Volume 12 Number 21 1984 Nucleic Acids Research

Preferential use of A- and U-rich codons for *Mycoplasma capricolum* ribosomal proteins S8 and L6

Akira Muto, Yasushi Kawauchi*, Fumiaki Yamao and Syozo Osawa

Laboratory of Molecular Genetics, Department of Biology, Faculty of Science, Nagoya University, Furo-Cho, Chikusa-Ku, Nagoya 464, Japan

Received 17 September 1984; Accepted 15 October 1984

ABSTRACT

The nucleotide sequence of the 1.3 kilobase-pair DNA segment, which contains the genes for ribosomal proteins S8 and L6, and a part of L18 of Mycoplasma capricolum, has been determined and compared with the corresponding sequence in Escherichia coli (Cerretti et al., Nucl. Acids Res. 11, 2599, 1983). Identities of the predicted amino acid sequences of S8 and L6 between the two organisms are 54% and 42%, respectively. The A + T content of the M. capricolum genes is 71%, which is much higher than that of E. coli (49%). Comparisons of codon usage between the two organisms have revealed that M. capricolum preferentially uses A- and U-rich codons. More than 90% of the codon third positions and 57% of the first positions in M. capricolum is either A or U, whereas E. coli uses A or U for the third and the first positions at a frequency of 51% and 36%, respectively. The biased choice of the A- and U-rich codons in this organism has been also observed in the codon replacements for conservative amino acid substitutions between M. capricolum and E. coli. These facts suggest that the codon usage of M. capricolum is strongly influenced by the high A + T content of the genome.

INTRODUCTION

The DNA of mycoplasmas is extremely rich in A and T (1,2). It is thus interesting to see how this characteristic feature effects on the structure of the mycoplasma genes. Recently, we have cloned a DNA segment containing a cluster of ribosomal protein genes of Mycoplasma capricolum (3). This paper reports the nucleotide sequence of the DNA segment coding for ribosomal proteins S8 and L6, demonstrating that the codon usage in M. capricolum genes is greatly deviated from that in E. coli.

MATERIALS AND METHODS

Plasmid DNA was prepared according to Oka et al. (4). DNA sequence determinations were performed by the chain-termination method of Sanger et al. (5) as described by Messing and Vieira (6).

Restriction-endonucleases, T4 DNA-ligase and DNA polymerase I (large fragment) were purchased from Takara-Shuzo Co. Ltd.(Kyoto). The enzyme reaction conditions were those recommended by the comercial supplier.

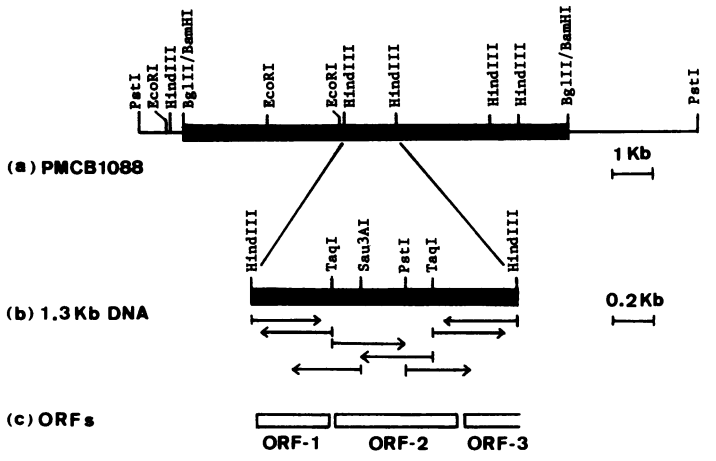

Fig. 1. (a) Restriction map of pMCB1088. The thick line shows the inserted M. capricolum DNA segment and the thin line represents the vector pBR322 DNA. (b) The strategy for DNA sequence determination of the 1.3 Kb HindIII-fragment from pMCB1088. The arrows indicate the direction of the sequence determinations. (c) The locations of open reading frames.

(α-^{32}P)dATP was purchased from Amersham Japan. The construction of plasmid pMCB1088 containing a DNA segment from M. capricolum ATCC27343(KID) was described elsewhere (3).

RESULTS

Plasmid pMCB1088 contained a 9 kilobase-pair (Kb) BglII-fragment of M. capricolum DNA that had been inserted into the BamHI-site of pBR322 (3). This DNA segment coded for at least eight ribosomal proteins (3). The restriction map of the insert is shown in Fig. 1(a). Deletion mapping experiments revealed that the 1.3 Kb HindIII-fragment derived from the central part of the insert contained the gene for at least a ribosomal small subunit protein of M. capricolum (MS5; see ref. 3).

The complete nucleotide sequence of the 1.3 Kb (1,251 base-pairs) fragment was determined according to the strategy shown in Fig. 1(b). More than 90% of the sequences was confirmed by sequencing both strands. Computer analyses for potential protein coding sequences revealed the presence of three open reading frames (ORF-1, ORF-2 and ORF-3) on the same DNA strand (Fig. 1(c); see also Fig. 2). The amino acid sequences of the ORF-1 and ORF-2 deduced from the nucleotide sequences were found to have 54% and 42% identities with those of E. coli ribosomal protein S8 (7) and L6 (8),

respectively. The reported N-terminal sequences of proteins S8 and L6 from Bacillus subtilis also showed high homologies with the corresponding sequences of the ORFs (9, 10). Thus, we considered the ORF-1 and ORF-2 as the genes for M. capricolum S8 or rpsH and L6 or rplF, respectively. The ORF-3 could encode 81 amino acid residues, the sequence of which resembled that from the N-terminus of the E. coli ribosomal protein L18 (31% identity). We tentatively identified the ORF-3 as a part of the L18 gene or rplR. The ORFs were interrupted by relatively short (19 and 28 base-pairs) spacers, that did not contain promoter- or terminater-like sequences, suggesting that these genes are a part of the operon for ribosomal protein genes. It is interesting that the order of the genes (rpsH-rplF-rplR) on this fragment was the same as that in the spc-operon of E. coli, where the genes for ten ribosomal proteins and two membrane proteins are clustered as a transcriptional unit (11). This indicates that the order of at least the three genes mentioned above has been conserved between M. capricolum and E. coli.

Since the complete sequence of the E. coli spc-operon was reported (11), we can directly compare the sequences of the ribosomal protein genes between M. capricolum and E. coli. Fig. 2 shows the nucleotide sequences of the 1.3 Kb DNA (mRNA-like strand) and the predicted amino acid sequences aligned with those of E. coli. The M. caprilolum S8 gene was 384 base-pairs long (128 codons including initiation codon), 6 base-pairs shorter than the E. coli gene. The L6 gene of M. capricolum was 540 base-pairs long (180 codons), 9 base-pairs longer than that of E. coli. The spacers between the genes of M. capricolum were longer than those of E. coli.

DISCUSSION

The A + T content of M. capricolum genomic DNA is 75% (12), one of the highest in prokaryotes so far known, whereas that of E. coli DNA is about 50%. Thus, the high A + T content of the genome may directly dictate the codon usage of the M. capricolum genes. In fact, the A + T content of the M. capricolum ribosomal protein genes is about 71%, which is much higher than that of E. coli (49%). In Table 1 is compared the codon usage in S8 and L6 genes between M. capricolum and E. coli. The choice of synonymous codons in M. capricolum is clearly different from that in E. coli and strongly biased. For example, UUA(Leu), AUU(Ile) and AGA(Arg) are the most frequently used codons among synonyms in M. capricolum, while they are rarely used in E. coli [CUG(Leu), AUC(Ile) and CGU(Arg) are predominant]. The most outstanding feature of the codon usage in M. capricolum can be seen in a very high

Fig. 2. The total nucleotide sequence of the 1.3 Kb HindIII-fragment and deduced amino acid sequences of the ORFs. The sequences are aligned with the corresponding E. coli sequences reported by Cerretti et al (11). M.c.: M. capricolum; E.c.: E. coli.

frequency of the codons ending in either A or U; 91% (280/308) of the M. capricolum codons has A or U at the third position, in contrast to only 51% (155/307) in E. coli (Table 2). The A + U content of the first position of the codons in M. capricolum is also significantly higher (57%;174/308) than that in E. coli (36%;110/307). It is thus evident that the choice of synonymous codon in M. capricolum is biased to the codons much richer in A and U as compared with E. coli.

To see further the preferential use of A- and U-rich condons in M. capricolum, individual codons for the S8 and L6 genes of the two organisms are compared. There exists 143 identical amino acids at the homologous positions between M. capricolum and E. coli S8 and L6 (Fig. 2). Among them,

Table 1. Codon usage in S8 and L6 genes of M. capricolum and E. coli

First		Second U			Second C			Second A			Second G			Third
			M.c.	E.c.*		M.c.	E.c.		M.c.	E.c.		M.c.	E.c.	
U	U	Phe	4	3	Ser	5	5	Tyr	5	3	Cys	0	1	U
		Phe	2	3	Ser	0	2	Tyr	3	5	Cys	0	1	C
		Leu	24	0	Ser	8	1	---	0	2	---	1	0	A
		Leu	0	1	Ser	0	0	---	1	0	Trp	0	1	G
	C	Leu	0	2	Pro	4	7	His	2	3	Arg	1	13	U
		Leu	0	1	Rro	0	1	His	1	0	Arg	0	5	C
		Leu	4	0	Pro	5	1	Gln	15	1	Arg	1	0	A
		Leu	0	18	Pro	1	4	Gln	0	11	Arg	0	0	G
	A	Ile	21	5	Thr	15	8	Asn	13	2	Ser	3	0	U
		Ile	5	12	Thr	0	9	Asn	4	9	Ser	1	2	C
		Ile	4	0	Thr	6	0	Lys	34	17	Arg	11	0	A
		Met	5	8	Thr	0	1	Lys	1	11	Arg	0	0	G
	G	Val	16	21	Ala	8	9	Asp	5	10	Gly	14	19	U
		Val	0	5	Ala	0	7	Asp	1	6	Gly	0	9	C
		Val	7	3	Ala	11	9	Glu	21	12	Gly	13	0	A
		Val	1	5	Ala	0	9	Glu	0	5	Gly	3	1	G

* Taken from Cerretti et al. (11)
M.c. : M. capricolum ; E.c. : E. coli

the corresponding codons for 46 amino acids are identical, while the rest 97 codons are different (synonymous codon replacement). All of these silent nucleotide substitutions in the 97 synonymous codons are listed in Table 3. Here, the preferential use of A- and U-rich codons in M. capricolum can clearly be seen. Sixty-six codons having G or C at the third position in E. coli are replaced by synonymous codons ending in A or U with M. capricolum.

Table 2. Nucleotide composition at three different positions of codons in S8 and L6 genes of M. capricolum and E. coli

	First		Second		Third		Total	
	M.c.	E.c.	M.c.	E.c.	M.c.	E.c.	M.c.	E.c.
U	51	26	93	87	116	111	260	224
C	34	67	63	73	17	77	114	217
A	123	84	105	95	164	44	392	223
G	100	130	47	52	11	75	158	257
A + U	174	110	198	182	280	155	625	447
(%)	(56.5)	(35.8)	(64.3)	(59.3)	(90.9)	(50.5)	(70.6)	(48.5)

M.c. : M. capricolum ; E.c. : E. coli

Table 3. Synonymous nucleotide substitutions in S8 and L6 genes between M. capricolum and E. coli

	M.c.	E.c.	(AU)[a]	No.[b]	a x b		M.c.	E.c.	(AU)[a]	No.[b]	a x b
Ala	GCA	GCG	(+1)	3	3	Lys	AAA	AAG	(+1)	8	8
	GCA	GCC	(+1)	2	2						
	GCU	GCG	(+1)	1	1	Phe	UUU	UUC	(+1)	1	1
	GCU	GCA	(0)	1	0		UUC	UUU	(-1)	1	1
Asn	AUU	AAC	(+1)	2	2	Pro	CCA	CCC	(+1)	1	1
	AAC	AAU	(-1)	1	-1		CCA	CCU	(0)	2	0
							CCG	CCU	(-1)	1	-1
Asp	GAU	GAC	(+1)	1	1						
	GAC	GAU	(-1)	1	-1	Ser	UCU	UCC	(+1)	1	1
							UCU	AGC	(+1)	1	1
Arg	AGA	CGC	(+2)	2	4		UCA	UCU	(0)	3	0
	AGA	CGU	(+1)	3	3		AGU	UCU	(0)	1	0
	CGA	CGC	(+1)	1	1		AGC	UCC	(0)	1	0
Gln	CAA	CAG	(+1)	5	5	Thr	ACA	ACG	(+1)	1	1
							ACA	ACC	(+1)	1	1
Glu	GAA	GAG	(+1)	3	3		ACU	ACC	(+1)	2	2
Gly	GGA	GGC	(+1)	4	4	Tyr	UAU	UAC	(+1)	2	2
	GGA	GGU	(0)	7	0		UAC	UAU	(-1)	1	-1
	GGU	GGC	(+1)	5	5						
	GGU	GGG	(+1)	1	1	Val	GUA	GUC	(+1)	1	1
	GGG	GGU	(-1)	2	-2		GUA	GUU	(0)	3	0
							GUU	GUA	(0)	2	0
Ile	AUU	AAC	(+1)	3	3						
	AUA	AUC	(+1)	2	2						
	AUC	AUU	(-1)	1	-1						
Leu	UUA	CUG	(+2)	9	18			Total		97	74[c]
	UUA	UUG	(+1)	1	1						
	CUA	CUG	(+1)	2	2						

M.c. : M. capricolum ; E.c. : E. coli
a : A + U gains in M. capricolum as compared with E. coli
b : Number of occurrence
c : Total A + U gains in 97 synonymously substituted codons (291 nucleotide residues) in M. capricolum as compared with E. coli

A striking example is that eight AAG condons for Lys in E. coli are substituted by AAA in M. capricolum.

Fifty conservative amino acid replacements (i.e., Lys/Arg, Ser/Thr, Leu/Ile/Val, etc.) can be seen at the homologous positions of the S8 and L6 proteins between the two organisms (Fig. 2). In Table 4 are collected all of these amino acid substitutions with their codon replacements. Most of the conservative amino acid substitutions take place so that the M. capricolum maintains much higher A + U content in the codons than E. coli. For example, seven Arg with codon CGU or CGC in E. coli are replaced at the corresponding

Table 4. Conservative amino acid substitutions in protein S8 and L6 between M. capricolum and E. coli with their codon replacements

Substitution	M. capricolum	E. coli	(AU-gain)[a]	Number of occurrence[b]	a x b
Lys/Arg	Lys(AAA)	Arg(CGU)	(+2)	6	12
	Lys(AAA)	Arg(CGC)	(+3)	1	3
Leu/Ile	Ile(AUU)	Leu(CUG)	(+2)	6	12
	Ile(AUU)	Leu(CUU)	(+1)	1	1
	Leu(UUA)	Ile(AUC)	(+1)	1	1
	Leu(UUA)	Ile(AUU)	(0)	3	0
	Leu(CUA)	Ile(AUC)	(0)	1	0
Leu/Val	Leu(UUA)	Val(GUC)	(+2)	1	2
	Leu(UAA)	Val(GUG)	(+2)	2	4
	Leu(UAA)	Val(GUU)	(+1)	1	1
	Val(GUA)	Leu(CUG)	(+1)	1	1
Ile/Val	Ile(AUU)	Val(GUU)	(+1)	4	4
	Ile(AUU)	Val(GUC)	(+2)	2	4
	Ile(AUU)	Val(GUA)	(+1)	1	1
	Ile(AUU)	Val(GUG)	(+2)	1	2
	Ile(AUA)	Val(GUU)	(+1)	2	2
	Ile(AUC)	Val(GUU)	(0)	1	0
	Val(GUA)	Ile(AUC)	(0)	1	0
	Val(GUU)	Ile(AUC)	(0)	1	0
	Val(GUU)	Ile(AUU)	(-1)	1	-1
Ser/Thr	Ser(AGU)	Thr(ACC)	(+1)	2	2
	Ser(UCA)	Thr(ACU)	(0)	1	0
	Thr(ACA)	Ser(AGC)	(+1)	1	1
Ala/Gly	Ala(GCU)	Gly(GGU)	(0)	1	0
	Gly(GGU)	Ala(GCU)	(0)	1	0
	Gly(GGG)	Ala(GCU)	(-1)	1	-1
Asn/Gln	Asn(AAC)	Gln(CAG)	(+1)	1	1
	Gln(CAA)	Asn(AAC)	(0)	1	0
Asp/Glu	Glu(GAA)	Asp(GAC)	(+1)	2	2
	Glu(GAA)	Asp(GAU)	(0)	1	0
		Total		50	54[c]

a : A + U gains in M. capricolum codons as compared with E. coli
b : Number of occurrence
c : Total A + U gains in 50 conservatively substituted codons (150 nucleotide residues) in M. capricolum as compared with E. coli

positions by Lys with AAA in M. capricolum. Thus, the choice of codons in M. capricolum seems to occur to discriminate against G and C and to use A and U wherever possible.

The genomic DNAs of M. capricolum as well as all the known species

belonging to the class Mollicute (Mycoplasma, Acholeplasma, Ureaplasma and Spiroplasma) are rich in A and T (61-78%), suggesting that the Mollicute has evolved with a constraint keeping the high A + T contents in their genomes. The obvious preference of A and T in M. capricolum is clearly seen not only in the codons, but also in the spacers. In fact, the spacers between rRNA genes of M. capricolum are extremely rich in A and T (13). Also, the A + U content of rRNAs from mycoplasmas is highest in prokaryotes (14). Thus, in M. capricolum, the constraint for the preferential use of A and T has operated at the DNA level as a selection force upon the codon choice as well as the construction of other parts of the genome.

ACKNOWLEDGEMENTS

This work was supported by grants from the Ministry of Education, Science and Culture, Japan (Nos. 57121003, 57121009 and 58220012) and the Naito Science Foundation Research Grant (81-106).

* On leave from Department of Biochemistry and Biophysics, Research Institute for Nuclear Medicine and Biology, Hiroshima University, Hiroshima 734, Japan.

REFERENCES

1. Maniloff, J. and Morowitz, H. J. (1972) Bacteriol. Rev. 36, 263-290.
2. Razin, S. (1978) Microbial. Rev. 42, 414-470.
3. Kawauchi, Y., Muto, A., Yamao, F. and Osawa, S. (1984) Mol. Gen. Genet. in press.
4. Oka, A., Sugisaki, H. and Takanami, M. (1981) J. Mol. Biol. 147, 217-226.
5. Sanger, F., Wicken, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
6. Messing, J. and Vieira, J. (1982) Gene 19, 269-276.
7. Allen, G. and Wittmann-Liebold, B. (1978) Hoppe-Seyler's Z. Physiol. Chem. 359, 1509-1525.
8. Chen, R., Afrsten, U. and Chen-Schmeisser (1977) Hoppe-Seyler's Z. Physiol. Chem. 358, 531-535.
9. Higo, K., Otaka, E. and Osawa, S. (1982) Mol. Gen. Genet. 185, 239-244.
10. Higo, K., Itoh, T., Kumazaki, T. and Osawa, S. (1980) in Genetics and Evolution of RNA polymerase, tRNA and Ribosomes. Osawa, S., Ozeki, H., Uchida, H. and Yura, T. eds. pp.655-666, Tokyo Univ. Press, Tokyo.
11. Cerretti, D. P., Dean, D., Davis, G. R., Bedwell, D. M. and Nomura, M. (1983) Nucl. Acids Res. 11, 2599-2616.
12. Neimerk, H. C. (1970) J. Gen. Microbiol. 63, 249-263.
13. Sawada, M., Muto, A., Iwami, M., Yamao, F. and Osawa, S. (1984) Mol. Gen. Genet. in press.
14. Ryan, J. L. and Morowitz, H. J. (1969) Proc. Natl. Acad. Sci. USA 63, 1282-1289.