ALTERNATIVE TRANSLATION AND FUNCTIONAL DIVERSITY OF RELEASE FACTOR 2 AND LYSYL-tRNA SYNTHETASE

Y. Nakamura, K. Kawakami, and O. Mikuni

Department of Tumor Biology
The Institute of Medical Science
The University of Tokyo
P.O. Takanawa, Tokyo 108, Japan

The *prfB* gene encodes peptide-chain-release factor 2 which catalyzes translation termination at nonsense codons UGA and UAA in bacteria. Mutations in *prfB* cause misreading of UGA, *i.e.*, increased frameshift or suppression. These translational errors were due to inefficient translation termination at UGA. As one extreme case, an opal (UGA) mutation in *prfB* is autogenously suppressed. The *prfB* gene is in the same operon with *herC*, a gene defined by a suppressor mutation of a replication-defective ColE1 plasmid. The *herC* gene coincides with *lysS*, a gene encoding a major species of lysyl-tRNA synthetase. Thus, the genetic organization of *prfB* and *lysS* in the same operon may suggest structural, functional and evolutional relevance of these gene products.

I. Release Factor 2 Gene

Polypeptide chain termination requires participation of two peptide-chain-release factors that recognize specific termination codons. Release factor 1 (RF1) catalyzes termination at UAG and UAA, and release factor 2 (RF2) catalyzes termination at UGA and UAA (Scolnick *et al.*, 1968). The RF1 gene has been cloned on the basis of competition between a nonsense suppressor tRNA and a translational release factor (Weiss *et al.*, 1984). The gene encoding RF1 has been named *prfA* and is located at 27 min on the *Escherichia coli* chromosome (Ryden *et al.*, 1986). The RF2 gene has been isolated from the Clarke-Carbon *E. coli* plasmid bank on the basis of RF2 overproduction detected by an anti-RF2 antibody (Caskey *et al.*, 1984). However, the chromosomal location of the RF2 gene had not been reported.

NATO ASI Series, Vol. H 49
Post-Transcriptional Control of Gene Expression
Edited by J. E. G. McCarthy and M. F. Tuite

A. *E. coli* RF2 Operon. The chromosomal location of the *prfB* gene was found by cloning and sequencing the *E. coli* DNA fragment carrying the *herC* gene mapped at the 62 min region (Kawakami *et al.*, 1988a; 1989). The *herC* gene has been defined by a suppressor mutation of a replication-deficient ColE1 plasmid. The deduced *herC* sequence encodes a 57,603-Da protein composed of 505 amino acids. It is overlapped with an open reading frame encoding a 54,626-Da protein (471 amino acids) in opposite orientation, whose gene product is not identified.

Nine bases upstream of the coding region of *herC*, we found another open reading frame which coincides with COOH-terminal part of the *prfB* gene (Kawakami, *et al.*, 1988a) except one guanine residue which had been lost in the published sequence (Craigen *et al.*, 1985). The complete *prfB* DNA was cloned in an intermediate-copy number plasmid pACYC184. This plasmid (pKK951) is able to complement *prfB* mutations (see below). However, we failed to reclone *prfB* into a high-copy number plasmid, such as pBR322 or pUC119. This is consistent with the previous observation that the pRF2 plasmid, an original isolate in the Clarke-Carbon colony bank (ColE1 replicon), exhibited a low activity of RF2 (Kawakami *et al.*, 1988a). We assume that a leakiness of termination at some UGA (and/or UAA) codons is required for cell growth for natural frameshift or suppression, and excess RF2 activity may be toxic.

The *prfB* and *herC* genes are cotranscribed in a 2,800-base transcript (Kawakami *et al.*, 1988a). The 5' end maps ~40 bp upstream of the initiator AUG codon of *prfB*. The location is consistent with the promoter site predicted by Craigen *et al.* (1985). The 3' end extends beyond the *herC* gene and maps ~40-50 bp downstream of the UAA stop codon of *herC*, where a potential ρ-independent termination signal is located. These results led us to conclude that *prfB* and *herC* genes are in a single operon.

B. *Salmonella typhimurium* RF2 Gene. *S. typhimurium prfB* DNA segments were cloned from λ phage libraries using the *E. coli prfB* probe. The entirety of the *prfB* gene was subcloned in pACYC184. The resulting *S. typhimurium prfB* plasmid (pSRF2) is able to complement both *E. coli* and *S. typhimurium* RF2 mutations. The deduced DNA sequence of *S. typhimurium prfB* was highly homologous to the *E. coli prfB* sequence; 95.6% identical in the amino acid sequence (349/365) and 87.0% identical in the nucleotide sequence (956/1099) (Kawakami & Nakamura, submitted). Both *prfB* genes contain an in-frame premature UGA codon

at position 26, indicating that the mature *S. typhimurium* RF2 protein is synthesized by a +1 frameshift as has been shown in *E. coli prfB* (Craigen *et al.*, 1985). The *S. typhimurium prfB* gene is followed by a DNA sequence highly homologous to *E. coli herC*. Therefore, we inferred that the *prfB-herC* operon is conserved in *S. typhimurium*.

II. Release Factor 2 Mutations

Near 27 min on the *E. coli* genetic map, *uar* and *sueB* mutations which affect RF1 have been located (Ryden *et al.*, 1986). These mutants misread UAA and/or UAG and increase the efficiency of suppression of these nonsense codons. On the other hand, no RF2 mutations had been isolated.

A. *E. coli* RF2 Mutations and Misreading UGA. Two rationales of selection were used to isolate RF2 mutations using localized mutagenesis of the *prfB* region. One was to directly select a suppressor mutation of a *lacZ* UGA allele (*lacZ659*), some of which may have a defect in the peptide chain releasing activity of RF2 at UGA. The other was to isolate a temperature-sensitive lethal mutant because RF2 function may be essential to *E. coli* growth. The first selection scheme yielded three RF2 mutations, *prfB1*, *prfB2* and *prfB3*, while the second selection scheme yielded one mutation, *prfB286* (Kawakami *et al.*, 1988b). Of the former mutants, the *prfB2* was also temperature-sensitive in growth. Likewise, the latter mutant *prfB286* was able to suppress UGA. The β-galactosidase levels in the cells harboring *lacZ659*(UGA) and *prfB* mutations were ~7% of the wild-type *lacZ* level and reduced to the background level in the presence of the plasmid pKK951 (*prfB*$^+$). These results indicated that these RF2 mutations are recessive UGA suppressors and the RF2 function is essential to *E. coli* growth.

These RF2 mutations did not suppress UAG and UAA when tested by using *lacI-lacZ* fusion plasmids containing nonsense mutations in *lacI* (Mikuni, unpub.). Although UAA is recognized both by RF1 and RF2 *in vitro*, it may be preferentially recognized *in vivo* by RF1 because the RF1 mutant exhibits a UAA-suppressor activity while the RF2 mutant does not. The efficiency of UGA suppression in *prfB* mutants varied depending on the allele of the UGA mutation; for instance, UGA positioned at 280 in *lacI* was not suppressed. This indicates that UGA suppression

is affected by the flanking RNA context (*i.e.*, tight RNA context; Mikuni, unpub.).

Another effect of the RF2 mutation on gene expression is on the RF2 gene itself. Translation of the *prfB* mRNA is autogenously controlled by a +1 frameshift at the premature UGA codon at position 26 (Craigen *et al.*, 1985). Taking advantage of the *prfB* mutant, we observed that a reduced level of RF2 activity leads to an increase in the +1 frameshift (Mikuni, unpub.). These results taken together with similar observations with the *S. typhimurium supK* RF2 mutant described below may yield the first genetic evidence for the autogenous control of RF2 synthesis.

It is not known how RF2 mutations generate UGA suppression. The only experimental relevance was shown in the basal level control of the tryptophan operon (Roesser *et al.*, 1989). The *prfB1* and *prfB3* mutations increased transcription termination two-fold at the *trp* operon attenuator. This was interpreted as indicating that a reduced level of RF2 activity leads to a stall of ribosome at an authentic UGA stop codon of the *trp* leader peptide, giving rise to an increase in formation of a terminating structure of the leader RNA. If the leader peptide contains an engineered UAG or UAA codon instead of UGA, the frequency of attenuation was not affected by *prfB* mutations. We assume that ribosomal pausing may be a general basis for UGA suppression and other translational errors in the mutant.

Mutational base changes in the *E. coli prfB* mutants were defined by cloning and sequencing the mutant RF2 genes or by direct sequencing of amplified DNAs by polymerase chain reaction. Two temperature-sensitive mutations, *prfB2* and *prfB286,* which have been isolated independently, coincidently caused identical amino acid substitutions at position 328. The *prfB1* and *prfB3* mutations caused a single amino acid change at position 89 and 143, respectively, both of which are to reduce acidic charge. These sequence studies aimed to investigate the structural and functional organization of RF2 will be published elsewhere (Mikuni, unpub.).

B. *S. typhimurium supK* Mutations. In *S. typhimurium*, a recessive UGA suppressor mutation, *supK*, had been isolated and mapped at the 62 min region of the chromosome (Reeves & Roth, 1971). The same authors have observed reduced levels of tRNA methyltransferase activities in several *supK* strains and have suggested that an unmodified tRNA causes UGA suppression (Reeves & Roth, 1975). However, we assumed,

on the basis of high homology between the *E. coli* and *S. typhimurium* genomes, that the *supK* mutation may affect the RF2 protein (Kawakami *et al.*, 1988a). Two lines of genetic evidence have been consistent with this assumption. First, the *supK* mutation of *S. typhimurium* is complemented by the *E. coli prfB* gene (Kawakami *et al.*, 1988a). Second, RF2 mutants of *E. coli* harbor a recessive UGA suppressor activity (Kawakami *et al.*, 1988b).

The above assumption was directly proven by cloning and sequencing the *S. typhimurium* RF2 gene from the *supK584* mutant. The nucleotide sequence of the mutant *prfB* gene contained a single base substitution of A for G at position 433. No other alteration was found within the coding sequence of *prfB*. This substitution generates a UGA stop codon for a UGG tryptophan codon at amino acid position 144. The pSRF2 plasmid encoding the wild-type RF2 protein eliminated the UGA-suppression activity of both *E. coli prfB* mutants and the *S. typhimurium supK584* mutant. On the other hand, the pSUPK plasmid carrying the above UGG-to-UGA change failed to complement these mutations. Of 10 temperature-resistant colonies isolated from the *supK584* strain, which is temperature-sensitive in growth, four revertants changed the UGA codon to UGG (*i.e.*, true reversion). These results led us to conclude that the *supK* gene encodes *S. typhimurium* RF2 and that the opal *supK584* substitution in the RF2 gene is solely responsible for the UGA suppression and the temperature-sensitivity (Kawakami & Nakamura, submitted).

C. Autogenous Suppression of UGA. It seems extraordinary that an opal mutation in the *prfB* gene generates an opal suppressor activity. The efficiency of termination at this mutational UGA codon was tested by using the *prfB-lacZ* fusion plasmid carrying the *supK584* opal substitution. The synthesis of the opal fusion protein, which is less than 4% of the wild-type fusion protein in the *supK*$^+$ strain, increased three-fold in the *supK584* strain (Kawakami & Nakamura, submitted). The +1 frameshift at the 26th UGA codon was also increased in *supK584*. These results can be interpreted as indicating that suppression is caused by a reduced cellular level of the mature RF2 protein in the opal mutant, presumably due to inefficient termination at UGA. Thus, the opal RF2 mutation is autogenously suppressed under these conditions, generating another feedback regulatory loop for the synthesis of RF2 in the mutant.

III. *herC* Mutation and Lysyl-tRNA Synthetase

E. coli has two forms of lysyl-tRNA synthetases (LysRS). A major form of LysRS is encoded by a gene named *lysS* (Emmerich & Hirshfield, 1987) and a minor form is encoded by *lysU* (Van Bogelen *et al.*, 1983). These genes are located on the distinct chromosomal loci; *lysU* at min 92 and *lysS* at min 62. However, the structural genes had not been isolated. In this article, we describe that the *herC* gene which is in the same operon with *prfB* encodes LysRS (*lysS*).

A. herC Mutation. The *herC180* mutation has been defined in a host mutant which restored maintenance of a replication-defective ColE1 plasmid carrying a primer RNA mutation named *cer114* (Kawakami *et al.*, 1989). Two modes of ColE1 DNA replication are known, one dependent on RNase H, and the other independent on RNase H (Dasgupta, 1987). The *cer114* mutant replicon is defective in both modes of replication and carries a single base pair alteration 95 bp upstream of the replication origin. It substitutes a G for an A at the 3'-terminus position of the AAA triplet in the structure IX loop of primer RNA. One of the revertants of *cer114* inserted an A to this region and regenerated an AAA triplet, thereby suggesting a crucial role of this AAA triplet in ColE1 DNA replication. An *E. coli* mutant which restored maintenance of the *cer114* replicon carried double mutations, one in the RNase H gene and the other in the *herC* gene. Complementation and reconstruction experiments revealed that the *herC180* mutation is recessive to its wild-type allele and supports maintenance of the mutant replicon in the absence of RNase H. The wild-type ColE1 replicon is also maintained in the double mutant. These data suggested that the wild-type *herC* protein (*i.e.*, LysRS as shown below) is not a factor essential for ColE1 DNA replication but a factor which eliminates *cer114* replication or prevents stable maintenance of the *cer114* replicon. The *herC180* mutation alone conferred cold-sensitivity in growth, suggesting that the *herC* function is essential to *E. coli* growth (Kawakami *et al.*, 1989). The DNA sequence analysis disclosed a single base substitution in *herC180* which causes a glycine-to-aspartate change at amino acid position 426 (Mikuni, unpub.).

B. Lysyl-tRNA Synthetase. We found that the *herC* protein shares a significant homology with *S. cerevisiae* cytoplasmic aspartyl-tRNA synthetase. Among several aminoacyl-tRNA synthetase activities tested, the LysRS activity in crude cell lysates increased three-fold in the presence

of pKK945 carrying *herC* and further increased in the presence of pKK990 where *herC* is fused to the overexpressing *tac* promoter (Kawakami, unpub.). The *herC* protein labeled with [^{35}S]methionine in the *in vitro* coupled transcription-translation system exactly comigrated with the purified LysRS in the O'Farrell 2D gel electrophoresis.

The overexpressed *herC* protein was purified from the pKK990-bearing cells (*lon$^{\Delta}$100, htpR16*) after induction by IPTG. The fractions containing the *herC* protein were monitored by staining proteins after SDS-PAGE electrophoresis and compared with those of the LysRS activity. Throughout the purification procedures which employed ammonium sulfate fractionation and several column chromatographies, the oversynthesized *herC* protein was exactly copurified with the LysRS activity (Nakamura, unpub.). These results led us to conclude that *herC* encodes LysRS.

IV. Summary and Perspectives

A. Comparative Aspects of Aminoacyl-tRNA Synthetases and Release Factors. We have proven that the RF2 and LysRS proteins are encoded in the same operon in *E. coli* and *S. typhimurium*. In addition to this chromosomal organization, LysRS and RF2 have several similarities in genetic and biochemical properties. First, *prfB* and *lysS* have structural and functional homologs, *prfA* and *lysU*, respectively. Second, they are essential factors in translation and both recognize RNA triplets, *i.e.*, codon UGA for RF2 and anticodon UUU or CUU for LysRS. Finally, LysRS and RF2 catalyze reverse reactions, formation of aminoacyl-tRNAs or hydrolysis of peptidyl-tRNAs. These genetic and functional relevance might suggest that they have evolved from a related family of proteins. From the point of view of UGA recognition and its evolution, it would be interesting to investigate the structural and functional organization of release factors and tryptophanyl-tRNA synthetase of *Mycoplsma capricolum*, in which UGA is used as a sense codon for tryptophan (Yamao *et al.*, 1985).

B. UGA Codon Recognition and Release Factor 2. The UGA codon is a leaky codon and is naturally misread at higher frequency than UAG and UAA. This leakiness is due to misdecoding by tryptophanyl-tRNA and also to limiting intracellular concentration of RF2. In fact, the leakiness

is eliminated upon introduction of the plasmid pKK951 which gives rise to an increase in the *prfB* gene dosage (Kawakami *et al.*, 1988b).

The UGA codon plays significant variability in translation. It has been known that several host mutations generate UGA misreading. In addition to this mutational variability, living organisms possess programmed alternatives in reading the UGA codon. In several instances, it is recognized as a signal for frameshift and suppression, or decoded as selenocysteine (reviewed by Parker, 1989). As a common feature, these programmed alternatives have a trick to stall ribosome before UGA and to eliminate proper functioning of the release factor at the specific UGA codon. In this article, we described that a reduced level of the RF2 activity or its cellular amount leads to an increase in translational errors such as suppression and frameshift. In addition, an increase in the *prfB* gene dosage reduces the efficiency of decoding of UGA to selenocysteine in the *E. coli fdhF* gene which encodes formate dehydrogenase (Nakamura & Böck, unpub.). All these results demonstrate that RF2 is directly involved in these alternative translation at UGA.

It is coming obvious that the RNA context plays a crucial role in ribosomal stalling. In the natural +1 frameshift in *prfB*, the in-frame UGA stop codon is preceded by leucine tRNA codon CUU (a *shifty* codon) and a Shine-Dalgarno like sequence AGGGGG three bases upstream of CUU. Weiss *et al.* (1988) have demonstrated that base-pairing between the 3' end of 16S rRNA and this upstream element is required for frameshift, presumably causing ribosomal stalling. Spontaneous high-frequency frameshift has been widely observed in eukaryotic organisms, such as retroviruses, coronaviruses and some yeast transposons (reviewed by Parker, 1989). These systems employ proper RNA contexts which produce RNA stem-loop and RNA pseudoknot structures, or encode minor codons, which cause ribosomal pausing. Development of the *in vitro* system to assay ribosomal stalling in dynamic reactions will certainly contribute to solve many problems in stop codon recognition.

C. Biological Relevance of Lysyl-tRNA Synthetase. LysRS is an exceptional case that it has two distinct forms of the synthetases encoded by *lysS* and *lysU*. The *lysS* gene is constitutively expressed, while *lysU* is heat-inducible though it remains almost silent under normal growth conditions (Hirshfield *et al.*, 1981). Independent of our studies, Gampel and Tzagoloff (1989) have also observed a similarity between yeast LysRS and *E. coli herC* protein. Recently Lévêque *et al.* (1990) have

cloned the *lysS* gene by using oligonucleotide probes directed from the NH_2-terminal peptide sequence and have found the identity of the *lysS* sequence to *herC*. These results, taken together with ours, have firmly established that *herC* encodes LysRS. The latter authors have also cloned the *lysU* gene and found 88% amino-acid sequence homology between *lysS* and *lysU*. Biological significance of two distinct LysRS species remains to be investigated.

A stretch of adenine, which includes the lysine codon, is recognized as a site for -1 frameshift in retroviruses, coronaviruses, *E. coli dnaX* and the transposase gene of insertion element IS*1* (reviewed by Parker, 1989; Tsuchihashi & Kornberg, 1990; Sekine & Ohtsubo, 1989). Most of these AAA lysine codons are followed, at some distance, by the stop codon UGA. Therefore, it is plausible that LysRS and the release factor may participate in regulation of these frameshifting.

It is not yet clarified if the *herC* mutation directly restores ColE1 DNA replication of the *cer114* mutant replicon or some other steps involved in the plasmid maintenance under *rnh*$^-$ conditions. However, the current studies suggested some relevance of LysRS to the mutant primer RNA. It is intriguing that the *cer114* mutation site is in the AAA triplet, which is a codon for lysine; thereby, one might speculate putative interaction between the anticodon loop of lysyl-tRNA and the primer RNA loop. Another possibility is that the suppression is mediated by 5',5'''-diadenosine tetraphosphate, possibly an important regulatory metabolite in cells synthesized by LysRS (Zamecnik, 1983). Diverse roles of tRNAs or aminoacyl-tRNA synthetases in global control of cell growth and gene expression are interesting topics for future work.

References

Caskey CT, Forrester WC, Tate W, Ward, CD (1984) Cloning of the *Escherichia coli* release factor 2 gene. J Bacteriol 158:365-368

Craigen WJ, Cook RG, Tate WP, Caskey CT (1985) Bacterial peptide chain release factors: conserved primary structure and possible frameshift regulation of release factor 2. Proc Natl Acad Sci USA 82:3616-3620

Dasgupta S, Masukata H, Tomizawa J (1987) Multiple mechanisms for initiation of ColE1 DNA replication: DNA synthesis in the presence and absence of ribonuclease H. Cell 51:1113-1122

Emmerich RV, Hirshfield IN (1987) Mapping of the constitutive lysyl-tRNA synthetase gene of *Escherichia coli* K-12. J Bacteriol 169:5311-5313

Gampel A, Tzagoloff A (1989) Homology of aspartyl- and lysyl-tRNA synthetases. Proc Natl Acad Sci USA 86:6023-6027

Garcia GM, Mar PK, Mullin DA, Walker JR, Prather NE (1986) The *E. coli dnaY* gene encodes an arginine transfer RNA. Cell 45:453-459

Hirshfield IN, Bloch PL, Van Bogelen RA, Neidhardt FC (1981) Multiple forms of lysyl-transfer ribonucleic acid synthetase in *Escherichia coli*. J Bacteriol 146:345-351

Kawakami K, Jönsson YH, Björk GR, Ikeda H, Nakamura Y (1988a) Chromosomal location and structure of the operon encoding peptide-chain-release factor 2 of *Escherichia coli*. Proc Natl Acad Sci USA 85:5620-5624

Kawakami K, Inada T, Nakamura Y (1988b) Conditionally lethal and recessive UGA-suppressor mutations in the *prfB* gene encoding peptide chain release factor 2 of *Escherichia coli*. J Bacteriol 170:5378-5381

Kawakami K, Naito S, Inoue N, Nakamura Y, Ikeda H, Uchida H (1989) Isolation and characterization of *herC*, a mutation of *Escherichia coli* affecting maintenance of ColE1. Mol. Gen. Genet. 219:333-340

Lévêque F, Plateau P, Dessen P, Blanquet S (1990) Homology of *lysS* and *lysU*, the two *Escherichia coli* genes encoding distinct lysyl-tRNA synthetase species. Nucl Acids Res 18:305-312

Parker J (1989) Errors and alternatives in reading the universal genetic code. Microbiol Rev 53:273-298

Reeves RH, Roth JR (1971) A recessive UGA suppressor. J Mol Biol 56:523-533

Reeves RH, Roth JR (1975) Transfer ribonucleic acid methylase deficiency found in UGA suppressor strains. J Bacteriol 124:332-340

Roesser JR, Nakamura Y, Yanofsky C (1989) Regulation of basal level expression of the tryptophan operon of *Escherichia coli*. J Biol Chem 264:12284-12288

Ryden M, Murphy J, Martin R, Isaksson L, Gallant J (1986) Mapping and complementation studies of the gene for release factor 1. J Bacteriol 168:1066-1069

Scolnick E, Tompkins R, Caskey T, Nirenberg M (1968) Release factors differing in specificity for terminator codons. Proc Natl Acad Sci USA 61:768-774

Sekine Y, Ohtsubo E (1989) Frameshifting is required for production of the transposase encoded by insertion sequence *1*. Proc Natl Acad Sci USA 86:4609-4613

Tsuchihashi Z, Kornberg A (1990) Translational frameshifting generates the γ subunit of DNA polymerase III holoenzyme. Proc Natl Acad Sci USA 87:2516-2520

Van Bogelen RA, Vaughn V, Neidhardt FC (1983) Gene for heat-inducible lysyl-tRNA synthetase (*lysU*) maps near *cadA* in *Escherichia coli*. J Bacteriol 153:1066-1068

Weiss RB, Dunn DM, Dahlberg AE, Atkins JF, Gesteland RF (1988) Reading frame switch caused by base-pair formation between the 3' end of 16S rRNA and the mRNA during elongation of protein synthesis in *Escherichia coli*. EMBO J 7:1503-1507

Weiss RB, Murphy JP, Gallant JA (1984) Genetic screen for cloned release factor genes. J Bacteriol 158:362-364

Yamao F, Muto A, Kawauchi Y, Iwami M, Iwagami Y, Azumi Y, Osawa S (1985) UGA is read as tryptophan in *Mycoplasma capricolum*. Proc Natl Acad Sci USA 82:2306-2309

Zamecnik PC (1983) Diadenosine-5',5'''-p^1,p^4-tetraphosphate (Ap_4A): its role in cellular metabolism. Anal Biochem 134:1-10