Only the last amino acids in the nascent peptide influence translation termination in *Escherichia coli* genes

Salim Mottagui-Tabar, Leif A. Isaksson*

Department of Microbiology, Stockholm University, 106 91 Stockholm, Sweden

Received 4 July 1997

Abstract Efficiency of translation termination is affected if the last two amino acids in the nascent peptide are changed [1,2]. By changing the corresponding codons upstream of the stop signal UGAA, we have analyzed if the −3 to −6 amino acids at the C-terminal region of the nascent peptide also affect termination. Lysine at position −3 gave increased readthrough, whereas a total of 28 variations at positions −4, −5, and −6 showed no significant effect on readthrough. The 3′-ends corresponding to the last six codons in 27 *Escherichia coli* genes were inserted upstream of a stop codon in the 3A′ translation assay gene [1]. Readthrough of the stop codon was measured and a possible correlation with the Codon Adaptation Index (*CAI*) [3] of the respective genes was investigated. Sequences from genes with low *CAI* do not give any such correlation, whereas sequences from genes with high *CAI* values are correlated with high termination efficiency. This correlation disappears if the −1 and −2 codons/amino acids are changed. The results suggest that mainly the terminal dipeptide of the terminal hexapeptide sequence has an influence on termination in the tested *E. coli* genes. This influence is dependent on the charge of the −2 amino acid and is correlated with the α-helix propensity of the −1 amino acid, in accordance with results obtained from synthetic gene constructs [1,2].

Key words: C-terminus; Hexapeptide; Dipeptide; Stop codon; Codon Adaptation Index; GHN bias; Translation termination

1. Introduction

Translation termination is influenced by the last two amino acids in the nascent peptide and the P-site tRNA [1,2,4]. Termination efficiency increases if the −2 amino acid is basic and/or if the −1 amino acid has a high propensity to participate in natural α-helix or β-strand structures. The C-terminal amino acid in the nascent peptide interacts with the elongation factor Tu during decoding of a non-sense codon by a near-cognate or suppressor tRNA [5]. In this case, the van der Waals volume of the −1 amino acid is the determinant. Thus, the C-terminal end of the nascent peptide is important for decoding stop codons.

Highly expressed genes avoid the use of rare codons, resulting in a bias in codon usage, and they terminate with an efficient stop codon [6]. Also, codon context around stop codons is not random [7]. Thus, both translation efficiency [8] and termination efficiency can be correlated to the codon usage (*CAI*; Codon Adaptation Index) [6]. Of the three stop codons, UAA gives the most efficient and UGA the least efficient translation termination. Nucleotides at the 3′ side of the stop codon influence the efficiency of decoding [9–11]. On the 5′ side, the nascent peptide and the P-site tRNA are the major influential factors [1,2,4]. Highly expressed genes would be expected to have been selected for codon/amino acid sequences that allow for efficient termination. Therefore, flanking sequences at the 3′ or 5′ sides of the stop codon could be biased if highly and lowly expressed genes are compared. In fact, statistical analysis of the nucleotide distribution in coding sequences in *E. coli* shows a bias (non-randomness) in the region immediately prior to the stop codon [12].

Almost 40% of the *E. coli* coding genes terminate with an opal (UGA) stop codon [7,13]. These genes have a *CAI* value ranging from 0.12 to 0.69. In order to analyze any correlation between the *CAI* value of the gene, the codon sequence upstream of its stop codon, and its termination efficiency, 38 sequences were selected from the TransTerm database [14]. The sequences corresponding to the terminal hexapeptides of genes representative of *CAI* values from 0.11 to 0.98 (for all three stop codons) were tested for their effects on termination efficiency. We also tested systematically any influence from upstream amino acids in the peptide. This was done by changing one codon at a time at positions −3, −4, −5, and −6 upstream of the stop codon and measuring the effect on termination/readthrough using the 3A′ translation assay system. The results suggest that the last two amino acids in the hexapeptide sequence of the C-terminal region of *E. coli* proteins, together influence termination efficiency.

2. Materials and methods

Protein A′ assays and procedures for cloning and sequencing were performed as described earlier [2]. The oligonucleotides used for cloning were obtained from SDS Diagnostics. The *E. coli* strain XAc, the *trpT*(Su9) opal suppressor and *tyrT*(Su3) amber suppressor strains CDJ64 and UB585, have been described elsewhere [15].

3. Results

The translation assay gene 3A′ codes for a protein of three identical domains derived from the antibody-binding B-domain of protein A from *Staphylococcus aureus*. The semisynthetic gene carries a linker region between the second and third coding domains. This linker carries the stop codon, giving a two-domain protein as the result of termination or a three-domain protein as the result of translation readthrough. The ratio between the three- and two-domain protein products (transmission) on a gel thus reflects stop-codon readthrough by some near-cognate or suppressor tRNA. Termination efficiency is inversely related to this transmission (*T*) value.

*Corresponding author. Fax: (46) 8-612-9552.
E-mail: leif.isaksson@mibi.su.se

PII S0014-5793(97)00978-2

Table 1
Peptide residues upstream of the stop signal and readthrough

Plasmid	Codon context and altered amino acid residues							Transmission	
	−6	−5	−4	−3	−2	−1	stop	Su⁻	Su9
pSM11	TTC (F)	TTC (F)	GAA (E)	GAC (D)	GAC (D)	CCA (P)	TGAA	0.91	4.60
pSMT120				**TGG** (W)				0.72	3.60
pSMT122				**GGC** (G)				0.82	4.55
pSMT121				**AAG** (K)				1.33	5.50
pAB118				**GAC** (D)	**CGC** (R)			0.05	0.24
pSMT123				**AAG** (K)	**CGC** (R)			0.11	0.65
pSMT164			**TGG** (W)					0.62	3.80
pSMT165			**GGA** (G)					0.70	4.20
pSMT163			**CCA** (P)					0.58	3.50
pSMT166			**CGC** (R)					0.72	4.40
pSMT125		**TGG** (W)						0.66	3.90
pSMT127		**GGC** (G)						0.62	3.30
pSMT153		**CCA** (P)						0.67	3.20
pSMT129		**CGC** (R)						0.56	3.55
pSMT159		**TGC** (C)						0.68	4.00
pSMT128		**GAC** (D)						0.72	4.20
pSMT161		**GTC** (V)						0.77	3.20
pSMT162		**CTC** (L)						0.80	4.40
pSMT126		**GCC** (A)						0.64	3.00
pSMT154		**AAA** (K)						0.77	4.00
pSMT155		**ATG** (M)						0.70	3.60
pSMT162		**GAG** (E)						0.68	3.30
pSMT161		**ATA** (I)						0.55	3.80
pSMT160		**CAT** (H)						0.58	4.10
pSMT137		**TGG** (W)			**AAA** (K)			0.06	0.28
pSMT131		**GGC** (G)			**AAA**			0.04	0.22
pSMT132		**CGC** (R)			**AAA**			0.07	0.25
pSMT135		**GAC** (D)			**AAA**			0.05	0.30
pSMT138		**GCC** (A)			**AAA**			0.07	0.30
pSMT136		**GAA** (E)			**AAA**			0.05	0.32
pSMT133		**AAG** (K)			**AAA**			0.05	0.25
pSMT134		**CAC** (H)			**AAA**			0.06	0.28
pSMT187	**CCA** (P)							0.65	3.20
pSMT167	**GAC** (D)							0.75	3.70
pSMT147	**CCA**	**CCT**	**CCC**	**CCA**	**CCT**	**CCC**	**TGAA**	0.50	4.30
pSMT148	**GGA**	**GGT**	**GGC**	**GGA**	**GGT**	**GGC**	**TGAA**	0.14	1.00
pSM08	TTC (F)	TTC (F)	GAA (E)	GAC (D)	**CCC** (P)	CCA (P)	TGAA	0.52	1.80
pMB50	TTC (F)	TTC (F)	GAA (E)	GAC (D)	GAC (D)	**GGC** (G)	TGAA	0.09	0.65
								Su⁻	Su9
pAG53	TTC (F)	TTC (F)	GAA (E)	GAC (D)	GAC (D)	CCA (P)	**TAGA**	<0.01	0.08
pSMT174		**GAC** (D)						<0.01	0.05
pSMT175		**CGC** (R)						<0.01	0.06
pSMT176		**GGC** (G)						<0.01	0.06
pSMT177		**GCC** (A)						<0.01	0.07

Readthrough at termination codons. Indicated codon sequences upstream of the UGA or UAG stop codons are inserted into the plasmid pSM11 [2]. Codons that are altered relative to the sequence in pSM11 are presented in bold letters. The corresponding amino acids are indicated by a one letter symbol in brackets after the codon. Readthrough by a near-cognate tRNA was measured in strain XAc and by *trpT*(Su9) suppressor tRNA in the mutant derivative CDJ64. Readthrough of UAG by *tyrT*(Su3) suppressor tRNA was measured using strain UB585 [15]. Readthrough is presented as a *T* value that is the ratio between the molar amounts of 3A′ (resulting from readthrough) and 2A′ proteins (resulting from termination).

Table 1 shows the changes made in the sequence upstream of UGA and UAG stop codons. If codons at position −3 are changed, the *T* values increase if AAG is at position −3, in both the wild-type *trpT*(Su⁻) strain and the suppressor *trpT*(Su9) strains, whereas the other codon alterations have no effect. If the residue at the −2 position also is changed (GAC to CGC), readthrough is decreased; but AAG still gives higher readthrough than GAC in −2 position. Thus, the codon at position −2 reduces the effect of the −3 codon but the relative effect of the −3 codon on readthrough is maintained. Codons representing four different amino acids which are know to give strong effects if located at positions −1 or −2 were analyzed at position −4 (proline, tryptophan, glycine and arginine) for their effect on readthrough. No significant change in the corresponding *T* values was observed.

A combination of 14 different codons at position −5 with CCA (proline) at −1, and 8 codons together with AAA (lysine) at position −1 were tested. All these different −5 codons had a similar effect on the readthrough values, in both the *trpT*(Su⁻) and *trpT*(Su9) strains. The data suggest that readthrough was associated with the nature of the −1 codon/amino acid but not with that of the −5 amino acid. A *prfB2* mutant (L328F) strain that was also analyzed showed similar results (data not shown).

Two constructs having different amino acids at position −6 (proline and aspartic acid; pSMT187 and pSMT167, respec-

Table 2
Influence of the terminal hexapeptide context on readthrough in *E. coli* genes

Plasmid	Gene AC no.	Codon positions							Transmission		*CAI*
		−6	−5	−4	−3	−2	−1	stop	Su^-	Su9	
pSMT140	X02613	GTT	GAG	GAG	ATG	CTG	GCA	TGAT	0.03	0.15	0.690
pSMT141	X70111	GAC	AGG	GTA	GTA	ATG	GCC	TGAT	0.10	1.32	0.163
pSMT142	L11241	AAG	CCG	TCA	CTC	GCG	ACC	TGAA	0.04	0.26	0.124
pSMT143	Y00720	AAA	TGC	CGA	TGT	GAA	GAC	TGAT	0.03	0.07	0.401
pSMT144	X68301	CGA	GAA	TCG	CGC	TAT	AAA	TGAA	0.04	0.14	0.462
pSMT145	U00008	TCG	TTC	GCG	CTT	AAT	TGC	TGAA	0.06	0.14	0.630
pSMT146	M63288	GCA	TTC	TCA	ACA	AGC	CGA	TGAA	0.03	0.06	0.573
pSMT220	M77129	CAT	GAG	CGT	AAA	GCC	AGC	TGAA	0.09	1.23	0.467
pSMT221	M29823	TTG	ATG	CTG	GCA	AGA	GCG	TGAA	0.10	1.31	0.405
pSMT222	D90227	GAA	TTC	ACT	ATG	GAG	CAC	TGAA	0.04	0.24	0.356
pSMT223	M11294	CGC	TGC	GTC	AGA	TCG	TAT	TGAA	0.05	0.19	0.528
pSMT224	X64395	TTG	AAA	GCC	ATC	AAG	GCC	TGAA	0.05	0.22	0.558
pSMT225	M64516	AAC	ACA	GGG	GAC	GTT	AAA	TGAA	0.13	0.17	0.329
pSMT130	Cat86	GGT	ATG	GTG	AAA	ACA	GAT	TGAA	0.10	1.66	n.a.
									Su^-	Su3	
pSMT214	U01159	GAT	TCA	ATT	AGA	TAT	CGG	TAGA	<0.01	0.33	0.118
pSMT215	M55661	TAT	TTG	ATT	TCA	AAA	ATT	TAGA	<0.01	0.41	0.132
pSMT216	L12007	AGA	AAG	CCA	AAT	CTA	TTT	TAGA	<0.01	0.27	0.267
pSMT217	M87049	CCT	TTT	ACC	TCT	TTC	TCA	TAGA	<0.01	0.83	0.393
pSMT218	X15371	CCT	TCA	TCG	CAG	AAC	GCC	TAGA	<0.01	0.69	0.201
pSMT219	M63145	GAT	GAA	CCT	ACC	GAA	GCT	TAGA	<0.01	0.43	0.376
pSMT197	M22044	AAA	CCG	GCT	AAA	TCT	GCT	TAAA	<0.01	0.01	0.988
pSMT198	X00415	ATT	GAA	GCC	CGT	GGT	AAA	TAAA	<0.01	0.01	0.743
pSMT199	X14436	GAA	GAG	CGC	GCT	AAG	AAG	TAAA	<0.01	0.01	0.726
pSMT200	D13187	GAC	CTG	GTT	GGT	AAA	ATC	TAAA	<0.01	0.01	0.797
pSMT210	M55609	ACT	TAT	TTT	AAA	GGT	GGA	TAAA	<0.01	<0.01	0.145
pSMT211	U01159	ACT	TAT	AAT	GAG	AGA	AAA	TAAA	<0.01	<0.01	0.157
pSMT212	L02373	ATT	CGT	AGT	GAA	GAG	AAA	TAAA	<0.01	<0.01	0.110
pSMT213	S56312	ATA	TCA	ATA	AAG	AAT	AAT	TAAA	<0.01	<0.01	0.112

Sequences of mRNA for 27 *E.coli* genes up to 18 nucleotides upstream relative to the stop codon. Their AC numbers and *CAI* values [15] are indicated. *T* values (see Table 1) for wild-type *trpT*(Su^-) and suppressor *trpT*(Su9) or *tyrT*(Su3) strains are shown. *CAI* values for the *cat*86 gene (pSMT130) is not available (n.a.).

tively) do not affect the readthrough of the stop codon differently.

A repetitive sequence coding for six proline (pSMT147) or glycine (pSMT148) residues were tested for their effect on readthrough. In a wild-type *trpT*(Su^-) strain, readthrough from the poly-proline (pSM147) context was similar in magnitude as that found earlier for Pro–Pro–UGA (pSM08) [1]. In the suppressor *trpT*(Su9) strain, however, the readthrough is almost 2.5-fold higher than the *T* value from pSM08. It has been shown earlier that the wild-type and suppressor forms of the tRNA interact differently with the nascent peptide [1]. It has also been shown that the elongating ternary complex (EFTu–aminoacyl-tRNA–GTP) is sensitive to the nascent peptide [5]. Thus, differently structured nascent peptides (poly-proline as compared to di-proline) probably interact differently with different forms of the same tRNA, resulting in varying *T* values. However, this does not apply to the difference in *T* values from the poly-glycine construct (pSMT148) compared to that found for Asp–Gly as the last two amino acids (pMB50) [2]. In both the *trpT*(Su^-) strain and the suppressor *trpT*(Su9) strain the difference between the two constructs is 1.5-fold.

Different codons at the −5 position were also tested for possible influences on termination by RF1 by changing the

Table 3
Effect of the last two amino acids in the nascent peptide on translation termination in natural contexts

Plasmid	Gene AC nr.	Peptide context	Transmission		*CAI*
			Su^-	Su9	
pSMT193	X07465	I-Q-A-E-W-P-TGA AGT	0.10	0.50	0.327
pSMT179	D10483	K-S-L-I-W-P-	0.13	0.52	0.247
pSMT192	M12868	Q-P-A-S-T-P-	0.10	0.42	0.248
pSMT178	U00008	Q-G-E-E-T-P-	0.10	0.52	0.382
pSMT194	L29404	H-E-E-A-L-P-	0.10	0.77	0.258
pSMT195	M32488	T-W-R-V-L-P-	0.08	0.58	0.304
pSMT188	M17317	M-V-L-S-F-P-	0.42	1.50	0.174
pSMT190	X54198	D-M-D-Y-F-E-	0.10	0.52	0.299
pSMT189	L08626	T-H-K-W-T-L-	0.04	0.14	0.336
pSMT191	U00007	T-A-K-S-T-R-	0.01	0.05	0.269

The last six amino acid residues of the protein products coded by 10 *E. coli* genes and derivatives. All genes terminate with the UGAA stop signal.

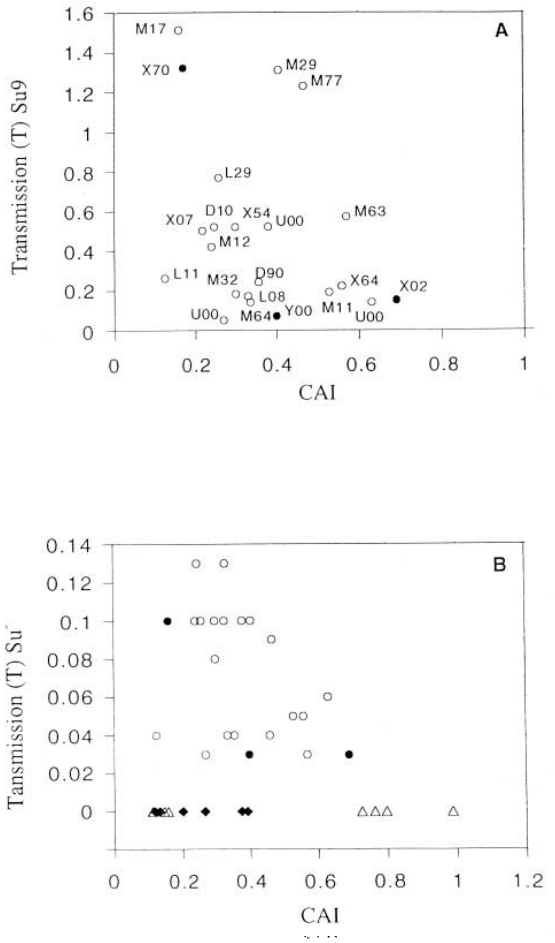

Fig. 1. *T* values plotted against the CAI of some high and low expressed genes in *E. coli*. The figure includes data from Table 1 and 2. A: *T* values from the suppressor (*trpT*(Su9)) strain CDJ64, used for measuring readthrough at the total of 22 UGAA (○) and UGAU (●) contexts. The gene AC numbers are abbreviated to only the first letter followed by two digits. The three genes with AC numbers U00008 (pSMT145), U00008 (pSMT178) and U00007 (pSMT191) are all abbreviated to U00. These three can be identified by their respective *CAI* values. B: Data for the suppressor-free strain XAc where readthrough is plotted against *CAI* values of genes terminating with TGAA (○), TGAT (●), TAGA (♦) and TAAA (△). The highest *T* value of 0.42 for M17317 is not included in the plot.

stop signal to UAGA. The last five constructs in Table 1 (bottom) show the codons analyzed at the −5 position together with this stop signal. We observed no significant difference in the effect on *T* values by these codon alterations. A similar lack of sensitivity to the −5 amino acid was found for a *prf*A1 mutant strain with altered RF1 (data not shown).

Table 2 summarizes the 27 termination sequences from *E. coli* coding regions that have been analyzed in our test system in an attempt to correlate their *T* values with their respective *CAI* values. The table shows that many highly expressed genes (*CAI* values >0.5) use UAA as a stop codon and therefore have very efficient termination, reflected in low *T* values. Genes with lower *CAI* values use all three stop codons, resulting in a greater variation in their *T* values.

Table 2 shows a construct (pSMT130) with the *cat*86 N-terminal hexapeptide (MVKTD) coding sequence cloned upstream of a UGAA stop signal. This sequence in its natural location in the *cat*-leader blocks translation. However, placed in our construct here, the readthrough is not dramatically different, in the suppressor-free strain, from other constructs shown in Tables 3 and 4 (pSMT192, pSMT178), even though there is only one residue (threonine at position −2) common to all three hexapeptides. Thus, the hexapeptide MVKTD does not have any extraordinary effect on translation termination if placed in the context shown here.

Table 3 shows the influence of the terminal dipeptide on termination efficiency for natural *E. coli* sequences. The 10 constructs are the terminal hexapeptides from *E. coli* genes terminating with UGAA, with proline as the last amino acid in 7 of them. These genes have terminal dipeptide combinations that have been tested earlier for their influence on efficiency of termination [1,2]. These earlier constructs had the same common four residues (FFED) at positions −3 to −6 but different −1 and −2 amino acids. Thus, the influence measured earlier was mainly from the last and the penultimate amino acids in the nascent peptide. In Table 3 the first two constructs (pSMT193 and pSMT179) have the same −1 and −2 amino acids but different residues in the preceding 4 positions. We observe that the *T* values for these constructs are similar to each other and to those measured earlier [2] for the same terminal dipeptide. The next two constructs, pSMT192 and pSMT178, coding for identical terminal dipeptides (threonine and proline), also give similar readthrough values. Also pSMT194 and pSMT195, coding for identical dipeptides (leucine–proline) with varied upstream residues show similar readthrough values. The construct pSMT188 codes for a terminal dipeptide Phe–Pro giving a *T* value of 0.42 that is comparable with earlier published results for the same terminal dipeptide (0.56) [2]. The transmission decreases when glutamic acid is at the −1 position (pSMT190). The terminal dipeptide combination threonine–leucine (pSMT189) gives lower readthrough as

Table 4
Effect of the last two amino acids in the nascent peptide on termination at TGAA in derivatives of natural contexts

Plasmid	Gene Ac no.	Context upstream of the stop signal TGAA							Transmission		*CAI*
		−6	−5	−4	−3	−2	−1	stop	Su^-	Su9	
pSMT188	M17317	ATG	GTT	CTG	AGT	TTC	CCA	TGAA	0.42	1.32	0.174
pSMT266	M17-low					**CGC**(R)			0.03	0.28	
pSMT178	U00008	CAA	GGT	GAG	GAA	ACA	CCA	TGAA	0.10	0.52	0.382
pSMT265	U00-low					**CGC**(R)			0.03	0.26	
pSMT264	U00-high					**GAC**(D)			0.82	5.33	
pSMT146	M63288	GCA	TTC	TCA	ACA	AGC	CGA	TGAA	0.03	0.57	0.571
pSMT263	M63-high					**GAC**(D)	**CCA**(P)		0.67	4.10	

Amino acid and codon sequences for three genes and derivatives. Alterations in the −1 and −2 positions in the derivatives are indicated with amino acid substitutions in parentheses. *T* values are shown.

compared to threonine–proline (pSMT178), as expected from earlier studies. Transmission is lowest when the −1 amino acid is arginine (pSMT191). Readthrough with leucine (pSMT189) or arginine (pSMT191) as the last amino acid in the nascent peptide has been determined earlier as being 0.05 (pMB88) [2] and <0.01 (pMB56) [2]. Thus, readthrough measured here (0.03 for the dipeptide tyr–leu and 0.01 for the dipeptide tyr–arg), is not significantly different from earlier measurements even though the other amino acids in the terminal hexapeptide are different [1,2]. Earlier values for constructs coding for different −3 to −6 amino acids are thus in agreement with the results found here and point to the importance of the last amino acid residue in translation termination.

Table 4 shows that three natural sequences with high (M17317), average (U00008), or low (M63288) readthrough values, respectively, are inversely proportional to their *CAI* values. In order to analyze these contexts in more detail, the last and/or last two amino acids were changed so as to raise or lower readthrough. The M17-low construct has a phenylalanine (TTC) to arginine (CGC) substitution at position −2. As a result of this substitution, transmission decreases from 0.42 to 0.03. The *T* value for the native construct U00008 is 0.1. A codon change at the −2 position (arginine in U00-low or aspartic acid in U00-high), changes the transmission accordingly; for U00-low it decreases to 0.03 and for U00-high it increases to 0.8, in agreement with our predictions. Finally, the low readthrough context for the native construct M63288 can be raised from 0.03 to 0.67 by substituting the last amino acids with the dipeptide Asp–Pro. Thus, the influence of the nascent peptide on efficiency of termination at natural stop codons is primarily connected with the final two amino acids.

Fig. 1 shows the correlations between readthrough values derived from data presented in Tables 2–4 and the Codon Adaptation Index. For genes terminating with UGAA/U, *T* values using a *trpT*(Su9) suppressor strain were plotted against *CAI* values of the corresponding genes (Fig. 1A). The 23 sequences that terminate with UGAA/U have *CAI* values ranging between 0.12 and 0.69. There is a wide range of *T* values for genes with low CAI values. However, for genes with *CAI* values greater than 0.45, termination is efficient as judged by the low *T* values. One gene with high *CAI* value was found (X02613) that uses the relatively strong terminator signal UGAU. For the other genes with *CAI* values higher than 0.45 that terminate with UGAA it appears that combinations of the last two amino acid residues in the nascent peptide that favour efficient termination at UGAA are selected.

Using the wild-type *trpT*(Su^-) strain we measured transmission for genes termination with any one of the three stop codons UGA, UAG, or UAA. In Fig. 1B the *T* values are plotted for 37 constructs having the C-terminal hexapeptide context of 37 *E. coli* genes, with a *CAI* range of 0.110–0.988. At higher *CAI* values, *E. coli* genes prefer the use of the strong stop codon UAA for efficient termination. For genes with low *CAI* values, any one of the three stop codons are used. However, if UGA is selected, efficient termination at *CAI* values above 0.45 can be achieved by using either UGAU or a peptide context (Fig. 1A) that favours better termination. Thus, *E. coli* genes that are highly to moderately expressed and terminate in UGAA tend to select a C-terminus that favours effective termination.

4. Discussion

The terminal dipeptide in the nascent polypeptide influences the efficiency of translation termination in *E. coli*, *S. typhimurium*, *B. subtilis* and *S. cerevisiae* (manuscript in preparation). The nascent peptide has been found to be involved in translation regulation in several procaryotes and also in lower eucaryotes [16]. We have altered the last amino acids in the nascent peptide, up to position −6 (counting from the C-terminal end), to systematically investigate the influence from the nascent peptide on translation termination. Except for a 1.5-fold increase associated with lysine at the −3 position, we see no effect from changes made at any of the other positions nor from any of the other tested amino acids. Thus, the nascent peptide exerts its main effect on termination efficiency through the terminal dipeptide. We further substantiate this observation by showing that the termination efficiency of natural *E. coli* genes can be altered by changing the terminal dipeptide sequence.

The coding regions of *E. coli* have a G–non-G-N sequence pattern [17,18] that has been proposed to be important for maintainance of the frame during triplet decoding. However, an in vivo experiment has failed to support that model [19]. The observation that the G–non-G-N sequence pattern in the coding regions is absent five codons prior to the stop codon [12] does not have any satisfactory explanation. The amino acid at position −5 corresponds to the position of the codon where the G–non-G-N sequence pattern breaks down. An α-helix requires the participation of a minimum of 4 amino acids; thus, a complete α-helix could be formed without the participation of the amino acid at the −5 position. Moreover, the −5 amino acid in an α-helix completes a turn and faces the same plane as the first one. The presence of proline at the −1 position would distort an α-helix structure at the C-terminal end of the nascent peptide, thus altering the structure and possibly also the interactions with RFs. Proline at the −1 position combined with 14 different amino acids at −5 were tested for their effect on termination. Eight of those were also tested together with lysine at the −1 position (lysine does not prevent the terminal pentapeptide from assuming an α-helix structure). A comparision of the two series of combinations reveals no significant alteration in readthrough. Thus, the −5 amino acid in the nascent peptide, whether a part of an α-helix structure or not, does not seem to influence termination efficiency.

We observe that a polyproline peptide gives readthrough values that agree with our earlier measurements for readthrough with a single (P-site) proline (pSMT147, compared to pSM08, in Table 1), only in a wild-type strain and not in a suppressor strain. The structure of the tRNA in the ternary complex influences the outcome of the competition between release factor activity (termination) and ternary complex (suppression) [2], a competition that is also influenced by the functional interaction between the nascent peptide and the elongation factor Tu [5]. Thus, a polyproline hexapeptide might interact differently with the ternary complex with wild-type *trpT*(Su^-) and suppressor *trpT*(Su9) forms of tRNA. There is 41% sequence homology and 31% amino acid identity between the two bacterial release factors RF1 and RF2 [20]. We found the −5 amino acid in the nascent peptide to have no influence on termination at UGA or UAG. Glycine forms a flexible peptide bond and allows the polypeptide to be less

constrained. For this reason, two peptides with different amino acids at position −5 were tested with glycine at −1. None of these alterations affected significantly the decoding efficiency of the nonsense codons. Therefore it appears that the deviations from the G–non-G-N sequence pattern immediately upstream of a stop codon cannot be correlated to any property of the nascent peptide coded by those codons. However, since many polycistronic messages require a Shine-Dalgarno-like sequence between the −2 and −4 codons upstream of the termination signal, some deviation might be attributed to this selection in polycistronic mRNAs.

Highly expressed genes have efficient initiation and termination signals [6,8]. As seen in Fig. 1A and B, genes with high *CAI* values code for terminal dipeptides or a stop codon (UAA) with efficient termination. There is a significant variation in termination efficiency for genes with lower *CAI* values (0.11–0.45). Such genes also show an increased usage of the relatively weak UGAA/U stop signals. It has been found that the UAA stop codon is preferred compared to UGA and UAG in prokaryotes [12] and that UAAU is the most efficient stop signal. Inefficient termination at the stop codon UGAA is improved by the nascent peptide in order to increase the efficiency of termination for genes with high *CAI* values. Besides the interactions with the release factors, the nascent peptide interacts with other components of the ribosome as well. There is a functional interaction between the nascent peptide and the elongation factor Tu [5]. As the nascent peptide is synthesized, it passes through the 50S subunit and can be chemically crosslinked to several nucleotides in the 23S rRNA [21]. The 24 residue product of the tnaC gene in the tryptophanase (tna) operon, in the presence of tryptophan, interacts with the ribosome and perhaps other factor(s) and prevents the release of the peptide at the stop codon [22]. The chloramphenicol acetyltransferase gene (*cat*86) has a pentapeptide (MVKTD) at the N-terminus that interacts with the ribosomal peptidyltransferase center in the presence of chloroamphenicol, thereby arresting translation [16]. This pentapeptide (Table 2; pSMT130), when cloned into our test system, in a suppressor-free strain, gave a readthrough value similar to other sequences (Table 3; M12868, U00008), even though the hexapeptide has only threonine in common at position −2. Thus, the N-terminal pentapeptide from *cat*86, when placed at the C-terminus of a long nascent peptide, does not interfere with the termination process.

In summary, our data suggest that the nascent peptide has an influence on termination efficiency of natural *E. coli* genes and that the major effect comes from the last two amino acids in the nascent peptide. Highly expressed genes avoid inefficient stop signals and code for terminal dipeptides that increase the efficiency of translation termination. Low expressed genes are less selective with respect to the terminal dipeptide sequence and its effect on translation termination.

Acknowledgements: This work has been supported by grants from the Swedish Natural Science Foundation (NFR) to Dr. L.A. Isaksson. We thank Dr. A. Björnsson for discussions in the early part of this work. Thanks are due to Maria Ohlsson and Patricia Raichura for assistance with running protein A′ assays and to Steven Muir for critical reading of the manuscript.

References

[1] Björnsson, A., Mottagui-Tabar, S. and Isaksson, L.A. (1996) EMBO J. 15, 1696–1704.
[2] Mottagui-Tabar, S., Björnsson, A. and Isaksson, L.A. (1994) EMBO J. 13, 249–257.
[3] Sharp, P.M. and Li, W.H. (1987) Nucl. Acids Res. 15, 1281–1295.
[4] Zhang, S.P., Rydén-Aulin, M. and Isaksson, L.A. (1996) J. Mol. Biol. 261, 98–107.
[5] Mottagui-Tabar, S. and Isaksson, L.A. (1996) Biochimie 78, 953–958.
[6] Sharp, P.M. and Li, W.H. (1986) Nucl. Acids Res. 1986, 7737–7749.
[7] Sharp, P.M. and Bulmer, M. (1988) Gene 63, 141–145.
[8] Robinson, M., Lilley, R., Little, S., Emtage, J.S., Yarranton, G., Stephens, P., Millican, A., Eaton, M. and Humphreys, G. (1984) Nucl. Acids Res. 12, 6663–6671.
[9] Tate, W.P., Poole, E.S., Dalphin, M.E., Major, L.L., Crawford, D.J.G. and Mannering, S.A. (1996) Biochimie 78, 945–952.
[10] Buckingham, R.H. (1990) Experientia 46, 1126–1133.
[11] Pedersen, W.T. and Curran, J.F. (1991) J. Mol. Biol. 219, 231–241.
[12] Brown, C.M., Stockwell, P.A., Trotman, C.N.A. and Tate, W.P. (1990) Nucl. Acids Res. 18, 2079–2086.
[13] P.M. Sharp, and C.J. Burgess, in: D.L. Hatfield, B.J. Lee, and R.M. Pirtle (Eds.), Transfer RNA in Protein Synthesis, CRC Press, Boco Raton, FL, 1992, pp. 397–425.
[14] Brown, C.M., Dalphin, M.E., Stockwell, P.A. and Tate, W.P. (1993) Nucl. Acids Res. 21, 3119–3123.
[15] Faxén, M., Kirsebom, A.L. and Isaksson, A.L. (1988) J. Bacteriol. 170, 3756–3760.
[16] Lovett, P.S. and Rogers, E.J. (1996) Microbiol. Rev. 60, 366.
[17] Trifonov, E.N. (1987) J. Mol. Biol. 194, 643–652.
[18] Trifonov, E.N. (1992) Biochimie 74, 357–362.
[19] Curran, J.F. and Gross, B.L. (1994) J. Mol. Biol. 235, 389–395.
[20] Craigen, J.W., Cook, R.G., Tate, W.P. and Caskey, C.T. (1985) Proc. Natl. Acad. Sci. USA 82, 3616–3620.
[21] Stade, K., Junke, N. and Brimacombe, R. (1995) Nucl. Acids Res. 23, 2371–2380.
[22] Konan, K.V. and Yanofsky, C. (1997) J. Bacteriol. 179, 1774–1779.