Plant Biotechnology 26, 53–56 (2009)

Minireview

Selection of synonymous codons for better expression of recombinant proteins in tobacco chloroplasts

Masayuki Nakamura[1], Masahiro Sugiura[1,2,*]

[1] Graduate School of Natural Sciences, Nagoya City University, Nagoya, Aichi 467-8501, Japan; [2] Sugiyama Human Research Center, Sugiyama Jogakuen University, Nagoya, Aichi 464-8662, Japan
* E-mail: sugiura@nsc.nagoya-cu.ac.jp Tel & Fax: +81-052-872-6021

Received November 15, 2008; accepted February 3, 2009 (Edited by H. Suzuki)

Abstract The 20 amino acids, except for methionine and tryptophan, are coded for by two to six codons called synonymous codons. Synonymous codons are used differently by different organisms. Hence, codon selection should be important to express recombinant proteins in tobacco chloroplasts. We present here the codon usage table from the entire 79 mRNAs from tobacco chloroplasts. As codons function on mRNAs, codon usage tables should be constructed from mRNAs but not from genes. We devised an *in vitro* assay for translation efficiencies of codons, and measured the translation efficiencies of several synonymous codon groups in tobacco chloroplasts. Our results indicated that translation efficiencies of individual codons are not always correlated with codon usage. Based on these data, we discuss on synonymous codon selection for expression of inserted genes in tobacco chloroplasts.

Key words: Chloroplast, codon usage, synonymous codon, tobacco, translation efficiency.

Chloroplast transformation provides a powerful tool to produce useful proteins in plants. After completion of the chloroplast genome sequencing from tobacco plants (Shinozaki et al. 1986), Pal Maliga's group developed the high-frequency chloroplast transformation system in tobacco (Svab and Maliga 1993). This method allows us to insert stably a foreign gene at a desired site by homologous recombination. Attempts have been made to express foreign proteins in tobacco chloroplasts, and reproducible yields of recombinant proteins of 5–25% of total soluble cellular protein in leaves have been achieved (Maliga 2003). However, the synthesis of some other foreign proteins was not detected though substantial amounts of their transcripts were accumulated. This observation suggested that the translation system of chloroplasts is different from that of bacteria and of eukaryotic cytoplasms.

The tobacco chloroplast genome includes 79 identified genes encoding polypeptides (Yukawa et al. 2005) which consist of 22,976 codons. Tobacco chloroplasts use the universal genetic code (all 64 codons including three stop codons). C-to-U RNA editing has been reported at 38 sites in the mRNA coding regions, and 37 of them cause amino acid conversion (Sasaki et al. 2003; Sugiura 2008). Though several genes contain multiple possible initiation codons, the longest open reading frames were generally annotated as protein-coding genes. The genuine start codon for *ndhD*, *psbC* and *ndhK* mRNAs has been determined experimentally (Hirose and Sugiura 1997; Kuroda et al. 2007; Yukawa and Sugiura 2008). Based on these data, we reexamined our previous codon usage table calculated from tobacco chloroplast genes (Wakasugi et al. 1986) and amended as shown in Table 1. This codon usage table is based exclusively on single mature chloroplast mRNAs because codons function on mRNAs but not on genomes (DNAs). Codon usage tables available in databases have been constructed by simple summation of collected gene sequences. One example for tobacco chloroplasts was constructed from only 11 genes while another table consisted of 137 genes, almost 1.5 times more than the existing genes. In the latter table, over one-third of the chloroplast genes were added two to three times whereas three chloroplast genes were missing and three nuclear genes were included.

The tobacco chloroplast genome contains 30 different tRNA genes (Sugiura and Wakasugi 1989). Seven of them are located in the large inverted repeat, indicating that these seven genes are duplicated and the other 23 genes are single-copy genes. This suggests that the contents of individual tRNA species are not significantly different each other in tobacco chloroplasts. The minimum number of tRNA species required for translation of all 61 codons is 33 (including the start codon AUG) if normal wobble base pairing occurs in codon-anticodon recognition. The tRNA species that

Table 1. Codon usage of the entire 79 tobacco chloroplast mRNAs.

	Codon	Fraction		Codon	Fraction		Codon	Fraction		Codon	Fraction
Phe	UUU	0.667	Ser	UCU	0.299	Tyr	UAU	0.804	Cys	UGU	0.752
Phe	UUC	0.333	Ser	UCC	0.152	Tyr	UAC	0.196	Cys	UGC	0.248
Leu	UUA	0.328	Ser	UCA	0.188	STOP	UAA	0.519	STOP	UGA	0.228
Leu	UUG	0.200	Ser	UCG	0.059	STOP	UAG	0.253	Trp	UGG	1.000
Leu	CUU	0.214	Pro	CCU	0.393	His	CAU	0.770	Arg	CGU	0.219
Leu	CUC	0.069	Pro	CCC	0.187	His	CAC	0.230	Arg	CGC	0.063
Leu	CUA	0.128	Pro	CCA	0.286	Gln	CAA	0.756	Arg	CGA	0.251
Leu	CUG	0.062	Pro	CCG	0.135	Gln	CAG	0.244	Arg	CGG	0.074
Ile	AUU	0.495	Thr	ACU	0.392	Asn	AAU	0.768	Ser	AGU	0.209
Ile	AUC	0.200	Thr	ACC	0.197	Asn	AAC	0.232	Ser	AGC	0.059
Ile	AUA	0.306	Thr	ACA	0.304	Lys	AAA	0.754	Arg	AGA	0.289
Met	AUG	0.853	Thr	ACG	0.107	Lys	AAG	0.246	Arg	AGG	0.104
fMet	AUG	0.141									
Val	GUU	0.372	Ala	GCU	0.448	Asp	GAU	0.797	Gly	GGU	0.322
Val	GUC	0.120	Ala	GCC	0.172	Asp	GAC	0.203	Gly	GGC	0.118
Val	GUA	0.380	Ala	GCA	0.280	G1 u	GAA	0.757	Gly	GGA	0.389
Val	GUG	0.128	Ala	GCG	0.100	Glu	GAG	0.243	Gly	GGG	0.171
fMet	GUG	0.006									

Fraction, the ratio of each codon in the family of synonymous codons. From Table 1 of Nakamura and Sugiura (2007) with extensive modifications.

Table 2. tRNA anticodons predicted from the entire 30 tobacco chloroplast tRNA genes.

	Codon	tRNA		Codon	tRNA		Codon	tRNA		Codon	tRNA
Phe	UUU		Ser	UCU		Tyr	UAU		Cys	UGU	
Phe	UUC	**GAA**	Ser	UCC	GGA	Tyr	UAC	GUA	Cys	UGC	GCA
Leu	UUA	**unkUAA**, **UmAA**	Ser	UCA	UGA	Tyr	UAA		STOP	UGA	
Leu	UUG	**CmAA**	Ser	UCG		STOP	UAG		Trp	UGG	**CmCA**, **CCA**
Leu	CUU		Pro	CCU		His	CAU		Arg	CGU	**ICG**
Leu	CUC		Pro	CCC		His	CAC	GUG	Arg	CGC	
Leu	CUA	**UAm7G**	Pro	CCA	**unkUGG**	Gln	CAA	**cmnm5UUG**	Arg	CGA	
Leu	CUG		Pro	CCG		Gln	CAG		Arg	CGG	
Ile	AUU		Thr	ACU		Asn	AAU		Ser	AGU	
Ile	AUC	**GAU**	Thr	ACC	**GGU**	Asn	AAC	GUU	Ser	AGC	GCU
Ile	AUA	**unkCAU**	Thr	ACA	UGU	Lys	AAA	**unkUUU**	Arg	AGA	UCU
Met	AUG	**CAU**	Thr	ACG		Lys	AAG		Arg	AGG	
fMet	AUG	CAU									
Val	GUU		Ala	GCU		Asp	GAU		Gly	GGU	
Val	GUC	GAC	Ala	GCC		Asp	GAC	**GUC**, **QUC**	Gly	GGC	GCC
Val	GUA	**unkUAC**	Ala	GCA	UGC	Glu	GAA	**mnm5s2UYC**	Gly	GGA	**UCC**
Val	GUG		Ala	GCG		Glu	GAG		Gly	GGG	
fMet	GUG	CAU									

tRNA, indicating anticodon. Box, codons read by a tRNA encoded in the chloroplast genome. From Table 2 of Sugiura and Wakasugi (1989) with extensive modifications. Boldface anticodons, from sequenced chloroplast tRNAs with modified nucleotides (unkU; unknown modified uridine, Urn; 2′-O-methyluridine, Cm; 2′-O-methylcytidine, m7G; 7-methylguanosine, unkC: unknown modified cytidine, cmnm5U; 5-carboxymethylaminomethyluridine, mnm5s2U; 5-methylaminomethyl-2-thiouridine, Y; Pseudouridine, I; Inosine, Q; Queuosine). From Sprinzl et al. (2005).

recognize leucine (CUU/C), proline (CCU/C), alanine (GCU/C) and arginine (CGG) codons (blanked in Table 2) are not encoded by the tobacco chloroplast genome. The first position (A) of $tRNA^{Arg}$(ACG)s is generally modified to inosine (I), and $tRNA^{Arg}$(ICG) reads not only CGU but also CGC and CGA (Curran 1995). As no evidence has been reported for tRNA import into chloroplasts, the above codons are probably read by the tRNAs with modified nucleotides (Table 2) and by the two-out-of-three (or the superwobble) mechanism (Rogalski et al. 2008).

To study mechanisms of translation unique to chloroplasts, we developed an *in vitro* translation system from isolated tobacco chloroplasts (Hirose and Sugiura 1996). We then improved extensively our original system using a gene for a modified green fluorescent protein (mGFP) instead of ^{35}S-methinine (Yukawa et al. 2007). The improved method is 100-fold more active than the original one, extremely low in background and requires no additional tRNAs and no micrococcal nuclease treatment. The rate of translation from a variety of mRNAs can be measured by monitoring the fluorescence intensity of synthesized mGFP. Based on this system, we devised an *in vitro* assay to measure translation

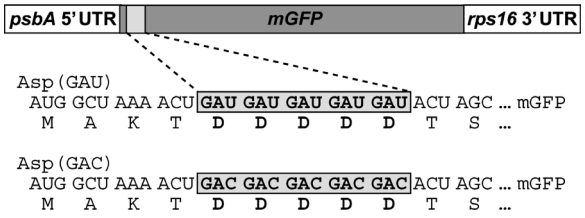

Figure 1. Schematic of test mRNAs to assay translation efficiencies of, for example, aspartic acid (D) codons (GAU and GAC). Partial mRNA sequences are shown below. Codons of interest are boxed. Test mRNAs were subjected to *in vitro* translation. Translation products were separated on native gels. Translation efficiency was quantified from the fluorescence intensity of the fused mGFP. From Figure 1 of Nakamura and Sugiura (2007) with extensive modifications.

Table 3. Codon usage and translation efficiency.

	Codon	Fraction	Translation
Asn	AAU	0.768	+++
	AAC	0.232	+
Asp	GAU	0.797	+++
	GAC	0.203	+
Ala	GCU	0.448	+++
	GCC	0.172	++
	GCA	0.280	++
	GCG	0.100	+
Tyr	UAU	0.804	+
	UAC	0.196	+++
Phe	UUU	0.667	+
	UUC	0.333	+++

Fraction, from Table 1. Translation, high (+++), medium (++) and low (+) efficiencies, tabled from Nakamura and Sugiura (2007).

efficiencies of synonymous codons (Nakamura and Sugiura 2007). As shown in Figure 1, we designed a template mRNA that contains the 5'untranslated region (UTR) from tobacco chloroplast *psbA*, the coding sequence for mGFP, and the 3′UTR from tobacco chloroplast *rps16* (Yukawa et al. 2007). The *psbA* mRNA is the most actively translated *in vitro* among chloroplast mRNAs examined (Hirose and Sugiura 1996). A codon of interest was repeated 5 times, and its codon block was inserted into the 3rd or 4th codons downstream from the initiation codon (1st codon) of test mRNAs. Then, the secondary structure of mRNA sequences surrounding the start codon was predicted *in silico* and adjusted by changing 3rd and 4th codons so that no significant difference was found between mRNAs to be compared, since translation efficiency is affected by mRNA secondary structure. Using the system, we measured the translation efficiency of five synonymous codons (Table 3). Unexpectedly, the translation efficiency of phenylalanine and tyrosine codons is opposite to their codon usage. Therefore, the translation efficiencies of synonymous codons are not always correlated with codon usage in tobacco chloroplasts.

To produce efficiently recombinant proteins in chloroplasts, codon optimization has often been performed according to codon usage tables. However, only 2- to 3-fold increase in protein accumulation has been obtained in transplastomic plants (Maliga 2004). Our results raise a question concerning the usefulness of the so-called codon optimization. At present, we suggest to refer Tables 1 and 3 for selection of synonymous codons. In addition, we do not recommend selecting CGA for arginine because the A:I pair was reported to be inefficient (Curran 1995). Finally, our *in vitro* translation system will be powerful for screening a number of designed mRNAs in a short time before chloroplast transformation.

Acknowledgements

We thank H. Kuroda, M. Yukawa and Y. Yukawa for advice. This work was partly supported by New Energy and Industrial Technology Development Organization (the Green Biotechnology Program) and by Grants-in-Aid from Ministry of Education, Culture, Sports, Science and Technology (No.19370021).

References

Curran JF (1995) Decoding with the A:I wobble pair is inefficient. *Nucl Acids Res* 23: 683–688

Hirose T, Sugiura M (1996) *Cis*-acting elements and *trans*-acting factors for accurate translation of chloroplast *psbA* mRNAs: development of an *in vitro* translation system from tobacco chloroplasts. *EMBO J* 15: 1687–1695

Hirose T, Sugiura M (1997) Both RNA editing and RNA cleavage are required for translation of tobacco chloroplast *ndhD* mRNA: a possible regulatory mechanism for the expression of a chloroplast operon consisting of functionally unrelated genes. *EMBO J* 16: 6804–6811

Kuroda H, Suzuki H, Kusumegi T, Hirose T, Yukawa Y, Sugiura M (2007) Translation of *psbC* mRNAs starts from the downstream GUG, not the upstream AUG, and requires the extended Shine-Dalgarno sequence in tobacco chloroplasts. *Plant Cell Physiol* 48: 1374–1378

Maliga P (2003) Progress towards commercialization of plastid transformation technology. *Trends Biotechnol* 21: 20–28

Maliga P (2004) Plastid transformation in higher plants. *Annu Rev Plant Biol* 55: 289–313

Nakamura M, Sugiura M (2007) Translation efficiencies of synonymous codons are not always correlated with codon usage in tobacco chloroplasts. *Plant J* 49: 128–134

Rogalski M, Karcher D, Bock R (2008) Superwobbling facilitates translation with reduced tRNA sets. *Nat Struct Mol Biol* 15: 192–198

Sasaki T, Yukawa Y, Miyamoto T, Obokata J, Sugiura M (2003) Identification of RNA editing sites in chloroplast transcripts from the maternal and paternal progenitors of tobacco (*Nicotiana tabacum*): comparative analysis shows the involvement of distinct trans-factors for *ndhB* editing. *Mol Biol Evol* 20: 1028–1035

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene

organization and expression. *EMBO J* 5: 2043–2049

Sprinzl M, Vassilenko KS (2005) Compilation of tRNA sequences and sequences of tRNA genes. *Nucl Acids Res* 33: D139–140

Sugiura M (2008) RNA editing in chloroplasts. In: Goringer HU (ed) RNA editing. Springer-Verlag, Berlin Heidelberg, pp 123–142

Sugiura M, Wakasugi T (1989) Compilation and comparison of transfer RNA genes from tobacco chloroplasts. *Critical Rev in Plant Sci* 8: 89–101

Svab Z, Maliga P (1993) High-frequency plastid transformation in tobacco by selection for a chimeric *aadA* gene. *Proc Natl Acad Sci USA* 90: 913–917

Wakasugi T, Ohme M, Shinozaki K, Sugiura M (1986) Structure of tobacco chloroplast genes for $tRNA^{Ile}$(CAU), $tRNA^{Leu}$(CAA), $tRNA^{Cys}$(GCA), $tRNA^{Ser}$(UGA) and $tRNA^{Thr}$(GGU): a compliation of tRNA genes from tobacco chloroplasts. *Plant Mol Biol* 7: 385–392

Yukawa M, Tsudzuki T, Sugiura M (2005) The 2005 version of the chloroplast DNA sequence from tobacco (*Nicotiana tabacum*). *Plant Mol Biol Rep* 23: 359–365

Yukawa M, Kuroda H, Sugiura M (2007) A new *in vitro* translation system for non-radioactive assay from tobacco chloroplasts: effect of pre-mRNA processing on translation *in vitro*. *Plant J* 49: 367–376

Yukawa M, Sugiura M (2008) Termination codon-dependent translation of partially overlapping *ndhC-ndhK* transcripts in chloroplasts. *Proc Natl Acad Sci USA* 105: 19550–19554