FEBS 24481 FEBS Letters 488 (2001) 105–109

Hypothesis

Molecular mechanism of stop codon recognition by eRF1: a wobble hypothesis for peptide anticodons

Tomonari Muramatsu[a,*], Klaus Heckmann[b], Chifumi Kitanaka[a], Yoshiyuki Kuchino[a]

[a] *Biophysics Division, National Cancer Center Research Institute, Tokyo 104-0045, Japan*
[b] *Institut für Allgemeine Zoologie und Genetik, Universität Münster, Schlossplatz 5, D-48149 Münster, Germany*

Received 15 November 2000; accepted 6 December 2000

First published online 21 December 2000

Edited by Horst Feldmann

Abstract We propose that the amino acid residues 57/58 and 60/61 of eukaryotic release factors (eRF1s) (counted from the N-terminal Met of human eRF1) are responsible for stop codon recognition in protein synthesis. The proposal is based on amino acid exchanges in these positions in the eRF1s of two ciliates that reassigned one or two stop codons to sense codons in evolution and on the crystal structure of human eRF1. The proposed mechanism of stop codon recognition assumes that the amino acid residues 57/58 interact with the second and the residues 60/61 with the third position of a stop codon. The fact that conventional eRF1s recognize all three stop codons but not the codon for tryptophan is attributed to the flexibility of the helix containing these residues. We suggest that the helix is able to assume a partly relaxed or tight conformation depending on the stop codon recognized. The restricted codon recognition observed in organisms with unconventional eRF1s is attributed mainly to the loss of flexibility of the helix due to exchanged amino acids.

Key words: Codon recognition mechanism; Conventional; Unconventional eukaryotic class I release factor; Eukaryotic release factor 1; *Tetrahymena thermophila*; *Euplotes octocarinatus*

1. Introduction

Termination of protein biosynthesis requires class I release factors (RFs) to recognize stop codons (RF1 and RF2 in Bacteria, aRF1 in Archaea and eRF1 in Eukarya). RFs are polypeptides which bind to the aminoacyl (A) site of the ribosome when they recognize a stop codon (for reviews, see [1–4]). They mimic tRNAs. This led to the suggestion that there exists a peptide anticodon in RFs which reads stop codons [1,3]. Recently Nakamura and his colleagues identified a tripeptide Pro-Ala-Thr in RF1, which recognizes UAG and UAA and a tripeptide Ser-Pro-Phe in RF2 which recognizes UGA and UAA [5]. It is known that the conventional eRF1s function as omnipotent RFs decoding all three stop codons UAG, UAA, and UGA. In this case some residues probably recognize G or A in the second position of a stop codon and other residues G or A in the third position. However, the two processes of recognition cannot be achieved independently of each other because the tryptophan codon UGG is not mistaken for a stop codon. This raises the question of the mechanism by which UGG is discriminated.

Ciliates have reassigned some of the canonical stop codons to sense codons in evolution and have developed unconventional eRF1s with restricted stop codon recognition. *Paramecium tetraurelia* and *Tetrahymena thermophila* use UAA and UAG to encode glutamine and only UGA as stop codon [6–11]. *Euplotes octocarinatus* uses UGA to encode cysteine [12] and mostly UAA but sometimes also UAG as stop codons [13]. The altered stop codon recognition is reflected in the primary sequence of the eRF1s of these organisms. They show an amino acid replacement in a region which is 100% conserved in conventional eRF1s. This allowed to identify the amino acid residues of eRF1 which most probably read the stop codons [14]. Here we propose a mechanism explaining the way in which stop codon recognition in eukaryotes might be achieved. The proposed mechanism explains why conventional eRF1s recognize UAG, UAA, and UGA but not UGG. It also explains the restricted stop codon recognition of three unconventional eRF1s.

2. The amino acid residues reading the stop codons

The overall shape and dimensions of human eRF1 resemble a tRNA. The molecule is organized into three domains and domain 1 was suggested to be involved in codon recognition [15]. Fig. 1 compares homologs of five conventional and three ciliate eRF1s of the region forming domain 1. It shows that the amino acids from position 57 to 68 are 100% conserved in the conventional eRF1s which recognize all three canonical stop codons. In the unconventional eRF1s [14,16] which are restricted in codon recognition, each one in a different way, the amino acids Gly57, Thr58, Ser60, Asn61, Ser64, Arg65 or Val66 are substituted. The positions of these amino acids in the three-dimensional structure of eRF1 are indicated in Fig. 2. Among the exchanged amino acids the doublets Gly57Thr58 and Ser60Asn61 occupy positions corresponding to tRNA anticodons. The side chains of Thr58, Ser60 and Asn61 point outward from the eRF1 molecule in a direction that corresponds with the direction in which the base moieties of the anticodons of tRNAs are directed. In addition, the perpendicular distance between the sets Gly57Thr58 and Ser60Asn61 is comparable to the distance between bases of

*Corresponding author. Fax: (81)-3-3546 1369.
E-mail: tmuramat@ncc.go.jp

Abbreviations: eRF, eukaryotic release factor

0014-5793/01/$20.00
PII: S0014-5793(00)02391-7

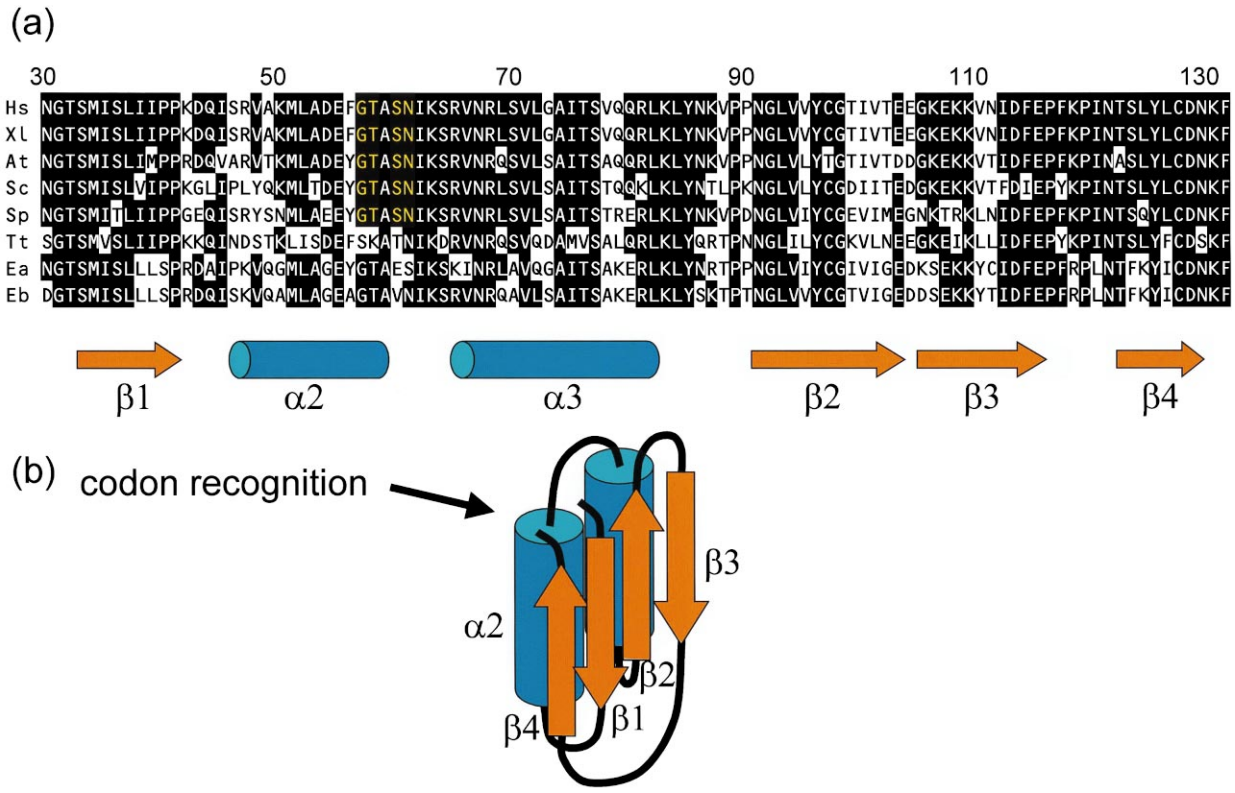

Fig. 1. Structure of domain 1 of eukaryotic class I RFs (eRF1s). (a) Comparison of the amino acid sequences of domain 1. The secondary structure for human eRF1 domain 1 [15] is also shown. Hs, *Homo sapiens* (accession number X81625); Xl, *Xenopus laevis* (Z14253); At, *Arabidopsis thaliana* (X69375); Sc, *Saccharomyces cerevisiae* (X04082); Sp, *Schizosaccharomyces pombe* (D63883); Tt, *T. thermophila* (AB026195); Ea, *E. octocarinatus* eRF1a (AJ272501); Eb, *E. octocarinatus* eRF1b (AF245454). Identical amino acids conserved in the majority of eRF1s are shown in black. The number of the amino acid position is counted from the N-terminal Met of the human eRF1 sequence. (b) Schematic drawing of the tertiary structure of domain 1 according to the crystal structure of human eRF1 [15].

the anticodon of tRNAs. Thus, by analogy with codon recognition by tRNA, we propose that the set Gly57Thr58 recognizes the second letter of a stop codon and that the set Ser60Asn61 recognizes its third letter. The amino acids Ser64, Arg65 and Val66 appear not to be directly involved in codon recognition. These residues are too distant from Gly57Thr58 and Ser60Asn61. Moreover, the side chains of Arg65 and Val66 are not directed towards the mRNA codon side. However, these residues may be involved in the regulation of the flexibility of the part of the helix on which Gly57Thr58 and Ser60Asn61 are located. Ser64, Arg65 and Val66 may interact with ribosomal RNA or other residues and may in this way contribute indirectly to codon recognition.

3. The codon recognition hypothesis

Conventional eRF1s recognize the three canonical stop codons UAG, UGA, and UAA but discriminate UGG, the codon for tryptophan. Conventional eRF1s therefore must be able to distinguish between the bases A and G in both, the second and the third position of a stop codon. This task could not be achieved by a recognition process based on size of the bases, i.e. in a kind of substrate recognition pocket, nor by hydrophobic interactions between bases and peptides. Therefore we propose that the adenine and the guanine bases are recognized via hydrogen bonds. This reflection together with the knowledge of amino acid replacements in unconventional eRF1s, and the knowledge of the dimensions of the tertiary structure in the region critical for codon recognition in human eRF1 leads us to propose the following hypothesis:

1. The C-terminal region of the α2-helix containing the amino acids critical for codon recognition is flexible and can assume tight and partly relaxed conformations.
2. Hydrogen bonds formed between an amino acid residue and the base at the second position of a stop codon restrict the bonds that can be formed between amino acids and the base at the third position.
3. The flexibility of the α2-helix is influenced by interactions of some of its amino acid residues with residues of the underlying β-sheets.

Thus, in conventional eRF1s recognition of UAG would occur by the formation of a hydrogen bond between the N_1 of the adenine and the hydroxyl group of the amino acid Thr58 (Fig. 3a), while the hydroxyl group of Ser60 and the amide group of Asn61 form bonds with the guanine base in the third position. Recognition of UGA occurs by the formation of a hydrogen bond of Thr58, this time with the guanine base (Fig. 3a), causing a shift of Thr58 which influences the recognition of the base in the third position. As a consequence only Asn61 and not Ser60 can form the hydrogen bond and therefore only adenine and not guanine can be recognized at the third position. Recognition of UAA is achieved by the formation of a hydrogen bond between the adenine in the

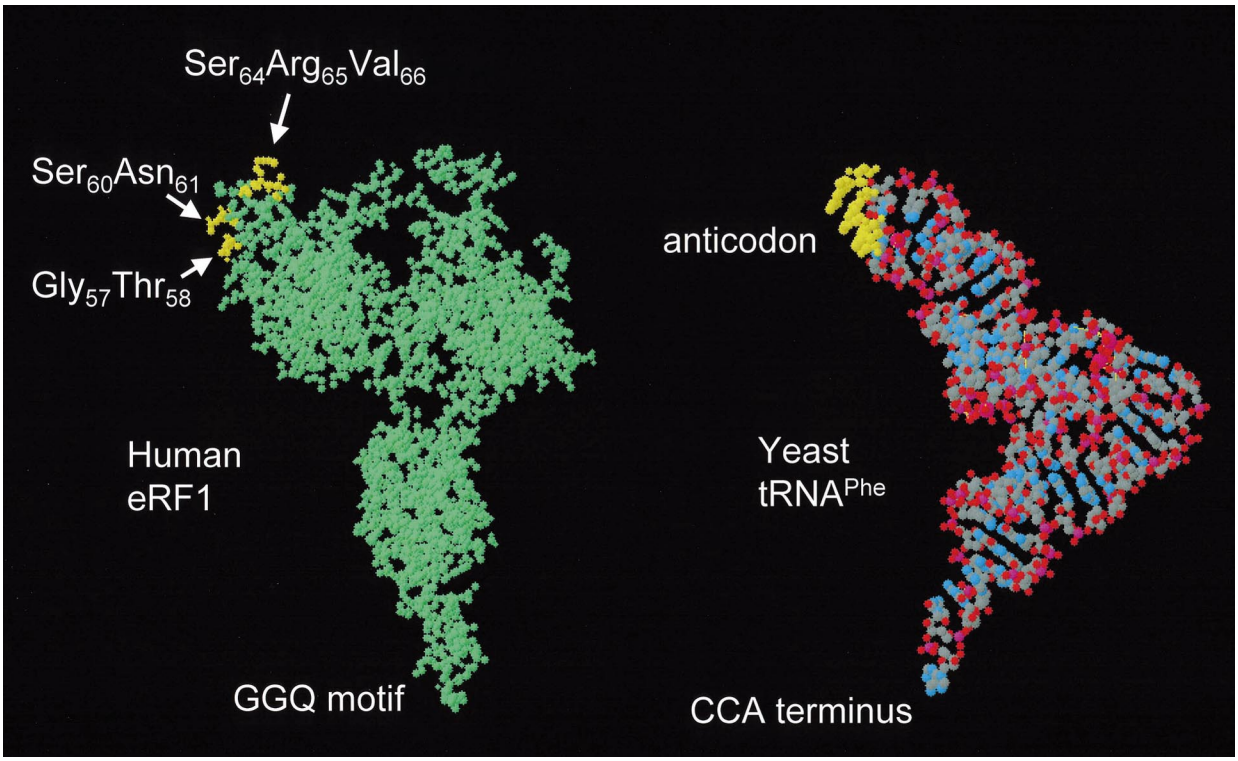

Fig. 2. Three-dimensional structures of human eRF1 and yeast $tRNA^{Phe}$. The coordinate data are from the Protein Data Bank (human eRF1: 1DT9, yeast $tRNA^{Phe}$: 1EVV) and displayed using the Cn3D program version 2.5 for Windows (http://www.ncbi.nlm.nih.gov). Gly57, Thr58, Ser60, Asn61, Ser64, Arg65, and Val66 of eRF1, and the anticodon of the tRNA are in yellow.

second position and Thr58 and by the formation of a hydrogen bond between the adenine in the third position and Asn61.

This proposal for codon recognition takes into account that the amino acids which interact with the stop codons are located towards the C-terminal part of the α2-helix which therefore should be flexible owing to a balance between two forces, one resulting from the tendency to form an α-helix (3.6 residues/turn), and the other coming from the tendency to form a 3_{10}-helix (3.0 residues/turn). The latter is supported by hydrophobic interactions among the periodically located hydrophobic residues (Ala53, Phe56, Ala59, and Ileu62), by the interaction between these hydrophobic residues and the β-sheets (β1 and β4) which lie behind the helix, and by the existence of glycine in position 57. In correspondence with the two forces, we assume two potential conformations: an α-helix-like 'relaxed' (R) conformation and a 3_{10}-helix-like 'tight' (T) conformation.

Although Gly57Thr58 and Ser60Asn61 are able to recognize both the adenine and the guanine bases by forming hydrogen bond(s) between their side chain(s) and the base moiety of the codon, codon recognition is restricted by the locations of these residues. Their location is regulated by the flexibility of the part of the helix which contains the two sets of amino acids (Gly57Thr58 and Ser60Asn61). The ability of the two parts of the helix to assume either tight (T) or relaxed (R) states allows to read all three stop codons: $T_{Gly57Thr58}T_{Ser60Asn61}$ for UAG, $R_{Gly57Thr58}T_{Ser60Asn61}$ for UGA, and $T_{Gly57Thr58}R_{Ser60Asn61}$ for UAA. Since a conformation tighter than the 3_{10}-helix cannot be assumed, UGG codon recognition is excluded. It would require $R_{Gly57Thr58}(\text{super tight})_{Ser60Asn61}$, which cannot be achieved.

4. Stop codon recognition in unconventional eRF1s

The hypothesis developed above to explain stop codon recognition in conventional eRF1s can also be applied to the three unconventional eRF1s present in ciliates (Fig. 1). To account for the situation in these ciliates, one has to assume that the flexibility of the part of the α2-helix bearing the position-recognition site for the second base in a stop codon has become restricted to the tight state probably due to the amino acid replacements in positions 64–66 but perhaps also by replacements in other positions.

In the eRF1 of *Tetrahymena* the hydroxyl group necessary for the formation of a hydrogen bond has shifted and is now provided by Ser57 instead of Thr58 (Fig. 3b). This causes a change in specificity at the second position (only guanine can now be recognized). In addition due to a methyl group caused by the substitution (Ser60 by Thr60) *Tetrahymena* eRF1 cannot recognize a guanine in the third position of the codon. Furthermore, the substitution of Gly57 by Ser57 causes a loss of flexibility, resulting in the fixation of the 'relaxed' conformation of the helix in the third position-recognition site. Thus, *Tetrahymena* eRF1 can only recognize UGA as a stop codon.

Although *Euplotes* eRF1a and eRF1b have conserved Gly57Thr58, amino acid residues outside of this region probably restrict the flexibility in the second position-recognition site. The residues Ile, Ile, Pro (38–40) and Ser, Leu, Tyr, Leu (123–126) which are well-conserved in conventional eRF1s, are altered in *Euplotes* to Leu, Leu, Ser and Phe, Lys, Tyr, Ile, respectively (Fig. 1), in both eRF1a and eRF1b. These residues are located in the β1 and β4 sheets, which lie behind the α2-helix (Fig. 1), and are assumed to influence the flexi-

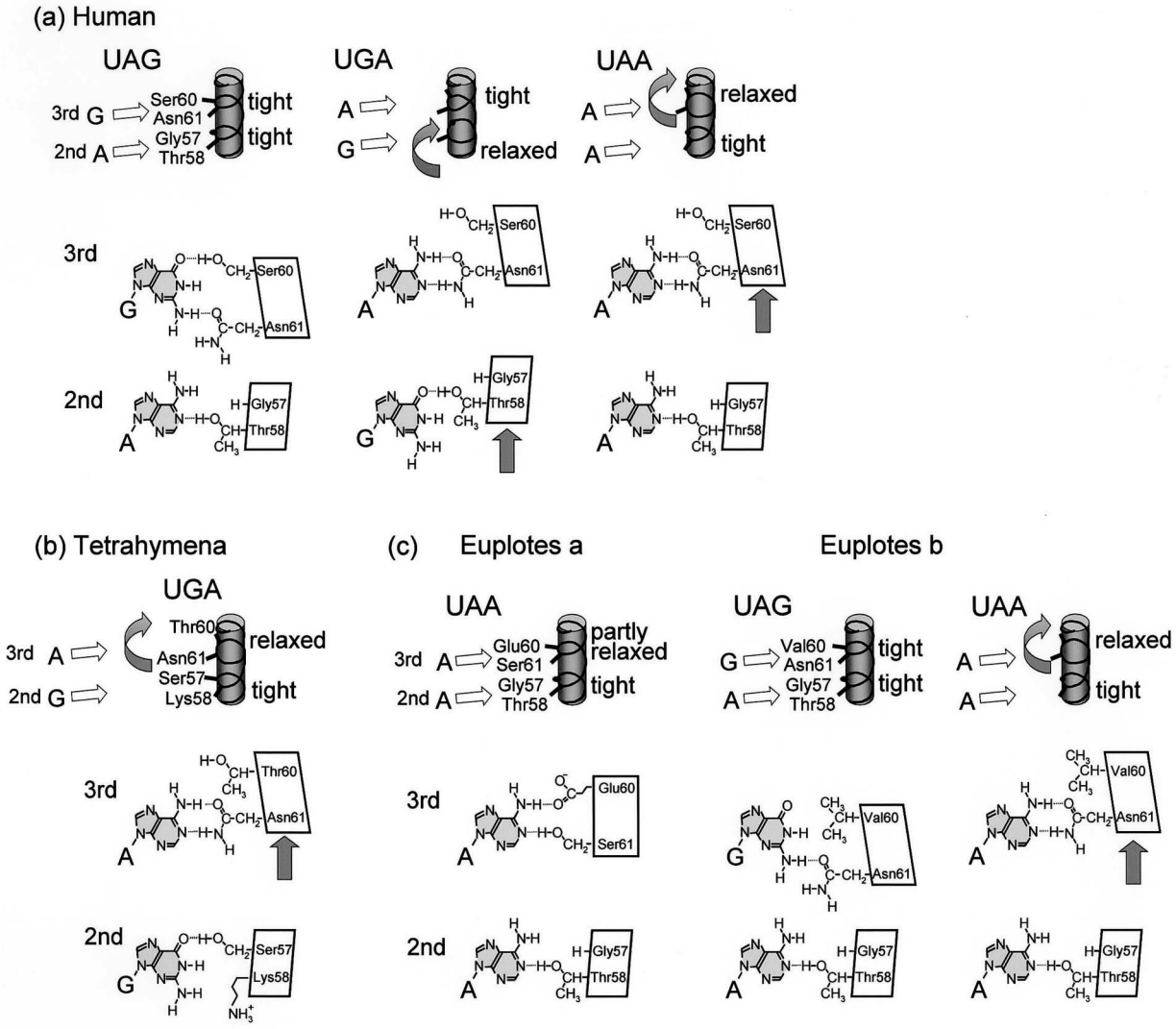

Fig. 3. Mechanism for stop codon recognition. (a) Human eRF1. (b) *Tetrahymena* eRF1. (c) *Euplotes* eRF1s. Shaded arrows in the lower panel indicate the direction of the shift necessary for assuming the relaxed conformation. These directions correspond to the rotation (shaded curved arrows) in the helix indicated in the upper panel. *Euplotes* eRF1a may recognize adenine in the third position of the stop codon by Glu60Ser61 which is located in an intermediate (partly relaxed) position between a 'tight' and a 'relaxed' state (c, left).

bility of helix 2. We propose that changes in these regions cause the rigidity in the second position-recognition site which allows only adenosine recognition. For the third position of stop codons, the replacement of Ser60Asn61 in eRF1a to Glu60Ser61 probably causes an adenine specific recognition (Fig. 3c). Although eRF1b can recognize both UAA and UAG because eRF1b has almost the same amino acid sequence as conventional eRF1s except for the position 61, the replacement of Ser60 by Val probably causes poor recognition of the guanine base in the third position of stop codons (Fig. 3c). This might be the reason why the stop codon UAG is only poorly used in *Euplotes*. Thus, we assume that *Euplotes* eRF1a is specific for UAA, and eRF1b is specific for UAA and UAG, however, the latter stop codon being only poorly recognized.

5. Testing the stop codon recognition hypothesis

There are three obvious ways for testing our hypothesis:

1. Biophysical studies such as nuclear magnetic resonance spectroscopy of domain 1 (residues 30–131) of human eRF1 should reveal whether the C-terminal portion of the α2-helix is indeed flexible enough to assume in solution a 3_{10}-helix in addition to an α-helix.
2. eRF1s of other organisms (other ciliates and the algae *Acetabularia cliftonii* and *A. mediteranea* [17]), where UAA and UAG encode glutamine should be sequenced. This would reveal whether they also show changes in amino acid residues claimed above to be critical in codon recognition.
3. Mutations introduced in residues of α2, α3, β1 and β4 might not only impair stop codon recognition but also cause alterations in the codon specificity by altering the flexibility of the α2-helix. This could perhaps be recognized in an in vitro system. Such a study has just been published [18]. All but one of the mutations selected for interference with stop codon recognition in yeast eRF1 map in the regions mentioned above.

6. Concluding remarks

In our hypothesis for stop codon recognition by eRF1s, Gly57Thr58 which can be regarded as 'the second letter of the peptide anticodon' recognizes either A or G in the second position of the stop codon, and Ser60Asn61 ('the first letter of the peptide anticodon') recognizes either A or G in the third position of the stop codon. Flexibility for shifts of Gly57Thr58 and Ser60Asn61 is required for differential recognition, which is analogous to codon recognition by tRNA molecules. In the wobbling concept devised by Crick [19], U in the first position of the anticodon of tRNA can pair with either A or G in the third position of a codon. To make GU pairs, U should be displaced from the location in an AU Watson–Crick type base pair. Yokoyama et al. showed that this movement is achieved by the flexibility of the ribose ring puckering, which is regulated by the post-transcriptional modification of uridine in the first position of the anticodon of tRNA [20]. For eRF1, flexibility required for stop codon recognition is conferred by the helical structure bearing the peptide anticodons. In conventional eRF1s Gly57Thr58 and Ser60Asn61 need to wobble in order to recognize A and G in the second and the third positions of stop codons.

While the wobbling in tRNA occurs only for the recognition of the third letter of a codon, wobbling at two sites of eRF1 is required for stop codon recognition: for the recognition of the second and for the recognition of the third letter of a stop codon. Moreover, eRF1 should not recognize these two positions of a stop codon independently from each other, in order to exclude UGG. This is an achievement of the helical nature of the 'peptide anticodon', which in this respect surpasses the 'nucleotide anticodon' of tRNAs. In this sense, eRF1, as a protein, not only mimics tRNA but has developed an even more sophisticated mechanism for codon recognition than tRNA.

The recognition mechanism of the uracil base in the first position of the stop codon is not addressed in our hypothesis. We assume that Glu55 is responsible for uracil recognition since Glu55 is well-conserved in all organisms, and Asp54Glu55, Gly57Thr58, and Ser60Asn61 are in line in the crystal structure of human eRF1. This assumption will be tested by mutagenesis or biophysical studies in the near future.

Acknowledgements: This work was supported by grants from the Ministry of Education, Science, Culture and Sports of Japan to Y.K. and T.M. We wish to thank Dr. Ito and Dr. Nureki of University of Tokyo for helpful discussions.

References

[1] Nakamura, Y., Ito, K. and Isaksson, L.A. (1996) Cell 87, 147–150.
[2] Buckingham, R.H., Grentzmann, G. and Kisselev, L. (1997) Mol. Microbiol. 24, 449–456.
[3] Nakamura, Y. and Ito, K. (1998) Genes Cells 3, 265–278.
[4] Nakamura, Y., Ito, K. and Ehrenberg, M. (2000) Cell 101, 349–352.
[5] Ito, K., Uno, M. and Nakamura, Y. (2000) Nature 403, 680–684.
[6] Caron, F. and Meyer, E. (1985) Nature 314, 185–188.
[7] Preer Jr., J.R., Preer, L.B., Rudman, B.M. and Barnett, A.J. (1985) Nature 314, 188–190.
[8] Horowitz, S. and Gorovsky, M.A. (1985) Proc. Natl. Acad. Sci. USA 82, 2452–2455.
[9] Kuchino, Y., Hanyu, N., Tashiro, F. and Nishimura, S. (1985) Proc. Natl. Acad. Sci. USA 82, 4758–4762.
[10] Cupples, C. and Pearlman, R.E. (1986) Proc. Natl. Acad. Sci. USA 83, 5160–5164.
[11] Hanyu, N., Kuchino, Y., Nishimura, S. and Beier, H. (1986) EMBO J. 5, 1307–1311.
[12] Meyer, F., Schmidt, H.J., Plümper, E., Hasilik, A., Mersmann, G., Meyer, H.E., Engstrom, A. and Heckmann, K. (1991) Proc. Natl. Acad. Sci. USA 88, 3758–3761.
[13] Jahn, C.L., Erbeznik, M., Jaraczewski, J.W., Melek, M. and Shippen, D. (1994) Gene 151, 231–235.
[14] Liang, A., Brünen-Nieweler, C., Muramatsu, T., Kuchino, Y., Beier, H. and Heckmann, K. (2000) Gene (in press).
[15] Song, H., Mugnier, P., Das, A.K., Webb, H.M., Evans, D.R., Tuite, M.F., Hemmings, B.A. and Barford, D. (2000) Cell 100, 311–321.
[16] Karamyshev, A.L., Ito, K. and Nakamura, Y. (1999) FEBS Lett. 457, 483–488.
[17] Schneider, S.U., Leible, M.B. and Yang, X.-P. (1989) Mol. Gen. Genet. 218, 445–452.
[18] Bertram, G., Bell, H.A., Ritchie, D.W., Fullerton, G. and Stansfield, I. (2000) RNA 6, 1236–1247.
[19] Crick, F.H.C. (1966) J. Mol. Biol. 19, 548–555.
[20] Yokoyama, S., Watanabe, T., Murao, K., Ishikura, H., Yamaizumi, Z., Nishimura, S. and Miyazawa, T. (1985) Proc. Natl. Acad. Sci. USA 82, 4905–4909.