inNovations

NEWSLETTER OF NOVAGEN, INC. • ADVANCED PRODUCTS AND PROTOCOLS FOR MOLECULAR BIOLOGY RESEARCH

NUMBER 12
JUNE 2001

FEATURE ARTICLE

Overcoming the codon bias of *E. coli* for enhanced protein expression

Robert Novy, Don Drott, Keith Yaeger and Robert Mierendorf — Novagen, Inc.

Most amino acids are encoded by more than one codon, and each organism carries its own bias in the usage of the 61 available amino acid codons. In each cell, the tRNA population closely reflects the codon bias of the mRNA population (1, 2). When the mRNA of heterologous target genes is overexpressed in *E. coli*, differences in codon usage can impede translation due to the demand for one or more tRNAs that may be rare or lacking in the population (3–5). Insufficient tRNA pools can lead to translational stalling, premature translation termination, translation frameshifting and amino acid misincorporation (4).

In practice with the pET System and other high-level *E. coli* expression systems, the presence of a small number of rare codons often does not severely depress target protein synthesis. However, heterologous protein expression can be very low when a gene encodes clusters of and/or numerous rare *E. coli* codons. The most severe effects on expression have been observed when multiple consecutive rare codons are near the N-terminus of a coding sequence (6).

Rare codons in *E. coli*

Examination of codon usage in all 4,290 *E. coli* genes reveals a number of codons that are underrepresented (Table 1). In particular, Arg codons AGA, AGG, and CGA, Ile codon AUA, and Leu codon CUA all represent less than 8% of their corresponding population of codons. The codon usage of abundantly expressed genes ("Class II", Table 1) demonstrates a more extreme bias in which the aforementioned low-usage codons are avoided, and codons for Gly (GGA), Arg (CGG) and Pro (CCC) fall to

continued on page 2

Table 1. Arg, Gly, Ile, Leu and Pro codon usage in *E. coli*

amino acid	codon	fraction in all genes	fraction in Class II
Arg	AGG	0.022	**0.003**
Arg	AGA	0.039	**0.006**
Arg	CGG	0.098	**0.008**
Arg	CGA	0.065	**0.011**
Arg	CGU	0.378	0.643
Arg	CGC	0.398	0.330
Gly	GGG	0.151	0.044
Gly	GGA	0.109	**0.020**
Gly	GGU	0.337	0.508
Gly	GGC	0.403	0.428
Ile	AUA	0.073	**0.006**
Ile	AUU	0.507	0.335
Ile	AUC	0.420	0.659
Leu	UUG	0.129	0.034
Leu	UUA	0.131	0.055
Leu	CUG	0.496	0.767
Leu	CUA	0.037	**0.008**
Leu	CUU	0.104	0.056
Leu	CUC	0.104	0.080
Pro	CCG	0.525	0.719
Pro	CCA	0.191	0.153
Pro	CCU	0.159	0.112
Pro	CCC	0.124	**0.016**

Codon usage is expressed as the fraction of all possible codons for a given amino acid. "All genes" is the fraction represented in all 4,290 coding sequences in the *E. coli* genome (6). "Class II" is the fraction represented in 195 genes highly and continuously expressed during exponential growth (7).

IN THIS ISSUE

ARTICLES

TECHNICAL NOTES

NEW PRODUCTS
10

Novagen®

A Brand of CN Biosciences, Inc.
An Affiliate of Merck KGaA, Darmstadt, Germany

continued from page 1

less than 2% of their respective populations. Under typical growth conditions, target genes are expressed in *E. coli* at levels similar to (or exceeding) those for abundantly expressed endogenous genes. Thus, it is likely in many cases that the resident tRNA population available for target protein synthesis would more closely resemble that of the "Class II" genes in Table 1. Theoretically, modification of culture conditions (e.g. lowering the temperature, changing media composition, etc.) might shift the codon usage bias enough to alleviate some codon usage-based expression problems. However, it has been reported that the levels of most of the tRNA isoacceptors corresponding to rare codons remain unchanged at different growth rates (2). Translation problems similar to those caused by codon usage bias can also be created by high-level expression of proteins having an abundant amino acid. In these cases, expression may be improved by supplying the limiting amino acid in the culture medium (3).

Supplying the demand

As shown in Table 1, a subset of the codons for Arg, Ile, Gly, Leu, and Pro are very rarely used in highly expressed *E. coli* genes. Several laboratories have shown that expression yields of proteins whose genes contain rare codons can be dramatically improved when the cognate tRNA is increased within the host (8–10). tRNA levels can be elevated by increasing the copy number of the respective tRNA gene. This is typically accomplished by inserting the wild type tRNA gene on a multiple copy plasmid. The tRNA gene is either inserted into the expression vector itself or placed on a compatible plasmid. Early studies focused on the effect of increasing the gene dosage of *argU* (also known as *dnaY*, which encodes a tRNA that recognizes the AGG/AGA codons) on expression yield, plasmid stability and cell viability (9). The yield of human tissue plasminogen activator was increased approximately 10-fold (from 3% of total cell protein to 30%) in a strain that carried the *dnaY* gene on a compatible plasmid. Numerous subsequent studies also reported a substantial increase in protein yield when employing *E. coli* hosts with enhanced *argU* expression. Increasing the amount of other rare tRNAs has also been shown to augment the yield and fidelity of heterologous proteins. tRNA genes for *ileX* (AUA), *leuW* (CUA), *proL* (CCC) or *glyT* (GGA) have all been used in this manner (reviewed in 3).

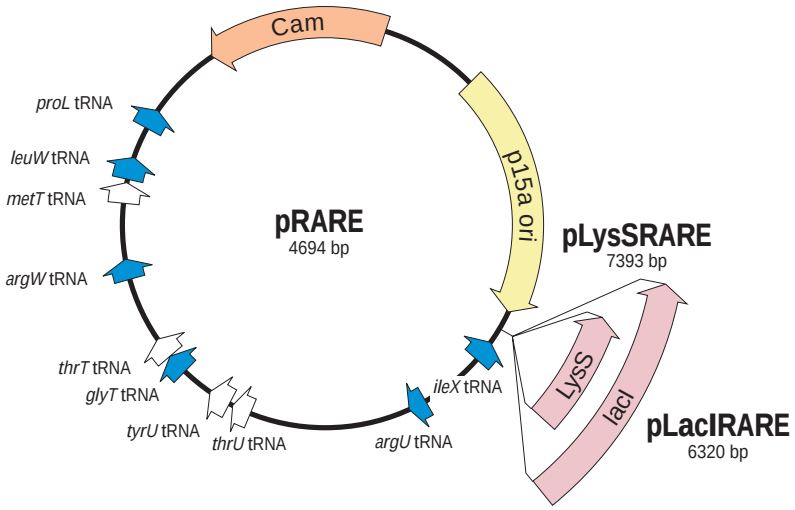

Figure 1. Map of pRARE plasmid family

The basic structure of pRARE is indicated. pLysSRARE and pLacIRARE contain the genes encoding T7 lysozyme (LysS) and *lac* repressor (lacI), respectively. Also indicated are chloramphenicol resistance gene (Cam), origin of replication (p15a ori) and tRNA genes. tRNA genes corresponding to rare codons in *E. coli* are indicated in blue. pRARE is derived from pRIG (11).

pRIG plasmid

More recently, various combinations of rare tRNA genes have been assembled to optimize the expression of genes isolated from organisms with AT or GC rich genomes that have corresponding codon usage bias. One such assembly, the pRIG plasmid (11), encodes tRNA genes *argU*, *ileX* and *glyT* under their native promoters on a pACYC backbone, which carries the p15a origin of replication. The presence of pRIG in the host strain was shown to significantly enhance the expression of several genes derived from an AT-rich *Plasmodium* genome.

pRARE and the Rosetta™ strains

To further extend the utility of pRIG for the expression of genes having rare *E. coli* codons, Novagen has added the *leuW* and *proL* tRNA genes to create pRARE (see Fig. 1). pRARE encodes tRNA genes for all of the "problematic" rarely used codons encoding Arg, Ile, Gly, Leu and Pro, except for Arg CGA/CGG. The entire RARE tRNA cassette was also added to the pLysS and pLacI plasmids (derived from pACYC184) to create pLysSRARE and pLacIRARE, respectively. The plasmids were transformed into various strains to create the Rosetta™ series of expression hosts (Table 2). These plasmids are compatible with Novagen's pET, pETBlue™ and pTriEx™ expression vectors, and with expression vectors driven by other *E. coli* promoters (Table 3). These host strains are well-suited to enhance protein expression from target genes containing rare *E. coli* codons that would otherwise impede translation.

Table 2. Rosetta host strains

Strain	Derivation	Key Feature(s)	Antibiotic Resistance	Available as Competent Cells
Rosetta	Tuner™ (B)	BL21 *lacYZ* deletion, Lacks *lon* and *ompT* proteases	Cam	yes
Rosetta(DE3)			Cam	yes
Rosetta(DE3)pLysS			Cam	yes
Rosetta(DE3)placI			Cam	yes
RosettaBlue™	NovaBlue (K-12)	*recA, endA, lacI*^{q} High transformation efficiency	Tet + Cam	yes
RosettaBlue(DE3)			Tet + Cam	yes
RosettaBlue(DE3)pLysS			Tet + Cam	yes
RosettaBlue(DE3)pLacI			Tet + Cam	yes
Rosetta-gami™	Origami™ (K-12)	*trxB*/*gor* mutant, greatly facilitates cytoplasmic disulfide bond formation	Kan + Tet + Cam	yes
Rosetta-gami(DE3)			Kan + Tet + Cam	yes
Rosetta-gami(DE3)pLysS			Kan + Tet + Cam	yes
Rosetta-gami(DE3)pLacI			Kan + Tet + Cam	yes

Table 3. Vector compatibility of Rosetta host strains

Strain	Compatible vectors*
Rosetta™ RosettaBlue™ Rosetta-gami™	*E. coli* promoter based vectors, e.g. *tac*, *trc*, T5, λ.
Rosetta(DE3) Rosetta(DE3)pLysS RosettaBlue(DE3) RosettaBlue(DE3)pLysS Rosetta-gami(DE3) Rosetta-gami(DE3)pLysS	T7 and T7*lac* promoter based pET vectors
Rosetta(DE3)pLacI RosettaBlue(DE3)pLacI Rosetta-gami(DE3)pLacI	T7*lac* promoter based pETBlue™, pTriEx™ vectors

* All vectors must carry a ColE1-based origin of replication (e.g. pBR, pUC) and lack chloramphenicol selection. RosettaBlue and Rosetta-gami hosts require the use of vectors that also lack tetracycline resistance, and tetracycline plus kanamycin resistance, respectively.

vtPA expression test

To evaluate the Rosetta host strains, a DNA fragment encoding vtPA (Δ6-175 deletion mutant of human tissue plasminogen activator) was cloned into the *Nco* I and *Hin*d III sites of both pET-21d(+) and pTriEx™-3 (12). Full length tPA was one of the original proteins whose expression was enhanced by additional copies of *argU* tRNA. The Δ6-175 vtPA construct encodes a 358 amino acid 40 kDa protein that contains 9 CCC, 8 AGA/AGG, 6 GGA, 6 CGG/CGA, 2 AUA, and 1 CUA for a total of 32 rare codons (8.9%). The pET construct was transformed into BL21(DE3), BL21(DE3)pLysS and the isogenic Rosetta strains, and the pTriEx plasmid was used with corresponding pLacI hosts. Fig. 2 shows an SDS-PAGE analysis of total cell extracts. In panel A, analysis of the pET constructs shows that expression was low but detectable in BL21(DE3) and undetectable in BL21(DE3)pLysS. In contrast, vtPA was the major protein present in the induced samples from the Rosetta strains. A similar strong expression band was obtained when vtPA was expressed from pTriEx-3 in a Rosetta(DE3)pLacI host (panel B) but not in the host lacking the RARE cassette (BL21(DE3)pLacI).

In summary, Novagen's new Rosetta host strains allow researchers to alleviate low protein expression yields caused by codon usage bias. The strains circumvent the need to synthesize codon optimized genes, and enable rapid evaluation of codon usage as a possible strategy to enhance target protein expression. The strains are derivatives of popular *E. coli* expression hosts and versions are available for use with all of Novagen's *E. coli* and multisystem expression vectors. In addition, non-λDE3 lysogens are available for compatibility with other ColE1 origin, *E. coli* promoter based expression vectors.

REFERENCES

1. Ikemura, T. (1981) *J. Mol. Biol.* **146**, 1–21.
2. Dong, H., Nilsson, L. and Kurland, C.G. (1996) *J. Mol. Biol.* **260**, 649–663.
3. Kane, J.F. (1995) *Curr. Opin. Biotechnol.* **6**, 494–500.
4. Kurland, C. and Gallant, J. (1996) *Curr. Opin. Biotechnol.* **7**, 489–493.
5. Goldman, E., Rosenberg, A.H., Zubay, G. and Studier, F.W. (1995) *J. Mol. Biol.* **245**, 467–473.
6. Nakamura, Y., Gojobori, T. and Ikemura, T. (2000) *Nucl. Acids Res.* **29**, 292.
7. Henaut, A. and Danchin, A. (1996) in *Escherichia coli and Salmonella typhimurium cellular and molecular biology, vol. 2*, pp. 2047–2066. Neidhardt, F., Curtiss III, R., Ingraham, J., Lin, E., Low, B., Magasanik, B., Reznikoff, W., Riley, M., Schaechter, M. and Umbarger, H. (eds.) American Society for Microbiology, Washington, DC
8. Rosenberg, A.H., Goldman, E., Dunn, J.J., Studier, F.W. and Zubay, G. (1993) *J. Bacteriol.* **175**, 716–722.
9. Brinkmann, U., Mattes, R.E. and Buckel, P. (1989) *Gene* **85**, 109–114.
10. Seidel, H.M., Pompliano, D.L. and Knowes J.R. (1992) *Biochemistry* **31**, 2598–2608.
11. Baca, A.M. and Hol, W.G.J. (2000) *Int. J. Parasitology* **30**, 113–118.
12. Novy, R., Yaeger, K., Monsma, S. and Scott, M. (1999) *in**Nova**tions* **10**, 1–5.

The authors thank Dr. Jon Beckwith for providing the Δ6-175 vtPA gene used in this article.

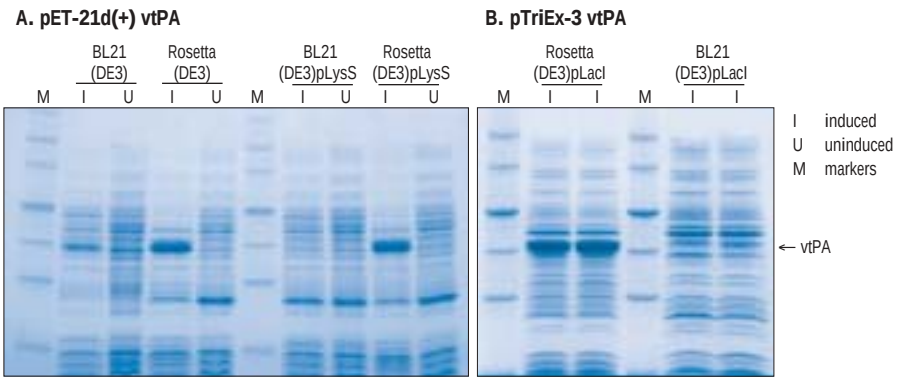

Figure 2. Expression of Δ6-175 vtPA in different host strains

vtPA constructs in pET-21d(+) and pTriEx-3 were transformed into the indicated host strains. Cultures were grown at 37°C in LB + 0.5% glucose to an OD_{600} of 0.6 to 1.0 and aliquots induced with 1 mM IPTG for 3 hours. Total cell protein samples were prepared and then analyzed by SDS-PAGE (4–20% gradient gels) and Coomassie blue staining. Panel A, pET-21d(+) vtPA; Panel B, pTriEx-3 vtPA. Duplicate induced cultures are shown in Panel B.

Product	Size	Cat. No.
Rosetta™ Competent Cells (guaranteed efficiency > 2 × 10⁶ cfu/µg)	0.4 ml 1 ml	70953-3 70953-4
RosettaBlue™ Competent Cells (guaranteed efficiency > 1 × 10⁸ cfu/µg)	0.4 ml 1 ml	71058-3 71058-4
Rosetta-gami™ Competent Cells (guaranteed efficiency > 2 × 10⁶ cfu/µg)	0.4 ml 1 ml	71054-3 71054-4
Rosetta(DE3) Competent Cells (guaranteed efficiency > 2 × 10⁶ cfu/µg)	0.4 ml 1 ml	70954-3 70954-4
Rosetta(DE3)pLysS Competent Cells (guaranteed efficiency > 2 × 10⁶ cfu/µg)	0.4 ml 1 ml	70956-3 70956-4
RosettaBlue(DE3) Competent Cells (guaranteed efficiency > 1 × 10⁸ cfu/µg)	0.4 ml 1 ml	71059-3 71059-4
RosettaBlue(DE3)pLysS Competent Cells (guaranteed efficiency > 1 × 10⁸ cfu/µg)	0.4 ml 1 ml	71034-3 71034-4
Rosetta-gami(DE3) Competent Cells (guaranteed efficiency > 2 × 10⁶ cfu/µg)	0.4 ml 1 ml	71055-3 71055-4
Rosetta-gami(DE3)pLysS Competent Cells (guaranteed efficiency > 2 × 10⁶ cfu/µg)	0.4 ml 1 ml	71057-3 71057-4
Rosetta(DE3)pLacI Competent Cells (guaranteed efficiency > 2 × 10⁶ cfu/µg)	0.4 ml 1 ml	70920-3 70920-4
RosettaBlue(DE3)pLacI Competent Cells (guaranteed efficiency > 1 × 10⁸ cfu/µg)	0.4 ml 1 ml	71060-3 71060-4
Rosetta-gami(DE3)pLacI Competent Cells (guaranteed efficiency > 2 × 10⁶ cfu/µg)	0.4 ml 1 ml	71056-3 71056-4