47. *Synonymous Codon Choice in Micrococcus luteus*

By Takeshi OHAMA, Akira MUTO, and Syozo OSAWA

Laboratory of Molecular Genetics, Department of Biology, Faculty of
Science, Nagoya University, Chikusa-ku, Nagoya 464

(Communicated by Motoo KIMURA, M. J. A., Oct. 12, 1989)

Introduction. It has been known that bacterial species differ from one another in the GC-content of their DNA. The GC-content of *Micrococcus luteus* DNA is one of the highest (74%) in eubacteria and that of *Mycoplasma capricolum* DNA is the lowest (25%). These variations have been suggested to result from directional biased mutation pressure on DNA towards AT predominating over GC (AT pressure), or towards GC over AT (GC-pressure).^{1)–3)}

We have proposed that the selective use of synonymous codons is largely determined by the directional mutation pressure.⁴⁾ However, it has been generally believed that the codon usage is mainly determined by tRNAs and that the selective use of synonymous codons is correlated with degree of expression.^{5), 6)} The latter phenomenon is seen in organisms with moderate genomic GC-content such as *Escherichia coli*.⁵⁾ In the present paper, we have analyzed the codon usage of highly expressed genes and weakly expressed genes from *M. luteus* and reached a conclusion that directional mutation is the primary factor for determining synonymous codon choice and that other factors act as modulators at least in this high GC bacterium.

Codon usage in highly and weakly expressed genes in *M. luteus*. The codon usage of the highly expressed genes in the S10-,⁷⁾ spc-,^{8), 9)} and str-operons,^{10), 11)} and that of the weakly expressed genes, uvrA,^{12), 13)} uvrB^{14), 15)} and dnaA¹⁶⁾ are compared between *M. luteus* and *E. coli*. The genes in the ribosomal and its related proteins in the above three operons and adk are expressed at very high level in the cell ($4\text{--}4.5 \times 10^4$ molecules/cell),¹⁷⁾ while those for uvrA, uvrB and dnaA are expressed much more weakly (150–200 molecules/cell).¹⁸⁾ Ikemura has shown that highly expressed genes of *E. coli* reveal strongly biased codon usage of alternative synonymous codons, while weakly expressed genes are less biased.⁵⁾ These observations have been interpreted as a result of the selection of “favorite” codons by abundant tRNAs or by preference of the anticodon pairing with codon in highly expressed genes. The less biased codon usage in the weakly expressed genes is due to a relaxation of selection by tRNAs, and largely affected by mutation pressure. For example, in all the NNU/C-type two-codon sets (N: U, C, A or G; Phe, Tyr, His, Asn, Asp and Cys), NNC codons are more “favorite” than NNU, because each set of NNU/C codons are translated by a single tRNA with anticodon GNN, which seems to translate NNC somewhat more efficiently than NNU. In *E. coli*, it is evident that the genes for ribosomal proteins and adk exhibit a stronger codon usage bias than uvrA, uvrB and dnaA. For example, in all NNU/C-type two-codon sets, except for Cys (UGU/C), NNC codons are preferentially used in the highly expressed genes, while NNC usage decreases in the weakly expressed genes (Fig. 1). On the contrary, in *M. luteus*, NNC codons are used almost exclusively and no difference in codon usage is shown between highly and weakly expressed genes in the all two-codon sets (Fig. 1).

Fig. 1. Synonymous codon ratios in highly and weakly expressed genes in *M. luteus* and *E. coli*. The data for *E. coli* family boxes are not shown to avoid the complexity of figures. H: Highly expressed genes (ribosomal and its related protein genes). W: Weakly expressed genes (*dnaA*, *uvrA* and *uvrB*). —: *M. luteus*;: *E. coli*; ----: GC-contents of codon third positions of *M. luteus*. Cys (UGY), Leu (UUR), Ser (AGY) and Arg (AGR) are not included in the figures of two-codon sets, because of their small number of occurrence.

This suggests that the codon usage in the two-codon sets is almost solely determined by GC-pressure, but not by tRNAs.

Not only in the NNU/C-type two-codon sets, but in all family boxes and NNA/G-type two-codon sets, *M. luteus* preferentially uses G or C at the third positions of codons both in highly (95–100% GC) and weakly expressed genes (90–100%) (Fig. 1). Isoleucine has three codons AUU, AUC and AUA. In both highly and weakly expressed genes, AUC is almost exclusively used. Preferential use of C both in highly and weakly expressed genes is also evident in codons CUN-UUA/G (Leu; CUN 99%) and CGN-AGA/G (Arg; CGN 98%), where the first position is silent. This would suggest that strong GC-biased mutations have primarily determined the extremely high GC-contents of the silent (mainly third) positions. In the NNU/C-type and NNA/G-type two-codon sets, and Ile three-codon set, choice of C or G in all the genes would represent almost purely the degree of GC-mutation pressure. Thus, it would be not possible for tRNAs acting

to shift up the NNC or NNG levels, because these levels have been already "saturated". On the other hand, in eubacteria with moderate GC-contents such as *E. coli*, tRNA can act to increase "favorite" codons along with degree of expression, since GC levels of the codon third positions, that had been affected by directional mutation pressure, are assumed to be not far from 50%. However, these facts do not mean that all the codon usages of *M. luteus* are solely determined by mutation pressure without any selective constraint by tRNAs. Three family boxes (ACN Thr, GCN Ala and UCN Ser) show relaxation of codon usage bias from highly to weakly expressed genes, decreasing NNC codons and increasing NNG codons without involvement of NNU and NNA codons (Fig. 1). This would mean that GC-pressure first determined the overall content of G plus C at the codon third positions. tRNA then acts as a constraint for the synonymous codon choice between NNG and NNC, without changing the GC-content of the codon third position. The ratio of NNG to NNC in three family boxes (CUN Leu, GUN Val and CCN Pro) does not differ between highly and weakly expressed genes, with practically no usage of NNA/U codons. Arg (CGN) and Gly (GGN) family boxes show somewhat complicated patterns. In both cases, NNU usage is exceptionally high, 20 and 14% respectively, in highly expressed genes, while NNG usage increases with apparent expenses of NNU in weakly expressed genes. Presumably, unusual anticodon compositions in these boxes (e.g., the presence of ICG anticodon for Arg) act as constraints to prefer U over G in highly expressed genes.

Six codons, AUA (Ile), AGA (Arg), UUA (Leu), GUA (Val), CUA (Leu), and CAA (Gln), all ending in A, are completely absent in 5516 codons so far sequenced. NNA/G-type codons, except AUA (Ile), are translated primarily by a single anticodon *UNN (*U: modified U). Anticodon *UNN pairs well with codon NNA, however only very weakly with codon NNG, especially in NNA/G-type two-codon sets. Thus, the number of CNN anticodons, pairing exclusively with NNG codons, increases with increasing GC-content of bacterial genomes, because *UNN anticodons cannot accommodate of a large amount of NNG codons resulted by GC-pressure in high GC bacteria. Thus, *M. luteus* would carry nearly the maximum numbers of tRNAs with CNN anticodons. By contrast, tRNAs with *UNN anticodons would have decreased in amount along with a decrease of NNA codons, and have become superfluous. It is possible that some of these tRNAs have already disappeared and consequently the corresponding NNA codons have become unassigned in *M. luteus*. We suggest that appearance/increase or disappearance/decrease of certain tRNA species is an adaptive phenomenon affected by directional mutation pressure. Once *UNN anticodons become trace in amount or non-existent, a few NNA codons that appear by mutations will be immediately selected out against, because of existence of little *UNN anticodons to translate them. This is probably the reason for almost complete absence of NNA codons in *M. luteus* genes.

By contrast, NNU and NNC codons are translated with a single anticodon GNN by wobble as noted above. Thus, a small amount of NNU codons that appeared by mutations may be read by GNN anticodons. In fact, a few NNU codons are used throughout. The absence of AUA (Ile) codon may be explained by the adaptive decrease or deletion of the corresponding anticodon *CAU (the modified nucleoside *C has been identified as lysidine (L) in *E. coli*¹⁹⁾) along with silent conversion of codon AUA to AUC by GC-pressure.

Initiation and termination codons. In *M. luteus*, GUG is used as initiation

codon in 10 out of 18 proteins,^{9),11)} whereas AUG much predominates over GUG in *E. coli*²⁰⁾ or exclusively used in *M. capricolum*.²¹⁾ Most of the termination codons in *M. capricolum*²¹⁾ and *E. coli*²⁰⁾ are UAA, whereas *M. luteus* uses 15 UGA, 2 UAA and 1 UAG. Although AUG is much more efficient initiation codon than GUG or UUG in *E. coli*,²²⁾ GUG must be the efficient initiation codon in *M. luteus*, because ribosomal proteins and elongation factors occur in a large amount in the cells. Massive use of GUG initiation codons instead of AUG, and UGA termination codons instead of UAA in *M. luteus* most probably has resulted from a strong preference of G usage over A, in accordance with the tendency of accumulating G/C at the neutral positions.

Acknowledgments. This work was supported by grants from the Ministry of Education, Science and Culture, Japan. We thank Hiroshi Yoshikawa and Takeshi Ogasawara for unpublished data of *M. luteus* dnaA codon usage.

References

- 1) Sueoka, N. (1962): Proc. Natl. Acad. Sci. USA., **48**, 582-592.
- 2) Jukes, T. H., and Bhushan, V. (1986): J. Mol. Evol., **24**, 39-44.
- 3) Muto, A., and Osawa, S. (1987): Proc. Natl. Acad. Sci. USA., **84**, 166-169.
- 4) Osawa, S. *et al.* (1988): *ibid.*, **85**, 1124-1128.
- 5) Ikemura, T. (1981): J. Mol. Biol., **51**, 389-409.
- 6) — (1982): *ibid.*, **158**, 573-579.
- 7) Zurawski, G., and Zurawski, S. M. (1985): Nucleic Acids Res., **13**, 4521-4526.
- 8) Cerretti, D. P. *et al.* (1983): *ibid.*, **11**, 2599-2616.
- 9) Ohama, T. *et al.* (1989): J. Mol. Evol. (in press).
- 10) Post, L. E., and Nomura, M. (1980): J. Biol. Chem., **255**, 4660-4666.
- 11) Ohama, T. *et al.* (1986): J. Bacteriol., **169**, 4770-4777.
- 12) Husain, I. *et al.* (1986): J. Biol. Chem., **261**, 4895-4901.
- 13) Shiota, S., and Nakayama, Y. (1989): Mol. Gen. Gent. (in press).
- 14) Arikan, K. *et al.* (1986): Nucleic Acids Res., **14**, 2637-2650.
- 15) Shiota, S., and Nakayama, Y. (1989): Mol. Gen. Gent., **213**, 21-29.
- 16) Hansen, F. G. *et al.* (1982): EMBO J., **1**, 1043-1048.
- 17) Brune, M. *et al.* (1985): Nucleic Acids Res., **13**, 2637-2650.
- 18) Sancar, A. *et al.* (1981): J. Mol. Biol., **148**, 63-76.
- 19) Muramatsu, T. *et al.* (1988): J. Biol. Chem., **263**, 9261-9267.
- 20) Aota, S. *et al.* (1988): Nucleic Acids Res., **16**, r315-r402.
- 21) Ohkubo, S. *et al.* (1987): Mol. Gen. Gent., **210**, 314-322.
- 22) Reddy, P. W. *et al.* (1985): Proc. Natl. Acad. Sci. USA., **82**, 5656-5660.