Vol. 211, No. 1, 1995 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
June 6, 1995 Pages 40-48

ANALYSIS OF TRANSLATIONAL TERMINATION OF RECOMBINANT HUMAN METHIONYL-NEUROTROPHIN 3 IN *ESCHERICHIA COLI*

Shi-Yuan Meng[1]*, John O. Hui[2], Mitsuru Haniu[2] and Larry B. Tsai[1]

[1]Department of Microbiology and Applied Microbial Genetics, and [2]Department of Protein Structure, Amgen Inc., Thousand Oaks, CA

Received April 26, 1995

SUMMARY: A highly efficient UGA stop codon readthrough event during the synthesis of human neurotrophin 3 in *E. coli* is described. The incorporation of a Trp residue at the UGA stop codon is confirmed combining both the chemical analyses and the molecular and genetic data in this report. The 3' adjacent nuleotide to the UGA stop codon plays a crucial role in determining the readthrough efficiency in the order of A > G > C > U. The replacement of UGA with UAA or UAG totally abolished this readthrough phenomenon and the use of Stp^R host cells also prevented the occurrence of UGA readthrough. Gene dosage (or plasmid copy number) effect was not indicated in this event; however, the titration of RF-2 by mRNA transcripts under over-expression conditions might explain why $tRNA^{trp}$ competes so well with RF-2 for UGA. Another apparently less produced readthrough product resulting from a transcript with no stop codon is also recorded, and the addition of a second in-frame stop codon increased the amount of the observed readthrough product.

Heterologous gene expression has become a common practice with the advent of modern molecular and genetic technologies. Such practice often casts concerns about the authenticity of the protein produced under over-expression conditions, especially when the protein is to be used for therapeutic purposes. *Escherichia coli* is the preferred organism for large-scale production of recombinant proteins for both technical and economical reasons. However, examination of the current literature shows there are an ample amount of cases concerning translational infidelity. These include mis-incorporation, frame shifting, ribosome hopping, stop codon readthrough, etc. Many of these phenomena are naturally evolved regulatory mechanisms in *E. coli* and have been well documented (1, 12). Understanding the signals for these events will help design synthetic gene to avoid producing undesirable novel proteins. Studies on stop codon readthrough and suppression on selected reporter genes strongly indicated codon context effect (2, 8, 10, 13). Genetic analyses on mutations affecting translational fidelity are also well documented (12). Neurotrophin 3 (NT-3) is one of the neurotrophic factors that are believed to play an essential role in the growth, survival, and differentiation of neurons in the nervous system (14). During the development of NT-3 as potential

* Correspondence to S.-Y. Meng, Amgen Inc., Mail-Stop 18S-1-A-377, Thousand Oaks, CA 91320-1789. Fax: 805-499-6819.

0006-291X/95 $5.00

therapeutics for neurodegenerative diseases, we observed the occurrence of an efficient UGA stop codon readthrough associated with the synthesis of human methionyl-NT-3 in *E. coli*. In this paper, the circumstances attributing to this phenomenon are discussed and compared with other reports.

RESULTS AND DISCUSSION

Construction of expression plasmids. Human *NT-3* gene was at first cloned into expression vectors pCFM1656 and pCFM1656.1 for direct expression in the cytoplasm of *E. coli*. Examination of the expressed proteins by SDS-PAGE showed the presence of two closely migrating but distinct polypeptides with pCFM1656-NT-3UGA (Figure 2A) but not with pCFM1656.1-NT-3UGA. Because pCFM1656.1 differs from pCFM1656 both in a deletion of 22 bp fragment and the plasmid copy number control, pCFM1656 was modified to generate pCFM1656(536) which had the same 22 bp deletion in pCFM1656.1 (Figure 1) but retained the plasmid copy number control of pCFM1656. Therefore, the comparison of NT-3 synthesis with pCFM1656 and pCFM1656(536) could not be interpreted as plasmid copy difference. Because pCFM1656(536) does not contain a second in-frame stop codon, it is possible that the readthrough product(s) is not stable enough to be observed; therefore, a second in-frame stop codon UAA was introduced into pCFM1656(536) resulting in pCFM1656(536UAA). As shown in Figure 1, all constructs cloned into pCFM1656 are either with different choices of stop codons or with changes 3' to the stop codon that make slightly different mRNA transcripts. None of the changes affect the coding sequence of the *NT-3* structural gene.

Stop codon usage and readthrough during the synthesis of NT-3. Automated Edman degradation indicated that the N-terminal sequence of the slower migrating NT-3 related polypeptide produced from pCFM1656-NT-3UGA was identical to that expected of the NT-3 polypeptide (7), suggesting the error(s) occurred at a later translational stage. Sequence information of the C-terminus was therefore crucial to identify the nature of the aberrant protein. While the protein chemistry analyses were going on to identify the NT-3 related polypeptide (7), molecular and genetic approaches were also used to analyze this phenomenon. Since the DNA coding sequence of NT-3 is correct, the co-expression of a protein of a higher MW is most likely to be a translational event. Several possibilities arise including frame-shift, mis-incorporation and readthrough. Replacement of UGA with UAG or UAA confirmed that it was a stop codon problem. Under the same expression condition, similar protein band of higher MW produced from pCFM1656-NT-3UGA was not detected when UAG or UAA was used as stop codon in pCFM1656-NT-3 UAG or pCFM1656-NT-3UAA (Figure 2). This stop codon evaluation is also consistent with the protein chemistry analysis that the NT-3 related polypeptide produced from pCFM1656-NT-3UGA is a UGA readthrough product by $tRNA^{trp}$. N-terminal analysis of the C-terminal peptide, together with mass spectrometry data, demonstrated that the formation of the higher molecular weight band was due to incorporation of a Trp residue (7)

Plasmid Construct	mRNA sequence 3' to stop codon
pCFM1656-NT-3UGA	**UGA** GGA UCC[GCG GAU AAA **UAA**GTAACGATCC]GGUCCAGUAAUGACCUC AGAACUCCAUCUGGAUUUGUUCAGAACGCUCGGUUGCCGCCGGGCGUUUUUUAU
pCFM1656.1-NT-3UGA pCFM1656(536)-NT-3UGA	**UGA** GGA UCC GGU CCA GUA AUG ACC UCA GAA CUC CAU CUG GAU UUG UUC AGA ACG CUC GGU UGC CGC CGG GCG UUU UUU AU
pCFM1656(536UAA)-NT-3UGA	**UGA** GGA UCC GGU CCA GUA AUG ACC UCA**UAA** CUCCAUCUGGAUUUGUU CAGAACGCUCGGUUGCCGCCGGGCGUUUUUUAU
pCFM1656-NT-3UAA	**UAA** GGA UCC GCG GAU AAA **UAA**GTAACGATCCGGUCCAGUAAUGACCUCA GAACUCCAUCUGGAUUUGUUCAGAACGCUCGGUUGCCGCCGGGCGUUUUUUAU
pCFM1656-NT-3UAG	**UAG** GGA UCC GCG GAU AAA **UAA**GTAACGATCCGGUCCAGUAAUGACCUCA GAACUCCAUCUGGAUUUGUUCAGAACGCUCGGUUGCCGCCGGGCGUUUUUUAU
pCFM1656-NT-3UGAA	**UGA** AGG AUC CGC GGA **UAA**AUAAGTAACGATCCGGUCCAGUAAUGACCUCA GAACUCCAUCUGGAUUUGUUCAGAACGCUCGGUUGCCGCCGGGCGUUUUUUAU
pCFM1656-NT-3UGAC	**UGA** CGG AUC CGC GGA **UAA**AUAAGTAACGATCCGGUCCAGUAAUGACCUCA GAACUCCAUCUGGAUUUGUUCAGAACGCUCGGUUGCCGCCGGGCGUUUUUUAU
pCFM1656-NT-3UGAU	**UGA** UGG AUC CGC GGA **UAA**AUAAGTAACGATCCGGUCCAGUAAUGACCUCA GAACUCCAUCUGGAUUUGUUCAGAACGCUCGGUUGCCGCCGGGCGUUUUUUAU
pCFM1656-NT-3UGAG	**UGA** GGG AUC CGC GGA **UAA**AUAAGTAACGATCCGGUCCAGUAAUGACCUCA GAACUCCAUCUGGAUUUGUUCAGAACGCUCGGUUGCCGCCGGGCGUUUUUUAU

Figure 1. Plasmid constructs for the production of NT-3. The gene encoding mature human NT-3 of 119 amino acids (14) was amplified using tailed oligonucleotides specific for the 5' and 3' ends of the gene. The 5' tail added an *Nde*I site containing the ATG start codon; the 3' tail provided a *Bam*HI site immediately following the native TGA sequence for the stop codon. The amplified product was cloned into the expression vector pCFM1656. Subsequently, similar site-directed mutagenesis and PCR techniques were use to generate different mRNA 3' tails. None of the modifications changed the structural gene sequence of *NT-3*. Expression vector pCFM1656 and its derivatives pCFM1656.1, pCFM1656(536) and pCFM1656(536UAA) are kanamycin resistant plasmids with multiple cloning sites downstream of the P_L promoter and a synthetic Shine-Dalgarno sequence and upstream of a ρ-independent transcription termination signal. There is a 22 bp deletion in pCFM1656.1 and pCFM1656(536) right after the *Bam*HI cloning site (underlined) compared to the sequence of pCFM1656. The sequence difference is indicated with a pair of brackets. All the listed plasmids except for pCFM1656.1 share the same plasmid copy control region. Expression vector pCFM1656.1 carries a point mutation that makes it a different copy number plasmid from the rest. The stop codons and the second in-frame stop codons used in each construct are indicated in bold type. The codons for additional aa resulting from potential stop codon readthrough are grouped in three letter code. All the plasmids were transformed into FM-5 which is an *E. coli* K-12 strain carrying the lambda *cI*857 repressor gene integrated in its chromosome. Oligonuleotide primers were chemically synthesized at Amgen Boulder. Plasmids were prepared using the Qiagen (Chatsworth, CA) Plasmid Midi Kit, and the double-stranded plasmids were sequenced using the Applied Biosystems (Foster City, CA) PRISM READY REACTION DyeDeoxy Terminator Cycle Sequencing Kit following the protocol supplied by the manufacturer.

Differences of expression patterns could be observed with each constructs after SDS-PAGE (Figure 2A). However, definite conclusion could not be drawn due to background obscurity and gel loading and running differences. Western blot analysis also faced the difficulties of separating two closely migrating polypeptides (Figure 2B). In spite of the separation difficulties, one previously unnoticed NT-3 related polypeptide produced from pCFM1656(536)-NT-3UGA was identified in the Western blot. It could also be a

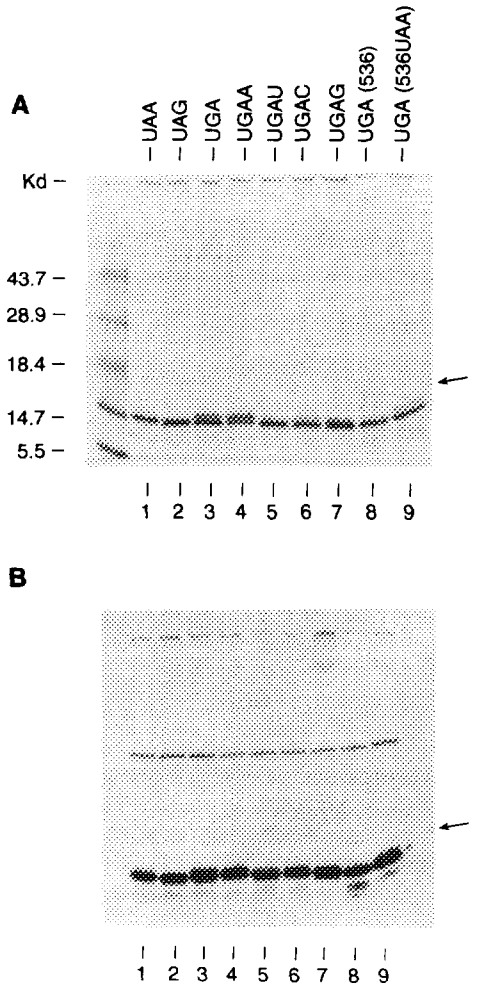


Figure 2. Comparison of NT-3 protein synthesis patterns from mRNA transcripts with different tails. Expression was conducted by diluting overnight LB + Km (50 μg/ml) culture grown at 30 °C twenty-five times into fresh 25 ml of LB with Km. The diluted culture was allowed to grow to an A_{600} of about 0.7 at 30 °C with aeration. The expression was initiated by increasing the temperature to 42°C. After four hours of induction, most of the NT-3 produced under such conditions aggregated in the cytoplasm in the form of inclusion bodies. To reduce the amount of background soluble *E. coli* proteins, cells were sonicated with Sonifier II Cell Disrupter purchased from Branson Ultrasonics Corp. (Danbury, CT) and the insoluble fraction was collected by centrifugation in a microcentrifuge. (A). Gel analysis. The pellet collected was resuspended and boiled in denaturing reducing buffer before being loaded onto an 18 % Tris-Glycine, SDS-PA gel. The gel, sample buffer, and running buffer were purchased from NOVEX (San Diego, CA). The gel was stained and destained according to Fast Stain (Zoion Research, Allston, MA) protocol. Lanes 1 through Lane 9 are insoluble fractions of induced NT-3 cultures. The stop codon usage and its adjacent nt are shown above each lane and correspond to the sequences listed in Figure 1. The less prominent band in Lane 8 is indicated by an arrow, which could be a readthrough product without a stop codon. (B). Western blot analysis. Samples for gel analysis were also blotted onto a nitrocellulose filter according to NOVEX protocol. The filter was probed with polyclonal anti-rat NT-3 antibodies using ECL Western Blotting Detection System purchased from Amersaham (Arlington Heights, IL). Lane 1 through Lane 9 correspond to the lane number in Figure 2A. Because the readthrough product migrated closely with the main product, they could not be well separated in this Western blot analysis. However, the much higher M.W. band in Lane 8 indicated by an arrow is visible.

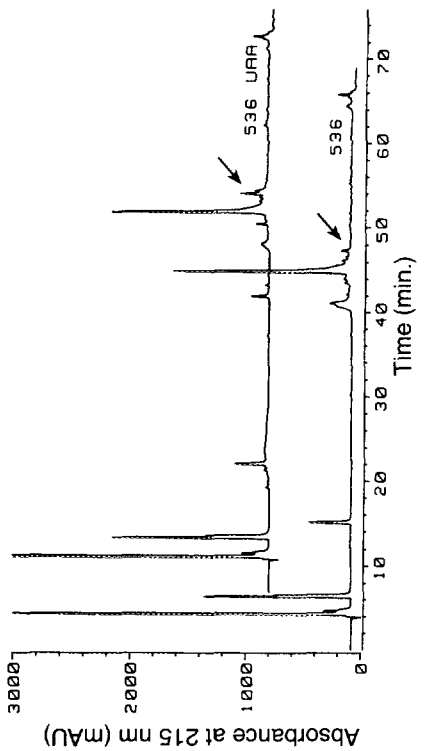
readthrough product from pCFM1656(536)-NT-3UGA which extended to a much longer length. It is unknown where this readthrough translation ends since there is no other in-frame stop codon after the original UGA stop codon in the transcript. Ribosomes might go all the way to the end of the transcript or they might fall off when stalled by the stem-loop structure of the transcriptional termination signal. To prove indeed there was readthrough occurring in pCFM1656(536)-NT-3UGA, a second in-frame stop codon UAA was introduced into pCFM1656(536)-NT-3UGA to generate pCFM1656(536UAA)-NT-3UGA by site-directed mutagenesis (Figure 1). This time a stronger signal of lower MW could be observed on an SDS- PA gel (Figure 2A). This indicates the same readthrough event could have different apparent results depending on what and how many amino acid residues are added on.

Although UGA is known to be the leakiest stop codon (5), the level of readthrough by normal $tRNA^{trp}$ is generally very low (1- 10 %) and thought to need a 3' adjacent A residue to promote its readthrough (6, 8). Most studies on leaky stop codons are conducted in strains with suppressors (*sup*) or mutations in tRNAs, releasing factors (*prf*) or small ribosomal proteins (*rps*) (2, 8, 13). On the contrary, the UGA readthrough in our case, with wild type *prfB* , *rpsL* and $tRNA^{trp}$ background, is a prominent event. This indicates the cellular concentration of $tRNA^{trp}$ is sufficient to occupy the four Trp codons in *NT-3* and to compete with RF-2 for UGA codon recognition. It is also possible that the concentration of RF-2 is titrated out under over expression conditions, and therefore the UGA stop codon is exposed to $tRNA^{trp}$ more readily. It would be interesting to see if the level of readthrough correlates with the amount of protein expressed. Stop codon replacement with either UAA or UAG did not result in detectable readthrough in NT-3 expression (Figure 2A and Figure 3), despite report on increased suppression efficiency on UAG if followed by a G residue (13).

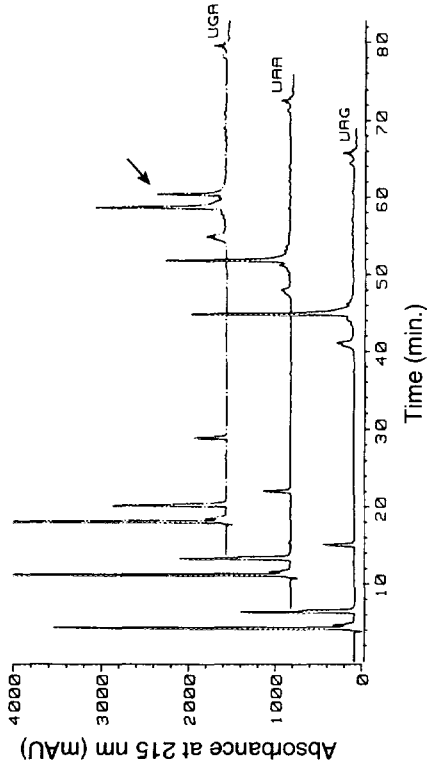
Context effect on UGA readthrough. The codon context effect has been implicated to play an important role in the fidelity and efficiency of codon recognition. Studies on the effects of nucleotide immediate 3' to stop codon UAG and UGA have been reported to influence suppression and readthrough (8, 13). The stop codon quadruplet theory was

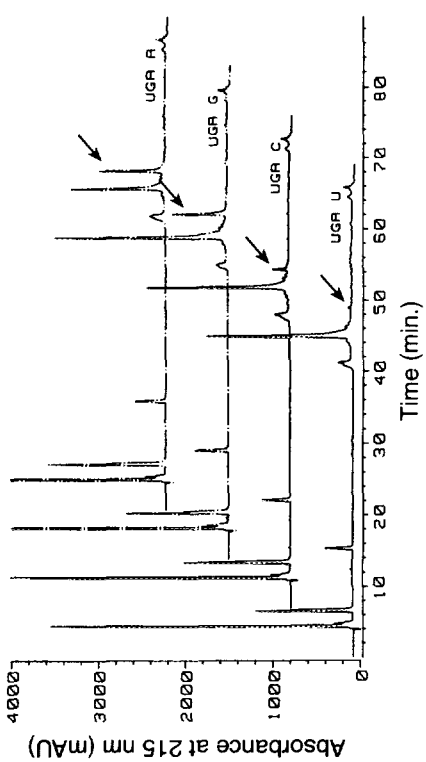
---

Figure 3. Quantification of various NT-3 stop codon readthrough products. To quantify the amount of readthrough from each expression system, the proteins were purified by RP-HPLC after prior reduction and S-carboxamidomethylation. In a typical experiment, the inclusion bodies from 10 ml of bacterial culture were solubilized in 100 µl of 6 M guanidine hydrochloride in 0.25 M Tris-HCl containing 1 mM EDTA at pH 8.5. Reduction was allowed to proceed with 10 mM DTT at 45 °C for 1 hr before the sulfhydryls were alkylated with 20 mM of iodoacetamide. The solution was diluted 10-fold with 0.1 % TFA in water and the precipitate was removed by centrifugation at 12,000 rpm for 20 min. The clear supernatant was chromatographed through a Vydac C4 column (0.46 x 25 cm). The conditions used have been described in details elsewhere (7). Each protein peak was collected manually and dried under vacuum prior to structural characterization. In each expression system, the ratio of the amount of termination readthrough to the total amount of protein expressed was determined by area integration. The readthrough products from each construct with different 3' mRNA tails are indicated by arrows and the percentages of readthrough are presented in the table on the lower right.



Plasmid Construct	Readthrough %
pCFM1656-NT-3UAA	Not Detected
pCFM1656-NT-3UAG	Not Detected
pCFM1656-NT-3UGA	30
pCFM1656-NT-3UGAA	37
pCFM1656-NT-3UGAU	1
pCFM1656-NT-3UGAC	7
pCFM1656-NT-3UGAG	18
pCFM1656-NT-3UGA(536)	3
pCFM1656-NT-3UGA(536UAA)	10





tested here on a heterologous protein, human NT-3, synthesis in *E. coli*. The UGA stop codon followed by an A, U, C or G was introduced into expression plasmid and the expression was compared (Figure 2A and Figure 3). Since we measured the amount of protein synthesized, any consequences associated with codon changes could affect expression level. For example, the obvious discrepancy between the 18 % readthrough in the case of pCFM1656-NT-3UGAG and the 30 % readthrough in the case of pCFM1656-NT-3UGA despite both had G residue following the UGA codon alerted us to the fact that we were comparing the synthesis of two different proteins that might have different translation efficiency and protein stability. Therefore the percentage of readthrough derived by measuring protein expression level in this paper should be interpreted with caution.

Despite the above mentioned concerns over the measurement of expression level, based on RP-HPLC analysis, we should still be able to draw an estimation that the effect of adjacent nucleotide 3' to the UGA codon on readthrough efficiency is in the order of A > G > C > U. This is in agreement with other published results (8). As for stop codon UAG, it has been shown that 3' nucleotide contribution to the suppression efficiency is greatest at G ≥ A > C > U (16), but we did not observe any readthrough at UAG, which is not a surprise since UAG readthrough (but not suppression) has been found in eukaryotes only (6). Buckingham and coworkers (4) also studied the codon context effect of 5' and 3' adjacent codons on UGA suppression and implicated that not only the two bases downstream of the nonsense codon are major determinants of suppression efficiency, the nature of the contiguous upstream codon can also affect nonsense suppression, as can the third base of the contiguous downstream codon. We have not systematically investigated the effects of the nucleotide at the second or third position 3' to the UGA codon and the contiguous codon 5' to the UGA codon.

Based on our unpublished data, when the stop codon UAG of brain-derived neurotrophic factor (BDNF), another neurotrophin in the same family of NT-3 (9), was substituted with UGA, similar readthrough was also observed in our *E. coli* expression system, although the 5' adjacent codon to UGA in BDNF gene is completely different from the one in NT-3 gene (Figure 4). In the recent paper published by Mottagui-Tabar

BDNF	AUU	AAA	AGG	GGA	CGU	**UAG**	GGATCC
	I	K	R	G	R	Z	
NT-3	AAA	AUC	GGA	AGA	ACA	**UGA**	GGATCC
	K	I	G	R	T	Z	

Figure 4. The last six codons and amino acid sequences of human BDNF and NT-3. Similar readthrough event also occured when UAG stop codon was replaced with UGA in *BDNF* . Shown here are the last six codons and their amino acid sequences used in natural *BDNF* and *NT-3*. genes. The stop codons (in bold type) are followed by a *Bam*HI site (underlined).

et al. (10), it suggests that high level of stop codon readthrough appears to be promoted by acidic, and low level of readthrough by basic amino acids in the second to last residue in the nascent peptide. The second to last amino acid of NT-3 is a basic arginine residue, but we have not investigated if the change to a acidic amino acid would further promoter readthrough. The curtailment of UGA readthrough in the Str^R derivative of FM-5 was observed in NT-3 readthrough (data not shown). To ensure host strain FM-5 does not contain some cryptic mutations, another *E. coli* K-12 strain W3110, with *cI857* integrated in the chromosome, known to have no genetic traits linked to translational fidelity, was tested for UGA readthrough. Similar readthrough efficiency was also observed in W3110 (data not shown), indicating this UGA readthrough was not caused by some unknown mutations in FM-5.

Effect of plasmid copy number on NT-3 translation fidelity. The conversion of pCFM1656-NT-3(UGA) to pCFM1656.1-NT-3(UGA) resulted in the apparent loss of the aberrant protein band observed for pCFM1656-NT-3(UGA) and the appearance of another NT-3 related protein band of much higher MW (Figure 2). Two possible causes could be accountable for the observation: plasmid copy difference resulted from single point mutation in the plasmid copy control region, and/or the mRNA message difference of 22 nt. Plasmid pCFM1656 and pCFM1656.1 both use R1 replication origin (11). While their copy numbers (about 10) are low at 30 °C, pCFM1656 has the potential to amplify more than 100 times when growth temperature is raised to above 36.5 °C, and thus characterized as a walkaway plasmid. In contrast, the magnitude of plasmid amplification of the non-walkaway plasmid pCFM1656.1 at higher temperature is very limited. Therefore, the copy numbers of pCFM1656-NT-3(UGA) and pCFM1656.1-NT-3(UGA) at induction temperature 42 °C could be very different. However, the mutation that changes copy number control is about 300 bp farther downstream from the mRNA message, it is unlikely to have any direct effect on translation. To test if there exists a copy number effect, pCFM1656(536) which has no point mutation in copy number control but retains 22 nt deletion was constructed. Experimental data indicated that it was the 22 nt deletion that affected NT-3 synthesis, as shown by comparing the expression patterns of pCFM1656-NT-3UGA and pCFM1656(536)-NT-3UGA (Figure 2). The exactly same expression patterns from pCFM1656.1-NT-3UGA and pCFM1656(536)-NT-3UGA (data not shown) also showed no influence of plasmid copy number in this event.

Effect of mRNA sequence and structure on NT-3 translation fidelity. The 22 nt sequence difference between pCFM1656-NT-3UGA and pCFM1656(536)-NT-3UGA could result in different local or general mRNA secondary structures. Although the 22 nt deletion in pCFM1656(536) had been shown to lead to the synthesis of a much longer polypeptide due to the lack of a second in-frame stop codon, the potential mRNA secondary structure change could lead to a different readthrough event other than Trp incorporation such as selenocysteine incorporation (3, 15, 17) . Detailed analysis of the

mRNA secondary structure of the message transcribed from pCFM1656(536)-NT-3UGA using GCG sequence analysis software package (Genetics Computer Group, Inc., Madison, Wisconsin) revealed no conspicuous stem-loop or pseudoknot structure that would signal selenocysteine incorporation at the UGA codon. Chemical analysis on the readthrough polypeptide also failed to detect the presence of any selenocysteine

ACKNOWLEDGMENTS: We are grateful to the Department of Chemistry, Amgen, Boulder, for the synthesis of oligonucleotides, and the DNA Sequencing group, Department of Developmental Biology, Amgen, for sequencing.

REFERENCES

1. Atkins, J.F., Weiss R.B. and Gestelan, R.F. (1990) Cell 62, 413-423.
2. Björnsson, A. and Isaksson, L.A. (1993) J. Mol. Biol. 232, 1017-1029.
3. Böck, A., Forchhammer, K.J., Heider Leinfelder, W., Sawers, G.,Veprek, B. and Zinoni, F. (1991) Mol-Microbiol. 5, 515-520.
4. Buckingham, R.H., Sorensen, P., Pagel, F.T., Hijazi, K.A., Mims, B.H., Brechemeir-Baey, D. and Murgola, E.J. (1990) Biochim. Biophy. Acta. 1050, 259-262.
5. Eggertsson, G. and Söll, D. (1988) Microbiol. Rev. 52, 354-374.
6. Engelberg-Kulka, H. and Schoulaker-Schwarz, R. (1988) J. Theor. Biol. 131,477-485.
7. Hui, J.O., Meng, S.-Y., Katta,V., Tsai, L., Rohde, M.F. and Haniu, M. (1995) Techniques in Protein Chemistry VI, 341-348.
8. Kopelowitz, J., Hampe, C., Goldman, R., Reches, M. and Engelberg-Kulka, H. (1992) J. Mol. Biol. 225, 261-269.
9. Maisonpierre, P.C., Le Beau, M.M., Espinosa III, R., Ip, N. Y., Belluscio, L., de la Monte, S.M., Squinto, S., Furth, M.E. and Yancopoulos, G.D. (1991) Genomics 10, 558-568.
10. Mottagui-Tabar, S., Biornsson, A. and Isaksson, L.A. (1994) EMBO J. 13, 249-257.
11. Nordström, K., Molin, S. and Light, J. (1984) Plasmid 12, 71-90.
12. Parker, J. (1989) Microbiol. Review 53, 273-298.
13. Pedersen, W.T. and Curran, J.F. (1991) J. Mol. Biol. 219, 231-241.
14. Rosenthal, A., Goeddel, D.V., Nguyen, T., Lewis, M., Shih, A., Laramee, G.R., Nikolics, K. and Winslow, J.W. (1990) Neuron 4, 767-773.
15. Stadtman, T.C., Davis, J.N., Ching, W.M., Zinoni, F. and Bock, A. (1991) Biofactors 3, 21-7.
16. Stormo, G.D., Schneider, T.D. and Gold, L. (1986) Nucl. Acids Res. 14, 6661-6679.
17. Zinoni, F., Heider, J. and Böck, A. (1990) Proc. Natl. Acad. Sci. 87, 4660-4664.