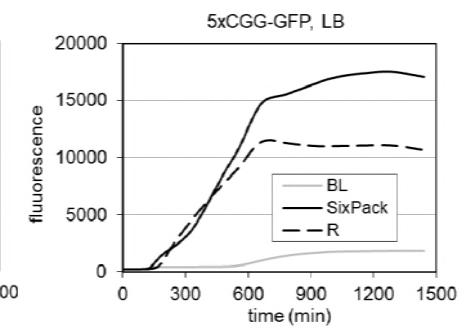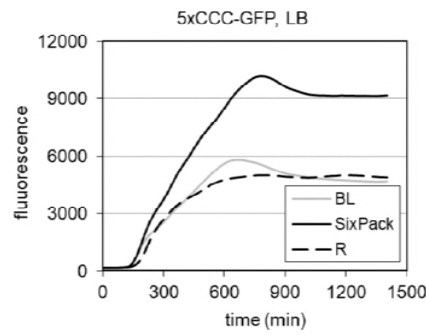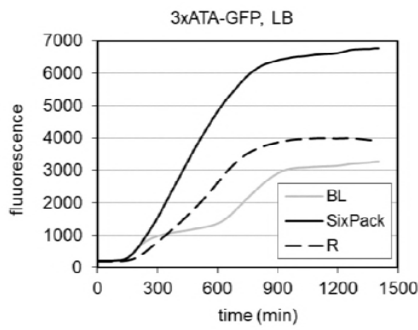
Figure 4.

A   B   C

D   E   F

G   H

24

553 Figure 5.

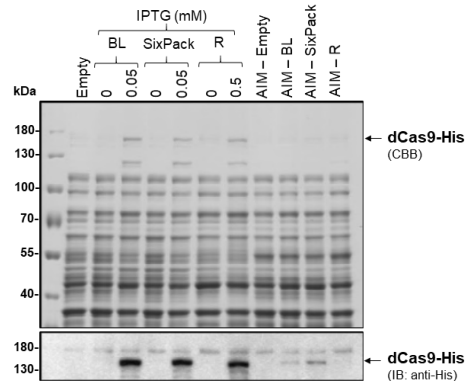554 A   B

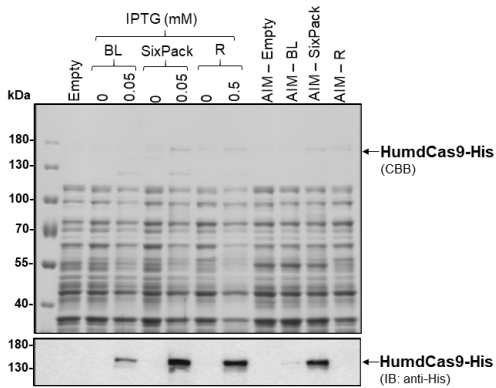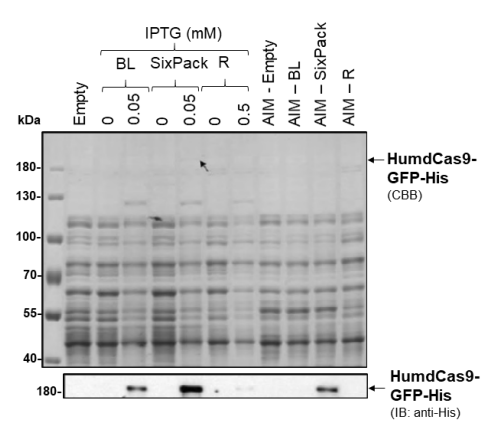561 C   D
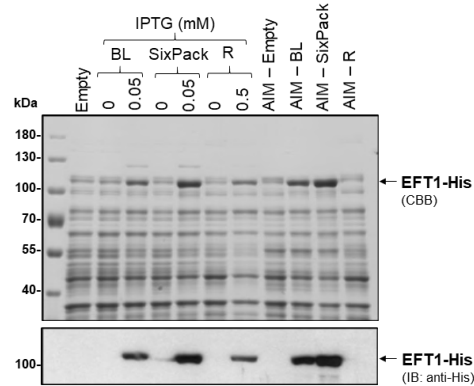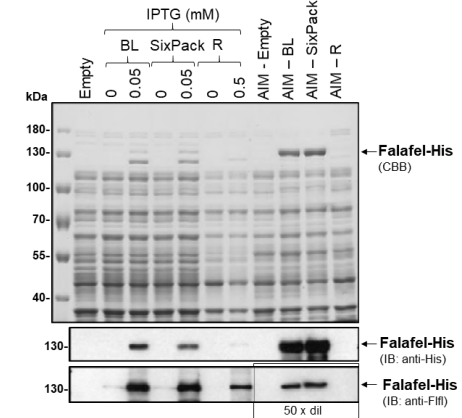

568 E   F

575

576

25

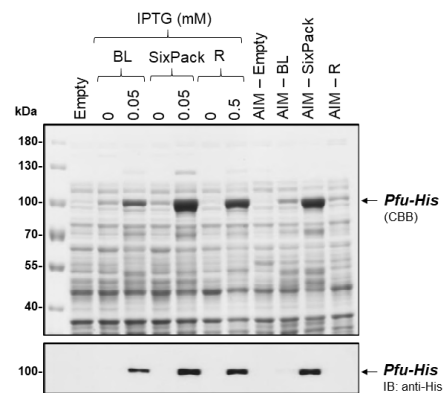577   G   H
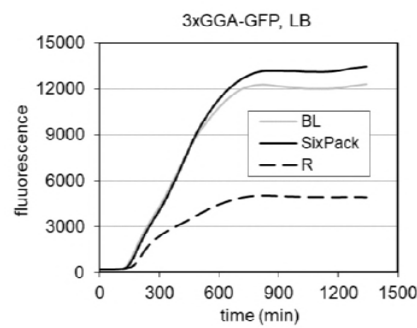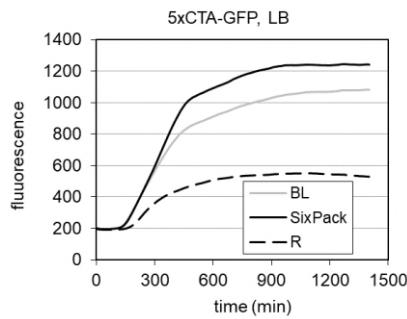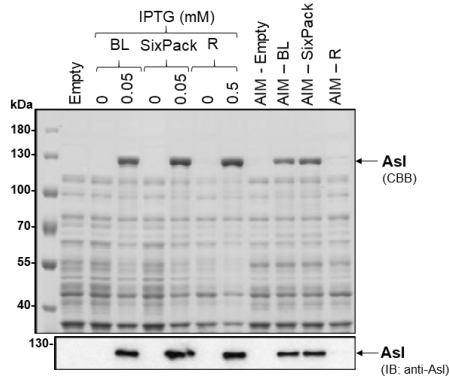
578

579

580

581

582

583

584

585   Figure 6.

586   A

587
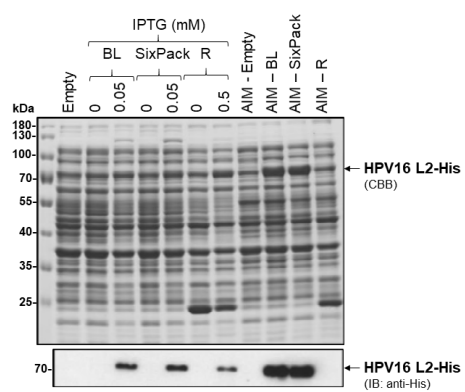
	LB, IPTG			AIM		
	BL	SixPack	R	BL	SixPack	R
GFP	++	+++	+	+++	+++	-
5xAGA-GFP	+	+++	++	+	+++	+
5xAGG-GFP	+	+++	++	+	+++	++
3xATA-GFP	+	+++	++	++	+++	++
5xCCC-GFP	++	+++	++	+	+++	++
5xCGG-GFP	+	+++	+++	+	+++	++
5xCTA-GFP	++	+++	+	++	+++	+
3xGGA-GFP	++	+++	+	+++	+++	+

593   B

594

	LB, IPTG			AIM		
	BL	SixPack	R	BL	SixPack	R
Asl	+++	+++	+++	++	++	-
dCas9	++	++	++	+	++	-
HumdCas9	+	+++	++	+	+++	-
HumdCas9-GFP	+	+++	-	-	++	-
EFT1	++	+++	+	++	+++	-
Flfl	++	++	+	+++	+++	-
HPV16 L2	++	++	+	+++	+++	-
Pfu	+	+++	++	-	+++	-

600

601

602

27

603 **Supplementary figure legends**

604

605 Figure S1. Schematic picture of the *E. coli* BL21(DE3) genome showing the location and

606 orientation of tRNA encoding genes (based on the BL21(DE3) complete genome sequence,

607 CP001509.3). Circles label the rare tRNA species genes whose copies were inserted into *rrnD*.

608

609 Figure S2. Nucleotide sequence of the insert containing the six rare tRNA species genes (bold).

610 Three incidental point mutations are underlined.

611

612 Figure S3. Main steps in the procedure for creating SixPack. The six tRNA genes and the 47-nt

613 segment (filled orange box) homologous to the 3' end of *rrnD* were cloned into pSG76A (**I**). The

614 plasmid was linearized by PCR using primers bearing overhanging regions (brown, orange lines)

615 homologous to *rrnD* (brown, orange open boxes) (**II**) and recombined into the genome via the

616 homologous overhangs (**III**). Plasmid sequences were then eliminated by CRISPR-Cas9

617 cleavage-stimulated homologous recombination between the 47-nt fragment (filled orange box)

618 and its downstream homologue (orange open box) (**IV**).

619

620 Figure S4. Growth curves of the strains expressing GFP (**A**) and different GFP variants (**B-H**) in

621 LB medium. (The curves represent the averages of four independent experiments.)

622

623 Figure S5. Growth curves of the strains expressing GFP (**A**) and different GFP variants (**B-H**) in

624 AIM. (The curves represent the averages of four independent experiments.)

625

28

626 Figure S6. Expression of GFP (**A**) and GFP variants (**B-H**) in BL21(DE3) (marked BL), SixPack,

627 and Rosetta2(DE3)pLysS (marked R) in AIM monitored by fluorescence measurements. (The

628 curves represent the averages of four independent experiments.)

629

630 Figure S7. Production of GFP (**A**) and GFP variants (**B-H**) after 10 h growth in LB or AIM

631 visualized on protein gels labeled by Coomassie Brilliant Blue staining.

632

633 Figure S8. Growth curves of the strains expressing various heterologous proteins in LB.

634 (The curves represent the averages of three independent experiments.)

635

636 Figure S9. Growth curves of the strains expressing various heterologous proteins in AIM. (The

637 curves represent the averages of three independent experiments.)

638

639 Figure S10. Overhanging primers (**A** and **B**) used for linearization of the plasmid and

640 recombination into the genome. The forward primer contained a 100-bp overhanging segment

641 homologous to *rrnD* downstream from the site of insertion (brown). The reverse primer contained

642 a 50-bp overhanging region homologous to *rrnD* upstream from the site of insertion (orange).

643 The CRISPR-Cas9 targeting sequence (**C**) was cloned into pCRISPR plasmid (Addgene #42875)

644 and used as crRNA targeting the Cas9 protein (expressed from Addgene #42876) to the

645 ampicillin resistance gene of pSG76A.

646

647 Table S1. Growth parameters of BL21(DE3) (marked BL), SixPack, and Rosetta2(DE3)pLysS

648 (marked R) strains grown in LB and AIM.

649

29

650   Table S2. Size and rare codon composition of the ORFs of test proteins. SixPack showed more

651   effective protein production than the parental strain BL21(DE3) when the ratio of rare codons

652   was about 8% or more (labeled with circles). The efficient translation of EFT1 by SixPack might

653   be due to the extremely high ratio of one rare codon species, AGA (labeled with rectangle).

654

655   Table S3. List of 16 protein-expressing plasmid (pETDuet-1) constructs and the cloning sites

656   used.

657

658   Table S4. Primers used for qRT-PCR. *argX*, *glyT*, *leuW*, *proL*, *argU,* and *ileX* correspond to rare;

659   *alaU*, *gltW* and *proK* correspond to abundant (control) tRNA species.