

GENE 07878

CCC.UGA: a new site of ribosomal frameshifting in *Escherichia coli*

(Accuracy of translation; termination of translation; shifty stop; heterologous gene expression; transferrin)

Maarten H. de Smit^a, Jan van Duin^a, Peter H. van Knippenberg^{a,†} and Hendrik G. van Eijk^b

^aDepartment of Biochemistry, Gorlaeus Laboratories, Leiden University, Leiden, The Netherlands; and ^bDepartment of Chemical Pathology, Erasmus University, Rotterdam, The Netherlands

Received by H. van Ormondt: 25 August 1993; Revised/Accepted: 17 December/19 December 1993; Received at publishers: 7 February 1994

SUMMARY

To activate expression of a human transferrin(Tf)-encoding cDNA in *Escherichia coli* by translational coupling, it was placed in an expression plasmid downstream from a 5'-terminal fragment from the replicase(R)-encoding gene of bacteriophage MS2. The resulting construct was found to produce, besides the desired Tf, a protein with the mobility of a fusion product (RTf) of the N-terminal R fragment and Tf. Analysis of available mutants showed that this fusion results from +1 ribosomal frameshifting at the end of the R reading frame. This region contains the sequence, CCC.UGA, suggesting that before termination occurs, tRNA^{Pro} may dislodge from the CCC codon and reassociate with the +1 triplet CCU. By further site-directed mutagenesis, we demonstrate that both the CCC codon and the termination codon are indeed required for the observed 2–4% frameshifting. When either triplet is changed, the frequency of frameshifting drops to 0.3% or less. These results classify CCC.UGA as a new '+1 shifty stop'.

INTRODUCTION

Escherichia coli ribosomes are generally quite proficient in maintaining the proper reading frame during translation of a mRNA. Although a certain level of inaccuracy is intrinsic to the translational machinery (Kurland, 1987), frameshifting on arbitrary RNA sequences normally occurs at a frequency of less than 0.01% per translocation (Atkins et al., 1991). However, many examples of far more efficient frameshifting events indicate that this frequency can be substantially enhanced by local properties of the mRNA sequence and structure

Correspondence to: Dr. J. van Duin, Department of Biochemistry, Gorlaeus Laboratories, University of Leiden, P.O. Box 9502, 2300 RA Leiden, The Netherlands. Tel. (31-71) 274-759; Fax (31-71) 274-340; e-mail: Smit_M@Rulgca.LeidenUniv.NL

[†]Deceased July 24, 1990.

Abbreviations: aa, amino acid(s); *E.*, *Escherichia*; nt, nucleotide(s); R, phage MS2 replicase; R, gene encoding R; RF2, release factor 2; RTf R-Tf fusion protein; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide-gel electrophoresis; Sm, streptomycin; sTf, serum Tf; Tf, human transferrin; Tf, gene encoding Tf.

(reviewed by Weiss et al., 1987; 1990a,b). Both prokaryotes and eukaryotes employ 'programmed' frameshifting at specific sites as a mechanism for translational regulation of gene expression (Atkins et al., 1990).

A well-known example of programmed frameshifting in *E. coli* is the +1 shift required for the synthesis of translational release factor RF2 (Craigie and Caskey, 1986). Here, one in three ribosomes slips at the sequence CUU.UGA that ends the zero frame. The single tRNA^{Leu} that has decoded the terminal CUU codon apparently shifts to the UUU triplet in the +1 frame, even though this requires the formation of a G-U-wobble interaction with the 5' nt of the codon. Extensive mutagenesis of the terminal sense codon has revealed that the frequency of frameshifting depends mainly on the stability of the shifted codon-anticodon interaction, but for unknown reasons the wild-type codon CUU is particularly efficient (Curran, 1993). Frameshifting at the RF2 site is strongly promoted by a base-pairing interaction between the 3' end of 16S rRNA and a complementary sequence upstream from the slip site, equivalent to the Shine-

Dalgarno region of translational initiation sites (Weiss et al., 1987; 1988).

Frameshifting without a biological function has so far been observed mainly at slip sites intentionally created in the laboratory. Sequences inducing frameshifting are presumably avoided in regular coding regions, but they may occur by chance in sequences that are not native to the host. For example, a synthetic insert in a coding region in *E. coli* was found to divert half the ribosomes to the +1 reading frame, because it contained two consecutive AGG codons (Spanjaard and Van Duin, 1988). Fortuitous shift sites can also be present in heterologous coding regions, because prokaryotic and eukaryotic ribosomes slip in response to different signals (Weiss et al., 1989; Tsuchihashi, 1991; Garcia et al., 1993). Sequences in the VP2 gene from infectious bursal disease virus have, for instance, been reported to induce +1 frameshifting in *E. coli* (Macreadie and Azad, 1991). We suspect that many similar observations have remained unpublished.

In this paper, we describe a +1 frameshifting event encountered during the construction of an expression plasmid for human transferrin (Tf). Using site-directed mutagenesis we show that the observed 2–4% frameshift-

ing efficiency depends on the stop codon of the zero frame and on the CCC codon immediately preceding it.

EXPERIMENTAL AND DISCUSSION

(a) A plasmid designed for translational coupling produces an unexpected fusion protein

As we will describe elsewhere, the aim of our work was high-level production of Tf in *E. coli*. For this purpose, we have cloned the Tf cDNA into an expression plasmid under control of the thermoinducible p_L -promoter from phage λ , and created an *E. coli*-type ribosome-binding site by site-directed mutagenesis. To improve expression by translational coupling, a 5' fragment from the replicase (R)-encoding gene of bacteriophage MS2 was inserted upstream. Fig. 1a shows that in the resulting construct (henceforth referred to as the R-reinitiation construct), Tf is coded in the +1 frame relative to the R fragment. Ribosomes that have initiated translation at the R start traverse the start codon of the Tf gene and terminate three triplets further downstream, where two consecutive nonsense codons (UGA.UAA) are present out-of-frame

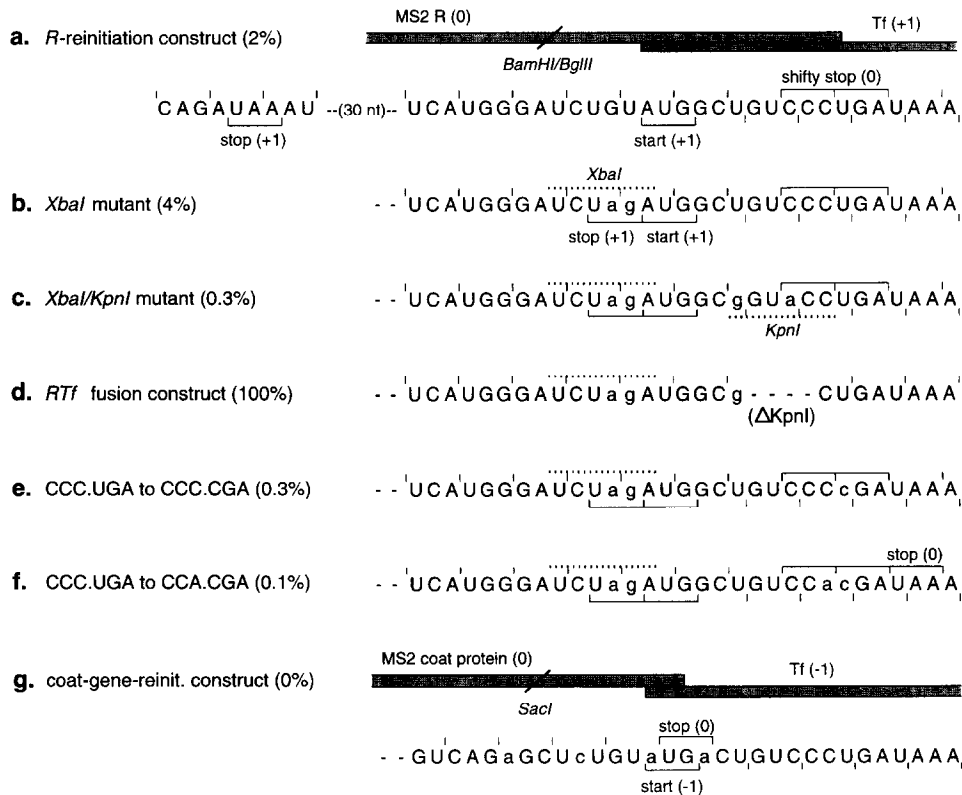


Fig. 1. Sequence of several relevant constructs at the junction between the upstream reading frame (either R or the MS2 coat-protein-encoding gene) and the Tf coding region. Substitutions introduced into the original R-reinitiation construct (a) by site-directed mutagenesis are shown in lower case. Ticks above the sequence denote the zero (= upstream) reading frame, ticks below the sequence denote the Tf frame (+1 or -1 as indicated). Relevant restriction sites and start and stop codons in both frames are indicated. $\Delta KpnI$ is a deletion of four nt, fusing the R and Tf reading frames. Percentages are frameshifting frequencies relative to the RTf fusion (d), as estimated from several Western blots. The Tf cDNA was obtained from ATCC as #53106. The expression vector is pPLc2833 (Remaut et al., 1983) and *E. coli* K-12 strain M5219 (Remaut et al., 1981) was used as host.

in the *Tf* coding sequence. The *Tf* start site can subsequently be reached by 'backscanning' over 8 nt (Adhin and Van Duin, 1990).

Induction of this construct in vivo turned out to result in the production of two proteins detectable by Western blotting with anti-Tf antibodies (Fig. 2, lane 3). Comparison with two Tf-derived fusion proteins of known size (Fig. 2, lanes 5 and 6) indicates that the smaller protein is the expected non-glycosylated form of Tf, while the other protein is about 100-aa larger. Realizing that the truncated *R* gene upstream from the *Tf* start codes for 101 aa, we suspected that the upper band could represent a RTf fusion protein. Accordingly, we fused the *R* and *Tf* frames artificially by deleting four nt (Fig. 1d). The resulting protein is indistinguishable from the original upper band by Western blotting (Fig. 2, lanes 2 and 3).

Synthesis of a fusion protein from the *R*-reinitiation construct implies that a fraction of the ribosomes translating the *R* sequence undergoes a +1 frameshift. This conclusion is supported by the disappearance of the upper band, when the upstream reading frame was replaced by part of the coat-protein-encoding gene of MS2. In the resulting construct, Tf is encoded in the -1 frame relative to the upstream gene, and ribosomes translating the zero frame terminate at a UGA codon overlapping the *Tf* start codon (Fig. 1g). This plasmid produces the same amount of Tf as the *R*-reinitiation construct, indicating that the upstream reading frame is efficiently translated. A second Tf-derived protein, however, is not observed (not shown).

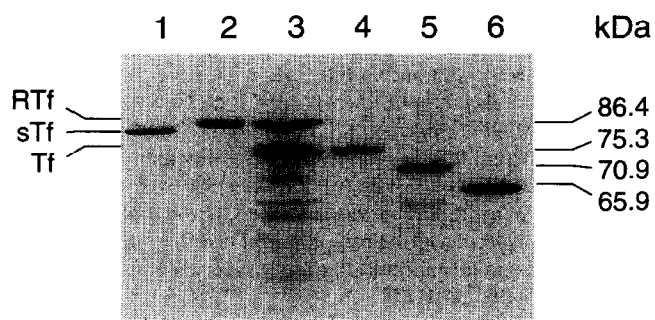


Fig. 2. Western blot showing various Tf-derived proteins. Lanes: 1, Tf isolated from human serum (sTf); 2, RTf fusion protein (see Fig. 1d); 3, proteins produced by the *R*-reinitiation construct (Fig. 1a); 4, Tf produced in *E. coli* (Tf, 75.3 kDa); 5 and 6, Tf-derived markers of 70.9 and 65.9 kDa, produced from appropriate fusion plasmids. The lower mobility of the serum Tf (lane 1) relative to the bacterial product (lane 4) can be ascribed to glycosylation of the former (Padma and Schryvers, 1990). Lanes 2 to 6 contain total cell extracts of cultures induced for 30 min and prepared as described (De Smit and Van Duin, 1990). Lane 3 contains 20-times more cell extract than the other lanes. Proteins were separated by 0.1% SDS-10% PAGE and detected using antibodies raised against human serum Tf (DAKO-Immunoglobulins, Glostrup, Denmark).

(b) Frameshifting occurs within a window of four codons

In the *R*-reinitiation construct, the frameshift window (i.e., the mRNA region in which the frameshift has to take place to produce the observed protein) is delimited by a UAA codon in the +1 frame of the *R* coding sequence, at position -48 relative to the *Tf* AUG, and the zero-frame UGA codon at position +12 terminating *R* translation (Fig. 1a). Fortunately, two substitutions originally designed to create an *Xba*I site for further cloning also change the triplet preceding the *Tf* AUG to a UAG nonsense codon in the +1 frame (Fig. 1b). Western blotting shows that this mutant still produces the fusion protein (Fig. 3, lane 4), reducing the frameshift window to only four codons. Note that the mutations do affect the efficiency of reinitiation at the *Tf* start, but that this is not related to frameshifting.

Since the remaining frameshift window does not contain an obvious shifty heptanucleotide, nor any consecutive rare codons, the only reasonable explanation for the formation of the fusion protein seemed to us a +1 shift at the CCC.UGA sequence ending the *R* reading frame. The GGG anticodon of a tRNA^{Pro} that has decoded the terminal CCC codon can be envisaged to undergo a +1 shift and rebind to the CCU triplet, analogous to the RF2 frameshift (see INTRODUCTION). Mutational analysis has shown that at similar sites, referred to as 'shifty stops', the stop codon forms an essential part of the frameshift signal (Weiss et al., 1987; 1990b; Hagervall et al., 1994). The shifted interaction in our system, consisting of two G-C pairs and a G-U-wobble pair, would be the strongest one possible for such a +1 shifty stop.

(c) Efficient frameshifting requires both a stop codon and the preceding triplet CCC

The terminal sense codon of the *R* reading frame, CCC, was altered to ACC, by a mutation originally introduced

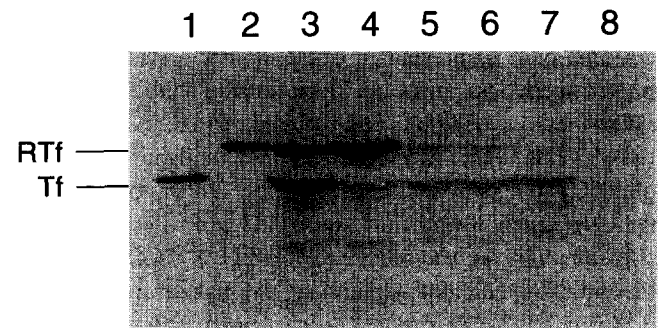


Fig. 3. Western blot showing proteins produced by several mutants of the *R*-reinitiation construct. Lanes: 1, reference plasmid producing only Tf; 2, *RTf* gene fusion (see Fig. 1d); 3, original *R*-reinitiation construct (Fig. 1a); 4, *Xba*I mutant (Fig. 1b); 5, *Xba*I/*Kpn*I mutant (Fig. 1c); 6, CCC.CGA mutant (Fig. 1e); 7, CCA.CGA mutant (Fig. 1f); 8, negative control (*Tf* deleted). Lanes 3 to 8 contain 20-times more cell extract than lanes 1 and 2. Protein were separated by 0.1% SDS-7.5% PAGE.

to create a *KpnI* site downstream from the AUG in addition to the *XbaI* site mentioned above (Fig. 1c). As expected, this change causes a substantial drop in frameshifting efficiency, from about 4% to 0.3% (Fig. 3, lane 5; percentages are relative to the fusion protein in lane 2, produced by the construct in Fig. 1d).

Because the *XbaI/KpnI* mutant contains a second substitution (3 nt upstream) that could influence the results, we constructed two more mutants by site-directed mutagenesis. In the first of them, the UGA codon was changed to CGA, resulting in termination at the adjacent UAA codon (Fig. 1e). Lane 6 of Fig. 3 shows that this substitution also causes a drop in frameshifting to 0.3%. While the resulting row of four Cs should be even more slippery than the original CCC.U sequence, the stop signal is clearly required for efficient frameshifting.

When CCC.CGA was further mutated to CCA.CGA (Fig. 1f), production of the RTf fusion protein dropped to a barely detectable level of 0.1% (Fig. 3, lane 7). Apparently, a few ribosomes still shift at the four C's in the previous mutant. We do not know whether the residual 0.1% originates at this same site or elsewhere within the frameshift window, but this percentage is close to the error level expected for four arbitrary codons (Atkins et al., 1991).

(d) The efficiency of frameshifting is relatively high

We estimate the frameshifting frequency at CCC.UGA to be about 2 to 4%, i.e. slightly higher than reported for other shifty stops not promoted by a Shine-Dalgarno-like interaction (Weiss et al., 1987; 1990b). For several reasons, this is somewhat unexpected. Firstly, our experiments were, for practical reasons, performed in a *Sm^R* strain carrying a mutation in ribosomal protein S12 (*rpsL*). This mutation causes a general reduction in translational errors, including frameshifting (Atkins et al., 1972; Weiss and Gallant, 1983), although some programmed frameshifts appear to be immune (Siple et al., 1991; Tsuchihashi, 1991). Secondly, consecutive stop codons, such as present in our constructs, have been reported to be less prone to frameshifting than single ones (Weiss et al., 1990b), and thirdly, the CCC-decoding tRNA^{Pro} is very inefficient at +1 shifting on sense oligo(C) strings (Weiss et al., 1990b). On the other hand, it was recently found that among 32 variants of the RF2 slip site carrying the sequence XYZ.UAG, CCC was the second most efficient triplet after the native codon CUU (Curran, 1993). This seems at least partly due to the high stability of the shifted codon-anticodon interaction, although properties of the tRNA may also contribute.

Hagervall et al. (1994) have provided evidence that undermodification of tRNA^{Pro} induces quadruplet reading at various sites, including CCC.UGA. This alternative mechanism for +1 frameshifting, however, does not

require the second codon to be a stop codon, while the frameshift observed by us clearly does. This indicates that the RTf fusion protein is not produced by a subpopulation of undermodified tRNAs.

(e) Conclusions

(1) We have detected a +1 frameshifting event that occurs during translation of a construct, designed for coupled expression of *Tf* in *E. coli*. This results in synthesis of a protein consisting of an N-terminal fragment of MS2 R fused to almost the entire *Tf* protein.

(2) Efficient frameshifting depends on the sequence CCC.UGA ending the zero frame. This site bears strong resemblance to a class of previously reported slip sites, known as shifty stops.

(3) The frameshifting efficiency is roughly 2–4%, i.e., rather high for a shifty stop not promoted by a Shine-Dalgarno-like interaction. This finding parallels the high efficiency of CCC.UAG relative to most other variants of the RF2 programmed frameshifting site.

ACKNOWLEDGEMENTS

We gratefully acknowledge W. Fiers and E. Remaut for providing the MS2 cDNA, the *p_L*-expression vector pPLc2833 and the host strain M5219. We thank J.F. Curran and G.R. Björk for helpful discussions and for providing data prior to publication.

REFERENCES

- Adhin, M.R. and Van Duin, J.: Scanning model for translational reinitiation in eubacteria. *J. Mol. Biol.* 213 (1990) 811–818.
- Atkins, J.F., Elseviers, D. and Gorini, L.: Low activity of β -galactosidase in frameshift mutants of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 69 (1972) 1192–1195.
- Atkins, J.F., Weiss, R.B. and Gesteland, R.F.: Ribosome gymnastics – degree of difficulty 9.5, style 10.0. *Cell* 62 (1990) 413–423.
- Atkins, J.F., Weiss, R.B., Thompson, S. and Gesteland, R.F.: Towards a genetic dissection of the basis of triplet decoding, and its natural subversion: programmed reading frame shifts and hops. *Annu. Rev. Genet.* 25 (1991) 201–228.
- Craigen, W.J. and Caskey, C.T.: Expression of peptide chain release factor 2 requires high-efficiency frameshift. *Nature* 322 (1986) 273–275.
- Curran, J.F.: Analysis of effects of tRNA: message stability on frameshift frequency at the *Escherichia coli* RF2 programmed frameshift site. *Nucleic Acids Res.* 21 (1993) 1837–1843.
- De Smit, M.H. and Van Duin, J.: Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proc. Natl. Acad. Sci. USA* 87 (1990) 7668–7672.
- Garcia, A., Van Duin, J. and Pleij, C.W.A.: Differential response to frameshift signals in eukaryotic and prokaryotic translational systems. *Nucleic Acids Res.* 21 (1993) 401–406.
- Hagervall, T.G., Tuohy, T.M.F., Atkins, J.F. and Björk, G.R.: Deficiency of 1-methylguanosine in tRNA from *Salmonella typhimurium* induces

- frameshifting by quadruplet translocation. *J. Mol. Biol.* 232 (1994) 756–765.
- Kurland, C.G.: Strategies for efficiency and accuracy in gene expression. *Trends Biochem. Sci.* 12 (1987) 126–128, 169–171 and 210–212.
- Macreadie, I.G. and Azad, A.A.: Internal initiation and frameshifting in infectious bursal disease virus sequence expressed in *Escherichia coli*. *Virology* 184 (1991) 773–776.
- Padda, J.S. and Schryvers, A.B.: N-linked oligosaccharides of human transferrin are not required for binding to bacterial transferrin receptors. *Infect. Immun.* 58 (1990) 2972–2976.
- Remaut, E., Stanssens, P. and Fiers, W.: Plasmid vectors for high-efficiency expression controlled by the p_L promoter of coliphage lambda. *Gene* 15 (1981) 81–93.
- Remaut, E., Tsao, H. and Fiers, W.: Improved plasmid vectors with a thermoinducible expression and temperature-regulated runaway replication. *Gene* 22 (1983) 103–113.
- Sipley, J., Dunn, J. and Goldman, E.: Bacteriophage T7 morphogenesis and gene 10 frameshifting in *Escherichia coli* showing different degrees of ribosomal fidelity. *Mol. Gen. Genet.* 230 (1991) 376–384.
- Spanjaard, R.A. and Van Duin, J.: Translation of the sequence AGG-AGG yields 50% ribosomal frameshift. *Proc. Natl. Acad. Sci. USA* 85 (1988) 7967–7971.
- Tsuchihashi, Z.: Translational frameshifting in the *Escherichia coli* dnaX gene in vitro. *Nucleic Acids Res.* 19 (1991) 2457–2462.
- Weiss, R. and Gallant, J.: Mechanism of ribosome frameshifting during translation of the genetic code. *Nature* 302 (1983) 389–393.
- Weiss, R.B., Dunn, D.M., Atkins, J.F. and Gesteland, R.F.: Slippery runs, shifty stops, backward steps, and forward hops: –2, –1, +1, +2, +5, and +6 ribosomal frameshifting. *Cold Spring Harbor Symp. Quant. Biol.* 52 (1987) 687–693.
- Weiss, R.B., Dunn, D.M., Dahlberg, A.E., Atkins, J.F. and Gesteland, R.F.: 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 (1988) 1503–1507.
- Weiss, R.B., Dunn, D.M., Shuh, M., Atkins, J.F. and Gesteland, R.F.: *E. coli* ribosomes re-phase on retroviral frameshift signals at rates ranging from 2 to 50 percent. *New Biol.* 1 (1989) 159–169.
- Weiss, R., Dunn, D., Atkins, J. and Gesteland, R.: The ribosome's rubbish. In: Hill, W., Moore, P., Garrett, R., Warner, J., Dahlberg, A. and Schlessinger, D. (Eds.), *Ribosomes: Structure and Function*. Am. Soc. Microbiol., Washington, DC, 1990a, pp. 534–540.
- Weiss, R.B., Dunn, D.M., Atkins, J.F. and Gesteland, R.F.: Ribosomal frameshifting from –2 to +50 nucleotides. *Progr. Nucleic Acid Res. Mol. Biol.* 39 (1990b) 159–183.