Utilization of rare codon-rich markers for screening amino acid overproducers

Bo Zheng1, Xiaoyan Ma1, Ning Wang1, Tingting Ding1, Liwei Guo1,2, Xiaorong Zhang3, Yu Yang1, Chun Li4 & Yi-Xin Huo1,2

The translation of rare codons relies on their corresponding rare tRNAs, which could not be fully charged under amino acid starvation. Theoretically, disrupted or retarded translation caused by the lack of charged rare tRNAs can be partially restored by feeding or intracellular synthesis of the corresponding amino acids. Inspired by this assumption, we develop a screening or selection system for obtaining overproducers of a target amino acid by replacing its common codons with the corresponding synonymous rare alternative in the coding sequence of selected reporter proteins or antibiotic-resistant markers. Results show that integration of rare codons can inhibit gene translations in a frequency-dependent manner. As a proof-of-concept, Escherichia coli strains overproducing l-leucine, l-arginine or l-serine are successfully selected from random mutation libraries. The system is also applied to Corynebacterium glutamicum to screen out l-arginine overproducers. This strategy sheds new light on obtaining and understanding amino acid overproduction strains.
The amino acids have a multi-billion-dollar market with applications in food, animal feed, pharmaceutical, and cosmetic industries. The worldwide market for amino acids represented overall $8.8 billion in 2007 and increased 3.47% per year to over $10 billion in 2015. Although large-scale microbial fermentation has satisfied most of the demands, production cost and yield remain suboptimal for most amino acids. Several amino acids, such as L-lysine and L-glycine, can only be enzymatically or chemically produced. High-performance fermentation strains are needed for overproducing the targeted amino acids.

The traditional strategy for screening amino acid overproducers took the advantage of toxic analogues, which has similar size, structure, and charge properties as the proteinogenic amino acids. An analogue would compete with its corresponding amino acid for the finite tRNAs in the process of protein biosynthesis. Once inserted into any polypeptide, the analogue could disrupt the synthesis or function of that polypeptide. The losing of functional proteins could result in growth retardation or even death. Cells overproducing an amino acid might produce enough functional proteins to survive the stresses from the analogue of that amino acid and could be selected. For instance, a high concentration of 4-azaleucine has been successfully applied to select L-leucine overproduction strains.

However, the use of analogues faces severe disadvantages that could compromise the selection results. First, an analogue could interfere with cellular activities beyond the protein synthesis. For example, it could jeopardize cell growth by disrupting the nucleus regions, affecting the structure of cellular membranes, inhibiting purine and pyrimidine biosynthesis or decreasing the level of ATP. Thus, mutants that have enhanced amino acid productivities may not survive these side effects. Second, cells may escape from the selection pressure of an analogue by developing detoxification mechanisms. Specifically, the analogues could be blocked by amino acid transporters with increased selectivity, or pumped outside of the cells by enhanced efflux (Fig. 1b). The analogues could also be degraded to nontoxic forms or be incorporated into the bacterial proteome after evolution. Therefore, it is necessary to test the selected mutations individually to verify their amino acid productivities. Third, proper analogues for specific amino acids are limited. Therefore, it is of great need to develop alternative approaches that provide accuracy, sensitivity, and high-throughput simultaneously.

The 20 proteinogenic amino acids are encoded by 61 codons, and with up to six different codons specifying the same amino acid according to standard code table. Many prokaryotes and eukaryotes display strong preference for certain codons over their synonymous alternatives. (Fig. 1c). A codon is categorized as common or rare codon for a specific species, depending on the frequency of its occurrence in the coding DNA of the whole genome. This codon usage bias affects translation phases such as translation initiation, elongation, and protein folding. Furthermore, the common and rare codons vary among different species, since a common codon for one species could be a rare one for another species. A codon could be translated into a corresponding peptide-forming amino acid by a charged tRNA isoacceptor which pairs the codon with the amino acid.

The heterologous protein expressions in host strains are always challenging. The codons of heterologous proteins are generally not the favored ones in the host, and some are paired with cognate tRNAs with low abundancy. Therefore, automated codon optimization algorithms have been developed to design coding sequences optimized for increasing expressions in certain hosts, which could lead up to 1000-fold increase in protein expressions. Opposite to the codon optimization strategy, the heterologous protein expressions in hosts could theoretically be reduced by replacing common codons with synonymous rare alternatives, especially under amino acid starvation conditions. The charging levels of rare isoacceptors approached zero immediately after the cells were challenged by the amino acid starvation, while the ones of common isoacceptors remained high for a few minutes. A rare isoacceptor is charged when its intracellular corresponding amino acid remains sufficient after charging the synonymous common isoacceptors. Therefore, an unexplored approach for screening amino acid overproducers, which is supposed to have high intracellular amino acid concentrations, could be developed based on the natural competitions between rare and common isoacceptors for intracellular free amino acids.

For a gene with numbers of rare codons, its translation under amino acid starvation conditions could be maintained if the amounts of intracellular amino acids are sufficient to support the easy charging of the rare tRNAs. Such strains are likely to be the amino acid high production strains. By linking the protein expressions with cell growth or color formation, we could easily distinguish the high production strains from a pool of candidates. This approach overcomes the drawbacks of analogues by only targeting the reporter genes without affecting other cellular systems and biosynthetic processes.

In this study, we establish and apply the above strategy to obtain Escherichia coli and Corynebacterium glutamicum amino acid overproducers. Three amino acids, L-leucine, L-arginine, and L-serine, are chosen because they are important fermentation products and have specific rare codons. The system is constructed by replacing defined numbers of leucine, arginine, and serine codons with the corresponding E. coli rare codon CTA, AGG or TCC in antibiotic resistance protein (KanR or SpecR), green fluorescent protein (GFP) or the chromogenic pranerpurple protein (PPG, from ATUM). The proteins encoded by rare codon-rich derivatives of the genes are expressed at low levels under amino acid starvation or growth restriction conditions. We show that the protein expressions from the rare codon-rich gene derivatives are dramatically increased by feeding or enhanced intracellular synthesis of the corresponding amino acid. Therefore, amino acid overproducers are readily screened out and the conditions for screening or selection are optimized. To evaluate its performance, the above strategy is successfully applied to screen for L-leucine, L-arginine, and L-serine overproducers from E. coli mutation libraries. Several C. glutamicum strains overproducing L-arginine are also successfully selected by the same strategy. This study proves that our rare codon-based strategy is a promising alternative for the high-throughput screening of amino acid overproducers.

Results
Rare codon-based selection and screening systems. Here, we establish two systems for the identification of amino acid overproducers. One is a selection system based on rare codon-rich antibiotic resistance genes. Any strain that survives the antibiotics is likely an amino acid overproducer. The other is a screening system that used colored proteins encoded by genes harboring the rare codons. Overproducers of the targeted amino acids are readily identified by visual screening. The original leucine, arginine or serine codons of the marker genes were replaced by the rare synonymous CTA (0.39%), AGG (0.11%) or TCC (0.86%), individually. Codon replacement was performed using PCR-based accurate synthesis and the generated genes were denoted RC (rare codon). The selection system was employed to pick out potential amino acid overproducers from E. coli mutation libraries derived by atmospheric room temperature plasma (ARTP) mutagenesis. For each candidate, titer of the targeted amino acid was verified by high-performance liquid
concentrations of the corresponding amino acids dramatically slow down protein expression (lower box). Theoretically, the retarded protein expression should be restored by increased intracellular corresponding amino acids, switching to the rare alternatives (e.g. leucine codon CTA for host would typically improve the expression of the desired protein (upper box). On the contrary, the rare tRNAs have lower chances to be charged with the 

For an exogenous gene, replacing its codons (e.g. leucine codon) with synonymous ones that are recognized by the most abundant tRNAs for a specific strain ZB-5.

Effects of rare codon frequency on the selection system. The \( \text{kan}^R \) gene encoding the aminoglycoside 3'-phosphotransferase type Ia contains 29 leucine codons and was employed for the construction of the l-leucine selection system (Fig. 2a). To investigate the effect of rare codon frequency on protein expressions, a series of \( \text{kan}^R \) genes with a gradient frequency of rare codon replacement were examined for kanamycin resistance. The codon replacement, if needed, was done by synthesizing \( \text{kan}^R \) containing 6, 16, 26, and 29 leucine rare codon CTA, generating \( \text{kan}^R-\text{RC6}, \text{kan}^R-\text{RC16}, \text{kan}^R-\text{RC26}, \) and \( \text{kan}^R-\text{RC29} \), respectively. The retarded Kan\(^R\) expression induced by rare codon would confer only limited resistance towards kanamycin, leading to arrested cell growth. These genes were introduced into \( E. \text{coli} \) strains DH5\( \alpha \) and TOP10, as well as an l-valine overproduction strain ZB-5.

The M9 media with different carbon and nitrogen sources and the Luria–Bertani (LB) media with different dilution factors were tested in this study. In the presence of kanamycin, the 0.2\( \times \) LB ensured significant differences in cell densities between strains harboring the wild-type \( \text{kan}^R \) and the \( \text{kan}^R-\text{RCs} \) (Fig. 2b and Supplementary Fig. 1), while the M9 media did not lead to differentiations in strain growth. Compared with the wild-type \( \text{kan}^R \), the cell densities kept decreasing as the leucine rare codon on \( \text{kan}^R \) became more frequent, in agreeing with the assumption that translation efficiency correlated negatively with rare codon frequency, especially under nutrient limitation conditions. The largest difference in OD\(_{600}\) between DH5\( \alpha \) strains containing wild-type \( \text{kan}^R \) and the \( \text{kan}^R-\text{RC29} \) was up to 8.5-fold and was observed in the 18-h culture. Under the same experimental conditions, the largest differences in OD\(_{600}\) were 2.27-fold for ZB-5 and 2.02-fold for TOP10, respectively (Fig. 2b). The results indicate that the rare codon-inhibited protein expression is in a frequency-dependent manner.

Feeding amino acids restored cell growth. The feeding assays were performed to investigate whether the increased intracellular concentrations of amino acids could alleviate the growth chromatography (HPLC) and the desired strains were identified. To test the potentials of this system, this strategy was also employed to \( C. \text{glutamicum} \) by replacing the original l-arginine common codons of a selection marker gene by AGG (0.32%), the rarest arginine codon in \( C. \text{glutamicum} \).
inhibition induced by the rare codons. The cell densities of DH5α strains containing KanR–RCs were partially restored by feeding l-leucine (1.0 g l⁻¹) or a mixture of three amino acids (3AA, 0.3 g l⁻¹ l-leucine, 0.3 g l⁻¹ l-isoleucine, and 0.3 g l⁻¹ l-valine) to the 0.2× LB medium (Fig. 3a). For strains containing kanR–RC16, the addition of l-leucine restored cell OD₆₀₀ by 36.44%, and the addition of 3AA restored cell OD₆₀₀ up to 16.72%. The restoration percentages of feeding l-leucine and 3AA were 49.41 and 40.58%, respectively, for the strains carrying kanR–RC26. The strongest responses toward feeding were observed in strains containing kanR–RC29. At 22 h, the addition of l-leucine led to a tenfold OD₆₀₀ increase from 0.024 to 0.246 and the addition of 3AA had a similar effect, leading to a ninefold OD₆₀₀ increase. These results indicate that the growth restoration become stronger as the leucine rare codon on kanR become more frequent.

To test the potential of the rare codon-based strategy, the kanR and specR genes were also employed for the construction of the rare codon-based selection systems for selecting l-arginine and l-serine overproducers, respectively (Supplementary Note 1). Changing the last eight arginine codons to AGG inhibited cell growth in both E. coli and C. glutamicum (Supplementary Note 1). Compared to the E. coli carrying pAKR-WT, the OD₆₀₀ of E. coli carrying pAKR-RC8 decreased 42.29-fold at 16 h (Fig. 3b). For C. glutamicum carrying pKan-CG-RC8, the OD₆₀₀ decreased 5.35-fold at 16 h compared to the one carrying pKan-CG (Fig. 3b). The growth restoration was also investigated with a gradient of l-arginine addition. At 16 h, the OD₆₀₀ of the E. coli strain carrying pAKR-RC8 was restored to 39.02% of the OD₆₀₀ of the strain carrying pAKR-WT by adding only 0.5 g l⁻¹ l-arginine (Supplementary Fig. 2a). Adding 1.0 g l⁻¹ or higher concentrations of l-arginine inhibited the growth of E. coli strains. The cell density of C. glutamicum strains carrying pKan-CG-RC8 was restored to 49.72 and 60.53% of the cell density of the strain carrying pKan-CG by adding 0.5 g l⁻¹ and 1.0 g l⁻¹ l-arginine, respectively (Supplementary Fig. 2b). The results indicated the sensitivity of the KanR expression from pKan-CG-RC8 to the intracellular concentration of l-arginine. The repressed growth by rare codons and the growth restoration by amino acid feedings in both E. coli and C. glutamicum strains suggest the potential of applying this strategy to different organisms. In addition, the inhibited cell growth in E. coli was observed after replacing all serine codons in specR gene by rare codon TCC. The cell OD₆₀₀ was restored to as much as 97.34% of the strains carrying pSSer-WT by adding only 1.0 g l⁻¹ l-serine (Supplementary Fig. 2c). Taken together, these results indicate that extra amino acid feeding could restore the rare codon repressed protein expression partially or completely in our growth-based selection system.

Modification of the selection stringency. To modify the selection stringency, the effects of gene copy number and promoter strength on our selection system were examined. The kanR gene was driven by various well-characterized promoters and was placed on plasmids containing different origins of replication (ORIs). Strains containing the engineered plasmids were grown in 0.2× LB and the OD₆₀₀ was measured. First, the original

![Fig. 2 Effects of the frequency of leucine rare codon CTA on protein expressions.](Image 357x436 to 472x552)
medium-copy-number ORI (pMB1) in pKan-RC29 or pAKR-RC8 was replaced by the low-copy-number ORI (p15A). The wild-type E. coli strains containing low-copy-number version of pKan-RC29 or pAKR-RC8 plasmid had 10.07- or 2.81-fold decrease in OD$_{600}$ than the same strains containing the corresponding medium-copy-number version of pKan-RC29 or pAKR-RC8 plasmid under amino acid starvation conditions (Fig. 4a). Second, the original constitutive promoter P$_{kan}$ was replaced by P$_{trc}$ (iGEM Part: BBa_J23100), P$_{lacO1}$ (iGEM Part: BBa_J23118), or P$_{lacO1}$ promoter, respectively. P$_{trc}$ and P$_{lacO1}$ are constitutive promoters and the former one is more active. P$_{lacO1}$ is the strongest promoter used in this study when fully induced by IPTG. Since the cell growth correlated positively with the Kan$^R$ level, significant differences in cell densities with different promoter replacements indicate that the expressions of Kan$^R$ protein could be tuned by the strength of the promoters (Fig. 4b). When P$_{lacO1}$ was used, 0.01, 0.1 or 1.0 mM of IPTG was added to the media to induce the expression of the promoter (Fig. 4b and Supplementary Fig. 3). The expressions of Kan$^R$ increased with the increase of IPTG added to the media, generating gfp-RC and ppg-RC (Fig. 2a). For the gfp system, difference in the fluorescence intensity between cells containing wild-type gfp and gfp-RC was measured (Fig. 2c). After 24 h, the fluorescence from cells containing the wild-type gfp was 52.16% higher than that from cells containing the gfp-RC and was 4.8 times as high as that of the control (empty vector) (Supplementary Fig. 4). Different amounts of l-arginine were added to the culture media to final concentrations of 0.5, 1.0 or 2.0 g l$^{-1}$ to study the influence of amino acid feedings on GFP expressions. As expected, the number of fluorescent cells harboring gfp-RC increased steadily with the increase of l-arginine concentrations. Contrary to the wild-type gfp, the protein expression from gfp-RC relied on the feeding of l-arginine in a concentration-dependent manner (Supplementary Fig. 2d).

The negative effects of rare codons on the expression of PPG were reflected by the retarded color development. Since the expression of ppg was driven by lac promoter, a high concentration of IPTG (e.g. 1.0 mM) led to strong induction of the ppg-RC and counteracted the negative effects of rare codons. Therefore, weak induction was preferred to maximize the difference between protein expressions from the wild-type gene and its rare codon-rich derivatives. A concentration of 0.1 mM IPTG achieved clear discrimination in color development 12 h after the induction (Fig. 2d). Compared to the purple color developed from the wild-type ppg, the protein expression from

**Effects of rare codons on color screening system.** The screening system used both the GFP and PPG. A total of 19 leucine codons in gfp, and 14 leucine codons in ppg were replaced by CTA, generating gfp-RC and ppg-RC (Fig. 2a). For the gfp system, difference in the fluorescence intensity between cells containing wild-type gfp and gfp-RC was measured (Fig. 2c). After 24 h, the fluorescence from cells containing the wild-type gfp was 52.16% higher than that from cells containing the gfp-RC and was 4.8 times as high as that of the control (empty vector) (Supplementary Fig. 4). Different amounts of l-arginine were added to the culture media to final concentrations of 0.5, 1.0 or 2.0 g l$^{-1}$ to study the influence of amino acid feedings on GFP expressions. As expected, the number of fluorescent cells harboring gfp-RC increased steadily with the increase of l-arginine concentrations. Contrary to the wild-type gfp, the protein expression from gfp-RC relied on the feeding of l-arginine in a concentration-dependent manner (Supplementary Fig. 2d).

The negative effects of rare codons on the expression of PPG were reflected by the retarded color development. Since the expression of ppg was driven by lac promoter, a high concentration of IPTG (e.g. 1.0 mM) led to strong induction of the ppg-RC and counteracted the negative effects of rare codons. Therefore, weak induction was preferred to maximize the difference between protein expressions from the wild-type gene and its rare codon-rich derivatives. A concentration of 0.1 mM IPTG achieved clear discrimination in color development 12 h after the induction (Fig. 2d). Compared to the purple color developed from the wild-type ppg, the protein expression from

![Fig. 3 Cell growth restored by feeding the corresponding l-amino acids. a Effects of feeding l-leucine or a mixture of three amino acids (3AA: l-leucine, l-valine and l-isoleucine) on the cell growth for C. glutamicum strains harboring kan$^R$ genes with 6–29 leucine rare codons (kan$^R$-RC6, kan$^R$-RC16, kan$^R$-RC26, kan$^R$-RC29). b Changes in cell ODs after feeding l-arginine to E. coli and C. glutamicum strains harboring kan$^R$ genes in which eight arginine codons were replaced by its rare alternatives (encoded by pAKR-RC8 or pKan-CG-RC8), and the growth restoration after feeding l-serine to E. coli strains carrying spec$^R$ which was rich in serine rare codon (encoded by pSSer-RC17). Values and error bars represent the mean and the s.d. (n = 3). **P < 0.01, ***P < 0.001 as determined by two-tailed t test.](image-url)
ppg-RC gave the medium a light pink, indicating the inhibition effect of rare codons on the expression of PPG. The accumulation of colored protein in the 24-h culture of strains containing ppg-RC could still be distinguished from the wild-type ppg when 0.1 mM IPTG was added. This demonstrates the stability and reliability of the rare codon-based screening system.

Obtaining amino acid overproducers by the selection system. Besides uptaking amino acids directly from the environment, high intracellular concentrations of amino acids could be achieved through enhanced de novo productions. Therefore, this selection system could be used to select strains enabling enhanced de novo biosynthesis of L-leucine, L-arginine or L-serine. To verify the efficiency of the rare codon-based selection system, the plasmids pKan-RC29, pAKR-RC8 or pSSer-RC17 (Supplementary Data 1) were applied to random mutation libraries generated by ARTP. The ARTP mutated strains that survived the rare codon-based selection were isolated and cultured individually in 0.2× LB medium. After 12 h of incubation, the mutants selected by the leucine rare codon-based strategy achieved an average OD_{600} of 0.148 while the wild-type strain only grew to an OD_{600} of 0.032 (Supplementary Fig. 5). Eight out of the top ten mutants with the highest OD_{600} values (0.243–0.285) showed increased productivities for L-leucine (Fig. 5a). Among these strains, a strain (LP-4) with a titer of 18.55 mg L-leucine per gram biomass was achieved, which was 2.91 times higher than that of the wild-type strain. L-arginine overproducers RP-1, RP-5, and L-serine overproducer SP-1 were also selected by similar strategies (Supplementary Fig. 5). The titers of 0.679, 0.698, and 0.535 mg corresponding amino acids per gram biomass were achieved respectively (Fig. 5b, c). The system was successfully employed in obtaining C. glutamicum L-arginine overproducers. Among them, a mutated strain CGL-4 achieved 3.7-fold increase in L-arginine production (2.742 mg g⁻¹ biomass) compared to that of the parent strain ATCC13032 (Fig. 5d). The selected overproducers could produce comparable amount of amino acid with and without the plasmid carrying the selection marker (Supplementary Fig. 6). Therefore, the presence of rare codon-rich marker genes has little effect on the production of the corresponding amino acid.

Mechanisms of increased amino acid productions. Significant differences in cell growth and amino acid productions were observed between the wild-type E. coli and the selected L-leucine overproducers. To figure out whether the increase in L-leucine production was due to the enhancement of biosynthetic pathways, the RNA sampled from cells in the exponential growth phase were sequenced for the wild-type strain and the selected mutants cultured in 0.2× LB. A total of over 27, 30, and 24 millions of raw reads were obtained for LP-4, LP-7, and the wild-type strains, respectively. The expression ratios of each gene involved in branched-chain amino acids (BCAAs) biosynthesis and transportations between the wild-type and LP-4 or LP-7 were calculated (Fig. 6a). A positive value denotes upregulated gene expression, while a negative value indicates downregulated gene expression.

L-leucine and L-valine biosynthesis share the same pathway from pyruvate to 2-keto-isovalerate (KIV) (Fig. 6b). LeuA catalyzes the first reaction in the L-leucine-specific biosynthesis pathway branching off from the L-valine pathway in E. coli. The coding gene leuA exhibited twofold higher expressions in both LP-4 and LP-7 compared to the wild-type strain, for which the activation was also verified by qRT-PCR analysis (Fig. 6c). The higher expression of leuA may enhance the conversion of KIV into 2-isopropylmalate and hence L-leucine. The expression of panB, which is responsible for the competing pathway that channels KIV to pantothenate, remained the same level in all three strains. In addition, the biosynthesis of pyruvate, the key precursor of BCAAs, increased in both LP-4 and LP-7. Reduced amounts of pyruvate into other amino acid biosynthetic pathways could facilitate the distribution of pyruvate to the BCAA biosynthetic pathway. As shown by the transcription profiles, the expressions of ilvM and ilvN that encode the two small subunits of the first common enzyme (acetohydroxyacid synthase II) in the super pathways of BCAA biosynthesis were upregulated by 3.26- and 2.95-fold in LP-4, respectively, and 2.86- and 2.92-fold in LP-7, respectively. An average increase of 5.3-fold in ilvN expression and a slight increase of 1.5-fold in ilvE expression were also detected for the two strains by qRT-PCR, suggesting an enhanced flux of pyruvate towards L-valine and L-leucine biosynthesis. Taken together, the activation of L-leucine biosynthetic pathway enables the accumulation of intracellular L-leucine in the mutants.

The gene expressions related to L-leucine transportation decreased at varied degrees (Fig. 6a, b). Most notably, the gene expression levels of the general L-amino acid transporters yhdW, yhdX, yhdY, and yhdZ were downregulated in both strains by
averages of 3.2-, 2.4-, 2.5-, and 2.4 fold, respectively. Gene expression levels of BCAA transporters livH, livM, livG, and livF were also downregulated in both strains by averages of 3.0-, 3.3-, 2.9-, and 2.1-fold, respectively. However, the expression of brnQ, a protein involved in amino acid uptake, was upregulated by 1.24- and 1.20-fold in LP-4 and LP-7, respectively, compared to the wild type. These results suggest that the intracellular accumulation of L-leucine in the selected overproducers could also be facilitated by increased L-leucine uptake and decreased L-leucine secretion, which is validated by the quantification of the mRNA levels of brnQ (Fig. 6c). Moreover, the expression of rpoS that encodes the stress response σ factor (σ38) was also increased twofold in both mutants, while its influence on the increased amino acid productions could not be determined.

To identify the specific mutations in L-leucine overproducers, the genomic DNA of the mutant strains LP-4 and LP-7 were sequenced by the illumine HiSeq PE150 platform, producing 8 million clean reads. A total of 56 and 62 scaffolds were generated for LP-4 and LP-7, respectively. The genomes were aligned among the mutants and the wild-type strain. The two mutants are highly similar and the largest difference between these two strains was a 7.5 kb duplicate of arcA-yaaX fragment that occurred only in LP-4. This indicates that the two L-leucine overproducers might derive from a common parent mutant but underwent slightly different DNA repairing processes. Among the mutations related to amino acid accumulation (Supplementary Table 1), two are the pathway genes related to L-leucine biosynthesis (ilvC and avtA), one is the gene (leuE) that encodes L-leucine transporter and one is a leader peptide (ivbL). Mutations also occurred at the upstream regions for two upregulated genes (brnQ and ptsG).

For any of the LP-4, LP-7, and the wild-type strains, the amount of detected tRNA_{3Leu} in 1× LB medium was significantly higher than the one in 0.2× LB medium (Supplementary Fig. 7). About 96–99% of the tRNA_{3Leu} from the three strains were in the deacylated (uncharged) state in 0.2× LB medium, while tRNA_{3Leu} from the three strains were predominately in the aminocylated (charged) state in 1× LB medium. In 0.2× LB medium, the amounts of detected tRNA_{3Leu} in LP-4 and LP-7, mostly in the deacylated state, were about 3.4- and 3.1-fold higher than the one of the wild-type strain, respectively. In 1× LB medium, the amounts of detected tRNA_{3Leu} in LP-4 and LP-7, mostly in the aminocylated state, were about 30–40% higher than the one of the wild-type strain, respectively.

### Discussion

In this study, we establish a rare codon-based high-throughput screening or selection system. The introduction of one of the three types of rare codons (CTA, AGG or TCC) in target genes decreases the translation efficiency of the mRNA and the expressions of the targeted proteins, which could be partially restored through the addition or increased intracellular production of the corresponding amino acid (Fig. 3 and Supplementary Fig. 2). The protein expressions from rare codon-modified genes rely mostly on the intracellular concentration of the corresponding amino acid. The rare codon-based strategy only depends on a single modified gene sequence and has less
influence on the expression of the proteome of the host. The efficiency of the strategy is rare codon frequency-dependent. When kanR or specR harboring rare codons is applied to select for amino acid overproducers, several desired overproduction strains for L-leucine, L-arginine, and L-serine are successfully obtained from mutation libraries.

Here, the cell growth could be significantly affected by the promoter strength and the copy number of the kanR genes in both rare codon-based selection systems for L-leucine, L-arginine, and L-serine are successfully obtained from mutation libraries.

Fig. 6 Transcription profiles of BCAA biosynthetic pathways in LP-4 and LP-7. a Transcriptome analysis of LP-4, LP-7, and the wild-type strains. Positive values (red) represent upregulated genes and negative values (blue) represent downregulated genes, which were calculated by the RPKM of LP-4 and LP-7 divided by that of the wild-type strain (in log2). The top, middle, and bottom panels contain genes in glycolysis, L-leucine biosynthesis and BCAA transportations, respectively. b The BCAA biosynthetic pathways in E. coli, red and blue arrows indicate the up- and downregulated genes, respectively. c qRT-PCR verification of the genes related to L-leucine biosynthesis and transportation. Values and error bars represent the mean and the s.d. (n = 3)
The false-positive strains could be identified by integrating a *gfp* gene into the plasmids. The overproduction strains containing the plasmid could express fluorescent proteins and could be screened out. Alternatively, a dual-resistance plasmid, which contains a wild-type antibiotic resistance gene (e.g. ampicillin resistance gene) and a rare codon-modified resistance gene (e.g. *kanR*-RC29), could be used to eliminate the false-positive strains. Any strain surviving the dual antibiotic selection should contain the original plasmid and is likely an amino acid overproducer. Enhanced expressions of the aaRS or the rare tRNAs induced by mutagenesis could be another way for strains to escape the selection system. However, based on the genomic sequences of the two selected *L*-leucine overproducers, the coding and regulatory regions for both the aaRS and the rare tRNAs remained unchanged. The odds of mutations in these regions could be low for a selected amino acid overproducer.

Besides false-positive mutants, the toxicity of certain amino acid could also affect the results of selection. For instance, serine is among the most toxic amino acids and *E. coli* cells stop growing in the presence of only 106 mg l\(^{-1}\) L-serine\(^{32}\). The toxicity of serine could be alleviated by the addition of L-isoleucine\(^{33}\). The 0.2× LB used in this study might be able to partially counteract the serine toxicity, and the selected mutants showed L-serine productivities up to 7.3-fold of that of the wild-type strain (Fig. 5c). However, the productivities of the selected L-serine overproducers were still relatively low when compared with the L-leucine or the L-arginine overproducers. This could be attributed to serine toxicity, as strains producing high levels of L-serine could suffer from growth inhibition and die out. The amino acid productivities of the selected mutants also depend on other factors (Supplementary Note 2). Even so, the successful selection of mutants with increased productions for L-leucine, L-arginine, and L-serine demonstrates the sensitivity and efficiency of the rare codon-based selection strategy. Other interferences on the selection system might involve changes in the mRNA secondary structures of the marker genes due to rare codon replacement (Supplementary Fig. 8; Supplementary Note 3). Nevertheless, we prove that our system is applicable to different amino acids, as well as to other organisms such as *C. glutamicum*.

Besides the *gfp*, *kanR*, *spec*\(^{8}\) and *ppg* used in this study, numerous available selective markers or toxin—antitoxin systems, such as *sacB*\(^{34}\), *tolC*\(^{35}\) and *cctB*\(^{36}\), could be used. If lethal genes would be used to select specific amino acid overproducer, the common codons in the corresponding antidote sequences should be replaced by rare codons. Thus, only strains with enhanced concentrations of intracellular amino acids could survive by maintaining the translation efficiency of the antidote genes to inhibit the lethal effects of the toxins. Furthermore, engineered tRNAs that recognize stop codons or carrying non-inherent amino acids might be needed to explore this strategy for amino acids without native rare codons\(^{37,38}\).

The transcriptome and qRT-PCR analysis show that the L-leucine overproducers have more active L-leucine biosynthetic pathway than the wild-type strain. The enhanced production of L-leucine enables the accumulation of intracellular L-leucine in the mutants, which could also be facilitated by increased L-leucine uptake and decreased L-leucine secretion. These strains could also survive the addition of L-leucine analogue (L-2-aminobutyric acid) of up to 1.5 g l\(^{-1}\) while the wild-type strain died at analogue concentrations >1.0 g l\(^{-1}\) (Supplementary Table 2). Among the mutations related to amino acid accumulation, the *lvbl* serves as a key component in the attenuation system that inhibits the expression of the *lvBN* operon when the intracellular L-leucine and L-valine are abundant. Mutations in both its coding and upstream regions might alleviate its attenuation effect, leading to increased expressions of *lvB* and *lvN* as reflected by the transcription profile (Fig. 6). Although exploring the mechanisms related to amino acid overproduction is not our main goal, the genetic information should provide some preliminary clues for understanding our selected mutants. Besides, the activity of alanine aminotransferase (ALT, encoded by *alaA*) was also measured (Supplementary Fig. 9; Supplementary Note 4). However, the ALT activities in the selected overproducers do not exceed that of the wild-type strain, suggesting that the ALT alone could not significantly affect the L-leucine biosynthesis.

Taken together, we develop screening or selection systems for L-leucine, L-arginine or L-serine overproducers, respectively. Results show that the mRNA translations are inhibited by the integration of rare codons in the target genes in a frequency-dependent manner. As a proof-of-concept work, *E. coli* and *C. glutamicum* mutants strain to overproduce a specific amino acid were successfully selected from random mutation libraries. This work provides an alternative strategy to obtain amino acid overproducers.

**Methods**

**Bacterial strains and culture conditions.** The *E. coli* ZB-5 was an L-valine overproducer which was obtained by selection with norvaline, the L-valine analogue, from a mutation library of the *E. coli* strain ICL16, which was a derivative of *E. coli* wild-type strain BW25113\(^{39}\). The *E. coli* DH5α, TOPO10 and ZB-5 used in this study were cultured in LB medium, and the *C. glutamicum* wild-type (ATCC13032) was cultured in CGIII medium. Strains employed in the selection system were cultured in 0.2× LB (2.0 g l\(^{-1}\) tryptone, 1.0 g l\(^{-1}\) yeast extract and 2.0 g l\(^{-1}\) NaCl) for *E. coli* and 0.3× CGIII (3.0 g l\(^{-1}\) peptone, 3.0 g l\(^{-1}\) yeast extract, 8.0 g l\(^{-1}\) NaCl and 0.6% glucose) for *C. glutamicum*. Solid media was prepared in all cases by adding 1.5% (w/v) agar. The *E. coli* strains were routinely grown in broth or on plates supplemented with appropriate antibiotics at the following concentrations: 50 μg ml\(^{-1}\) of kanamycin (25 μg ml\(^{-1}\) for *C. glutamicum*), 100 μg ml\(^{-1}\) of ampicillin, 25 μg ml\(^{-1}\) of chloramphenicol or 50 μg ml\(^{-1}\) of spectinomycin. The *E. coli* and *C. glutamicum* strains were grown at 37 and 30 °C, respectively.

**Rare codon introductions.** The codon usages of the *E. coli* and *C. glutamicum* genomes were generated from protein coding sequences according to Refseq\(^{40}\). The rare codon CTA and TCC were chosen for the selection of L-leucine and L-serine overproducers in *E. coli*, respectively. The rare codon AGG could be used for selection of L-arginine overproducers in both *E. coli* and *C. glutamicum*. All these rare codon-modified genes were synthesized by PCR-based accurate synthesis\(^{41}\) (Supplementary Data 2).

**Plasmid constructions.** All the DNA manipulations and cloning for system constructions were performed by standard cloning techniques\(^{42}\). The plasmids and primers used in this study are listed in Supplementary Data 1 and 3, respectively. To construct pKan-RC29, pCM-GFP-RC, and pCPB-37-441-RC (encoding *ppg*), leucine codons in these marker genes were all replaced by rare codon CTA. The genes were synthesized by ~60 bp oligonucleotides with 20 bp overlaps and cloned into the corresponding vectors (pKan-F/pKan-R for pKan-RC9, pCM-F/pCM-R for pCM-GFP-RC, and pCPB-F/pCPB-R for pCPB-37-441-RC) by Gibson assembly\(^{43}\). To construct pKan-RC6, pKan-RC16, and pKan-RC26, the fragments containing *ppg* and 26 rare codons *CAs* were amplified using primers 6RCF-6RCR, 16RCF-16RCR, and 26RCF-26RCR, and cloned into pKan. The *E. coli* Corynebacterium shuttle vector pKan-CG-RC8 was derived from pXMJ19, of which the cat marker was replaced by *kan*\(^{8}\). The different promoters were replaced by in vivo assembly\(^{44}\).

**Rare codon-based selection system.** The pKan, pKan-RC6, pKan-RC16, pKan-RC26, and pKan-RC29 were transformed into *E. coli* DH5α, TOPO10 and ZB-5. For each strain, three colonies were randomly picked and inoculated respectively into 15 ml 0.2× LB medium supplemented with kanamycin (50 μg ml\(^{-1}\)) and incubated at 37 °C in a shaker at 200 rpm. The OD\(_{600}\) were measured at defined time points. The DH5α strains harboring pKan, pKan-RC6, pKan-RC16, pKan-RC26, and pKan-RC29 were inoculated respectively into 10 ml 0.2× LB medium with *L*-leucine (1 g l\(^{-1}\)) or 3AA (*L*-leucine, *L*-arginine, and *L*-valine, 0.3 g l\(^{-1}\) for each amino acid) for each group. The *L*-arginine and *L*-serine selections using pAKR-RC8 and pSser-RC17 were processed in the same way, while the *L*-arginine and *L*-serine were...
replenished at 0.5, 1.0, and 2.0 g l⁻¹ respectively. The OD₆₀₀ (OD₆₀₀ for C. glutamicum) were measured in triplicate at defined time points.

**Rare codon-based screening system.** The pCm-GFP and pCm-GFP-RC were transformed into E. coli DH5α. For each strain, three colonies were randomly picked and inoculated respectively into 10 ml LB medium and incubated at 37 °C in a shaker at 200 rpm, while the background fluorescence was measured using DH5α harboring pCm. The fluorescence (excitation: 470 nm; emission: 510 nm) and OD₆₀₀ were measured at 12 h. The fluorescence microscopy images were captured at 12, 18, and 24 h from drop of culture solution containing fluorescent cells by a Nikon Eclipse 80i system equipped with DS-Ri1 camera. The ratio of GFP to OD₆₀₀ (GFP/OD₆₀₀) was used to represent fluorescence intensity. The DH5α strain harboring plasmids pCPB-37-441 or pCPB-37-441-RC was inoculated into 10 ml LB medium containing 0.1 mM IPTG and incubated in a shaker at 200 rpm at 37 °C.

**Screening amino acid overproducers from mutation libraries.** The ARTP mutation system [Wuxi Tmantree Biotechnology Co., Ltd.] that could cause greater gene damage than traditional mutagenesis was employed to generate the mutation libraries 45. The wild-type strain grown to OD₆₀₀ between 0.5 and 0.8 was selected for ARTP treatment. Ten microliters of the culture were dropped onto the stainless steel minidisc and then exposed to ARTP jet for 0, 20, 40, and 60 s with fatal rates of 0, 42, 65, and 79%, respectively. After the treatment, mutants were washed and sterilized 1.5 ml Eppendorf tube with 200 μl LB medium and incubated in a shaker at 200 rpm at 37 °C for 1 h. Then 100 μl of the culture was inoculated into 5 ml LB medium and incubated at 250 rpm at 37 °C until the OD₆₀₀ reached 0.4–0.6. The E. coli ARTP mutants were made into competent cells and transformed with pKan-RC29, pKan-RC17, pCm-RC17, pCm-I-serine, and i-serine overproducers selection, respectively. While the C. glutamicum mutants were transformed with pKan-CG-RC8 via electroporation 46. The fast growing cells in 0.2× LB medium were collected by centrifugation, and i-serine, and i-serine were measured at 24 h using a HPLC system (Shimadzu, Tokyo, Japan).

**Analysis of aminoacylation levels of tRNA.** Total tRNA was isolated using TRizol reagent (Invitrogen) and redissolved in 10 mM sodium acetate (pH 4.5).

The aminoacylation levels of tRNA were evaluated for the L-leucine rare codon TRIzol reagent (Invitrogen) and redissolved in 10 mM sodium acetate (pH 4.5). The aminoacylation levels of tRNA were evaluated for the L-leucine rare codon

The aminoacylation levels of tRNA were evaluated for the L-leucine rare codon with pKan-RC29, pKan-RC17, pCm-RC17, pCm-I-serine, and i-serine overproducers selection, respectively. While the C. glutamicum mutants were transformed with pKan-CG-RC8 via electroporation 46. The fast growing cells in 0.2× LB medium were collected by centrifugation, and i-serine, and i-serine were measured at 24 h using a HPLC system (Shimadzu, Tokyo, Japan).

**References**

1. Tomouchi, N. & Ito, H. Present global situation of amino acids in industry. In Amino Acid Fermentation (eds Yokota, A. & Ikeda, M.) (Springer, New York, 2017).

2. Tatsunami, N. & Inui, M. Corynebacterium Glutamicum: Biology and Biotechnology (Springer, New York, 2012).

3. Becker, J. & Wittmann, C. Systems and synthetic metabolic engineering for amino acid production—the heart of industrial strain development. Curr. Opin. Biotech. 23, 718–726 (2012).

4. Shive, W. & Skinner, C. G. Amino acid analogues. In Metabolic Inhibitors: A Comprehensive Treatise (eds Hochster, R. & Quastel, J.) (Academic Press, Cambridge, MA, 2012).

5. Rodgers, K. J. & Shiozawa, N. Misincorporation of amino acid analogues into proteins by biosynthesis. Int. J. Biochem. Cell Biol. 40, 1452–1466 (2008).

6. Norris, R., & Lea, P. The use of amino acid analogues in biomedical studies. Sci. Prog. 63, 65–85 (1976).

7. Bach, T. M. & Takagi, H. Properties, metabolisms, and applications of L-proline analogues. Appl. Microbiol. Biotechnol. 97, 6623–6634 (2013).

8. Park, J. H. & Lee, S. Y. Fermentative production of branched chain amino acids: a focus on metabolic engineering. Appl. Microbiol. Biotechnol. 85, 491–506 (2010).

9. Park, J. H. & Lee, S. Y. Towards systems metabolic engineering of microorganisms for amino acid production. Curr. Opin. Biotech. 19, 454–460 (2008).

10. Guseytiner, M., Lunts, M., Kozlov, Y., Ivanovskaya, L. & Voroshilova, E. DNA coding for mutant isoprOPylimate synthase, i-leucine-producing microorganism and method for producing L-leucine. US Patent 2008/0281251, 122, 964 (2005).

11. Cowie, D. B., Cohen, G. N., Bolton, E. T. & De Robichon-Szulmajster, H. Amino acid analog incorporation into bacterial proteins. Biochim. Biophys. Acta 39, 39–46 (1959).

12. Richmond, M. The effect of amino acid analogues on growth and protein synthesis in microorganisms. Biotechnol. Rev. 26, 398 (1962).

13. Hoel, M. G. et al. Chemical evolution of a bacterial proteome. Angew. Chem. Int. Ed. 54, 10030–10034 (2015).

14. Crick, F. H. C. On the genetic code. Science 139, 461–464 (1963).

15. Hershberg, R. & Petrov, D. A. Selection on codon bias. Annu. Rev. Genet. 42, 287–299 (2008).

16. Plotkin, J. B. & Kadla, G. Synonymous but not the same: the causes and consequences of codon bias. Nat. Rev. Genet. 12, 32–42 (2011).
