JOURNAL OF BACTERIOLOGY, Oct. 1987, p. 4770–4777 Vol. 169, No. 10
0021-9193/87/104770-08$02.00/0

Organization and Codon Usage of the Streptomycin Operon in *Micrococcus luteus*, a Bacterium with a High Genomic G+C Content

TAKESHI OHAMA,* FUMIAKI YAMAO, AKIRA MUTO, AND SYOZO OSAWA

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

Received 27 March 1987/Accepted 14 July 1987

The DNA sequence of the *Micrococcus luteus str* operon, which includes genes for ribosomal proteins S12 (*str* or *rpsL*) and S7 (*rpsG*) and elongation factors (EF) G (*fus*) and Tu (*tuf*), has been determined and compared with the corresponding sequence of *Escherichia coli* to estimate the effect of high genomic G+C content (74%) of *M. luteus* on the codon usage pattern. The gene organization in this operon and the deduced amino acid sequence of each corresponding protein are well conserved between the two species. The mean G+C content of the *M. luteus str* operon is 67%, which is much higher than that of *E. coli* (51%). The codon usage pattern of *M. luteus* is very different from that of *E. coli* and extremely biased to the use of G and C in silent positions. About 95% (1,309 of 1,382) of codons have G or C at the third position. Codon GUG is used for initiation of S12, EF-G, and EF-Tu, and AUG is used only in S7, whereas GUG initiates only one of the EF-Tu's in *E. coli*. UGA is the predominant termination codon in *M. luteus*, in contrast to UAA in *E. coli*.

In eubacteria, the G+C content of genomic DNA varies from about 25 to 74% (27). The phylogenetic tree of 5S rRNA has indicated that the genomic G+C contents are closely related to phylogeny (10). This suggests that G+C content is influenced by mutation pressure, the direction and magnitude of this pressure varying among the bacterial phylogenetic lines (21). Biased AT/GC pressure seems to have been exerted on the entire bacterial genome, directionally increasing or decreasing the G+C contents of various parts of the genome (21). Thus, the codon usage pattern in a bacterium seems to have been affected by the AT/GC pressure. For example, in the extremely G+C-poor bacterium *Mycoplasma capricolum* (G+C, 25%), the codon choice is strongly biased toward use of A and U in silent positions (20).

The G+C content of *Micrococcus luteus* DNA is one of the highest (74%) in eubacteria. All species belonging to the *Micrococcus* group so far reported have genomes with high G+C contents (65 to 74%). Thus, in this phylogenetic line, a strong GC pressure may have been affecting composition of the DNA during evolution. In the present study, we have cloned and sequenced the *M. luteus* streptomycin (*str*) operon, which includes the genes for ribosomal proteins S7 and S12, and elongation factors (EF) EF-G and EF-Tu to examine the effect of GC pressure on codon usage in this high-G+C-content bacterium.

MATERIALS AND METHODS

Preparation of DNA. Chromosomal DNA of *M. luteus* IFO3333 was prepared by lysozyme lysis and phenol extraction method according to the method of Godson (8) and further purified by CsCl gradient centrifugation. The genomic G+C content of this strain is 73.6% as measured by direct liquid chromatographic analysis of the component nucleotides (28).

Isolation of Sm^r mutant of *M. luteus*. A spontaneous streptomycin-resistant (Sm^r) mutant was obtained from a streptomycin-sensitive (Sm^s) strain of *M. luteus* (IFO3333) as follows. The cells were cultured in brain heart infusion medium (Difco Laboratories, Detroit, Mich.) at 37°C with vigorous shaking until the early stationary phase. A 0.3-ml portion of the culture was spread on the brain heart infusion agar plate containing 100 μg of streptomycin sulfate per ml. After 3 days of incubation at 37°C, four to five Sm^r colonies per plate were obtained. Chromosomal DNA was prepared from one of these colonies and used as the donor DNA.

Transformation. A Sm^s strain of *M. luteus* (ATCC 27141) was used as a recipient because of its high transformation efficiency (15, 26). The transformation was done by the method of Kloos (14) with the following modifications. Cells were cultured, harvested by centrifugation, and dispersed in 0.5% monosodium glutamate–0.01 M $CaCl_2$ buffer. The competent cells were mixed with glycerol at a final concentration of 15%. A 200-μl portion was frozen quickly in liquid nitrogen and stored at −70°C. These cells were active in transformation for at least 4 months. The competent cells (200 μl) were mixed with digested chromosomal DNA (0.05 to 0.1 μg) from strain IFO3333 (Sm^r) (or recombinant plasmid DNA containing the DNA fragment from strain IFO3333 [Sm^r]) and shaken for 45 min at 30°C, followed by the addition of 2 volumes of brain heart infusion medium. The mixture was shaken at 37°C with vigorous aeration (180 rpm with a rotary shaker) for more than 7 h. A 0.3-ml portion of the culture medium was spread on a brain heart infusion agar plate containing 50 μg of streptomycin per ml and incubated for 3 days at 37°C.

Cloning and screening of *M. luteus str* operon. Chromosomal DNA from Sm^r *M. luteus* IFO3333 was digested with restriction enzyme *Bam*HI and fractionated according to size by sucrose gradient centrifugation (10 to 40%) (17). A transformation assay for each fraction showed that a fraction containing 15- to 20-kilobase-pair (kbp) DNA fragments had the highest activity. The DNA fragments in this fraction were randomly ligated to plasmid vector pUC18 DNA and transfected to *Escherichia coli* HB101 cells. About 120 independent plasmids from the transformed cells were ob-

* Corresponding author.

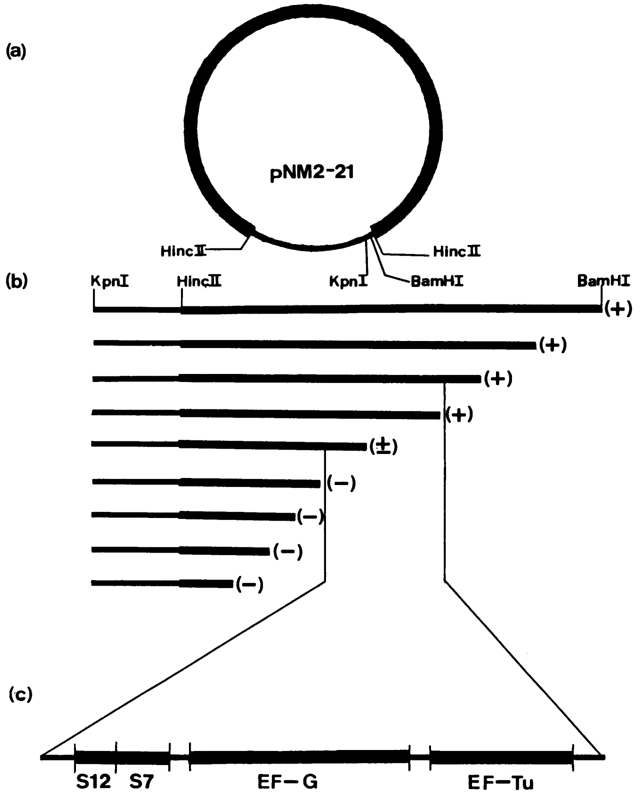

FIG. 1. (a) Construction of plasmid pNM2-21 containing *M. luteus str* operon. A 15-kbp DNA fragment of *M. luteus* was ligated to the *Hin*cII site of the vector plasmid pUC18. *Bam*HI and *Kpn*I sites in the pUC18 multicloning site are also shown. (b) Plasmid DNA pNM2-21 was deleted stepwise from the *Bam*HI site to the *Kpn*I site by exonucleases III and VII (32) and tested for transformation activity. Symbols: +, more than 100 transformants per plate; ±, 20 to 100 transformants per plate; −, fewer than 20 transformants per plate. (c) Organization of the sequenced 5.3-kbp region that includes the *M. luteus str* operon.

tained and assayed for transformation of Sm^s *M. luteus* cells to Sm^r phenotype.

Isolation of plasmids and sequencing of DNA. The *E. coli* cells containing the recombinant plasmids (see above) were inoculated in 2 ml of 2× YT medium (0.8% tryptone [Difco], 0.5% yeast extracts, and 0.25% NaCl [wt/vol]) and shaken for 12 to 16 h at 37°C. Crude plasmid DNA isolated by the quick boiling method (9) was used for the transformation assay. Further restriction of the DNA fragment containing Sm^r character was made by stepwise deletion from one end by *Exo*III (Takara Shuzo, Kyoto, Japan) and *Exo*VII (Bethesda Research Laboratories, Gaithersburg, Md.) nucleases (32). The dideoxy-chain termination method was used for DNA sequencing by these pUC plasmids (5, 16). Deoxy-7-deazaguanosine triphosphate was used in place of dGTP (18) to resolve the ladder rich in G. The sequence data were analyzed by the computer program DNASIS (Hitachi SK, Yokohama, Japan). The program was also used to align the deduced amino acid sequences with those of *E. coli*.

RESULTS AND DISCUSSION

Cloning of DNA containing the *str* operon. Since the Sm^r phenotype is caused by mutational change in ribosomal protein S12 in many bacterial species, transformation of the Sm^s cells by Sm^r DNA was used to screen for DNA fragments that contain the *str* operon. The transformation assay for *Bam*HI-digested total Sm^r DNA fractionated by sucrose gradient centrifugation revealed that the fraction containing 15- to 20-kbp DNA fragments had the highest activity. The DNA fraction was then ligated to vector plasmid pUC18 and transfected into *E. coli* HB101. One of the 50 plasmid DNAs tested that had a very high Sm^r transformation activity was selected. The plasmid included an approximately 15-kbp fragment of *M. luteus* DNA. About 200 bp were deleted from both ends of this inserted DNA by nuclease *Bal* 31, and the shortened DNA fragment was recloned to the *Hin*cII site of vector plasmid pUC18 (pNM2-21) (Fig. 1a). The Sm^r transformation activity of pNM2-21 DNA was more than 100 times higher than that of unfractionated total DNA. The plasmid DNA was deleted stepwise from one end of the insert by *Exo*III and *Exo*VII digestion by using *Bam*HI and *Kpn*I sites, and the DNA was used for the Sm^r transformation assay to map an approximate position of the gene for S12 (Fig. 1b). To sequence the complementary strand, the inserted DNA fragment was recloned to the *Hin*cII site of pUC19 and deleted from the other end by the method used with pNM2-21. More than 98% of the sequences was determined for both strands. The results suggested that the location of this gene was at about the middle of the inserted DNA.

Organization of the *str* operon. The DNA sequence of a part (about 5.3 kbp) of the pNM2-21 insert, including the putative *str* locus (see below), revealed the presence of four open reading frames in the same strand (Fig. 2). Their amino acid sequences deduced from the DNA sequences showed high homologies with those of *E. coli* ribosomal proteins S12 (70%) and S7 (58%) and elongation factors EF-G (60%) and EF-Tu (71%). Thus, we conclude that this gene cluster is homologous to the *str* operon of *E. coli* (23, 33, 35). The genes were arranged in the order (5′) S12, S7, EF-G, EF-Tu (3′) (Fig. 1c), in agreement with the gene order in the *E. coli str* operon (23).

In Fig. 2 are shown the total DNA sequence and the deduced amino acid sequence of the 5.3-kbp fragment of pNM2-21, which included about 350 bp upstream from the 5′ end of the S12 gene. The sequence of the above noncoding region was very rich in G+C (68%), and no typical "*E. coli*-type" promoter sequences were found, suggesting that the *M. luteus* promoter sequence that is supposed to exist in this region is G+C rich and different from that of *E. coli*. In another G+C-rich bacterium, *Streptomyces coelicolor* (G+C, 73%), two types of promoter, the A+T-rich "*E. coli* type" and the G+C-rich type, have been predicted (11). The boxed sequences, AGGATTGT (−190 to −197) and GGCAAATTG (−172 to −180), in Fig. 2 could be promoter-like sequences, because of their considerable similarity to the G+C-rich promoter like sequences of the *Streptomyces plicatus endoH* gene (25, 31) and also to the "sigma^{37}" promoter sequences of *Bacillus subtilis* (19).

The G+C content of the spacer regions was very high (68 to 70%). There was no spacer between genes for S7 and S12, where the third letter of the S7 termination codon (TAA) overlapped with the first letter of the S12 initiation codon (ATG) as $-\mathrm{TA}\overline{\mathrm{A}}\underline{\mathrm{A}}\mathrm{TG}-$. The spacers between the S7 and EF-G and between the EF-G and EF-Tu were 214 and 275 bp, respectively, which were much longer than the corresponding parts in *E. coli* (27 and 70 bp, respectively).

Since no "*E. coli*-type" promoter like sequences have been recognized in *M. luteus*, we could not decide whether

```
-350        -340        -330        -310        -300        -290        -280
GCCGTCTCGC  TGGACGACAT  CGACCGCGGG  GCGAACTTTC  CGACGCCTGA  TCCCGTCCAG  GGCCGCCGTT  CCGGCCGCTC  ACGGCGGGGG
-260        -250        -240        -230        -220        -210        -200        -190        -180
AGCGGCGGCC  CCTTGGCGTG  CGGGGCCGGC  TCGTCGACGG  CGCCGGGGCG  TGCTATGCTG  TCAAGGATTG  TTTTTCGTGT  GGCAAATTGG
-170        -160        -150        -140        -130        -120        -110        -100        -90
ACGCTGTCCG  GGTCGCGGGG  TCATTGCGCG  GCAGATGCCA  CACTTTTGCA  CGCCTGCGGA  CGTCTGCTCG  TTGCGGAGCC  CCCACAGGGA
-80         -70         -60         -50         -40         -30         -20         -10         S12 Start
GAGGCAAAGC  AGAGGAACGA  GCAGGCCCGG  GCGCCGGTCG  ATCACGTCAG  CGATTCACCG  AAACACAGGA  AGGCTGAAGA  GTG CCT ACG
                                                                                                Met Pro Thr
ATT CAG CAG CTG GTC CGC AAG GGC CGC TCA CCC AAG GTC GTC AAT ACG AAC GGG CCT GCC CTG CAG GGG AAC CCC ATG CGT
Ile Gln Gln Leu Val Arg Lys Gly Arg Ser Pro Lys Val Val Asn Thr Asn Gly Pro Ala Leu Gln Gly Asn Pro Met Arg
                105                 120                 135                 150                 165
CGT GGC GTG TGC ACC CGT GTG TAC ACC ACC ACG CCC ACG AAG CCG AAC TCG GCC GTG CGC AAG GTC GCC CGT GTG CGC CTG
Arg Gly Val Cys Thr Arg Val Tyr Thr Thr Thr Pro Thr Lys Pro Asn Ser Ala Val Arg Lys Val Ala Arg Val Arg Leu
        180                 195                 210                 225                 240
AAC GGC GGC ATC GAG GTC ACC GCC TAC ATC CCC GGC GAG GGC CAC AAC CTG CAG GAG CAC TCG ATC GTG CTC GTC CGC GGC
Asn Gly Gly Ile Glu Val Thr Ala Tyr Ile Pro Gly Glu Gly His Asn Leu Gln Glu His Ser Ile Val Leu Val Arg Gly
255                 270                 285                 300                 315                 330
GGT CGT GTG AAG GAC CTG CCC GGC GTG CGC TAC AAG ATC GTG CGC GGC GCC CTC GAC ACC CAG GGC GTG AAG AAC CGC GGC
Gly Arg Val Lys Asp Leu Pro Gly Val Arg Tyr Lys Ile Val Arg Gly Ala Leu Asp Thr Gln Gly Val Lys Asn Arg Gly
            345                 360                     S7 Start                        405
CAG GCC CGC TCC CGC TAC GGC GCC AAG AAG GAG AAG AAG TA  ATG CCT CGT AAG GGT CCT GCG CCG AAG CGC CCC CTC GTC
Gln Ala Arg Ser Arg Tyr Gly Ala Lys Lys Glu Lys Lys  *  Met Pro Arg Lys Gly Pro Ala Pro Lys Arg Pro Leu Val
    420                 435                 450                 465                 480                 495
GTC GAT CCC GTC TAC GGC TCC CCG CTG GTC ACG CAG CTG ATC AAC AAG GTG CTC GTC GAC GGC AAG AAG TCC ACC GCC GAG
Val Asp Pro Val Tyr Gly Ser Pro Leu Val Thr Gln Leu Ile Asn Lys Val Leu Val Asp Gly Lys Lys Ser Thr Ala Glu
                510                 525                 540                 555                 570
CGC ATC GTG TAC GGC GCG CTC GAG GGC GCC CGT GCC AAG AAC GGC GCC CGA TCC CGT GGC CAC CCC ATC AAG AAG GCC ATG
Arg Ile Val Tyr Gly Ala Leu Glu Gly Ala Arg Ala Lys Asn Gly Ala Arg Ser Arg Gly His Pro Ile Lys Lys Ala Met
        585                 600                 615                 630                 645
GAC AAC ATC AAG CCG GCC CTC GAG GTG CGC TCC CGC CGC GTC GGC GGC GCC ACC TAC CAG GTG CCC GTC GAG GTC AAG CCG
Asp Asn Ile Lys Pro Ala Leu Glu Val Arg Ser Arg Arg Val Gly Gly Ala Thr Tyr Gln Val Pro Val Glu Val Lys Pro
                    675                 690                 705                 720                 735
GGC CGC TCC ACG GCC CTG GCC CTG CGC TGG CTG GTC GGC TTC TCC AAG GCC CGC CGT GAG AAG ACC ATG ACC GAG CGT CTG
Gly Arg Ser Thr Ala Leu Ala Leu Arg Trp Leu Val Gly Phe Ser Lys Ala Arg Arg Glu Lys Thr Met Thr Glu Arg Leu
            750                 765                 780                 795                 810
ATG AAC GAG ATC CTG GAC GCC TCG AAC GGC CTG GGC GGC GCC GTG AAG CGT CGC GAG GAC ACC CAC AAG ATG GCC GAG GCC
Met Asn Glu Ile Leu Asp Ala Ser Asn Gly Leu Gly Gly Ala Val Lys Arg Arg Glu Asp Thr His Lys Met Ala Glu Ala
    825                 840                              860         870         880         890         900
AAC AAG GCC TTC GCC CAC TAC CGC TGG TGA      TC  CGTGTGGCCC  GGGGCGCGCG ATCCGCCTCG  GCGGCAGCGC  TCGCGTCCTC
Asn Lys Ala Phe Ala His Tyr Arg Trp  *
        910         920         930         940         950         960         970         980         990
TCGGACGCGC  CACCACGCAT  CCACCTCACG  AAGTCCCCCA  TCTCCATCCC  AGACCAAGGG  AGACCAACGT  CGGCAAGGAC  GTGCCTGACT
        1000        1010        1020        1030        1040        1050        1060    EF-G Start
GACCTGCACA  CGGTCCGCAA  CATCGGCATC  ATGGCTCACA  TCGACTCGCC  GGAGACACCG  TGCGCACAGG  AC  GTG CTG ACT GAC CTG
                                                                                        Met Leu Thr Asp Leu
1080                1095                1110                1125                1140                1155
CAC AAG GTC CGC AAC ATC GGC ATC ATG GCT CAC ATC GAT GCC GGC AAG ACC ACC ACC ACC GAG CGC CAT CTG TTC TAC ACG
His Lys Val Arg Asn Ile Gly Ile Met Ala His Ile Asp Ala Gly Lys Thr Thr Thr Thr Glu Arg His Leu Phe Tyr Thr
            1170                1185                1200                1215                1230
GGT GTC AAC CAC AAG CTG GGC GAG ACC CAC GAC GGC GGC GCG ACG ACG GAC TGG ATG GAG CAG GAG AAG GAG CGC GGC ATC
Gly Val Asn His Lys Leu Gly Glu Thr His Asp Gly Gly Ala Thr Thr Asp Trp Met Glu Gln Glu Lys Glu Arg Gly Ile
    1245                1260                1275                1290                1305                1320
ACC ATC ACC TCC GCC GCG GTG ACC TGC TTC TGG AAC GAC CAC CAG ATC AAC ATC ATC GAC AAC CCC GGC CAC GTG GAC TTC
Thr Ile Thr Ser Ala Ala Val Thr Cys Phe Trp Asn Asp His Gln Ile Asn Ile Ile Asp Asn Pro Gly His Val Asp Phe
                1335                1350                1365                1380                1395
ACG GTC GAG GTG GAG CGC TCG CTG CGC GTG CTC GAC GGC GCC GTC GCC GTG TTC GAC GGC AAG GAG GGC GTG GAG CCC CAG
Thr Val Glu Val Glu Arg Ser Leu Arg Val Leu Asp Gly Ala Val Ala Val Phe Asp Gly Lys Glu Gly Val Glu Pro Gln
        1410                1425                1440                1455                1470
TCG GAG ACC GTG TGG CGT CAG GCG GAC AAG TAC GAC GTC CCC CGC ATC TGC TTC GTC AAC AAG ATG GAC AAG CTG GGC GCG
Ser Glu Thr Val Trp Arg Gln Ala Asp Lys Tyr Asp Val Pro Arg Ile Cys Phe Val Asn Lys Met Asp Lys Leu Gly Ala
1485                1500                1515                1530                1545                1560
GAC TTC TAC TTC ACC GTC GAC ACG ATC GTG AAG CGC CTC GGC GCG CGT CCG CTT GTG ATG CAG CTG CCG ATC GGC GCC GAG
Asp Phe Tyr Phe Thr Val Asp Thr Ile Val Lys Arg Leu Gly Ala Arg Pro Leu Val Met Gln Leu Pro Ile Gly Ala Glu
            1575                1590                1605                1620                1635
AAC GAC TTC GTG GGC GTC GTC GAC CTG ATC TCC ATG AAG GCC TTC GTG TGG CCG GGC GAC GCC AAT GGG ATC GTC ACC ATG
Asn Asp Phe Val Gly Val Val Asp Leu Ile Ser Met Lys Ala Phe Val Trp Pro Gly Asp Ala Asn Gly Ile Val Thr Met
    1650                1665                1680                1695                1710
GGC GCC TCC TAC GAG ATC GAG ATC CGC CAG CTC CAG GAG AAG GCC GAG GAG GAG TAC CGC AAC GAG CTC GTC GAG GCC GTC
Gly Ala Ser Tyr Glu Ile Glu Ile Arg Gln Leu Gln Glu Lys Ala Glu Glu Glu Tyr Arg Asn Glu Leu Val Glu Ala Val
                1740                1755                1770                1785                1800
GCC GAG ACT TCC GAG GAG CTC ATG GAG AAG TAC CTC GAG GGC GAG GAG CTC ACC GTC GAG GAG ATC CAG GCC GGC GTG CGT
Ala Glu Thr Ser Glu Glu Leu Met Glu Lys Tyr Leu Glu Gly Glu Glu Leu Thr Val Glu Glu Ile Gln Ala Gly Val Arg
        1815                1830                1845                1860                1875
CAG CTG ACC GTG AAC GCC GAG GCT TAC CCG GTG TTC TGC GGC TCC GCG TTC AAG AAC CGT GGC GTG CAG CCG ATG CTG GAC
Gln Leu Thr Val Asn Ala Glu Ala Tyr Pro Val Phe Cys Gly Ser Ala Phe Lys Asn Arg Gly Val Gln Pro Met Leu Asp
1890                1905                1920                1935                1950                1965
GCC GTG GTC GCC TAC CTG CCG AAC CCG CTG GAC GCT GGC CCC GTC AAG GGC CAC GCC GTG AAC GAC GAG GAG GTG GTG CTG
Ala Val Val Ala Tyr Leu Pro Asn Pro Leu Asp Ala Gly Pro Val Lys Gly His Ala Val Asn Asp Glu Glu Val Val Leu
            1980                1995                2010                2025                2040
GAG CGC GAG GTG TCG AAG GAG GCC CCG TTC TCG GCG CTC GCC TTC AAG ATC GCC ACG CAC CCT TTC TTC GGC ACG CTG ACC
Glu Arg Glu Val Ser Lys Glu Ala Pro Phe Ser Ala Leu Ala Phe Lys Ile Ala Thr His Pro Phe Phe Gly Thr Leu Thr
    2055                2070                2085                2100                2115                2130
TTC ATC CGC GTG TAC TCC GGC CGC CTG GAG TCC GGT GCG CAG GTC CTC AAC GCC ACC AAG GGC AAG AAG GAG CGC ATC GGC
Phe Ile Arg Val Tyr Ser Gly Arg Leu Glu Ser Gly Ala Gln Val Leu Asn Ala Thr Lys Gly Lys Lys Glu Arg Ile Gly
                2145                2160                2175                2190                2205
AAG CTG TTC CAG ATG CAC GCC AAC AAG GAG AAC CCG GTG GAC GAG GTG GTC GCC GGC CAC ATC TAC GCC GTG ATC GGC CTC
Lys Leu Phe Gln Met His Ala Asn Lys Glu Asn Pro Val Asp Glu Val Val Ala Gly His Ile Tyr Ala Val Ile Gly Leu
        2220                2235                2250                2265                2280
AAG GAC ACC ACC ACG GGC GAC ACC CTG TGC GAT CCC GCG AAC CCG ATC ATC CTC GAG TCG ATG ACC TTC CCG GAG CCC GTG
Lys Asp Thr Thr Thr Gly Asp Thr Leu Cys Asp Pro Ala Asn Pro Ile Ile Leu Glu Ser Met Thr Phe Pro Glu Pro Val
2295                2310                2325                2340                2355                2370
ATC TCC GTG GCC ATC GAG CCG AAG ACC AAG GGT GAC CAG GAG AAG CTC TCC ACC GCC ATC CAG AAG CTC GTC GCC GAG GAC
Ile Ser Val Ala Ile Glu Pro Lys Thr Lys Gly Asp Gln Glu Lys Leu Ser Thr Ala Ile Gln Lys Leu Val Ala Glu Asp
```

FIG. 2. DNA sequence of *M. luteus str* operon. Boxed sequences, Promoter like sequences (see the text); underlined sequences, Shine-Dalgarno-like sequences; sequences marked with arrows, two dyad-symmetrical sequences, a probable transcriptional termination signal. The third letter A of the S7 termination codon (TAA) was overlapped with the first letter of the S12 initiation codon (ATG).

```
            2385                2400                2415                2430                2445
CCC ACG TTC CGC GTG AAC CTC AAC GAG GAG ACC GGT CAG ACC GAG ATC GGC GGC ATG GGC GAG CTC CAC CTG GAC GTG TTC
Pro Thr Phe Arg Val Asn Leu Asn Glu Glu Thr Gly Gln Thr Glu Ile Gly Gly Met Gly Glu Leu His Leu Asp Val Phe
    2460                2475                2490                2505                2520                2535
GTG GAC CGC ATG AAG CGC GAG TTC AAG GTC GAG GCC AAC GTG GGC AAG CCC CAG GTC GCG TAC CGC GAG ACC ATC AAG CGC
Val Asp Arg Met Lys Arg Glu Phe Lys Val Glu Ala Asn Val Gly Lys Pro Gln Val Ala Tyr Arg Glu Thr Ile Lys Arg
                2550                2565                2580                2595                2610
AAG GTC GAC AAG GTC GAC TAC ACG CAC AAG AAG CAG ACC GGC GGC TCC GGC CAG TTC GCC AAG GTG CAG CTG TCC TTC GAG
Lys Val Asp Lys Val Asp Tyr Thr His Lys Lys Gln Thr Gly Gly Ser Gly Gln Phe Ala Lys Val Gln Leu Ser Phe Glu
        2625                2640                2655                2670                2685
CCC CTG GAC ACG CCG AGG GGC ACG GTC TAC GAG TTC GAG AAC GCC ATC ACC GGC GGT CGC GTG CCC CGC GAG TAC ATC CCC
Pro Leu Asp Thr Pro Arg Gly Thr Val Tyr Glu Phe Glu Asn Ala Ile Thr Gly Gly Arg Val Pro Arg Glu Tyr Ile Pro
2700                2715                2730                2745                2760                2775
TCG GTG GAC GCG GGC ATC CAG GAC GCC ATG AAG TTC GGC GTG CTG GCC GGC TAC CCG ATG GTC CGC GTG AAG GCA ACC TCC
Ser Val Asp Ala Gly Ile Gln Asp Ala Met Lys Phe Gly Val Leu Ala Gly Tyr Pro Met Val Arg Val Lys Ala Thr Ser
            2790                2805                2820                2835                2850
CTC GAC GGT GCG TAC CAC GAC GTC GAC TCC TCG GAG ATG GCG TTC AGG ATC GCC GGC TCC CAG GCC TTC AAG GAG GGT GTC
Leu Asp Gly Ala Tyr His Asp Val Asp Ser Ser Glu Met Ala Phe Arg Ile Ala Gly Ser Gln Ala Phe Lys Glu Gly Val
    2865                2880                2895                2910                2925                2940
CGC AAG GCC ACC CCG ATC ATC CTC GAG CCG CTG ATG GCC GTG GAG GTC CGC ACC CCC GAG GAG TTC ATG GGC GAC GTC ATC
Arg Lys Ala Thr Pro Ile Ile Leu Glu Pro Leu Met Ala Val Glu Val Arg Thr Pro Glu Glu Phe Met Gly Asp Val Ile
                2955                2970                2985                3000                3015
GGC GAC CTG AAC TCC CGC CGC GGC CAG ATC CAG ATC CAG TCC ATG GAG GAC GCC ACC GGC GTG AAG GTG GTC AAC GCC CTC
Gly Asp Leu Asn Ser Arg Arg Gly Gln Ile Gln Ile Gln Ser Met Glu Asp Ala Thr Gly Val Lys Val Val Asn Ala Leu
        3030                3045                3060                3075                3090
GTG CCG CTG TCG GAG ATG TTC GGC TAC ATC GGC GAC CTG CGT TCC AAG ACG CAG GGC CGC GCG GTG TAC TCG ATG ACC TTC
Val Pro Leu Ser Glu Met Phe Gly Tyr Ile Gly Asp Leu Arg Ser Lys Thr Gln Gly Arg Ala Val Tyr Ser Met Thr Phe
3105                3120                3135                3150                3165                3180
CAC TCC TAC GCC GAG GTC CCC AAG GCC GTG GCG GAC GAG ATC GTC CAG AAG TCC CAG GGC GAG TGA CC      GAGGTCACTG
His Ser Tyr Ala Glu Val Pro Lys Ala Val Ala Asp Glu Ile Val Gln Lys Ser Gln Gly Glu *
      3190        3200        3210        3220        3230        3240        3250        3260        3270
CTCTGAACGA  CCTGCCCCGG  GTCGGCCGCC  GCAGCCGCGC  GCCGGGCGAC  CGGACCGCCC  TCCCCGGCGC  GGCGGTTCGC  GATCAGGACG
      3280        3290        3300        3310        3320        3330        3340        3350        3360
GGTCAGGTCC  GGTCCGCTTC  CGGCGGTCCC  GACGGCCTGG  AACCGGACTG  GCCCTCGTGA  GGCCGGGATT  TCCCCAACCC  CGTACCATCC
      3370        3380        3390        3400        3410        3420        3430        3440  EF-Tu Start
GAGTAGACTC  AGTCCAAGTT  GTCAGCACGT  CCCAGGCGGC  TGAGCAGTCT  TCTTGACGAT  GAACCAGTTC  TTAGGAGGAA  CTA     GTG
                                                                                                        Met
GCA AAG GCA AAG TTC GAG CGG ACG AAG GCG CAC GTC AAC ATC GGC ACC ATC GGC CAC GTT GAC CAC GGC AAG ACC ACG CTG
Ala Lys Ala Lys Phe Glu Arg Thr Lys Ala His Val Asn Ile Gly Thr Ile Gly His Val Asp His Gly Lys Thr Thr Leu
3530                3545                3560                3575                3590                3605
ACC GCC GCC ATC TCG AAG GTC CTG TAC GAC AAG TAC CCG GAC CTG AAT GAG GCC CGT GAC TTC GCG ACG ATC GAT TCC GCC
Thr Ala Ala Ile Ser Lys Val Leu Tyr Asp Lys Tyr Pro Asp Leu Asn Glu Ala Arg Asp Phe Ala Thr Ile Asp Ser Ala
            3620                3635                3650                3665                3680
CCC GAG GAG CGT CAG CGC GGC ATC ACC ATC AAC ATC TCC CAC GTG GAG TAC CAG ACC GAG AAG CGT CAC TAC GCC CAC GTG
Pro Glu Glu Arg Gln Arg Gly Ile Thr Ile Asn Ile Ser His Val Glu Tyr Gln Thr Glu Lys Arg His Tyr Ala His Val
    3695                3710                3725                3740                3755                3770
GAC GCC CCC GGT CAC GCC GAC TAC ATC AAG AAC ATG ATC ACC GGC GCC GCT CAG ATG GAC GGC GCG ATC CTC GTG GTC GCC
Asp Ala Pro Gly His Ala Asp Tyr Ile Lys Asn Met Ile Thr Gly Ala Ala Gln Met Asp Gly Ala Ile Leu Val Val Ala
                3785                3800                3815                3830                3845
GCT ACC GAC GGC CCG ATG GCC CAG ACC CGT GAG CAC GTG CTC CTG GCC CGC CAG GTC GGC GTG CCG GCC CTG CTC GTG GCC
Ala Thr Asp Gly Pro Met Ala Gln Thr Arg Glu His Val Leu Leu Ala Arg Gln Val Gly Val Pro Ala Leu Leu Val Ala
        3860                3875                3890                3905                3920
CTG AAC AAG TCG GAC ATG GTG GAG GAC GAG GAG CTC CTC GAG CGT GTC GAG ATG GAG GTC CGG CAG CTG CTG TCC TCC AGG
Leu Asn Lys Ser Asp Met Val Glu Asp Glu Glu Leu Leu Glu Arg Val Glu Met Glu Val Arg Gln Leu Leu Ser Ser Arg
3935                3950                3965                3980                3995
AGC TTC GAC GTC GAC GAG GCC CCG GTC ATC CGC ACC TCC GCT CTG AAG GCC CTC GAG GGC GAC CCC CAG TGG GTC AAG TCC
Ser Phe Asp Val Asp Glu Ala Pro Val Ile Arg Thr Ser Ala Leu Lys Ala Leu Glu Gly Asp Pro Gln Trp Val Lys Ser
        4025                4040                    4055                4070                4085
GTC GAG GAC CTC ATG GAT GCC GTG GAC GAG TAC ATC CCG GAC CCG GTG CGC GAC AAG GAC AAG CCG TTC CTG ATG CCG ATC
Val Glu Asp Leu Met Asp Ala Val Asp Glu Tyr Ile Pro Asp Pro Val Arg Asp Lys Asp Lys Pro Phe Leu Met Pro Ile
    4100                4115                4130                4145                4160
GAG GAC GTC TTC ACG ATC ACC GGC CGT GGC ACC GTG GTG ACC GGT CGC GCC GAG CGC GGC ACC CTG AAG ATC AAC TCC GAG
Glu Asp Val Phe Thr Ile Thr Gly Arg Gly Thr Val Val Thr Gly Arg Ala Glu Arg Gly Thr Leu Lys Ile Asn Ser Glu
            4190                    4205                4220                4235                4250
GTC GAG ATC GTC GGC ATC CGC GAC GTG CAG AAG ACC ACT GTC ACC GGC ATC GAG ATG TTC CAC AAG CAG CTC GAC GAG GCC
Val Glu Ile Val Gly Ile Arg Asp Val Gln Lys Thr Thr Val Thr Gly Ile Glu Met Phe His Lys Gln Leu Asp Glu Ala
        4265                4280                4295                4310                4325
TGG GCC GGC GAG AAC TGC GGT CTG CTC GTG CGC GGT CTG AAG CGC GAC GAC GTC GAG CGC GGC CAG GTG CTG GTG GAG CCG
Trp Ala Gly Glu Asn Cys Gly Leu Leu Val Arg Gly Leu Lys Arg Asp Asp Val Glu Arg Gly Gln Val Leu Val Glu Pro
4340                4355                4370                4385                4400                4415
GGC TCC ATC ACC CCG CAC ACC AAC TTC GAG GCG AAC GTC TAC ATC CTG TCC AAG GAC GAG GGT GGG CGT CAC ACC CCG TTC
Gly Ser Ile Thr Pro His Thr Asn Phe Glu Ala Asn Val Tyr Ile Leu Ser Lys Asp Glu Gly Gly Arg His Thr Pro Phe
            4430                4445                4460                4475                4490
TAC TCG AAC TAC CGC GCG CAG TTC TAC TTC CGC ACC ACC GAC GTC ACC GGC GTC ATC ACG CTG CCC GAG GGC ACC GAG ATG
Tyr Ser Asn Tyr Arg Pro Gln Phe Tyr Phe Arg Thr Thr Asp Val Thr Gly Val Ile Thr Leu Pro Glu Gly Thr Glu Met
    4505                4520                4535                4550                4565                4580
GTC ATG CCC GGC GAC ACC ACC GAG ATG TCG GTC GAG CTC ATC CAG CCG ATC GCC ATG GAG GAG GGC CTC GGC TTC GCC ATC
Val Met Pro Gly Asp Thr Thr Glu Met Ser Val Glu Leu Ile Gln Pro Ile Ala Met Glu Glu Gly Leu Gly Phe Ala Ile
                4595                4610                4625                  4640        4650        4660
CGC GAG GGT GGC CGC ACC GTG GGC TCC GGC CGC GTC ACC AAG ATC ACC AAG TGA     TCTGAC  CGCATCCGTC  CTTGTGGCTC
Arg Glu Gly Gly Arg Thr Val Gly Ser Gly Arg Val Thr Lys Ile Thr Lys *
      4670        4680        4690        4700        4710        4720        4730        4740        4750
CGTTGAGCCG  CTGACGTCGA  CGCCCGTCTC  GGGTCTCCGG  GAGGCGGGGG  CGTCCCGCTG  CTCCGACCGA  CGGCGCGGCG  GCGGCCCGTA
      4760        4770        4780        4790        4800        4810        4820        4830        4890
GGCTGAGGAC  AATGCCTAAC  TGCTCATCCC  AGTGGATGCT  CGCCGTGGCC  GCCGTGGCCG  TGCGCTGACG  CTCGCGGACC  TTCCGTCCCG
      4850        4860        4870        4880        4890        4900        4910        4920        4930
AGATTCGACC  GTGCCCGGCC  GATCCGGCGG  GGCGAATGCG  GCCGGGACGC  TGAACGATGA  GCTTGCCAGC  CGGCGACTGT  TCGCATAGCG
      4940
ATTAGGCAGAC
```

these four genes belong to one transcriptional unit or whether the long spacers mentioned before contain the promoter sequences. The sequences AGGACGTG (976 to 983) and CGCAACATC (1005 to 1114) in the spacer region between S7 and EF-G could be promoter sequences, because of some similarities to the promoter like sequences AGGATTGT and GGCAAATTG in the 5′-upstream region of S12 (see above). There existed two dyad-symmetrical

```
                      15                                      45                                      75
EF-G                  GTG CTG ACT GAC CTG CAC AAG GTC CGC AAC ATC GGC ATC ATG GCT CAC ATC GAT GCC GGC AAG ACC ACC
                      MET Leu Thr Asp Leu His Lys Val Arg Asn Ile Gly Ile Met Ala His Ile Asp Ala Gly Lys Thr Thr
EF-Tu GTG GCA AAG GCA AAG TTC GAG CGG ACG AAG GCG CAC GTC AAC ATC GGC ACC ATC GGC CAC GTT GAC CAC GGC AAG ACC ACG
      Met Ala Lys Ala Lys Phe Glu Arg Thr Lys Ala His Val Asn Ile Gly Thr Ile GLy His Val Asp His Gly Lys Thr Thr

                        105                                         135                                     165
ACC ACC GAG CGC CAT CTG TTC TAC ACG GGT GTC AAC CAC AAG CTG GGC GAG ACC CAC GAC GGC GGC GCG ACG ACG GAC TGG ATG GAG
Thr Thr Glu Arg His Leu Phe Tyr Thr Gly Val Asn His Lys Leu Gly Glu Thr His Asp Gly Gly Ala Thr Thr Asp Trp Met Glu
CTG ACC GCC GCC ATC TCG AAG GTC CTG TAC GAC AAG TAC CCG GAC CTG AAT GAG GCC CGT GAC TTC GCG ACG ATC GAT TCC GCC CCC
Leu Thr Ala Ala Ile Ser Lys Val Leu Tyr Asp Lys Tyr Pro Asp Leu Asn Glu Ala Arg Asp Phe Ala Thr Ile Asp Ser Ala Pro

                                195                                     225                                     255
CAG GAG AAG GAG CGC GGC ATC ACC ATC ACC TCC GCC GCG GTG ACC TGC TTC TGG AAC GAC CAC CAG ATC AAC ATC ATC GAC AAC CCC
Gln Glu Lys Glu Arg Gly Ile Thr Ile Thr Ser Ala Ala Val Thr Cys Phe Trp Asn Asp His Gln Ile Asn Ile Ile Asp Asn Pro
GAG GAG CGT CAG CGC GGC ATC ACC ATC AAC ATC TCC CAC GTG GAG TAC CAG ACC GAG AAG CGT CAC TAC GCC CAC GTG GAC GCC CCC
Glu Glu Arg Gln Arg Gly Ile Thr Ile Asn Ile Ser His Val Glu Tyr Gln Thr Glu Lys Arg His Tyr Ala His Val Asp Ala Pro

                                    285                                     315
GGC CAC GTG GAC TTC ACG GTC GAG GTG GAG CGC TCG CTG CGC GTG CTC GAC GGC GCC GTC GCC GTG TTC GAC GGC AAG GAG GGC GTG
Gly His Val Asp Phe Thr Val Glu Val Glu Arg Ser Leu Arg Val Leu Asp Gly Ala Val Ala Val Phe Asp Gly Lys Glu Gly Val
GGT CAC GCC GAC TAC ATC AAG AAC ATG ATC ACC GGC GCC GCT CAG ATG GAC GGC GCG ATC CTC GTG GTC GCC GCT ACC GAC GGC CCG
Gly His Ala Asp Tyr Ile Lys Asn Met Ile Thr Gly Ala Ala Gln Met Asp Gly Ala Ile Leu Val Val Ala Ala Thr Asp Gly Pro

345                                     375                                     405
GAG CCC CAG TCG GAG ACC GTG TGG CGT CAG GCG GAC AAG TAC GAC GTC CCC CGC ATC TGC --- TTC GTC AAC AAG ATG GAC AAG CTG
Glu Pro Gln Ser Glu Thr Val Trp Arg Gln Ala Asp Lys Tyr Asp Val Pro Arg Ile Cys     Phe Val Asn Lys Met Asp Lys Leu
ATG GCC CAG ACC CGT GAG CAC GTG CTC CTG GCC CGC CAG GTC GGC GTG CCG GCC CTG CTC GTG GCC CTG AAC AAG TCG GAC ATG GTG
Met Ala Gln Thr Arg Glu His Val Leu Leu Ala Arg Gln Val Gly Val Pro Ala Leu Leu Val Ala Leu Asn Lys Ser Asp Met Val
```

FIG. 3. Comparison of the amino acid sequences of *M. luteus* EF-G and EF-Tu. One gap was introduced to the sequence of EF-G. The sites for identical amino acids and those for conservative amino acid substitutions were boxed with solid and dotted lines, respectively.

sequences, probable transcriptional termination signals, in the 3′-downstream region of the EF-Tu gene (Fig. 2).

The amino acid sequence homology between *M. luteus* and *E. coli* was 70% in S12, 58% in S7, 60% in EF-G, and 71% in EF-Tu. The S7 protein from *E. coli* K-12 is 178 amino acids long, whereas that from the *E. coli* B strain is 154 (30), where 24 amino acids from the C terminus are deleted. The *M. luteus* S7 also lacked 21 amino acids from the C terminus, suggesting that these C-terminal region are not essential for ribosomal function. A partial amino acid sequence homology was observed between *M. luteus* EF-G and EF-Tu in the N-terminal regions (Fig. 3), as in the case of *E. coli* (35).

The mean G+C content of the entire DNA sequence determined in this study was 67%, which was still a few percent lower than the total genomic G+C content of this bacterium (74%). This may reflect amino acid content of the proteins in the streptomycin operon. The proteins are somewhat higher in amino acids coded by "AU" codons such as lysine (AAN) and lower in those coded by "GC" codons such as alanine (GCN) and glycine (GGN), compared with the average bacterial proteins (12). Therefore, there must be other higher-G+C regions in the *M. luteus* genome.

The facts described above show that the structure and organization of the *str* operon are well conserved between the two species. Since, phylogenetically, the gram-positive *M. luteus* and the gram-negative *E. coli* separated more than a billion years ago (6, 10), the fundamental organization of the *str* operon was established at an early time, before the separation of the gram-positive and gram-negative bacteria.

Codon usage. Since the DNA sequence of most parts of the *E. coli str* operon, except for a part of the gene for S7, has been reported (23, 33, 35), we compared the codons used in the *str* operon in *M. luteus* and *E. coli*. Reflecting the high G+C content of the genomic DNA, the G+C content of the protein-coding region of *M. luteus* was much higher (66%) than that of *E. coli* (52%). In Table 1, the codon usage in the *str* operon of *M. luteus* (total, 1,382 codons) and *E. coli* analyzed (1,339 codons) are compared. An outstanding feature is a very high frequency, 95% (1,309 of 1,382), of the codons that have G or C at the third codon position in *M. luteus*, in contrast to only 52% (698 of 1,339) in *E. coli*. The G+C content of the first and the second positions in *M. luteus* were 64.1 and 40.2%, respectively, which were almost the same as the values of 63 and 40.0% in *E. coli* (Table 2). The published data on G+C content of three codon positions in A+T-rich bacterium, *Mycoplasma capricolum* (G+C, 25%) (20) is included in Table 2 for comparison (see below). The different levels of G+C content in different codon positions can be the consequence of selective constraints that eliminate functionally deleterious mutations, as discussed previously (21). The third positions of the codons are the most variable, because most synonymous codon substitutions occur in the third position. These results support the view that the strong GC pressure replacing AT pairs with GC pairs has been exerted on the *M. luteus* genome during evolution. G and C bias in codon usage was also reported for several genes of *Streptomyces* species (2, 25, 29, 34), *Pseudomonas aeruginosa* (3), and *Thermus thermophilus* (11) which are also high in genomic G+C.

To show more details in the preferential use of G- and C-rich codons in *M. luteus*, the codon substitutions at 1,291 homologous sites in four genes of the *str* operon were compared between *M. luteus* and *E. coli* (Table 3). About 100 sites in *M. luteus* that include a large part of the S7 gene could not be compared, because of a lack of information on the DNA sequence in *E. coli*. Some gaps and deletions or additions between the two species were also omitted for comparisons. Table 3 summarizes the codon substitution patterns between the two species. A total of 375 codons were identical. Among 916 substitutions, there existed 468 synonymous (silent) codon substitutions, 112 conservative amino acid substitutions (see Table 3, footnote *e*), and 336 other substitutions. Almost all the synonymous codon substitutions (99%; 464 of 468) occurred at the third position, where 82.7% (387 of 468) of them resulted in G or C richness in *M. luteus*; 15% (71 of 468) were without a gain or loss of G or C and only 2.3% (10 of 468) accompanied a loss of G or C. For example, 51 GGA or GGU codons for glycine in *E. coli* were replaced by GGC in *M. luteus*, 41 CGU for arginine were replaced by CGG/C, 38 AAA for lysine were replaced by AAG, and 38 GAA for glutamic acid were replaced by GAG. A few synonymous codon substitutions occurred at the first position of the codon (three instances) and the first, second, and third positions (one instance), all of which gained G or C, compared with *E. coli*. By these synonymous substitutions, *M. luteus* gained 377 bases of G and C, compared with *E. coli*. Among 112 conservative amino acid substitutions, which occurred primarily by changing the codon at the first and third positions (Table 3), 60 codon changes occurred so as to cause higher G+C content in *M. luteus*. For example, many UCU codons for serine and AUC codons for isoleucine in *E. coli* were replaced by ACC for threonine and CUG for leucine, respectively, in *M. luteus*. Thirty-six substitutions were without a gain or loss of G+C, and sixteen substitutions accompanied a loss of G+C.

Among 336 other codon substitutions, a gain of G+C was observed in 156 codons of *M. luteus*, whereas a loss of G+C was observed in 75 codons.

Among 916 substituted codons of a total of 1,291 codons compared, 603 codons (66%, 603 of 916) gained G+C, and 101 (11%, 101 of 916) lost G+C. The total G+C gain and loss in all nucleotide sites (2,748) in the substituted codons was 670 and 113, respectively.

In *M. luteus*, many codons were completely absent among 1,382 codons examined. These were UUU (Phe), UUA (Leu), UUG (Leu), CUA (Leu), AUA (Ile), GUA (Val), UCU (Ser), CCA (Pro), ACA (Thr), UAU (Tyr), CAA (Gln), AAA (Lys), GAA (Glu), UGU (Cys), AGU (Ser), AGA (Arg), and GGA (Gly), all of which have U or A at the third position of codons, with one exception of UUG leucine (Table 1). A complete absence of UUA (and UUG) leucine codons leads us to speculate the deletion of the corresponding tRNA having anticodon UAA and CAA. Codon UUA and UUG were probably converted by GC pressure to CUC or CUG, for the mutation of the first nucleotide U to C is silent. The above unused codons are used in high frequencies in an A+T-rich bacterium, *Mycoplasma capricolum* (G+C, 25%; see below).

In contrast to the GC-biased codon usage in *M. luteus*, a strongly AU-biased codon usage has been shown in *M. capricolum*, in which the codons for ribosomal protein genes in the *spc* operon were compared with the corresponding codons in *E. coli* (4). Since the sequence of the *str* operon in *Mycoplasma capricolum* and that of the *spc* operon in *M. luteus* are not available at present, the homologous codon sites between these two bacteria could not be directly compared. Even so, the comparison of the codon usage tables between the two species (for the codon usage table of *Mycoplasma capricolum*, see reference 20) clearly shows a much sharper contrast than the comparison between *M. luteus* and *E. coli*. This is because of intermediary G+C content of *E. coli* (51.6%) between that of *M. luteus* (74%) and *Mycoplasma capricolum* (25%). For example, in *M. luteus* 97% (92 of 95) of threonine codons are ACC and ACG, whereas in *Mycoplasma capricolum* 99.5% of threonine codons are ACA and ACU; all codons for lysine are AAG in *M. luteus*, whereas 90% are AAA in *Mycoplasma capricolum*; the UAU (isoleucine) codon was not detected in *M.*

TABLE 1. Codon usage in *str* operon of *M. luteus* and *E. coli*

Codon	No. used by:	
	M. luteus	E. coli
Phe (UUU)	0	7
Phe (UUC)	43	37
Leu (UUA)	0	2
Leu (UUG)	0	2
Leu (CUU)	1	4
Leu (CUC)	35	5
Leu (CUA)	0	0
Leu (CUG)	54	78
Ile (AUU)	1	14
Ile (AUC)	73	72
Ile (AUA)	0	0
Met (AUG)	38 (1)[a]	36 (3)[a]
Val (GUU)	1	70
Val (GUC)	67	1
Val (GUA)	0	39
Val (GUG)	74 (3)[a]	12 (1)[a]
Ser (UCU)	0	27
Ser (UCC)	36	18
Ser (UCA)	1	3
Ser (UCG)	16	0
Pro (CCU)	5	6
Pro (CCC)	27	2
Pro (CCA)	0	5
Pro (CCG)	35	52
Thr (ACU)	3	34
Thr (ACC)	68	40
Thr (ACA)	0	5
Thr (ACG)	24	0
Ala (GCU)	6	48
Ala (GCC)	83	6
Ala (GCA)	3	25
Ala (GCG)	21	35
Tyr (UAU)	0	8
Tyr (UAC)	37	28
Stop (UAA)	1	3
Stop (UAG)	0	0
His (CAU)	1	7
His (CAC)	30	24
Gln (CAA)	0	1
Gln (CAG)	47	39
Asn (AAU)	3	2
Asn (AAC)	43	41
Lys (AAA)	0	69
Lys (AAG)	89	20
Asp (GAU)	5	16
Asp (GAC)	71	55
Glu (GAA)	0	90
Glu (GAG)	113	21
Cys (UGU)	0	4
Cys (UGC)	6	6
Stop (UGA)	3	1
Trp (UGG)	8	8
Arg (CGU)	23	67
Arg (CGC)	61	16
Arg (CGA)	1	0
Arg (CGG)	2	1
Ser (AGU)	0	6
Ser (AGC)	1	3
Arg (AGA)	0	0
Arg (AGG)	3	0
Gly (GGU)	15	76
Gly (GGC)	100	38
Gly (GGA)	0	2
Gly (GGG)	4	2

[a] Initiation codon.

TABLE 2. Nucleotide composition at three positions of codons in *M. luteus*, *E. coli*, and *Mycoplasma capricolum*

Codon position	*M. luteus*			*E. coli*			*Mycoplasma capricolum*[a]		
	No.[b]	%	G+C%[c]	No.[b]	%	G+C%[d]	No.[b]	%	G+C%[e]
First			64.1			63.0			43.5
U	151	10.9		154	11.5		51	16.6	
C	322	23.3		307	23.0		34	11.0	
A	346	25.0		342	25.5		123	39.9	
G	563	40.8		536	40.0		100	32.5	
Second			40.2			40.0			35.7
U	387	28.0		379	28.3		93	30.2	
C	328	23.7		306	22.8		63	20.5	
A	440	31.8		424	31.7		105	34.1	
G	227	16.5		230	17.2		47	15.2	
Third			94.7			52.2			9.1
U	64	4.6		396	29.5		116	37.7	
C	781	56.5		392	29.3		17	5.5	
A	9	0.7		245	18.3		164	53.2	
G	528	38.2		306	22.9		11	3.6	

[a] *Mycoplasma capricolum* ribosomal proteins S8 and L6 (20).
[b] Number of occurrences.
[c] Total G+C%, 66.3%.
[d] Total G+C%, 51.7%.
[e] Total G+C%, 29.4%.

luteus, whereas this codon is used abundantly in *Mycoplasma capricolum*. A similar sharp contrast can be seen in most of other codons between these two species.

In *E. coli* or *B. subtilis* (G+C, 45%), AUG is the regular initiation codon, although GUG is used rarely. In *M. luteus*, S12, EF-G and EF-Tu used GUG as initiation codon, although S7 starts with AUG (Fig. 2). In contrast, GUG was used for only one of the EF-Tu's (1) in *E. coli*, showing that the AUG initiation codon for S12 and EF-G are replaced by GUG in *M. luteus*. In other G+C-rich bacteria belonging to the genus *Streptomyces*, three genes start with GUG among seven genes so far analyzed (34). Thus, the G+C-rich bacteria including *M. luteus* seems to use GUG initiation codons much more frequently than *E. coli* or *B. subtilis*. Ghosh et al. (7) have shown that GUG can be an initiation codon but its efficiency is only about one-third that of AUG in *E. coli* in an in vitro system. Reddy et al. (24) showed that UUG can also be used for an initiation codon and the replacement TTG (UUG) of the *E. coli* adenylate cyclase gene (*cya*) by GTG doubled the *cya* gene expression; the replacement by ATG greatly enhanced the expression of this gene so that the cells become lethal. These facts suggest that AUG is a much more efficient initiation codon than GUG or UUG in *E. coli*. Since ribosomal proteins and elongation factors occur in large amounts in the cell, the initiation codon GUG must be the efficient initiation codon in *M. luteus*, in contrast to *E. coli*. In *Mycoplasma capricolum* and *E. coli*, UAA is a predominant termination codon for ribosomal protein genes, whereas *M. luteus* seems to use UGA predominantly as seen for S7, EF-G, and EF-Tu; only S12 terminated with UAA. The biased uses of GUG as initiation codons and UGA as termination codons in *M. luteus* suggest the conversion of AUG to GUG and UAA to UGA, respectively, by GC pressure.

All these facts indicate that the strong GC pressure has caused *M. luteus* to discriminate against A and U and to use

TABLE 3. Codon substitutions in *str* operon (916 codons) between *M. luteus* and *E. coli*

Type of substitution	No. of substituted codons at:				No. of changed codons	Net G+C gain[a]
	Codon position			Two or three positions		
	1	2	3			
Synonymous[b]						
Gain[c]	3	0	383	1	387	+387
+/−[d]	0	0	71	0	71	0
Loss[c]	0	0	10	0	10	−10
Conservative[e]						
Gain[c]	7	0	8	45	60	+72
+/−[d]	11	6	5	14	36	0
Loss[c]	1	0	0	15	16	−20
Other						
Gain[c]	12	8	2	134	156	+211
+/−[d]	12	5	5	83	105	0
Loss[c]	13	5	0	57	75	−83
Total	59	24	484	349	916	+557

[a] Net G+C gained in *M. luteus* as compared with *E. coli*.
[b] Synonymous (silent) codon substitution.
[c] Number of codons that gained or lost G or C, as compared with *E. coli*.
[d] Number of codons without a gain or loss of G or C.
[e] Conservative amino acid substitution includes Lys/Arg, Leu/Ile, Leu/Val, Ser/Thr, Ala/Gly, Gln/Asn, Glu/Asp, and Phe/Tyr.

G and C preferentially, in sharp contrast to the preferential use of A and U in *Mycoplasma capricolum*. Thus, we conclude that the codon choice pattern in a given bacterium is largely influenced by the GC/AT-biased pressure exerted on the entire genome. The high G+C content (69%) of a thermophile bacterium, *T. thermophilus*, has been explained as an adaptation to high temperature (82°C) (13). However, *Micrococcus* spp. and other G+C-rich bacteria such as *Streptomyces* spp. and *Pseudomonas* spp. are not thermophiles.

It should be noted that the codon choice pattern of the *Acanthamoeba castellanii* actin gene (22) is very similar to that of *M. luteus*, including the absence of codons UUA (Leu), UUG (Leu), CUA (Leu), AUA (Ile), GUA (Val), UCA (Ser), CCA (Pro), ACA (Thr), AAA (Lys), GAA (Glu), AGU (Ser), and AGA (Arg). This would suggest that the GC pressure similar to that in eubacteria is also exerted on the high genomic G+C eucaryotes.

ACKNOWLEDGMENTS

We thank T. H. Jukes for valuable comments and critical reading of the manuscript. Thanks are also due to the Takara Shuzo Co., Ltd. for supply of deoxy-7-deazaguanosine triphosphate.

This work was supported by grants from the Ministry of Education, Science and Culture, Japan.

LITERATURE CITED

1. **An, G., and F. D. Jriensen.** 1980. The nucleotide sequence of *tuf*B and four nearby tRNA structural genes of *Escherichia coli*. Gene **12**:33–39.
2. **Bibb, M. J., M. J. Bibb, L. M. Word, and S. N. Cohen.** 1985. Nucleotide sequences of three antibiotic resistance genes indigenous to *Streptomyces*. Mol. Gen. Genet. **199**:26–36.
3. **Brown, N. L., S. J. Ford, R. D. Primore, and D. C. Fritzinger.** 1983. Nucleotide sequence of a gene from the *Pseudomonas* transposon Tn*501* encoding mercuric reductase. Biochemistry **22**:4089–4095.
4. **Cerretti, D. P., D. Dean, G. R. Davis, D. M. Bedwell, and M. Nomura.** 1983. The *spc* ribosomal protein operon of *Escherichia coli*: sequence and cotranscription of the ribosomal protein genes and a protein export gene. Nucleic Acids Res. **11**:2599–2616.
5. **Chen, E. Y., and P. H. Seeburg.** 1985. Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. DNA **4**:165–170.
6. **Dekio, S., R. Yamasaki, J. Jidoi, H. Hori, and S. Osawa.** 1984. Secondary structure and phylogeny of *Staphylococcus* and *Micrococcus* 5S rRNAs. J. Bacteriol. **159**:233–237.
7. **Ghosh, H. P., D. Söll, and H. G. Khorana.** 1967. Initiation of protein synthesis *in vitro* as studied by using ribonucleotides with repeating sequences as messengers. J. Mol. Biol. **25**: 275–298.
8. **Godson, G. N., and D. Vapnek.** 1973. A simple method of preparing large amounts of ϕ χ 174 RSI supercoiled DNA. Biochim. Biophys. Acta **299**:516–520.
9. **Holmes, D. S., and M. Quigley.** 1981. A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. **114**: 193–197.
10. **Hori, H., and S. Osawa.** 1986. Evolutionary change in 5S rRNA secondary structure and a phylogenic tree of 352 5S rRNA species. BioSystems **19**:163–172.
11. **Jaurin, B., and S. N. Cohen.** 1985. *Streptomyces* contain *Escherichia coli*-type A+T-rich promoters. Gene **39**:191–201.
12. **Jukes, T. H., and V. Bhushan.** 1986. Silent nucleotide substitutions and G+C content of some mitochondrial and bacterial genes. J. Mol. Evol. **24**:39–44.
13. **Kagawa, Y., H. Nojima, N. Nukiwa, M. Ishizuka, T. Nakajima, T. Yasuhara, T. Tanaka, and T. Oshima.** 1984. High guanine plus cytosine content in the third letter of codons of an extreme thermophile. J. Biol. Chem. **259**:2956–2960.
14. **Kloos, W. E.** 1969. Factors affecting transformation of *Micrococcus lysodeikticus*. J. Bacteriol. **98**:1397–1399.
15. **Kloos, W. E., and L. M. Schultes.** 1969. Transformation in *Micrococcus lysodeikticus*. J. Gen. Microbiol. **55**:307–317.
16. **Korneluk, R. G., F. Quan, and R. A. Gravel.** 1985. Rapid and reliable sequencing of double-stranded DNA. Gene **40**:317–323.
17. **Maniatis, T., R. C. Hardison, E. Lacy, J. Lauer, C. O'Connell, D. Quon, D. K. Sim, and A. Efstratiadis.** 1978. The isolation of structural genes from libraries of eucaryotic DNA. Cell **15**: 687–701.
18. **Mizusawa, S., S. Nishimura, and F. Seela.** 1986. Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. Nucleic Acids Res. **14**:1319–1324.
19. **Moran, C. P., W. C. Johnson, Jr., and R. Losick.** 1982. Close contacts between sigma^{37}-RNA polymerase and a *Bacillus subtilis* chromosomal promoter. J. Mol. Biol. **162**:709–713.
20. **Muto, A., Y. Kawauchi, F. Yamao, and S. Osawa.** 1984. Preferential use of A- and U-rich codons for *Mycoplasma capricolum* ribosomal proteins S8 and L6. Nucleic Acids Res. **12**: 8209–8217.
21. **Muto, A., and S. Osawa.** 1987. The guanine and cytosine content of genomic DNA and bacterial evolution. Proc. Natl. Acad. Sci. USA **84**:166–169.
22. **Nellen, W., and D. Gallwitz.** 1982. Actin genes and actin messenger RNA in *Acanthamoeba castellanii*: nucleotide sequence of the split actin gene I. J. Mol. Biol. **159**:1–18.
23. **Post, L. E., and M. Nomura.** 1980. DNA sequences from the *str* operon of *Escherichia coli*. J. Biol. Chem. **255**:4660–4666.
24. **Reddy, P., A. Peterkofsky, and K. McKenney.** 1985. Translational efficiency of the *Escherichia coli* adenylate cyclase gene: mutating the UUG initiation codon to GUG or AUG results in increased gene expression. Proc. Natl. Acad. Sci. USA **82**: 5656–5660.
25. **Robbins, P. W., R. B. Trimble, D. F. Wirth, C. Hering, F. Maley, G. F. Maley, R. Das, B. W. Gibons, N. Royal, and K. Biemann.** 1984. Primary structure of *Streptomyces* enzyme endo-β-acetylglucosaminidase H. J. Biol. Chem. **259**:7577–7583.
26. **Smith, H. O., and D. B. Danner.** 1981. Genetic transformation. Annu. Rev. Biochem. **50**:41–68.
27. **Sueoka, N.** 1962. On the genetic basis of variation and heterogeneity of DNA base composition. Proc. Natl. Acad. Sci. USA **48**:582–592.
28. **Tamaoka, J., and K. Komagata.** 1984. Determination of DNA base composition by reversed-phase high-performance liquid chromatography. FEMS Microbiol. Lett. **25**:125–128.
29. **Thompson, C. J., and G. S. Gray.** 1983. Nucleotide sequence of a streptomycete aminoglycoside phosphotransferase gene and its relationship to phosphotransferases encoded by resistance plasmids. Proc. Natl. Acad. Sci. USA **80**:5190–5194.
30. **Tritsch, D., J. Reinbolt, and B. Wittmann-Liebold.** 1977. The primary structure of ribosomal proteins S7 from *Escherichia coli* strains K and B. FEBS Lett. **77**:89–93.
31. **Westpheling, J., M. Ranes, and R. Losick.** 1985. RNA polymerase heterogeneity in *Streptomyces coelicolor*. Nature (London) **313**:22–27.
32. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.
33. **Yokota, T., H. Sugisaki, M. Takanami, and Y. Kaziro.** 1980. The nucleotide sequence of the cloned *tuf*A gene of *Escherichia coli*. Gene **12**:25–31.
34. **Zalacain, M., A. González, M. C. Guerrero, R. J. Mattaliano, F. Malpartida, and A. Jiménez.** 1986. Nucleotide sequence of the hygromycine B phosphotransferase gene from *Streptomyces hygroscopicus*. Nucleic Acids Res. **14**:1565–1581.
35. **Zengel, J. M., R. H. Arch, and L. Lindahal.** 1984. The nucleotide sequence of *Escherichia coli fus* gene, coding for elongation factor G. Nucleic Acids Res. **12**:2181–2192.