Volume 10 Number 22 1982 Nucleic Acids Research

Codon usage in bacteria: correlation with gene expressivity

M.Gouy and C.Gautier

Equipe Evolution moléculaire, Laboratoire de Biométrie, Université Lyon I, 43 Bd du 11 Nov 1918, 69622 Villeurbanne, France

Received 9 August 1982; Accepted 19 October 1982

ABSTRACT

The nucleic acid sequence bank now contains over 600 protein coding genes of which 107 are from prokaryotic organisms. Codon frequencies in each new prokaryotic gene are given. Analysis of genetic code usage in the 83 sequenced genes of the Escherichia coli genome (chromosome, transposons and plasmids) is presented, taking into account new data on gene expressivity and regulation as well as iso-tRNA specificity and cellular concentration. The codon composition of each gene is summarized using two indexes: one is based on the differential usage of iso-tRNA species during gene translation, the other on choice between Cytosine and Uracile for third base. A strong relationship between codon composition and mRNA expressivity is confirmed, even for genes transcribed in the same operon. The influence of codon use on peptide elongation rate and protein yield is discussed. Finally, the evolutionary aspect of codon selection in mRNA sequences is studied.

INTRODUCTION

Bacteria, and especially Escherichia coli, furnish the best documented example of the effect of natural selection on codon usage. The resulting bias in genetic code usage has two main components: correlation with tRNA availability in the cell and non random choices between pyrimidine ending codons. We present here two simple indexes to quantify these components and characterize all 107 bacterial sequences in the Lyon sequence bank ACNUC.

The E. coli sample (83 sequences) can now be considered as representative of the whole genome since it contains sequences from the chromosome as well as from plasmids and transposons. These sequences code for structural proteins (membrane and ribosomal proteins), for enzymes (amino acid biosynthesis operons) and for regulatory proteins (repressors, lexA and recA genes). Also, copious data on primary structure, codon recognition and relative abundance of E. coli tRNA species are available. Finally, regulating mechanisms and cellular concentrations of translation products are known for many genes in our sample, the latter, however, often on a qualitative basis. These data permit a more precise approach to the relationship between codon usage and translation efficiency and a somewhat new evolutionary point of view.

0305-1048/82/1022-7055$ 2.00/0

All genes studied here can be identified with Table 1. Sequence sources appear in the list of sequence references at the end of the paper. Table 2 shows the codon composition of 85 new genes not already presented in our last two compilations (1,2). As previously, codon compositions are expressed per 1000 and both initiator and stop codons are excluded from the data. E. coli ilvG gene (symbol ECOILVG) is given again because this nucleotide sequence has been redetermined and another reading frame has been proposed (3,4). Genes for elongation factor G (ECOEFG) and ribosomal protein S7 (ECOS7) are also re-included because more of their codons have been sequenced.

CODON USAGE AND tRNA AVAILABILITY

The first relationship between codon occurrences in mRNAs and cellular tRNA abundancies described in the literature is known as the tRNA adaptation theory (5,6). This phenomenon has been observed in many specialized cells mainly involved in the translation of very few mRNA species. This is not the case of E. coli cell. However, the high use of codons corresponding to major tRNAs was remarked by Post et al. upon sequencing ribosomal protein (r-protein) genes (7,8). Since then, various workers have observed the same tendency in many other highly expressed genes, i.e. genes coding for abundant proteins (see the r-proteins, elongation factors and membrane proteins items in the list of sequence references). Ikemura quantified the relative frequencies of tRNA species in E. coli and presented an elegant demonstration of the correlation between codon use in messengers and tRNA cellular quantities (9,10). The linear correlation coefficient computed on several genes discriminates between high and low expressivity genes with exceptions that seem mainly due to mathematical artifacts (10). Following the work of one of us (11,12) we construct here a measure of the influence of tRNA availability based on a model of protein synthesis dynamics

At the beginning of each polypeptide elongation cycle, a codon is found at the ribosomal A-site. The ternary complexes (aminoacyl-tRNAs bound with elongation factor Tu and GTP) diffusing in the cytoplasm interact with the codon and the ribosome at the A-site. Most often the codon does not belong to the tRNA recognition spectrum and therefore the aminoacyl-tRNA dissociates from the ribosome. When the specificity condition is fulfilled, the elongation cycle starts: transpeptidation and translocation occur. Thus, each codon can be characterized by the average number of codon-tRNA interactions at the A-site during one elongation cycle. The relative concentration of the codon-cognate tRNA is equivalent to its probability of colliding with the A-site

Table 1: mRNA sequence portfolio

Symbol	Species and gene	No Codons
ECOLPP	E. coli outer membrane lipoprotein	77
ECOOMPA	E. coli ompA gene	345
ECOS2	E. coli ribosomal protein S2	240
ECOS4	E. coli ribos. prot. S4 3' partial seq.	60
ECOS7	E. coli ribos. prot. S7 5' & 3' partial seq.	112
ECOS10	E. coli ribosomal protein S10	102
ECOS11	E. coli ribos. prot. S11 5' partial seq.	55
ECOS12	E. coli ribosomal protein S12	123
ECOS13	E. coli ribos. prot. S13 partial sequence	35
ECOS17	E. coli ribos. prot. S17 3' partial seq.	24
ECOS20	E. coli ribosomal protein S20 UUG initiator	86
ECOL1	E. coli ribosomal protein L1	233
ECOL3	E. coli ribos. prot. L3 5' partial seq.	79
ECO712	E. coli ribosomal proteins L7/L12	120
ECOL10	E. coli ribosomal protein L10	164
ECOL11	E. coli ribosomal protein L11	141
ECOL14	E. coli ribos. prot. L14 5' partial seq.	19
ECOTUFA	E. coli elong. factor Tu (tufA) GUG initiator	393
ECOTUFB	E. coli elong. factor Tu (tufB)	393
ECOEFG	E. coli elongation factor G 5' & 3' partial seq.	114
ECOEFTS	E. coli elongation factor Ts (tsf)	282
ECORECA	E. coli recA	352
ECORPOA	E. coli RNA polymerase α-subunit 5' partial seq.	158
ECORPOB	E. coli RNA polymerase β-subunit	1341
ECORPOC	E. coli RNA polym. β'-subunit 5' partial seq.	510
ECORPOD	E. coli RNA polymerase σ-subunit	612
ECOALRS	E. coli alanyl-tRNA-synthetase	875
ECOUNC1	E. coli hypoth. gene uncl GUG initiator	129
ECOUNC2	E. coli a-subunit of ATP-synthase (unc2)	270
ECOUNC3	E. coli c-subunit of ATP-synthase (unc3)	78
ECOUNC4	E. coli b-subunit of ATP-synthase (unc4)	155
ECOUNC5	E. coli δ-subunit of ATP-synthase (unc5)	176
ECOUNCA	E. coli α-subunit of ATP-synthase (uncA)	512
ECOUNCG	E. coli γ-subunit of ATP-synthase (uncG)	286
ECOUNCD	E. coli β-subunit of ATP-synthase (uncD)	459
ECOUNCC	E. coli ε-subunit of ATP-synthase (uncC)	132
ECOLACI	E. coli lacI lac operon repressor	359
ECOLACZ	E. coli lacZ β-galactosidase 3' partial seq.	24
ECOLACY	E. coli lacY lactose permease	416
ECOLACA	E. coli lacA transacetylase 5' partial seq.	25
ECOTRPL	E. coli trpL attenuation of trp operon	13
ECOTRPE	E. coli trpE anthranilate synthetase component I	519
ECOTRPD	E. coli trpD	530
ECOTRPC	E. coli trpC	451
ECOTRPB	E. coli trpB	397
ECOTRPA	E. coli trpA	267
ECOTRPR	E. coli trpR trp & aroH operons repressor	107
ECOILVL	E. coli ilvL attenuation of ilv operon	31
ECOILVG	E. coli ilvG (mutant Valine resistant)	520
ECOILVE	E. coli ilvE 5' partial sequence	80
ECOTHRL	E. coli thrL attenuation of thr operon	20
ECOTHRA	E. coli thrA	819
ECOPHEL	E. coli pheL attenuation of phe operon	14

Symbol	Species and gene	No Codons
ECOHISL	E. coli hisL attenuation of his operon	15
ECOAMPC	E. coli ampC cephalosporinidase	376
ECOARAC	E. coli araC ara operon regulation	291
ECOAROH	E. coli aroH 5' & 3' partial sequence	152
ECOASNA	E. coli asnA asparagine synthetase	329
ECOFOL	E. coli dihydrofolate reductase	158
ECOLEXA	E. coli lexA	201
ECOLTA	E. coli heat-labile toxin, A-subunit partial seq.	124
ECOLTB	E. coli heat-labile toxin, B-subunit	123
ECONDH	E. coli ndh NADH dehydrogenase	433
ECOPHOA	E. coli phoA alkaline phosphatase	76
ECOR1EN	E. coli restriction endonuclease R1	276
ECOR1ME	E. coli restriction methylase R1	325
ECOTNAA	E. coli tnaA tryptophanase	470
ECOUVRB	E. coli uvrB	28
ECOTN3L	E. coli transposon TN3 β-lactamase	285
ECOTN3T	E. coli transposon TN3 tnpA transposase	1014
ECOTN3R	E. coli transposon TN3 repressor	184
ECOTN9	E. coli transposon TN9 chloramphenicol resistance	218
EC0903K	E. coli transposon TN903 kanamycin resistance	270
ECPFOL	plasmid R388 dihydrofolate reductase	77
CLOIMMU	plasmid CLO-DF13 immunity gene	84
CLOH	plasmid CLO-DF13 gene H	48
R1PORI1	plasmid R1 hypoth. gene ori1 10500 d protein	85
R1PORI2	plasmid R1 hypoth. gene ori2 7000 d protein	60
SM1RA1	plasmid SM1 repA1 gene GUG initiator	284
SM1RA2	plasmid SM1 repA2 gene	102
SAULRER	plasmid PE194 erythromycin resist. regulator	18
SAURERY	plasmid PE194 erythromycin resistance	243
SAUBLAC	Staphylococcus aureus β-lactamase	46
STYTRPL	Salmonella typhimurium trpL attenuation trp operon	13
STYTRPE	Salmonella typhimurium trpE	519
STYTRPD	Salmonella typhimurium trpD 5' partial seq.	193
STYTRPB	Salmonella typhimurium trpB	396
STYTRPA	Salmonella typhimurium trpA	267
STYILVL	Salmonella typhimurium ilvL attenuation ilv operon	31
STYHISL	Salmonella typhimurium hisL attenuation his operon	15
STYHISJ	Salmonella typhimurium hisJ	259
STYLEUL	Salmonella typhimurium leuL attenuation leu operon	27
STYARGT	Salmonella typhimurium argT	259
SALHIN	Salmonella hin protein	189
SMALPP	Serratia marcescens outer membrane lipoprotein	76
SMATRPE	Serratia marcescens trpE 3' partial sequence	26
SMATRPG	Serratia marcescens trpG	192
SDYTRPE	Shigella dysenteriae trpE 3' partial sequence	26
SDYTRPD	Shigella dysenteriae trpD 5' partial sequence	193
EAMLPP	Erwinia amylovora outer membrane lipoprotein	77
KAETRPB	Klebsiella aerogenes trpB 3' partial sequence	19
KAETRPA	Klebsiella aerogenes trpA	268
KPNNIFH	Klebsiella pneumoniae nifH nitrogenase 2	292
ANANIFH	Anabaena nifH nitrogenase reductase	298
RHINIFH	Rhisobium melitoti nifH nitrogenase	296
BLIPEN	Bacillus licheniformis penicillinase	306
HALRHO	Halobacterium halobium bacteriorhodopsin	261

SEQUENCE REFERENCES (Table 1 order)

ECOLPP	Nakamura, K., Inouye, M.: Cell, 18, 1109-1117(1979)
ECOOMPA	Beck, E., Bremer, E.: Nucleic Acids Res., 8, 3011-3027(1980) Movva, N.R., Nakamura, K., Inouye, M.: Proc. Natl. Acad. Sci. USA, 77, 3845-3949(1980)
ECOS2	An, G., Bendiak, D.S., Mamelak, L.A., Friesen, J.D.: Nucleic Acid Res., 9,4162-4172(1981)
ECOS4	Post, L.E., Nomura, M.: J. Biol. Chem., 254, 10604-10606(1979)
ECOS7	Post, L.E., Nomura, M.: J. Biol. Chem. 255, 4660-4666(1980)
ECOS10	Olins, P.O., Nomura, M.: Cell, 26, 205-211(1981)
ECOS11	Post, L.E., Arfsten, A.E., Davis, G.R., Nomura, M.: J. Biol. Chem., 255, 4653-4659(1980)
ECOS12	same reference as ECOS7
ECOS13	same reference as ECOS11
ECOS17	Post, L.E., Arfsten, A.E., Reusser, F., Nomura, M.: Cell, 15, 215-229(1978)
ECOS20	Mackie, G.A.:J. Biol. Chem., 256, 8177-8182(1981)
ECOL1	Post, L.E., Strycharz, G.D., Nomura, M., Lewis, H., Dennis, P.P.: Proc. Natl. Acad. Sci. USA, 76, 1697-1701(1979)
ECOL3	same reference as ECOS10
ECO712	same reference as ECOL1
ECOL10	same reference as ECOL1
ECOL11	same reference as ECOL1
ECOL14	same reference as ECOS17
ECOTUFA	Yokota, T., Sugisaki, H., Takanami, M., Kaziro, Y.: Gene, 12, 25-31(1980)
ECOTUFB	An, G., Friesen, J.D.: Gene, 12, 33-39(1980)
ECOEFG	Post, L.E., Nomura, M.:J. Biol. Chem., 255, 4660-4666(1980) Yokota, T., Sugisaki, H., Takanami, M., Kaziro, Y.: Gene, 12, 25-31(1980)
ECOEFTS	same reference as ECOS2
ECORECA	Sancar, A., Stachelek, C., Konigsberg, W., Rupp, W.D.:Proc. Natl. Acad. Sci. USA,77,2611-2615(1980)
ECORPOA	same reference as ECOS4
ECORPOB	Ovchinnikov, Y.A., Monastyrskaya, G.S., Gubanov, V.V., Guryev, S.O., Chertov, O.Y., Modyanov, N.N., Grinkevich, V.A., Makarova, I.A., Marchenko,T.V., Polovnikova, I.N., Lipkin, V.M., Sverdlov,E.D.: Eur. J. Biochem., 116, 621-629(1981)
ECORPOC	same reference as ECORPOA Squires, C., Krainer, A., Barry, G., Shen, W.F., Squires, C.L.: Nucleic Acids Res., 9, 6827-6840 (1981)
ECORPOD	Burton, Z., Burgess, R.R., Lin, J., Moore, D., Holder, S., Gross, C.A.: Nucleic Acids Res.,9, 2889-2903(1981)
ECOALRS	Putney, S.D., Royal, N.J., De Vegvar, H.N., Herlihy, W.C., Biemann, K., Schimmel, P.:Science, 213, 1497-1500(1981)

ECOUNC1	Gay, N.J., Walker, J.E.: Nucleic Acids Res., 9, 3919-3926(1981)
ECOUNC2	same reference as ECOUNC1
ECOUNC3	same reference as ECOUNC1
ECOUNC4	same reference as ECOUNC1
ECOUNC5	same reference as ECOUNC1
ECOUNCA	Gay, N.J., Walker, J.E.: Nucleic Acids Res., 9, 2187-2194(1981)
ECOUNCG	Saraste, M., Gay, N.J., Eberle, A., Runswick, M.J., Walker, J.E.: Nucleic Acids Res., 9, 5287-5296(1981)
ECOUNCD	same reference as ECOUNCG
ECOUNCC	same reference as ECOUNCG
ECOLACI	Farabaugh, P.: Nature,274,765-769(1978)
ECOLACZ	Buchel, D.E., Gronenborn, B., Muller-Hill, B.:Nature, 283,541-545(1980)
	Maizels, N.:Proc. Natl. Acad. Sci. USA, 70, 3585-3592(1973)
ECOLACY	Buchel, D.E., Gronenborn, B., Muller-Hill, B.: Nature, 283, 541-545(1980)
ECOLACA	same reference as ECOLACY
ECOTRPL	Platt, T.: Cell, 24, 10-23(1981)
ECOTRPE	Nichols, B.P., Van Cleemput, M., Yanofsky, C.: J. Mol. Biol., 146,45-54(1981)
ECOTRPD	Nichols, B.P., Miozzari, G.F., Van Cleemput, M., Bennett, G.N., Yanofsky, C.: J. Mol. Biol., 142, 503-517(1980)
	Horowitz, H., Christie, G.E., Platt, T.: J. Mol. Biol., 156, 245-256(1982)
ECOTRPC	Christie, G.E., Platt, T.: J. Mol. Biol., 142,519-530, (1980)
ECOTRPB	Crawford, I.P., Nichols, B.P., Yanofsky, C.: J. Mol. Biol., 142, 489-502(1980)
ECOTRPA	Nichols, B.P., Yanofsky, C.: Proc. Natl. Acad. Sci. USA, 76, 5244-5248(1979)
ECOTRPR	Gunsalus, R.P. Yanofsky, C.: Proc. Natl. Acad. USA, 77, 7117-7121(1980)
ECOILVL	Nargang, F.E., Subrahmanyam, C.S., Umbarger, H.E.: Proc. Natl. Acad. Sci. USA,77,1823-1827(1980)
	Lawther, R.P., Calhoun, D.H., Adams, C.W.,Hauser,C.A., Gray, J., Hatfield, G.W.: Proc. Natl. Acad. Sci. USA, 78, 922-925(1981)
ECOILVG	Lawther, R.P., Calhoun, D.H., Adams, C.W., Hauser, C.A., Gray, J., Hatfield, G.W.: Proc. Natl. Acad. Sci. USA, 78, 922-925(1981)
ECOILVE	Lawther, R.P., Nichols, B., Zurawski, G., Hatfield, G.W.: Nucleic Acids Res., 7, 2289-2301(1979)
ECOTHRL	Gardner, J.F.: Proc. Natl. Acad. Sci. USA, 76, 1706-1710(1979)
ECOTHRA	Katinka, M., Cossart, P., Sibilli, L., Saint-Girons, I., Chalvignac, M.A., Le Bras, G., Cohen, G.N., Yaniv, M.:Proc. Natl. Acad. Sci. USA,77, 5730-5733(1980)
ECOPHEL	Zurawski, G., Brown, K., Killingly, D., Yanofsky, C.: Proc. Natl. Acad. Sci. USA,75,4271-4275(1978)

ECOHISL	Di Nocera, P.P., Blasi, F., Di Lauro, R., Frunzio, R. Bruni, C.B.: Proc. Nat. Acad. Sci. USA, 75, 4276-4280(1878)
ECOAMPC	Jaurin, B., Grundstrom, T.: Proc. Natl. Acad. Sci. USA, 78, 4897-4901(1981)
ECOARAC	Miyada, C.G., Horwitz, A.H., Cass, L.G., Timko, J., Wilcox, G.: Nucleic Acids Res.,8,5267-5274(1980) Wallace, R.G., Lee, N., Fowler, A.V.: Gene, 12, 179-190(1980) Stoner, C.M., Schleif, R.:J. Mol. Biol., 154, 649-652(1982)
ECOAROH	Zurawski, G., Gunsalus, R.P., Brown, K.D., Yanofsky, C.:J. Mol. Biol.,145,47-73(1981)
ECOASNA	Nakamura, M., Yamada, M., Hirota, Y., Sugimoto, K., Oka, A., Takanami, M.:Nucleic Acids Res., 9, 4669-4676(1981)
ECOFOL	Smith, D.R., Calvo, J.M.: Nucleic Acids Res., 8, 2255-2274(1980)
ECOLEXA	Miki, T., Ebina, Y., Kishi, F., Nakazawa, A.: Nucleic Acids Res., 9,529-543(1981) Horii, T., Ogawa, T., Ogawa, H.: Cell,23,689-697(1981) Markham, B.E., Little, J.W., Mount, D.W.: Nucleic Acids Res., 9, 4149-4161(1981)
ECOLTA	Spicer, E.K., Kavanaugh, W.M., Dallas, W.S., Falkow, S., Konigsberg, W.H., Schafer, D.E.: Proc. Natl. Acad. Sci. USA, 78, 50-54(1981)
ECOLTB	Dallas, W.S., Falkow, S.: Nature, 288,499-501(1980)
ECONDH	Young, I.G., Rogers, B.L., Campbell, H.D., Jaworowski, A., Shaw, D.C.: Eur. J. Biochem., 116, 165-170(1981)
ECOPHOA	Kikuchi, Y., Yoda, K., Yamasaki, M., Tamura, G.: Nucleic Acids Res., 9, 5671-5678(1981)
ECOR1EN	Greene, P.J., Gupta, M., Boyer, H.W., Brown, W.E., Rosenberg, J.M.: J. Biol. Chem. , 256, 2143-2153 (1981) Newman, A.K., Rubin, R.A., Kim, S.H., Modrich, P.: J. Biol. Chem., 256, 2131-2139(1981)
ECOR1ME	same reference as ECOR1EN
ECOTNAA	Deeley, M.C., Yanofsky, C.:J. Bact., 147,787-796(1981)
ECOUVRB	Van Den Berg, E., Zwetsloot, J., Noordermeer, I., Pannekoek, H., Dekker, B., Dijkema, R., Van Ormondt, H.: Nucleic Acids Res.,9,5623-5643(1981)
ECOTN3L	Heffron, F., Mc Carthy, B.J., Ohtsubo, H., Ohtsubo, E.: Cell, 18, 1153-1163(1979) Sutcliffe, G.: Proc. Natl. Acad. Sci. USA, 75, 3737-3741(1978)
ECOTN3R	Heffron, F., Mc Carthy, B.J., Ohtsubo, H., Ohtsubo, E.:Cell,18, 1153-1163(1979) Chou, J., Lemaux, P.G., Casabadan, M.J., Cohen, S.N.: Nature, 282, 801-806(1979)
ECOTN3T	same reference as ECOTN3R
ECOTN9	Alton, N.K., Vapnek, D.: Nature, 282, 864-869(1979) Marcoli, R., Iida, S., Bickle, T.A.: FEBS Lett., 110, 11-14(1980)

ECO9O3K	Oka, A., Sugisaki, H., Takanami, M.:J. Mol. Biol., 147, 217-226(1981) Heidekamp, F., Baas, P.D., Van Boom, J.H., Veeneman, G.H., Zipursky, S.L., Jansz, H.S.: Nucleic Acids Res., 9, 3335-3354(1981)
ECPFOL	Zolg, J.W., Hanggi, U.J.: Nucleic Acids Res.,9, 698-709(1981) Swift, G., Maccarthy, B.J., Heffron, F.: Mol. Gen. Genet., 181,441-447(1981)
CLOIMMU	Elzen, P.J.M., Gaastra, W., Spelt, C.E., De Graaf, F.K., Veltkamp,E., Nijkamp, H.J.J.: Nucleic Acids Res.,8,4349-4363(1980)
CLOH	Stuitje, A.R., Spelt, C.E., Veltkamp, E., Nijkamp, H.J.J.:Nature,290,264-267(1981)
R1PORI1	Stougaard, P., Molin, S., Nordstrom, K., Hansen, F.G.: Mol. Gen. Genet.,181,116-122(1981)
R1PORI2	same reference as R1PORI1 Stougaard, P., Molin, S., Nordstrom, K.: Proc. Natl. Acad. Sci. USA,78,6008-6012(1981)
SM1RA1	Rosen,J., Ryder, T., Inokuchi, H., Ohtsubo, H., Ohtsubo,E.: Molec. Gen. Genet.,179,527-537(1980) Rosen, J., Ryder, T., Ohtsubo,H., Ohtsubo, E.:Nature, 290,794-797(1981)
SM1RA2	same reference as SM1RA1
SAULRER	Horinouchi, S., Weisblum, B.: Proc. Natl. Acad. Sci. USA,77,7079-7083(1980)
SAURERY	same reference as SAULRER
SAUBLAC	Mc Laughlin, J.R., Murray, C.L., Rabinowitz, J.C.: J. Biol. Chem.,256,11283-11291(1981)
STYTRPL	Yanofsky, C., Van Cleemput, M.: J. Mol. Biol.,154, 235-246(1982)
STYTRPE	same reference as STYTRPL
STYTRPD	Nichols, B.P., Miozzari, G.F., Van Cleemput, M., Bennett, G.N., Yanofsky, C.: J. Mol. Biol., 142, 503-517(1980)
STYTRPB	same reference as ECOTRPB
STYTRPA	same reference as ECOTRPA
STYILVL	Taillon, M.P., Gotto, D.A., Lawther, R.P.:Nucleic Acids Res.,9,3419-3432(1981)
STYHISL	Barnes, W.M.:Proc. Natl. Acad. Sci. USA,75,4281-4285 (1978)
STYHISJ	Higgins, C.F., Ames, G.F.L.: Proc. Natl. Acad. Sci. USA,78,6038-6042(1981)
STYLEUL	Gemmill, R.M., Wessler, S.R., Keller, E.B., Calvo, J.M.: Proc. Natl. Acad. Sci. USA,76,4941-4945(1979)
STYARGT	same reference as STYHISJ
SALHIN	Zieg, J., Simon, M.: Proc. Natl. Acad. Sci. USA,77, 4196-4200(1980)
SMALPP	Nakamura, K., Inouye, M.: Proc. Natl. Acad. Sci. USA, 77,1369-1373(1980)
SMATRPE	same reference as STYTRPD
SMATRPG	same reference as STYTRPD
SDYTRPE	same reference as STYTRPD
SDYTRPD	same reference as STYTRPD
EAMLPP	Yamagata, H., Nakamura, K., Inouye, M.: J. Biol. Chem.,256,2194-2198(1981)

KAETRPB	Nichols, B.P., Blumenberg, M., Yanofsky, C.: Nucleic Acids Res., 9,1743-1755(1981)
KAETRPA	same reference as KAETRPB
KPNNIFH	Sundaresan, V., Ausubel, F.M.: J. Biol. Chem., 256, 2808-2812(1981)
ANANIFH	Mevarech, M., Rice, D., Haselkorn, R.: Proc. Natl. Acad. Sci. USA,77,6476-6480(1980)
RHINIFH	Torok, I., Kondorosi, A.: Nucleic Acids Res.,9, 5711-5723(1981)
BLIPEN	Neugebauer, K., Sprengel, R., Schaller, H.: Nucleic Acids Res.,9,2577-2588(1981)
HALRHO	Dunn, R., Mc Coy, J., Simsek, M., Majumdar, A., Chang, S.H., Rajbhandary, U.L., Khorana, H.G.: Proc. Natl. Acad. Sci. USA,78,6744-6748(1981)

codon. Hence, if this probability is f, the mean number of codon-tRNA interactions necessary for the elongation cycle to occur is 1/f. The rarest tRNA species in E. coli is tRNA$_2^{Ile}$ decoding codon AUA (0.3% of the total tRNA population, (9) and legend to fig. 1); the most abundant species is tRNA$_3^{Gly}$ decoding GGPy codons (6.5% of the tRNA population). Consequently, 1/0.003 = 333 non specific codon-tRNA interactions are necessary, on the average, to translate the AUA codon, whereas an average of 1/0.065 = 15 such interactions occur when translating GGU or GGC codons. For a given gene, we weight the mean number of non specific codon-tRNA interactions for each codon by the relative frequency of the codon in the sequence. The value thus obtained is the average number of tRNA discriminations per elongation cycle (P1 index). This index is scaled on the abcissa of fig. 1.

All highly expressed genes appear to the left of the figure. These are genes coding for major membrane proteins, for the 3 elongation factors Tu, Ts and G, for r-proteins, and the recA gene, which is not constitutively highly expressed as are the preceding messengers. These genes are therefore highly optimized for a small number of tRNA discriminations. We calculate that the best theoretical messenger for the L1 r-protein (ECOL1) would require 22.3 mean discriminations instead of the observed value of 25.5. Translation of a codon whose cognate tRNA is rare in the cell requires more non specific codon-tRNA interactions at the A-site than do codons decoded by major tRNAs. Consequently, translation of the latter codons might be faster and/or less error-prone than translation of other synonymous codons. A quantitative assessment of this phenomenon is not yet possible since the duration of a non specific codon-tRNA interaction is unknown (11). Therefore, given the available data, P1 is the best possible index for quantifying the effect of codon use on translation rate.

		ECO	ECO	ECO	ECO	ECO	ECO	ECO	ECO	ECO	ECO	ECO	ECO	ECO	ECO	ECO	ECO	ECO	ECO	ECO	ECO	ECO	ECO	ECO
		S2	S4	S7	S10	S17	S20	L3	L14	TUFA	TUFB	EFG	EFTS	RPOA	RPOB	RPOC	RPOD	ALRS	UNC1	UNC2	UNC3	UNC4	UNC5	UNCA
ARG	CGA	0	0	0	0	0	0	0	0	0	0	0	0	6	1	2	0	0	8	4	0	6	6	2
	CGC	21	17	45	49	83	23	13	53	5	8	18	14	19	21	24	29	27	0	0	0	13	28	10
	CGG	0	0	9	0	0	0	0	0	0	0	0	0	0	0	4	0	1	16	0	0	0	6	0
	CGU	33	50	36	59	42	58	51	53	53	51	44	18	25	45	47	46	35	8	11	26	45	23	47
	AGA	0	0	0	10	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0
	AGG	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
LEU	CUA	0	0	0	0	0	0	0	0	0	0	0	0	13	0	0	2	0	0	0	0	0	0	2
	CUC	4	0	9	0	0	0	0	0	0	0	0	0	13	11	4	5	5	23	4	13	6	23	8
	CUG	75	83	63	88	125	35	25	53	69	69	18	57	57	74	76	69	61	78	130	141	45	45	74
	CUU	8	17	9	0	0	0	0	0	3	3	0	0	13	4	2	8	7	8	7	0	32	28	6
	UUA	0	0	18	0	0	0	13	0	0	0	0	0	6	1	6	2	6	23	11	0	6	0	0
	UUG	0	0	9	10	0	0	0	0	0	0	0	0	6	4	2	3	9	47	15	0	0	6	4
SER	UCA	0	0	9	0	0	12	0	0	0	0	0	0	6	0	6	3	3	0	11	0	0	6	0
	UCC	13	17	18	10	83	0	0	53	8	8	9	14	0	23	20	11	9	0	7	0	13	6	31
	UCG	0	0	0	0	0	0	0	0	0	0	0	0	13	2	12	3	5	8	11	0	6	6	2
	UCU	17	50	45	0	0	23	13	0	18	18	9	14	25	17	14	15	14	16	11	0	0	28	21
	AGC	17	0	9	10	0	12	0	0	0	3	0	7	13	11	10	13	22	16	22	0	13	6	4
	AGU	0	0	27	0	0	0	0	0	0	0	18	0	6	1	0	2	5	16	0	0	0	17	4
THR	ACA	0	0	0	10	0	0	13	0	3	5	18	0	6	2	0	0	3	16	7	0	6	0	4
	ACC	25	17	9	29	0	0	63	0	41	38	53	14	44	25	29	47	21	0	11	0	6	11	33
	ACG	0	17	0	0	42	0	0	0	0	3	0	0	6	4	2	11	11	8	26	0	0	11	0
	ACU	29	0	18	39	42	23	13	53	33	31	44	25	13	13	16	3	14	0	0	13	6	0	8
PRO	CCA	0	17	9	10	0	0	13	0	3	3	0	7	0	7	10	2	5	8	15	0	6	6	6
	CCC	0	0	9	0	0	0	0	0	0	0	9	0	0	0	0	0	1	8	4	0	0	6	0
	CCG	29	17	27	39	42	12	0	0	48	48	26	11	25	28	33	28	16	16	26	13	6	6	27
	CCU	0	0	0	0	0	0	13	0	0	0	0	0	13	6	2	2	3	8	11	26	0	0	2
ALA	GCA	21	0	18	10	0	70	13	53	13	15	26	50	13	16	25	13	22	39	7	26	39	34	35
	GCC	0	17	0	20	0	12	0	53	3	3	26	0	13	7	6	16	18	16	22	13	52	40	16
	GCG	17	33	27	39	42	0	0	0	20	23	26	28	19	22	33	33	40	39	19	51	52	40	23
	GCU	67	50	54	20	0	140	101	0	33	28	44	60	6	14	18	18	24	23	15	77	58	40	31
GLY	GGA	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	5	8	4	0	0	0	2
	GGC	21	17	18	0	0	12	63	0	53	51	9	53	38	26	25	20	37	31	22	26	19	17	31
	GGG	0	0	0	0	0	0	0	0	3	3	9	0	6	1	2	5	7	23	4	13	0	0	2
	GGU	50	17	36	39	0	12	63	53	48	48	61	32	44	51	49	15	43	0	33	90	0	28	57
VAL	GUA	21	0	45	10	42	12	38	53	25	28	18	18	6	23	24	5	15	16	22	13	13	34	18
	GUC	0	17	9	29	0	0	13	53	0	3	0	0	13	10	14	11	13	16	11	13	13	23	4
	GUG	13	17	9	10	0	12	38	0	8	8	0	14	51	18	14	16	22	85	33	38	19	6	12
	GUU	38	17	18	39	125	0	51	0	61	56	26	71	38	31	31	23	27	62	19	13	19	28	47
LYS	AAA	58	33	45	39	83	93	76	0	46	43	26	82	38	42	41	34	38	39	26	13	65	34	37
	AAG	21	117	36	18	42	70	13	0	13	15	18	18	13	18	24	21	16	16	11	0	32	11	10
ASN	AAC	50	17	18	20	0	58	38	105	18	18	35	28	19	36	25	29	24	39	22	13	19	23	27
	AAU	0	0	18	0	0	12	0	0	0	0	0	0	6	2	0	3	14	0	22	13	6	11	4
GLN	CAA	0	0	0	20	0	12	0	53	0	0	0	0	6	6	2	11	8	0	7	13	13	23	10
	CAG	29	33	36	39	0	47	25	53	20	20	35	35	32	37	35	38	29	23	19	13	52	28	47
HIS	CAC	21	17	18	20	0	23	13	0	25	23	18	11	25	13	14	8	16	8	15	0	0	11	14
	CAU	4	0	0	10	0	12	0	0	3	5	26	4	13	1	2	7	9	8	11	0	6	0	0
GLU	GAA	50	67	63	20	42	35	63	53	76	76	53	85	32	66	63	95	51	8	19	26	84	51	53
	GAG	17	83	9	39	42	0	0	0	18	18	26	28	51	25	22	21	22	0	11	0	32	45	8
ASP	GAC	54	17	0	29	0	23	25	53	51	51	35	43	6	45	27	39	33	8	19	0	39	28	31
	GAU	17	33	18	49	0	0	13	0	10	10	9	7	44	22	27	47	35	0	11	38	13	23	23
TYR	UAC	8	17	27	10	0	12	13	0	20	23	18	7	6	22	24	15	11	0	15	26	6	6	21
	UAU	0	0	0	0	0	0	0	0	5	3	18	4	6	10	4	7	10	0	7	0	0	0	8
CYS	UGC	4	0	0	0	42	0	0	0	5	5	0	7	13	1	10	5	3	0	0	0	6	6	0
	UGU	0	0	0	0	0	0	0	0	3	3	0	0	0	4	4	0	3	0	0	0	0	6	6
PHE	UUC	42	0	27	10	0	12	25	0	33	33	26	32	0	25	22	18	29	39	63	38	13	0	25
	UUU	8	17	18	10	0	12	0	0	3	3	0	4	13	8	10	7	13	31	11	13	6	34	2
ILE	AUA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	31	4	0	0	0	0
	AUC	46	17	18	88	42	70	38	53	66	66	79	50	44	49	45	52	27	8	44	90	65	23	59
	AUU	13	17	18	0	0	12	25	0	8	8	26	18	38	13	14	18	19	8	41	13	19	34	20
MET	AUG	29	17	9	10	0	35	13	105	25	25	26	32	13	27	16	39	23	23	41	90	13	34	21
TRP	UGG	13	17	9	0	42	0	13	0	3	3	18	0	0	3	10	7	9	31	26	0	6	6	2

		ECO UNCG	ECO UNCD	ECO UNCC	ECO LACZ	ECO LACA	ECO TRPL	ECO TRPE	ECO TRPD	ECO TRPC	ECO TRPB	ECO TRPR	ECO ILVL	ECO ILVG	ECO THRL	ECO THRA	ECO PHEL	ECO HISL	ECO AMPC	ECO ARAC	ECO AROH	ECO ASNA	ECO LEXA	ECO LTA
ARG	CGA	0	0	8	0	80	0	2	2	2	3	0	65	8	0	4	0	0	3	10	0	0	0	0
	CGC	14	11	15	42	0	77	40	30	29	23	37	0	8	50	23	0	67	19	34	33	21	25	8
	CGG	0	0	8	0	0	0	0	0	2	5	0	0	6	0	6	0	0	0	14	7	3	0	0
	CGU	31	46	8	0	0	0	33	13	18	13	47	0	15	0	22	0	0	5	17	26	33	45	16
	AGA	0	0	0	0	80	0	2	2	0	3	0	0	0	0	0	0	0	3	0	7	0	0	81
	AGG	0	0	0	0	40	0	0	0	2	0	0	0	0	0	2	0	0	0	0	7	0	5	8
LEU	CUA	0	0	0	0	40	0	8	4	4	5	0	32	2	0	4	0	0	5	0	7	0	0	8
	CUC	7	7	15	0	0	0	19	15	4	5	19	0	8	0	16	0	0	11	17	26	0	5	8
	CUG	66	76	61	83	0	77	67	70	47	60	75	32	50	0	53	0	0	27	45	39	85	60	8
	CUU	7	7	8	0	40	0	12	8	13	10	9	65	8	0	10	0	0	16	14	0	21	10	16
	UUA	3	0	8	0	0	0	13	17	22	5	37	0	31	0	12	0	0	19	14	13	6	10	32
	UUG	17	2	0	42	0	0	10	11	18	13	37	0	17	0	20	0	0	13	17	0	9	25	0
SER	UCA	3	4	0	42	0	0	8	4	9	0	19	0	8	0	7	0	0	5	10	0	6	5	32
	UCC	17	31	0	0	0	77	10	9	2	18	0	0	10	0	13	0	0	8	0	7	6	15	0
	UCG	17	0	8	0	0	0	10	6	18	8	0	0	6	0	11	0	0	8	21	0	9	5	8
	UCU	7	15	30	0	0	0	12	8	2	8	9	0	4	0	12	0	0	8	0	13	18	0	8
	AGC	24	4	30	42	0	0	25	19	16	8	28	65	10	50	15	0	0	16	34	13	27	20	8
	AGU	3	0	8	0	0	0	8	13	7	5	0	0	6	0	4	0	0	8	0	13	0	0	8
THR	ACA	0	0	0	0	0	0	8	9	2	8	0	32	4	50	4	0	67	24	0	0	0	5	32
	ACC	35	39	23	42	80	0	27	28	11	33	0	0	38	350	22	71	0	21	17	26	18	10	24
	ACG	3	2	8	42	0	0	6	8	16	10	28	0	4	0	7	0	0	16	7	33	12	10	0
	ACU	7	11	8	0	0	77	8	2	4	8	9	0	10	0	5	0	0	11	0	20	6	5	24
PRO	CCA	0	0	0	0	40	0	10	4	4	10	9	32	12	0	2	0	0	16	7	0	3	20	24
	CCC	3	0	0	0	0	0	12	6	4	3	19	0	4	0	6	71	0	19	3	13	3	0	16
	CCG	17	35	8	0	0	0	29	43	20	23	9	32	37	0	22	71	0	16	45	20	30	20	0
	CCU	10	9	23	0	0	0	6	2	2	10	0	0	10	0	5	0	67	19	3	13	6	5	0
ALA	GCA	21	17	30	0	40	77	12	19	31	8	28	32	35	0	18	71	0	32	7	7	9	20	32
	GCC	42	13	30	42	0	0	35	36	40	43	28	32	27	0	44	0	0	24	34	39	15	10	0
	GCG	24	28	45	0	0	0	31	58	35	38	37	0	31	50	33	0	0	24	21	33	43	35	8
	GCU	7	15	15	0	0	0	23	11	22	15	0	65	17	0	17	0	0	27	10	7	15	15	8
GLY	GGA	0	4	0	0	40	0	6	9	7	5	9	65	2	0	11	0	0	8	14	13	6	0	24
	GGC	42	31	45	0	40	0	23	34	27	43	28	0	37	50	27	0	0	29	0	39	43	25	16
	GGG	7	0	8	0	0	0	4	15	13	15	9	32	12	0	12	0	0	13	27	0	21	30	0
	GGU	21	63	38	42	0	77	21	28	22	45	0	0	31	100	27	0	0	27	31	13	18	30	32
VAL	GUA	7	41	15	0	0	77	15	6	11	0	0	0	6	0	6	0	0	16	10	20	24	10	0
	GUC	21	13	23	42	0	0	12	15	16	8	28	32	15	0	22	0	0	13	14	26	12	30	16
	GUG	31	22	23	0	40	0	27	28	33	25	19	161	25	0	31	0	0	27	14	13	21	15	0
	GUU	28	26	8	0	0	0	10	13	24	15	0	0	27	0	23	0	67	3	14	7	15	40	8
LYS	AAA	59	37	30	42	40	154	23	21	31	40	28	0	27	50	27	71	67	40	17	26	36	40	24
	AAG	21	7	15	0	80	0	6	6	9	5	9	32	6	0	15	0	0	19	7	7	9	15	0
ASN	AAC	31	20	8	0	40	0	13	28	20	30	28	0	19	50	22	0	0	21	14	13	9	15	0
	AAU	10	4	0	0	0	0	19	15	20	5	19	0	10	0	27	0	0	21	34	13	0	15	56
GLN	CAA	7	4	15	42	0	0	15	23	22	10	28	0	25	0	13	0	67	37	21	26	21	15	32
	CAG	42	39	45	83	0	0	25	26	40	30	37	0	37	0	22	0	0	29	31	33	40	50	8
HIS	CAC	10	15	45	0	0	0	8	21	13	20	9	0	15	0	7	71	200	0	14	20	21	5	8
	CAU	3	0	0	42	0	0	13	23	11	15	9	0	17	0	10	0	267	13	27	20	18	20	24
GLU	GAA	35	59	61	0	40	0	58	42	27	60	56	0	33	0	46	0	0	24	31	46	40	65	16
	GAG	17	26	30	0	0	0	10	13	24	15	56	0	15	0	18	0	0	16	31	26	30	15	40
ASP	GAC	28	46	30	0	0	0	27	17	24	13	9	0	29	0	16	0	67	19	17	13	27	20	8
	GAU	31	13	15	42	0	0	40	23	40	30	28	0	35	0	35	0	0	19	31	53	46	40	40
TYR	UAC	24	24	23	83	40	0	10	9	11	10	9	0	13	0	10	0	0	13	17	7	6	0	16
	UAU	14	9	8	0	40	0	17	15	24	20	9	0	10	0	15	0	0	27	17	13	9	5	32
CYS	UGC	7	0	0	0	0	0	13	2	11	8	0	32	8	0	11	0	0	5	10	13	0	0	0
	UGU	0	2	0	42	0	0	10	8	4	5	0	0	10	0	4	0	0	0	7	20	6	0	16
PHE	UUC	10	24	8	42	0	77	23	13	16	15	0	0	17	0	23	286	0	11	21	20	21	20	40
	UUU	3	7	8	0	40	0	15	15	20	18	9	0	15	0	13	214	67	19	38	7	9	10	24
ILE	AUA	3	0	0	0	40	0	0	2	0	0	0	0	2	0	2	71	0	5	7	0	3	0	32
	AUC	21	50	38	0	0	0	10	11	27	35	9	32	37	50	20	0	0	24	21	20	21	20	0
	AUU	21	9	38	42	0	77	23	38	20	23	19	129	0	150	35	0	0	32	27	33	18	45	40
MET	AUG	42	31	30	42	80	0	21	23	9	38	28	0	35	0	26	0	0	16	14	13	15	20	16
TRP	UGG	7	2	0	42	0	154	0	2	7	3	19	0	12	0	5	0	0	35	21	7	24	5	8

		ECO LTB	ECO NDH	ECO PHOA	ECO RIEN	ECO RIME	ECO TNAA	ECO UVRB	ECO 993K	ECP FOL	CLO IMMU	CLO H	RIP ORI1	RIP ORI2	SMI RA1	SMI RA2	SAU LRER	SAU RERY	SAU BLAC	STY TRPL	STY TRPE	STY TRPD	STY TRPB	STY ILVL
ARG	CGA	0	5	0	7	3	0	36	19	0	0	0	0	33	7	10	0	0	0	0	4	10	3	65
	CGC	8	21	13	0	3	21	0	11	26	0	0	12	17	39	0	0	8	0	77	42	16	39	0
	CGG	0	0	26	0	0	0	0	11	0	12	21	12	17	18	0	0	0	0	0	10	0	3	0
	CGU	0	25	26	4	0	28	36	11	13	0	0	12	0	74	10	0	0	0	0	21	21	13	0
	AGA	16	0	0	22	31	2	0	4	0	0	0	47	0	0	39	0	25	0	0	2	0	0	0
	AGG	8	0	0	18	3	0	0	4	0	0	0	0	17	4	0	0	4	0	0	0	0	0	0
LEU	CUA	33	5	0	14	15	2	0	7	0	0	0	12	0	0	0	0	0	0	0	10	26	8	32
	CUC	0	5	13	11	3	4	36	15	13	0	0	12	0	11	10	0	0	0	0	6	16	13	0
	CUG	8	61	66	0	9	45	71	15	13	12	42	35	0	74	39	0	0	0	0	64	73	56	32
	CUU	0	7	13	22	15	6	0	22	26	12	42	12	67	11	29	0	21	0	0	6	5	8	65
	UUA	33	5	39	40	28	15	0	22	13	36	0	0	0	0	20	0	62	152	77	17	10	5	0
	UUG	0	16	13	11	15	4	0	26	0	0	21	12	17	0	10	0	8	0	0	13	0	8	0
SER	UCA	16	5	0	25	0	4	0	26	13	12	21	12	0	11	20	0	25	43	0	8	0	3	0
	UCC	16	9	0	4	0	6	36	4	26	12	42	12	17	18	20	0	8	0	77	8	16	23	0
	UCG	16	12	0	7	12	0	0	7	0	12	21	12	0	4	10	0	0	0	0	8	10	3	0
	UCU	8	5	13	29	25	6	36	7	13	12	21	12	17	7	20	0	8	0	0	8	0	5	0
	AGC	8	12	26	14	12	19	0	11	13	0	0	0	67	7	0	56	12	0	0	27	10	13	65
	AGU	16	5	0	7	12	4	36	4	26	24	0	12	17	4	29	56	21	65	0	6	5	3	0
THR	ACA	49	5	26	14	15	9	0	11	13	0	0	35	0	14	20	56	16	0	77	4	16	8	32
	ACC	24	21	13	0	0	36	0	7	13	12	0	0	0	25	10	0	4	0	0	31	16	28	0
	ACG	24	21	13	4	3	9	0	7	0	0	42	0	0	28	0	0	4	0	0	17	10	16	0
	ACU	8	14	39	18	9	6	0	11	13	0	0	24	17	11	20	0	12	22	77	4	5	0	0
PRO	CCA	0	12	13	4	15	4	36	7	13	24	21	0	17	7	10	56	4	0	0	12	0	3	32
	CCC	16	0	13	0	0	0	0	7	26	0	0	0	17	7	0	0	4	0	0	13	16	10	0
	CCG	8	25	13	0	0	36	36	11	0	0	0	0	33	18	10	0	0	0	0	35	31	25	32
	CCU	0	2	39	18	18	2	36	30	26	12	0	24	0	0	0	0	16	0	0	4	10	8	0
ALA	GCA	24	12	53	14	9	23	0	15	26	0	63	24	17	11	59	0	12	22	154	6	10	5	32
	GCC	8	16	26	4	3	21	0	7	52	0	0	12	0	49	20	0	4	22	0	48	26	35	32
	GCG	16	44	0	4	6	36	36	22	26	0	21	0	17	35	10	0	0	0	77	46	62	53	0
	GCU	8	16	53	33	12	11	36	11	39	0	21	24	17	14	29	0	8	65	0	13	10	15	65
GLY	GGA	8	9	0	18	9	4	0	7	0	0	21	0	0	18	0	0	8	0	0	8	5	13	65
	GGC	24	39	39	4	9	49	36	7	39	12	21	0	17	7	10	56	4	0	0	33	41	48	0
	GGG	0	12	26	22	15	2	36	19	26	24	21	12	17	14	20	0	4	0	0	0	16	18	32
	GGU	0	37	26	33	18	26	0	30	52	60	21	12	0	7	10	0	4	22	77	17	36	30	0
VAL	GUA	16	16	0	18	9	15	0	7	13	24	0	0	17	14	20	56	8	22	0	13	5	5	0
	GUC	16	21	0	7	3	9	0	22	39	12	21	12	0	7	0	0	0	0	0	12	21	20	32
	GUG	0	16	13	4	6	26	0	4	39	0	21	24	0	28	20	0	8	0	0	27	21	5	161
	GUU	16	9	13	29	40	17	0	26	13	36	63	47	17	7	29	56	29	43	0	13	16	16	0
LYS	AAA	73	44	53	62	68	55	107	33	26	63	21	94	67	25	49	111	107	109	0	25	10	43	0
	AAG	16	14	0	18	40	6	0	11	13	24	42	35	33	25	49	0	21	22	0	2	10	5	32
ASN	AAC	24	42	13	11	28	23	0	22	13	12	21	0	17	14	10	56	33	22	0	15	36	20	0
	AAU	33	12	13	62	56	9	36	33	0	48	0	24	17	11	20	0	78	65	0	12	16	8	0
GLN	CAA	24	7	13	25	12	9	0	19	39	0	0	24	50	7	29	56	25	0	0	10	10	18	0
	CAG	24	25	26	7	3	32	36	19	26	12	63	47	17	28	29	0	4	0	0	29	31	25	0
HIS	CAC	8	23	0	7	6	6	0	0	13	12	0	12	0	7	0	0	12	0	77	10	0	18	0
	CAU	8	16	0	14	15	11	0	26	0	24	0	0	0	7	10	56	25	43	0	13	47	18	0
GLU	GAA	49	42	26	36	31	47	71	26	52	48	21	24	17	35	59	0	41	22	0	25	10	33	0
	GAG	8	12	0	25	31	26	71	22	13	12	0	71	0	35	10	0	8	22	0	19	26	38	0
ASP	GAC	24	12	13	11	6	26	0	11	0	12	0	12	0	11	0	0	8	0	0	35	10	13	0
	GAU	8	30	66	54	65	28	36	61	26	95	21	35	33	32	49	0	41	43	0	44	31	33	0
TYR	UAC	0	14	0	7	15	21	0	4	0	36	0	12	33	14	10	0	8	0	0	13	16	10	0
	UAU	49	14	0	22	49	32	0	22	39	0	21	0	33	14	10	56	33	43	0	15	5	20	0
CYS	UGC	0	9	0	0	12	13	0	7	13	0	0	12	33	14	0	0	4	0	0	19	5	10	32
	UGU	24	0	0	4	9	2	0	11	0	0	21	0	17	4	20	0	4	22	0	6	10	3	0
PHE	UUC	0	18	0	4	22	30	36	19	0	12	0	12	67	25	20	0	21	0	0	19	16	20	0
	UUU	24	7	13	36	49	19	36	41	26	63	63	24	0	18	10	111	49	22	77	19	10	13	0
ILE	AUA	33	0	0	43	22	0	0	4	0	12	42	35	33	7	10	0	58	22	0	8	0	3	0
	AUC	24	18	0	14	6	28	0	22	0	12	42	24	0	18	10	56	12	0	0	13	31	35	32
	AUU	41	42	66	25	43	34	36	22	13	48	21	0	17	25	10	111	29	65	0	12	36	23	129
MET	AUG	41	32	26	22	3	32	0	26	0	12	0	35	0	14	29	0	21	0	0	19	31	35	0
TRP	UGG	8	5	0	7	6	4	0	22	26	48	0	0	33	7	0	0	4	0	154	0	10	3	0

		STY HISL	STY HISJ	STY LEUL	STY ARGT	SMA TRPE	SMA TRPG	SDY TRPE	SDY TRPD	EAN LPP	KAE TRPB	KAE TRPA	KPN NIFH	ANA NIFH	RHI NIFH	BLI PEN	HAL RHO
ARG	CGA	0	0	0	0	0	0	0	0	0	53	0	0	0	3	13	0
	CGC	67	19	37	15	38	42	77	31	13	53	45	31	17	24	3	11
	CGG	0	0	0	0	0	0	0	0	0	0	0	0	0	7	10	4
	CGU	0	15	37	15	38	21	38	16	52	0	0	14	37	3	10	11
	AGA	0	0	37	0	0	0	0	0	0	0	0	0	3	7	7	0
	AGG	0	0	0	0	0	0	0	0	0	0	0	0	3	0	7	0
LEU	CUA	0	4	111	8	0	0	0	0	0	0	0	0	13	3	0	8
	CUC	0	0	37	15	0	16	0	26	0	0	7	24	7	44	10	46
	CUG	0	58	0	42	38	94	38	62	117	53	82	38	7	34	10	57
	CUU	0	12	0	12	0	5	0	10	0	0	7	3	7	17	49	8
	UUA	0	12	0	8	0	0	0	16	0	0	4	0	23	0	26	8
	UUG	0	4	0	8	0	0	0	10	0	0	0	0	23	3	13	23
SER	UCA	0	0	37	19	0	0	0	0	26	0	0	3	3	3	7	0
	UCC	0	27	0	15	0	10	0	0	0	0	19	21	17	27	3	11
	UCG	0	8	0	8	0	21	0	10	0	0	11	3	0	3	10	19
	UCU	0	15	0	12	0	5	0	16	39	0	4	0	13	3	3	0
	AGC	0	8	0	8	0	10	0	31	26	0	11	3	7	3	7	19
	AGU	0	0	0	4	0	0	0	0	0	0	0	0	0	0	16	4
THR	ACA	67	8	0	4	0	0	0	5	0	0	0	0	3	10	26	8
	ACC	0	23	0	31	38	31	77	26	13	53	30	41	47	20	10	27
	ACG	0	15	0	19	0	0	38	10	0	0	7	14	0	7	23	38
	ACU	0	8	37	8	38	0	38	5	52	0	0	0	3	3	16	0
PRO	CCA	0	4	0	4	0	10	0	10	0	0	4	3	10	3	10	11
	CCC	0	4	0	8	38	5	0	0	0	0	7	10	7	3	3	11
	CCG	0	23	37	8	38	26	0	47	0	0	60	14	0	17	16	23
	CCU	67	4	0	8	0	10	0	0	0	0	4	0	13	7	10	0
ALA	GCA	0	19	37	8	0	0	38	10	52	0	4	3	27	24	42	19
	GCC	0	23	0	54	154	52	77	21	13	0	90	48	7	47	26	31
	GCG	0	50	0	54	38	47	0	52	13	53	67	45	3	30	16	61
	GCU	0	23	0	12	0	5	115	16	65	0	4	3	54	14	26	4
GLY	GGA	0	8	0	8	0	0	0	5	0	0	15	3	0	7	23	15
	GGC	0	35	74	31	0	99	0	31	13	53	45	68	10	51	10	38
	GGG	0	15	37	19	0	0	0	16	0	53	7	3	0	17	3	31
	GGU	0	23	37	23	0	5	0	47	26	0	11	17	87	24	16	15
VAL	GUA	0	4	0	0	0	0	38	0	39	53	0	0	37	14	16	11
	GUC	0	23	0	12	0	31	0	10	0	0	26	27	0	27	10	31
	GUG	0	23	74	23	77	42	0	36	13	0	30	41	0	24	20	34
	GUU	67	4	37	19	38	0	0	5	26	0	0	7	20	7	10	11
LYS	AAA	67	62	0	73	0	5	38	10	52	105	19	38	20	7	62	11
	AAG	0	35	0	27	77	5	0	5	13	0	11	17	27	51	33	15
ASN	AAC	0	23	37	8	0	16	0	26	65	0	7	38	47	30	23	11
	AAU	0	8	0	12	38	16	38	26	13	0	11	3	7	14	33	0
GLN	CAA	67	8	37	15	0	10	0	5	0	0	4	7	30	14	16	0
	CAG	0	39	0	27	0	52	38	41	78	0	45	27	3	17	13	15
HIS	CAC	200	4	0	0	38	36	0	16	13	0	15	7	23	17	3	0
	CAU	267	0	74	4	38	5	77	31	0	53	7	0	0	10	0	0
GLU	GAA	0	35	0	23	38	36	77	41	0	53	30	62	74	34	42	15
	GAG	0	31	0	15	77	21	38	0	0	0	30	38	10	41	29	27
ASP	GAC	67	23	0	19	38	21	38	16	91	158	26	38	54	41	16	31
	GAU	0	46	0	77	0	21	0	21	13	0	26	17	3	17	62	8
TYR	UAC	0	19	0	27	0	16	0	10	13	0	4	27	23	24	3	31
	UAU	0	12	0	15	0	5	0	16	0	0	22	3	7	10	16	11
CYS	UGC	0	4	0	0	0	10	0	10	13	0	11	24	13	17	7	0
	UGU	0	4	0	8	0	5	0	5	0	0	0	7	7	0	0	0
PHE	UUC	0	23	37	12	38	16	38	10	0	53	30	14	20	10	10	50
	UUU	67	31	37	31	0	5	0	16	0	0	15	7	0	7	20	0
ILE	AUA	0	0	0	4	0	0	0	0	0	0	0	0	3	3	7	0
	AUC	0	27	37	12	38	57	0	10	39	105	56	58	50	57	13	57
	AUU	0	31	74	31	0	10	38	67	0	53	15	24	20	10	26	0
MET	AUG	0	12	0	12	0	36	0	31	0	0	15	51	47	30	16	34
TRP	UGG	0	4	0	8	0	5	0	5	0	0	0	3	0	0	13	31

		HIGH. EXPR.
ARG	CGA	0
	CGC	18
	CGG	0
	CGU	44
	AGA	1
	AGG	0
LEU	CUA	0
	CUC	2
	CUG	60
	CUU	3
	UUA	2
	UUG	3
SER	UCA	1
	UCC	15
	UCG	0
	UCU	21
	AGC	8
	AGU	3
THR	ACA	4
	ACC	24
	ACG	2
	ACU	28
PRO	CCA	6
	CCC	1
	CCG	25
	CCU	2
ALA	GCA	33
	GCC	8
	GCG	23
	GCU	64
GLY	GGA	1
	GGC	31
	GGG	1
	GGU	45
VAL	GUA	27
	GUC	8
	GUG	15
	GUU	38
LYS	AAA	62
	AAG	25
ASN	AAC	35
	AAU	2
GLN	CAA	5
	CAG	31
HIS	CAC	13
	CAU	6
GLU	GAA	51
	GAG	17
ASP	GAC	35
	GAU	16
TYR	UAC	15
	UAU	3
CYS	UGC	4
	UGU	1
PHE	UUC	20
	UUU	6
ILE	AUA	0
	AUC	46
	AUU	12
MET	AUG	21
TRP	UGG	6

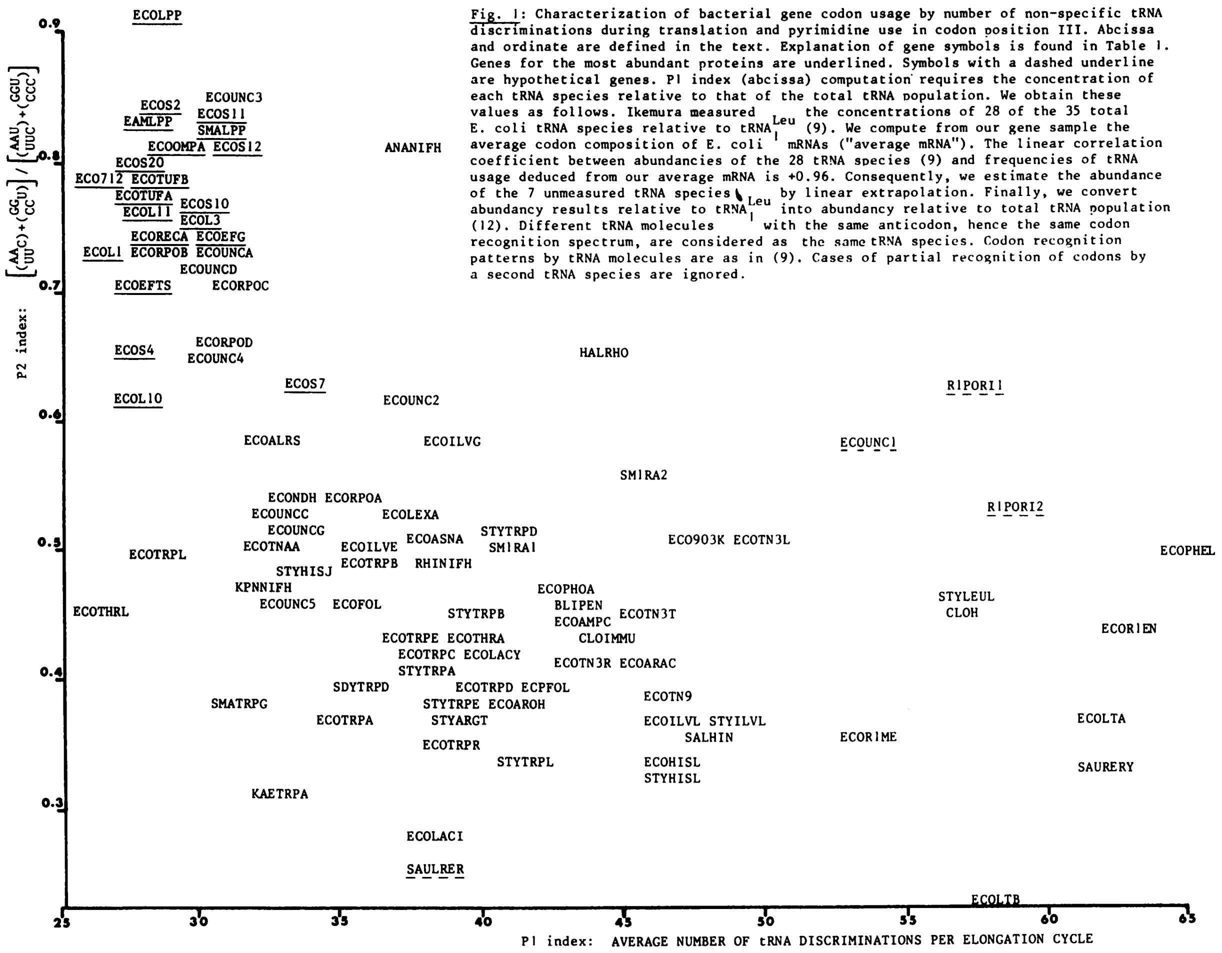

Fig. 1: Characterization of bacterial gene codon usage by number of non-specific tRNA discriminations during translation and pyrimidine use in codon position III. Abcissa and ordinate are defined in the text. Explanation of gene symbols is found in Table 1. Genes for the most abundant proteins are underlined. Symbols with a dashed underline are hypothetical genes. PI index (abcissa) computation requires the concentration of each tRNA species relative to that of the total tRNA population. We obtain these values as follows. Ikemura measured the concentrations of 28 of the 35 total E. coli tRNA species relative to $\text{tRNA}_\text{I}^\text{Leu}$ (9). We compute from our gene sample the average codon composition of E. coli mRNAs ("average mRNA"). The linear correlation coefficient between abundancies of the 28 tRNA species (9) and frequencies of tRNA usage deduced from our average mRNA is +0.96. Consequently, we estimate the abundance of the 7 unmeasured tRNA species \ by linear extrapolation. Finally, we convert abundancy results relative to $\text{tRNA}_\text{I}^\text{Leu}$ into abundancy relative to total tRNA population (12). Different tRNA molecules with the same anticodon, hence the same codon recognition spectrum, are considered as the same tRNA species. Codon recognition patterns by tRNA molecules are as in (9). Cases of partial recognition of codons by a second tRNA species are ignored.

The results in fig. 1 suggest that translation of highly expressed mRNAs is faster than translation of others. The complex relationship between translation rate and mRNA protein yield has been analyzed by Bergmann and Lodish (13). Briefly, these authors show that the main factor determining protein yield is initiation rate, but that elongation and initiation rates can interact and modulate protein yield.

Plasmid and transposon genes generally have a P1 index value greater than 43, indicating that the cellular tRNA population is poorly adapted to their codon usage. Conversely, few chromosomal genes are located in the same region of fig. 1. E. coli mRNAs falling there are very weakly expressed: toxin genes (ECOLTA and ECOLTB), the Eco RI restriction enzyme system (ECOR1EN endonuclease and ECOR1ME methylase), two leader genes of amino acid biosynthesis operons (ECOPHEL and STYLEUL) and an hypothetical regulatory gene in unc operon (ECOUNC1).

Genes from enterobacteriaceae other than E. coli (Salmonella, Serratia, Shigella, Klebsiella and Erwinia (14)) closely follow E. coli codon usage as seen with genes sequenced in both species, namely lpp and trp operon genes. Genes from non enteric bacteria (Anabaena, Bacillus, Rhizobium, Staphylococcus and the archaebacterium Halobacterium) have been included for completeness, but the P1 index, based on the E. coli tRNA population,may be meaningless for them.

CHOICE BETWEEN PYRIMIDINES IN CODON POSITION III

Grosjean et al. first noted in the MS2 phage genome a bias in the choice between C and U bases in codon position III (15). They found that nucleotides in the degenerate position consistently yield a codon-anticodon binding energy of intermediate strength. For example, AAC appears more often than AAU, and GCU more often than GCC. This bias is independent of amino acid composition (whatever the bases X_1 and X_2, codons X_1X_2C and X_1X_2U code for the same amino acid) and of tRNA availability (X_1X_2C and X_1X_2U are mainly decoded by the same tRNA (9)). Our group showed by statistical analysis of 29 bacterial genes that this bias is not present in all sequences, as already recognized by Fiers and Grosjean in the lac I gene (16), but strongly depends on the expressivity level of the gene (2). Ikemura later stated the same conclusion (9,10). If the first two bases of a codon are both A or U, C in third position will give a codon-anticodon binding energy nearer the mean than would U. Likewise, if the first two bases are both C or G, the "right choice" for third base is U, for C would give a strong binding energy.

Consequently, we characterize each messenger by the frequency of "right choices" between the pyrimidines among codons beginning with AA, AU, UA, UU, CC, CG, GC or GG. This frequency, called P2 index, is scaled on the ordinate of fig. 1. As with P1 index, a discrimination between messengers for abundant and rare proteins occurs. The highly expressed genes appear in the upper left part of the figure.

In order to summarize the optimal choice between synonymous codons, we compute the mean frequency of each codon after pooling all highly expressed genes, that is genes with underlined symbols in fig. 1. The results, HIGH. EXPR. column of Table 2, clearly show the two tendencies in codon usage discussed (the codon recognition patterns of E. coli tRNA species are given in (9)).

The interpretation of pyrimidine selection given by Grosjean et al. involves the energy level of the codon-anticodon interaction (15) and is connected to translational fidelity (29). However, the nature of the effect of U/C choice on mRNA translation remains unclear. This effect cannot be restricted to fidelity because we show here that pyrimidine choice is related to gene expressivity.

AN INTRA-OPERON RELATIONSHIP BETWEEN GENE EXPRESSIVITY AND CODON USAGE

To examine more precisely the relationships between codon usage and gene expressivity, we reviewed the available quantitative data (Table 3). Although these data are not numerous, a general agreement with fig. 1 results: clearly, all genes coding for the most abundant proteins are clustered to the top left of fig. 1. The nine unc (or atp) operon genes and the four RNA-polymerase subunit genes of E. coli merit more comment, however.

The completely sequenced unc operon comprises 8 structural genes coding for the 8 polypeptides that form the ATP-synthase complex and a hypothetical control gene unc1 (17-19). All these genes are controlled by the same promoter. Products of genes uncA, uncD, unc5, uncG and uncC form the F_1-part of the ATP-synthase complex with the stoichiometry 3:3:1:1:1 (18-21). The unc2, unc3 and unc4 gene products, respectively the a, c and b polypeptides, form together the F_o-part of the same complex (17). The stoichiometry of the F_o-part is unclear, but the c peptide appears in more than 5 copies and the a and b peptides in 1 or 2 copies in the complex (21). The positions in fig. 1 of these 8 structural genes correlate with the gene product copy numbers in the ATP-synthase complex. Genes with high copy numbers (unc3, uncA and uncD) have a higher P2 index value and generally a lower P1 value than genes with low copy numbers (unc5, uncG and uncC). As stressed by Fillingame, the coordinate expression of the 8 unc operon structural genes probably involves unidentified post-transcriptional mechanisms (21).

Three of the four RNA-polymerase genes (rpoB, C and D) are located in fig. 1 with the highly expressed genes while rpoA is not. Table 3 shows that these genes correspond to proteins of medium abundance and that rpoA is more highly expressed than the other three genes because its product, the α-subunit appears in 2 copies in the holoenzyme (22). Note however that this is constitutive expressivity level; it is therefore difficult to compare with the lac and trp genes whose levels are for maximum derepression. The rpoA gene is cotranscribed with other r-protein genes in this order: rpsM (S13), rpsK (S11), rpsD (S4), rpoA and rplQ (L17) (23). No regulatory feature in the operon primary structure is apparent between the first 3 r-protein genes and rpoA, although the expressivities of these genes are markedly different (24). The locations in fig. 1 of rpsK (ECOS11),rpsD (ECOS4) and rpoA genes show that the same correlation between codon usage and gene expressivity holds for this operon as for the unc operon. However, only part of these last 3 genes has been sequenced, hence we cannot be sure of their overall coding strategy. On the other hand, rpoD is part of a single-gene operon (25) while rpoB and C are cotranscribed with rplJ (L10) and rplL (L7/L12, symbol ECO712) but a transcription attenuator site, which has been shown to function actively in vivo (26), is present in the nucleotide sequence between rplL and rpoB, C genes (27).

Table 3: Cellular contents of various E. coli gene products.

Symbol, gene and product	No. molecules per genome	Ref.
ECOLPP lpp (outer membrane lipoprotein)	330,000	31
ECOTUF- tufA,B (elongation factor Tu)	89,000	11[§], 32
ECORECA recA	38,000[&]	33
ECOOMPA ompA (outer membrane protein)	36,800	34
ECO712 rplL (L7 and L12 r-proteins)	25,000	34
other r-proteins	9,200	11[§]
ECOEFTS tsf (elongation factor Ts)	9,200	11[§]
ECOEFG fusA (elongation factor G)	9,200	11[§]
ECORPOA rpoA (α-subunit of RNA-polymerase)	4,000	34
ECOLACY lacY (lactose permease)	3,300[&]	35
ECOTRP- trpA-E (tryptophan biosynthesis)	3,100[&]	36
ECORPOB rpoB (β-subunit of RNA-polymerase)	1,400	34
ECORPOC rpoC (β'-subunit of RNA-polymerase)	1,400	34,37
ECOALRS (aminoacyl-tRNA-synthetases)	500-1,300	34
ECOARAC araC (ara operon regulation)	100	38
ECOTRPR trpR (trp and aroH operons repressor)	30	39
ECOLACI lacI (lac operon repressor)	10	40

Data have been computed for a cell growth rate of 1.5 doublings per hour when possible.

[&] variable with growth conditions, value for maximum derepression.

[§] see references therein.

In addition to the RNA-polymerase genes and unc operon, other cases exist where several genes in one operon are sequenced. Three operons have been completely sequenced in E. coli. 1) trp operon: trpL, trpE, trpD, trpC, trpB and trpA genes (also partly sequenced in Salmonella typhimurium, Serratia marcescens, Shigella dysenteriae and Klebsiella aerogenes); 2) rpl operon: rplK (L11) and rplA (L1) genes; 3) elongation factor Ts operon: rpsB (S2) and tsf (EF-Ts) genes. Examples of partly sequenced operons are rpsL (S2), rpsG (S7), fusA (EF-G) and tufA (EF-Tu) genes; rpsJ (S10) and rplC (L3) genes and finally genes for the Eco RI restriction and modification enzymes (symbols ECORIEN and ECORIME). In all these cases, both the expressivity level and codon composition of genes from each operon (fig. 1) are similar. We cannot interpret Fig. 1 location of ilv operon genes (ilvL, G and E) since data on their expressivity are lacking. Moreover, ilvG is a peculiar cryptic gene (4).

The leader genes for amino acid biosynthesis operons (hisL, ilvL, pheL, thrL and trpL in E. coli and S. typhimurium) are exceptions to this observation, being widely scattered throughout fig. 1. We believe they are not adequately represented by this method because peculiar constraints are exerted on their structure (28).

This analysis of genes in the same operon argues for a relationship between P2 index (U/C choice in degenerate position) and the relative level of gene expressivity. Thus, codon usage may be involved in expressivity tuning of genes in the same operon according to cell needs. Nevertheless, direct experimentation is needed to confirm this hypothesis, and other mechanisms may exist to regulate translation rates. One alternate hypothesis is that transcriptional efficiency is involved. However, differences between expressivity of genes in the same transcription unit correspond to variation of P2. This indicates that transcriptional efficiency is not involved.

DARWINIAN EVOLUTION AND mRNA TRANSLATION

Whatever the process, codon usage optimization implies the existence of strong evolutionary pressures. This point of view has been largely developped by Ikemura (9,10). His main idea is a balance between selection for optimal codons and the mutational process. When a gene does not have a high translation rate, selective pressure is weak and the mutational process may blur its effect. For highly expressed genes the pressure is stronger and optimization clearly occurs.

We first want to point out that codon usage bias is the clearest example known of overall optimization of gene sequences. Selective pressure does not act on one particularly important codon, but on the relative frequency of each codon. This is an argument in favour of Darwinian evolution as a process of cumulative small variations. Although we have no means to measure the phenotypic effect of one silent nucleotide substitution, it would be surprising if the fitness were greatly affected. Direct competition experiments between strains varying in codon usage would be of great evolutionary interest.

Our second remark derives from comparisons among P2 index values. Table 2 in (2) indicates that weakly expressed genes are not free of constraints on C/U choice in codon position III but that the rule is opposite that for highly expressed genes. Ikemura's evolutionary scheme, wherein each gene independently evolves toward an optimum, does seem valid for tRNA considerations, but may not be true for C/U choice. Our study indicates that the sellection target also includes harmonization of expressivity and C/U choice, particularly when the genes are in the same operon. This would explain the positions of unc genes in fig. 1. Although we do not find data on ATP-synthase (unc operon product) cell content, this enzyme is most probably not as abundant as r-proteins or elongation factors. Hence, the high use of optimal codons in unc3, uncA and uncD genes has to be explained by their expressivity relative to that of remaining genes in the unc operon, rather than by their absolute expressivity.

Finally, the relationship between codon usage and gene expressivity demonstrated in E. coli may be valid for eukaryotic organisms as well. Indeed, a similar correlation between codon usage, tRNA population and gene expressivity has been recently described in the eukaryotic, undifferentiated yeast cell (30). However, the yeast catalog of "optimal" codons is different from that in E. coli probably because anticodon sequences, codon recognition patterns and relative concentrations of iso-tRNA species differ widely between these two organisms.

REFERENCES

1. Grantham, R., Gautier, C. and Gouy, M. (1980) Nucleic Acids Res. 8, 1893-1912.
2. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. and Mercier, R. (1981) Nucleic Acids Res. 9, r43-r74.
3. Lawther, R.P., Nichols, B., Zurawski, G. and Hatfield, G.W. (1979) Nucleic Acids Res. 7, 2289-2301.
4. Lawther, R.P., Calhoun, D.H., Adams, C.W., Hauser, C.A., Gray, J. and Hatfield, G.W. (1981) Proc. Natl. Acad. Sci. USA 78, 922-925.
5. Garel, J.-P., Mandel P., Chavancy, G. and Daillie, J. (1970) FEBS lett. 7, 327-329.
6. Chavancy, G. and Garel, J.-P. (1981) Biochimie 63, 187-195.
7. Post, L.E., Strycharz, G.D., Nomura, M., Lewis, H. and Dennis, P.P. (1979) Proc. Natl. Acad. Sci. USA 76, 1697-1701.
8. Post, L.E. and Nomura, M. (1980) J. Biol. Chem. 255, 4660-4666.

9. Ikemura, T. (1981) J. Mol. Biol. 146, 1-21.
10. Ikemura, T. (1981) J. Mol. Biol. 151, 389-409.
11. Gouy, M. and Grantham, R. (1980) FEBS lett. 115, 151-155.
12. Gouy, M. (1981) Thèse de troisième cycle, Université Lyon I.
13. Bergmann, J. E. and Lodish, H.F. (1979) J. Biol. Chem. 254, 11927-11937.
14. Sanderson, K.E. (1976) Ann. Rev. Microbiol. 30, 327-349.
15. Grosjean, H., Sankoff, D., Min Jou, W., Fiers, W. and Cedergren, R.J. (1978) J. Mol. Evol., 12, 113-119.
16. Fiers, W. and Grosjean, H. (1979) Nature 277, 328.
17. Gay, N.J. and Walker, J.E. (1981) Nucleic Acids Res. 9, 3919-3926.
18. Gay, N.J. and Walker, J.E. (1981) Nucleic Acids Res. 9, 2187-2194.
19. Saraste, M., Gay, N.J., Eberle, A., Runswick, M.J. and Walker, J.E. (1981) Nucleic Acids Res. 9, 5287-5296.
20. Futai, M. and Kanazawa, H. (1980) Curr. Topics Bioenerg. 10, 181-215.
21. Fillingame, R.H. (1981) Curr. Topics Bioenerg. 11, 35-106.
22. Chamberlin, M.J. (1974) Ann. Rev. Biochem. 43, 721-775.
23. Lindahl, L., Post, L., Zengel, J., Gilbert, S.F., Strycharz, W.A. and Nomura, M. (1977) J. Biol. Chem. 252, 7365-7383.
24. Post, L.E. and Nomura, M. (1979) J. Biol. Chem. 254, 10604-10606.
25. Burton, Z., Burgess, R.R., Lin, J., Moore, D., Holder, S. and Gross, C.A. (1981) Nucleic Acids Res. 9, 2889-2903.
26. Little, R., Fiil, N.P. and Dennis, P.P. (1981) J. Bact. 147, 25-35.
27. Post, L.E., Strycharz, G.D., Nomura, M., Lewis, H. and Dennis, P.P. (1979) Proc. Natl. Acad. Sci. USA 76, 1697-1701.
28. Keller, E.B. and Calvo, J.M. (1979) Proc. Natl. Acad. Sci. USA 76, 6186-6190.
29. Grosjean, H.J., de Hénau, S. and Crothers, D.M. (1978) Proc. Natl. Acad. Sci. USA 75, 610-614.
30. Bennetzen, J.L. and Hall, B.D. (1982) J. Biol. Chem. 257, 3026-3031.
31. di Rienzo, J.M., Nakamura, K. and Inouye, M. (1978) Ann. Rev. Biochem. 47, 481-532.
32. Furano, A.V. (1975) Proc. Natl. Acad. Sci. USA 72, 4780-4784.
33. Gudas, L.J. and Pardee, A.B. (1976) J. Mol. Biol. 101, 459-477.
34. Pedersen, S., Bloch, P.L., Reeh, S. and Neidhardt, F.C. (1978) Cell 14, 179-190.
35. Zabin, I. and Fowler, A.W. (1980) in The Operon, Miller, J.H. and Reznikoff, W.S. Eds., Cold Spring Harbor Laboratory.
36. Morse, D.E., Mosteller, R.D. and Yanofsky, C. (1969) Cold Spring Harbor Symp. Quant. Biol. 34, 725-739.
37. Dennis, P.P. (1977) J. Mol. Biol. 115, 603-625.
38. Casabadan, M.J. (1976) J. Mol. Biol. 104, 557-566.
39. Rose, J.K. and Yanofsky, C. (1974) Proc. Natl. Acad. Sci. USA 71, 3134-3138.
40. Gilbert, W. and Müller-Hill, B. (1966) Proc. Natl. Acad. Sci. USA 56, 1891-1898.