

Translational Features of Human Alpha 2b Interferon Production in *Escherichia coli*

C. A. Valente, D. M. F. Prazeres, J. M. S. Cabral, and G. A. Monteiro*

Centro de Engenharia Biológica e Química, Instituto Superior Técnico, 1049-001 Lisbon, Portugal

Received 21 November 2003/Accepted 29 April 2004

The yield of human alpha 2b interferon in *Escherichia coli* was optimized by replacement of low-usage arginine codons located in the mRNA 5' end. The differences observed among the various gene variants suggest that codon usage, Shine-Dalgarno-like sequences, and mRNA secondary structure contribute to the performance of *E. coli* translation machinery.

Human alpha interferons (IFN- α) belong to a family of homologous proteins that are known to display antiviral, anti-proliferative, and immunomodulatory activities (3, 11). IFN- α 2b is one of the most studied IFN- α , and its intronless sequence encodes a nonglycosylated protein containing 165 amino acid residues (3, 10).

The expression of heterologous proteins in *Escherichia coli* is often hampered by the presence of arginine low-usage codons, AGG and AGA. This is probably due to the low abundance of the cognate tRNA in bacteria (2, 4, 6, 9, 16). The inhibition of translation is more pronounced when these low-usage codons are located in clusters near the mRNA 5' end (1, 5, 7, 13), which can mimic the Shine-Dalgarno (SD) sequence even in the absence of an associated cognate initiation codon (1, 8, 12). The presence in mRNA of stable secondary structures that can block initiation has also been correlated with preventing translation in *E. coli* (14, 15).

This work aims to study the effect of rare arginine codons, single and clustered, located at the mRNA 5' end, as well as the effect of the inherent secondary structures, on the yield of human IFN- α 2b (hu-IFN- α 2b) in *E. coli*. Site directed mutagenesis was used to construct a series of hu-IFN- α 2 gene variants. Some variants comprised the replacement of arginine clusters (Arg¹²Arg¹³ and Arg²²Arg²³), while others included the replacement of an isolated codon (Arg³³). The evaluation of hu-IFN- α 2b yield from these gene variants suggests the contribution of several factors, namely codon usage, SD-like sequences, and mRNA secondary structure, to gene expression in *E. coli*.

A 520-bp fragment coding for hu-IFN- α 2b was amplified from human genomic DNA by PCR, using the sense, 5'-TTG AATTCATATGTGTGATCTGCCTCAAACCCACAGC-3', and the antisense, 5'-CCTAGGTAATAAGGAAGAAGAAT TTGAAAGAACG-3', oligonucleotides. This fragment was restricted and cloned into BamHI/EcoRI-digested pUC19. Subsequently, the hu-IFN- α 2b fragment was cut from pUC-IFN α and cloned into a BamHI/NdeI-digested pET-9a expression vector (Promega). A series of hu-IFN- α 2b gene variants (Fig.

1) was constructed by site-directed mutagenesis (QuickChange kit; Stratagene). Both IFN wild-type and variant genes were sequenced by Replicon (Berlin, Germany).

E. coli JM109(DE3) cells harboring either pET9-IFN α or pET9-IFN-MRx were grown at 37°C and at 250 rpm in a shake flask containing 100 ml of Luria-Bertani medium plus kanamycin (30 μ g/ml). The host strain contains, integrated into its chromosome, a single copy of the gene for the T7 RNA polymerase under the control of the isopropyl- β -D-thiogalactopyranoside (IPTG)-inducible *lacUV5* promoter. Consequently, upon IPTG induction, large amounts of T7 RNA polymerase are produced, which leads to the expression of genes positioned downstream of the T7 promoter, as is the case of the hu-IFN- α 2b gene. Protein expression was induced at mid-exponential phase (optical density [OD] at 600 nm of \sim 1) with 1 mM IPTG. No significant differences were observed between maximal growth rates (0.57 ± 0.07 h⁻¹) of cells expressing the different gene variants.

Total bacterial proteins were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis and transferred to a nitrocellulose membrane. Immunodetection was performed through an overnight incubation with a monoclonal antibody against hu-IFN- α 2b, produced by MT₄/E₄ hybridoma cells (European Collection of Cell Cultures), followed by incubation with a goat anti-mouse immunoglobulin G alkaline phosphatase conjugate and then with an alkaline conjugate substrate (Bio-Rad). Western blots were photographed in the Eagle Eye System (Stratagene) and analyzed by densitometry with the TotalLab v1.11 software (Phoretix, Newcastle upon Tyne, United Kingdom). Total hu-IFN- α was quantified using a calibration curve on each Western blot constructed with 50, 100, 200, and 400 ng of a pharmaceutical IFN- α 2b standard, Intron A (Schering-Plough Corp.).

Both amino-terminal and internal amino acid sequencing (carried out by Eurosequence, Groningen, The Netherlands) confirmed, respectively, the hu-IFN- α 2b sequences Asp²-Thr⁶ and Ser¹⁵⁰-Leu¹⁶¹, indicating that the correct protein was expressed.

Recombinant protein expression was detected through sodium dodecyl sulfate-polyacrylamide gel electrophoresis, and the hu-IFN- α 2b yield was evaluated by performing densitometry on Western blots. Yields were compared at 5 h postinduction (mid-exponential phase). At this stage, hu-IFN- α 2b yields

* Corresponding author. Mailing address: Centro de Engenharia Biológica e Química, Instituto Superior Técnico, Av. Rovisco Pais, 1049-001 Lisbon, Portugal. Phone: 351-218419065. Fax: 351-218419062. E-mail: gabmonteiro@ist.utl.pt.

	Cys ¹ ... Ser ¹¹ Arg ¹² Arg ¹³ Thr ¹⁴ Leu ¹⁵ Met ¹⁶ ... Met ²¹ Arg ²² Arg ²³ Thr ²⁴ ... Lys ³¹ Asp ³² Arg ³³ His ³⁴ ... Glu ¹⁶⁵
hu-IFN α 2	TGT ... AGC <u>AGG AGG</u> ACC TTG ATG ... ATG <u>AGG AGA</u> ATC ... AAG GAC AGA CAT ... GAA
hu-IFN α 2-MR1	TGT ... AGC <u>CGC CGC</u> ACC TTG ATG ... ATG <u>AGG AGA</u> ATC ... AAG GAC AGA CAT ... GAA
hu-IFN α 2-MR2	TGT ... AGC <u>AGG AGG</u> ACC TTG ATG ... ATG <u>CGC CGC</u> ATC ... AAG GAC AGA CAT ... GAA
hu-IFN α 2-MR12	TGT ... AGC <u>CGC CGC</u> ACC TTG ATG ... ATG <u>CGC CGC</u> ATC ... AAG GAC AGA CAT ... GAA
hu-IFN α 2-MR3	TGT ... AGC <u>AGG AGG</u> ACC TTG ATG ... ATG <u>AGG AGA</u> ATC ... AAG GAC <u>CGC</u> CAT ... GAA
hu-IFN α 2-MR13	TGT ... AGC <u>CGC CGC</u> ACC TTG ATG ... ATG <u>AGG AGA</u> ATC ... AAG GAC <u>CGC</u> CAT ... GAA
hu-IFN α 2-MR123	TGT ... AGC <u>CGC CGC</u> ACC TTG ATG ... ATG <u>CGC CGC</u> ATC ... AAG GAC <u>CGC</u> CAT ... GAA

FIG. 1. Nucleotide sequences of hu-IFN- α 2b gene variants. The nucleotide substitutions are shown underlined, and SD-like sequences are highlighted in grey. (For amino acid position numbering we have not considered the first methionine, which is found only in the recombinant protein.)

were maximal in all constructs except for MR13 and MR123 (~75% of maximal yield).

The gene variants constructed by replacing the low-usage arginine codons (AGG and AGA) with the major-usage codon (CGC) were found to express higher levels of hu-IFN- α 2b than the nonmutated clone. However, a correlation between the number of replaced codons and protein yield was not found.

Hu-IFN- α 2b yields of variants MR1 and MR2 increased 3.6-fold and 2.0-fold, respectively, relative to the control gene (Fig. 2), which has two rare tandem arginine codons, Arg¹²Arg¹³ (AGGAGG) and Arg²²Arg²³ (AGGAGA) (Fig. 1). Due to the similarity with the SD sequence, these dual clusters of arginine codons often have been called artificial translation initiators. The presence of those sequences in a gene which is to be expressed in *E. coli* is known to originate

incorrect translation events, such as frameshifting and dissociation of the translation complexes (13). This results from the competition with the functional SD sequence (8, 12) and thus leads to a decrease in protein yield. The improved hu-IFN- α 2b yield obtained upon replacement of the artificial translational initiators (MR1 and MR2 variants) can be associated to a decrease in the frequency of these incorrect translational events, corroborating previous findings (17). However, since the number of rare arginine codons replaced in these two constructs was the same (two codons), the differences in protein yield observed between them was unexpected. Apparently the potential role of these tandem arginine codons as translational initiators, due to their similarity with the SD sequence, is less important the farther away they are from the initiation codon. In variant MR2, which has the tandem arginine 63

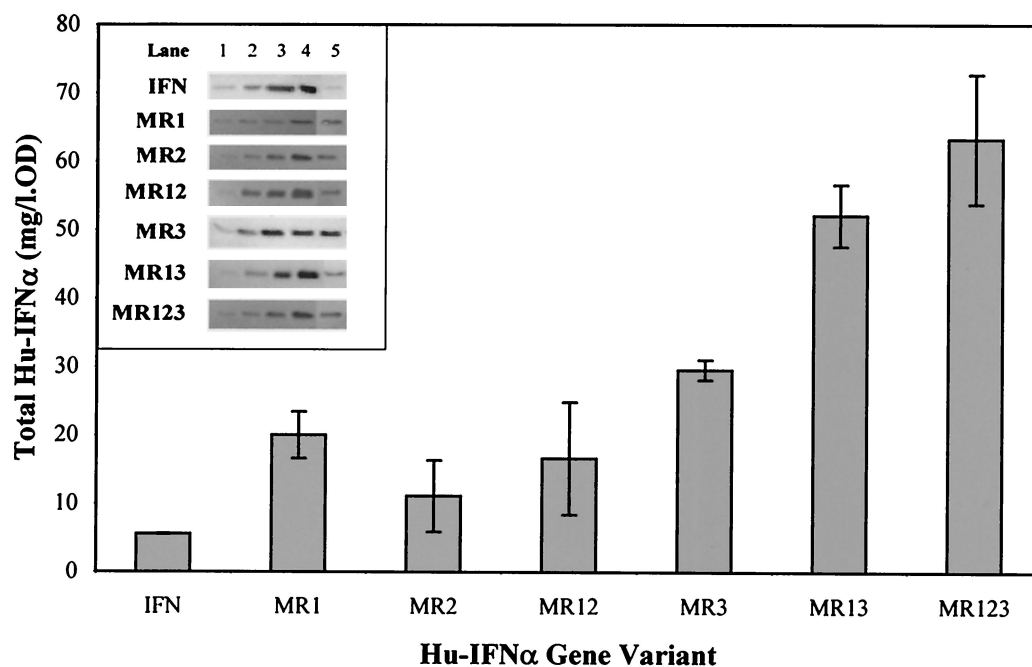


FIG. 2. Hu-IFN- α 2b yield obtained from each gene variant at 5 h postinduction. In the insert, the first four lanes in the Western blots correspond to the standard hu-IFN- α , 50, 100, 200, and 400 ng, respectively. Lane 5 corresponds to the 5-h sample of each gene variant: IFN (5 μ l; OD = 1.96), MR1 (2 μ l; OD = 2.58), MR2 (10 μ l; OD = 1.50), MR12 (4 μ l; OD = 1.37), MR3 (5 μ l; OD = 1.50), MR13 (1 μ l; OD = 1.80) and MR123 (1 μ l; OD = 2.72) l., liter.

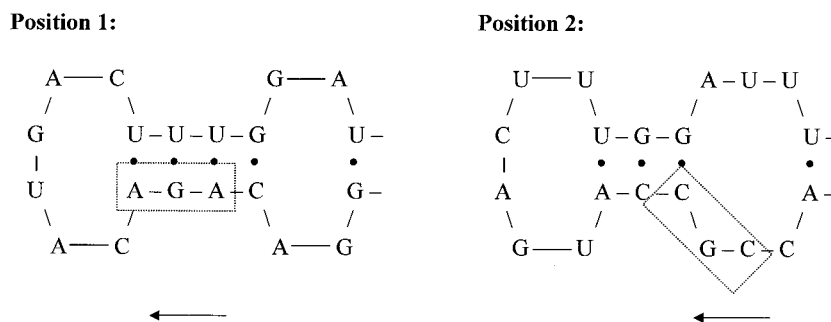


FIG. 3. Close-up of the positions of the Arg³³ codon found in the mRNA secondary structures. The small arrows indicate the direction of translation.

nucleotides downstream from the initiation codon, the replacement effect is weaker than in MR1, which contains the tandem arginine (33 nucleotides) closer to the initiation codon, and thus, hu-IFN- α 2b yield increases less than with the nonmutated gene.

Gene variant MR12 has both rare tandem arginine codons (Arg¹²Arg¹³ and Arg²²Arg²³) replaced and expressed 16.6 mg of hu-IFN/liter of a bacterial culture with an OD of 1, a figure that corresponds to a threefold increase relative to the nonmutated clone. Therefore, hu-IFN- α 2b yield from MR12 is probably regulated not only by codon usage but also by a particular feature in the mRNA secondary structure which will be discussed later.

The more intriguing result was the comparatively high hu-IFN- α 2b yield, 29.6 mg of hu-IFN/liter of a bacterial culture with an OD of 1 (Fig. 2), attained by the single replacement of Arg³³ (gene variant MR3). This was further confirmed by introducing additional mutations, which resulted in the MR13 and MR123 variants (Fig. 1). These clones displayed high levels of hu-IFN- α 2b yield which correspond to a 9.5-fold and 11.5-fold increase relative to the nonmutated clone (Fig. 2).

The reason behind such a strong replacement effect was thought to be related to some feature in the mRNA secondary structure. An *in silico* analysis of the mRNA secondary structures of each transcript variant (677 nucleotides), as predicted by MFOLD (17) (<http://www.bioinfo.rpi.edu/applications/mfold/old/rna/> [18 February 2003, last date accessed]), showed no significant differences between folding energies in all gene variants (E value of ≥ -231.6 kcal/mol and ≤ -227.5 kcal/mol). Regions with a high number of base pairs, such as stems, are less favorable for ribosome progression than unpaired regions, such as loops, and thus decrease the translation rate. However, a relationship between the total number of base pairs in the transcript and the protein yield was not found. Nevertheless, a detailed analysis of the predicted mRNA secondary structures of each construct shows a distinct location for the Arg³³ codon. Figure 3 revealed that the nonreplaced Arg³³ codon (IFN, MR1, MR2, and MR12) is placed in a stem (position 1). The highest hu-IFN- α 2b producers (MR3, MR13, and MR123) contain the replaced Arg³³ codon in a 3-3 nucleotide loop (position 2). Regardless of the codon replacement effect, the hu-IFN- α 2b yield increases with the decrease in the number of base pairs located in the neighboring regions of the Arg³³ codon. This suggests that the easiness of the ribosome progres-

sion through a specific secondary and tertiary structure of the mRNA might have some contribution to the efficiency of the translational machinery.

Independently of the regulating effects of translation, the described replacement of arginine codons showed a cumulative effect upon hu-IFN- α 2b yield: MR13 \approx MR1 + MR3; MR123 \approx MR1 + MR2 + MR3 \approx MR2 + MR13. However, this cumulative replacement effect is not clearly seen in the MR12 variant, probably because the mRNA secondary and tertiary structure negative effect overrides the benefits of replacing the less-rare codons.

In conclusion, this work highlights several aspects of protein translation in *E. coli*. Codon usage, SD-like sequences, and mRNA secondary and tertiary structures were found to contribute to the overall efficiency of the bacterial translational system.

This work was supported by the Portuguese Ministry of Science and Technology FCT/ADI PRAXIS XXI project Protexpress and a Ph.D. grant, BD/16962/97.

We thank A. Kelly for his assistance in the design of oligonucleotides.

REFERENCES

- Alexandrova R., M. Eweida, F. Georges, B. Dragulev, M. AbouHaidair, and I. Ivanov. 1995. Domains in human interferon alpha-1 gene containing tandem arginine codons AGG play the role of translational initiators in *Escherichia coli*. *Int. J. Biochem. Cell Biol.* **27**:469–473.
- Bonekamp, F., and K. F. Jensen. 1988. The AGG codon is translated slowly in *E. coli* even at very low expression levels. *Nucleic Acids Res.* **16**:3013–3024.
- Bordens, R., S. E. Grossberg, P. P. Trotta, and T. L. Nagabushan. 1997. Molecular and biologic characterization of recombinant interferon- α 2b. *Semin. Oncol.* **24**:S9–41–S9–51.
- Brinkmann, U., R. E. Mattes, and P. Buckel. 1989. High-level expression of recombinant genes in *Escherichia coli* is dependent on the availability of the *dnaY* gene product. *Gene* **85**:109–114.
- Chen, G. F., and M. Inouye. 1990. Suppression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of the *Escherichia coli* genes. *Nucleic Acids Res.* **18**:1465–1473.
- Garcia, O. L., B. González, A. Menéndez, A. E. Sosa, J. R. Fernández, H. Santana, and N. Meneses. 1996. The *argU* gene product enhances expression of the recombinant human alpha 2-interferon in *Escherichia coli*. *Ann. N. Y. Acad. Sci.* **782**:79–86.
- Goldman, E., A. H. Rosenberg, G. Zubay, and F. W. Studier. 1995. Consecutive low-usage leucine codons block translation only when near the 5' end of a message in *Escherichia coli*. *J. Mol. Biol.* **245**:467–473.
- Ivanov, I. G., R. Alexandrova, B. Dragulev, A. Saraffova, and M. AbouHaidair. 1992. Effect of tandemly repeated AGG triplets on the translation of CAT-mRNA in *Escherichia coli*. *FEBS Lett.* **307**:173–176.
- Jeong, W., and H. C. Shin. 1998. Supply of the ArgU gene product allows high-level expression of recombinant human interferon-alpha-2a in *Escherichia coli*. *Biotechnol. Lett.* **20**:19–22.

10. Kaluz, S., A. Gibadulinová, and P. Kontsek. 1993. Interferon alpha 2b but not interferon alpha 2a detected in human genomic DNA. *Acta Virol.* **37**:97–100.
11. Lydon, N. B., C. Favre, S. Bove, O. Neyret, S. Benureau, A. M. Levine, G. F. Seelig, T. L. Nagabhushan, and P. P. Trotta. 1985. Immunochemical mapping of α -2 interferon. *Biochemistry* **24**:4131–4141.
12. Mawn, M. V., M. J. Fournier, D. A. Tirrell, and T. L. Mason. 2002. Depletion of free 30S ribosomal subunits in *Escherichia coli* by expression of RNA containing Shine-Dalgarno-like sequences. *J. Bacteriol.* **184**:494–502.
13. Rosenberg, A. H., E. Goldman, J. J. Dunn, F. W. Studier, and G. Zubay. 1993. Effects of consecutive AGG codons on translation in *Escherichia coli*, demonstrated with a versatile codon test system. *J. Bacteriol.* **175**:716–722.
14. Tessier, L. H., P. Sondermeyer, T. Faure, D. Dreyer, A. Benavente, D. Villeval, M. Courtney, and J. P. Lecocq. 1984. The influence of mRNA primary and secondary structure on human IFN- γ . *Nucleic Acids Res.* **12**:7663–7675.
15. Wang, H., S. Miao, D. Chen, L. Wang, and S. S. Koide. 1999. Assignment of chromosomal locus and evidence for alternatively spliced mRNAs of a human sperm membrane protein (hSMP-1). *Biochim. Biophys. Acta* **1447**:119–124.
16. Zahn, K. 1996. Overexpression of an mRNA dependent on rare codons inhibits protein synthesis and cell growth. *J. Bacteriol.* **178**:2926–2933.
17. Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**:3406–3415.