The optimal gene sequence for optimal protein expression in *Escherichia coli*: principle requirements

(A review article)

(Received: 14 .04 .2006; Accepted: 16 .05 .2006)

EL-Rashdy M. Redwan

Protein Research Department, Genetic Engineering and Biotechnology Research Institute, Mubarak City for Scientific Research and Technology Applications, New Borg EL Arab 21934, Alexandria, Egypt.

e-mail: redwan1961@yahoo.com

ABSTRACT

As an essential element in the drug discovery, development process and post-genomic structural biology, there remains a need for expression of large numbers of recombinant protein agents and targets from diverse genomes. Escherichia coli remain the host of choice for many of these applications due to its speed, ease of genetic manipulation, and high level of recombinant synthesis. Expression can often be achieved at levels higher than 30% of total cellular protein and following induction greater than 65% of the protein synthetic machinery of the cell may be dedicated to synthesis of the target protein. However, there is the risk of translational errors related to the presence of high frequencies or clusters of rare codons in the foreign recombinant mRNA. In the case of therapeutic protein biopharmaceuticals, even at low levels these microheterogeneities have the potential to manifest themselves in adverse immunogenic responses. At higher levels, translational errors may impact overall tertiary structure and functional activity. During gene design for optimal expression, the codon usage into the codons bias of E. coli genes should be adapted. Further, regions of very high (>80%) or very low (<30%) GC content should be avoided where possible. Moreover, the following cis-acting sequence motifs are recommended to be avoided; internal TATA boxes, chi-sites stretches, and internal ribosomal entry sites, AT-rich or GC-rich sequence stretches, repeat sequences and RNA secondary structures, arg/UAG or arg/UGA, and urgently recommend using prolonged UAAU or connective UAAUAA stop codon.

Keywords: *Escherichia coli, recombinant protein expression, rare codons, translation initiation, frameshift, stop codons.*

INTRODUCTION

Escherichia coli was the first host used to produce a recombinant DNA (rDNA) biopharmaceutical, enabling the approval of Eli Lilly's rDNA human insulin in 1982. This is especially noteworthy, because insulin was already a 'mature' biopharmaceutical. The marketing of Monsanto's bovine growth hormone (bGH) product in 1994 set a new standard for pharmaceutical protein production from *E. coli*. Even more impressive is the fact that both insulin and bGH require oxidative protein folding and that insulin is a heterodimer. Both examples attest to the versatility and economic potential of *E. coli*-based production. This review will be focusing on the encountered knowledges (gene optimization) gained during work on some

biopharmaceutical proteins projects. It is intended to set up the basic requirements to get overexpression of the protein of interest. There are several other avenues to increase and stabilize gene expression (i.e., expression vector, host competent cell kinds, etc.), but these are beyond the scope of this review article.

Rare codon interference in recombinant protein biosynthesis

E. coli carries its own specific bias in the usage of the 61 amino acid codons encoding the 20 naturally occurring amino acids relative to other organisms (Table 1&2). Codon usage in *E. coli* is reflected by the level of cognate amino-acylated tRNAs available in the cytoplasm. Major codons occur in highly expressed genes whereas the minor or rare codons tend to be in genes expressed at low levels. Codons rare in *E. coli* are often abundant in heterologous genes from sources such as eukaryotes (Table 3), archaea bacteria and other distantly related organisms with different codon frequency preferencies (Kane, 1995). Expression of genes containing rare codons can lead to translational errors, as a result of ribosomal stalling at positions requiring incorporation of amino acids coupled to minor codon tRNAs (McNulty *et al.,* 2003).

Codon bias problems become highly prevalent in recombinant expression systems, when transcripts containing rare codons in clusters, such as doublets and triplets accumulate in large quantities. Translational errors arising from rare codon bias include mistranslational amino acid substitutions, frameshifting events or premature translational termination (Kurland and Gallant, 1996; Sørensen *et al.,* 2003a). In-frame two amino acids "hops" have been reported at a single disfavoured AGA codon (Kane et al., 1992). Protein quality is influenced by codon bias through the insertion of lysine for arginine at AGA codons (Calderone et al., 1996; Seetharam et al., 1988). Therefore, expression of full-length protein at high levels is not equivalent with translational integrity. The most problematic codons are decoded by products of the genes argU (AGA and AGG), argX (CGG), argW (CGA and CGG), ileX (AUA), glyT (GGA), leuW (CUA), proL (CCC) and lys (AAG). AAG is a major E. coli codon decoded by tRNALys/UUU, which is enabled to wobble to G by the xm5s2U34 modification (Yarian et al., 2000). Since UUU reads AAG less efficient by, there is a problem when a target sequence contains consecutive AAG codons. Most focus has been on the rare arginine codons AGG and AGA, occurring in E .coli at frequencies of 0.14 and 0.21%, respectively (Kane, 1995). In this regard, we note that 24 protein genes in E. coli end with AGG and 28 end with AGA, another rare arginine codon. The overall frequencies of usage for these rare codons in E. coli are 0.12% and 0.21%, respectively. On the basis of these frequencies, only about 16 genes would be expected to end with AGG or AGA. This overrepresentation of rare codons at the 3' ends of E. coli genes has been noted and attributed, in part, to the requirements of closely spaced downstream genes for translation initiation signals (Kane, 1995).

However, how can this problem be overcome? three alternative strategies are utilized to remedy codon bias. One approach is the total gene synthesis which is rapidly becoming the preferred method for applications requiring the assembly of DNA sequences, both of natural and engineered. The major drawback of total gene synthesis is the high associated cost. Second, the site-directed mutagenesis of the target sequence for the generation of codons reflecting the tRNA pool in the host system is expected to solve some of these problems, the process may become tedious and costly if too many nucleotides

need to be changed. Third, is the co-transformation (as above mentioned) of the host competent with a plasmid harboring (Fig. 1) a gene encoding the tRNA cognate to the problematic codons (Dieci et al., 2000). By increasing the copy number of the limiting tRNA species, E. coli can be controlled to match the codon usage frequency in heterologous genes. Several plasmids are available for rare tRNA co-expression, most of which are based on the p15A replication origin. This enables maintenance in the presence of the ColE1 replication origin present in most expression plasmids. Numerous reports confirm the concept of plasmid mediated tRNA complementation (Baca and Hol, 2000; Kim et al., 1998; Sørensen et al., 2003b). Commercially available tRNA complementation plasmids include pRARE (Fig. 1) and that implemented in the Codon Plus system from Stratagene (http://www.stratgene.com).

Table (1): Escherichia coli total codons 323059, and total proteins 968 (Wada et al., 1992).

	U	C	A	G	
	6089 (Phe)	3382 (Ser)	4875 (Tyr)	1552 (Cys)	U
	5838 (Phe)	3048 (Ser)	2495 (Tyr)	1961 (Cys)	C
U	3397 (Leu)	2105 (Ser)	744 (Stp)	259 (Stp)	A
	3659 (Leu)	2549 (Ser)	65 (Stp)	4169 (Trp)	G
	3204 (Leu)	2123 (Pro)	3666 (His)	7981 (Arg)	U
C	3135 (Leu)	1354 (Pro)	3471 (His)	6946 (Arg)	C
	960 (Leu)	2623 (Pro)	4223 (Gln)	989 (Arg)	A
	17477 (Leu)	7724 (Pro)	9589 (Gln)	1491 (Arg)	G
	8809 (Ile)	3499 (Thr)	5265 (Asn)	2380 (Ser)	U
A	8712 (Ile)	7873 (Thr)	7868 (Asn)	4831 (Ser)	C
	1273 (Ile)	2108 (Thr)	12104 (Lys)	692 (Arg)	A
	8507 (Met)	4050 (Thr)	3857 (Lys)	428 (Arg)	G
	6716 (Val)	5770 (Ala)	10382(Asp)	9202 (Gly)	U
G	4552 (Val)	7490 (Ala)	7116 (Asp)	9824 (Gly)	C
	3896 (Val)	6760 (Ala)	14134 (Glu)	2246 (Gly)	A
	7972 (Val)	10641 (Ala)	6149 (Glu)	3110 (Gly)	G

Table (2): Escherichia coli gross base composition for sum of reading frames (numbers, U: 226989, C: 238372, A: 238380, G: 265465) (Wada et al., 1992; Boycheva et al., 2003).

	U	C	A	G	
	1.47 (Phe)	0.78 (Ser)	1.12 (Tyr)	0.32 (Cys)	U
	1.34 (Phe)	0.67 (Ser)	0.94 (Tyr)	0.38 (Cys)	C
U	0.78 (Leu)	0.46 (Ser)	0.14 (Stp)	0.05 (Stp)	A
	0.75 (Leu)	0.50 (Ser)	0.01 (Stp)	0.73 (Trp)	G
	0.74 (Leu)	0.46 (Pro)	0.80 (His)	1.57 (Arg)	U
C	0.69 (Leu)	0.28 (Pro)	0.72 (His)	1.30 (Arg)	C
	0.21 (Leu)	0.55 (Pro)	0.88 (Gln)	0.18 (Arg)	A
	0.43 (Leu)	1.44 (Pro)	1.79 (Gln)	0.25 (Arg)	G
	2.02 (Ile)	0.76 (Thr)	1.15 (Asn)	0.47 (Ser)	U
A	1.90 (Ile)	1.64 (Thr)	1.64 (Asn)	0.90 (Ser)	C
	0.28 (Ile)	0.44 (Thr)	2.52 (Lys)	0.13 (Arg)	A
	1.67 (Met)	0.76 (Thr)	0.72 (Lys)	0.07 (Arg)	G
	1.38 (Val)	1.13 (Ala)	2.04 (Asp)	1.62 (Gly)	U
G	0.89 (Val)	1.40 (Ala)	1.33 (Asp)	1.65 (Gly)	C
	0.76 (Val)	1.26 (Ala)	2.64 (Glu)	0.38 (Gly)	A
	1.40 (Val)	1.79 (Ala)	1.03 (Glu)	0.47 (Gly)	G

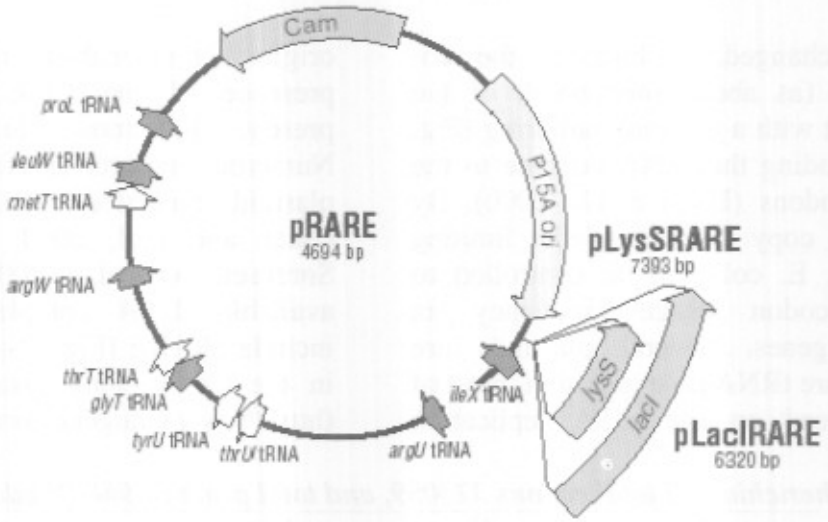

Fig. (1): Map of pRARE plasmid family: The basic structure of pRARE is indicated, plysRARE and placIRARE contain the genes encoding T7 lysozyme (plyeS) and lac repressor (lac), respectively. Also indicated are chloramphenicol resistance gene (cam), replicon (P15A ori) and tRNA genes. tRNA genes corresponding to rare codons in E.coli are indicated (http://www.novagen.com).

The industrial applications of these approaches were clear in human recombinant interferon production. The yield of human interferon alpha 2b expression was increased 9.5-11.5 fold by replacement of the low-usage rare arginine codons (Table 4), into abundant-usage codons (Valente et al., 2004). While, co-transformed the E. coli BL21 (DE3) with the gene of human interferon-α2a and the argU gene that codes for rare tRNAArg(AGG/AGA) produced IFN-α2a in yield of more than 25% of the total cellular proteins (Jeong and Shin, 1998).

Translation initiation

Translation initiation from the translation initiation region (TIR) of the transcribed messenger RNA requires a ribosomal binding site (RBS) including the Shine-Dalgarno (SD) sequence and a translation initiation codon (Sørensen et al., 2002). The Shine-Dalgarno sequence is located 7±2 nucleotides upstream from the initiation codon, which is the canonical AUG in efficient recombinant expression systems (Ringquist et al., 1992). Optimal translation initiation is obtained from mRNAs with the SD sequence UAAGGA GG. The RBS secondary structure is highly important for translation initiation and efficiency is improved by high contents of adenine and thymine (Laursen et al., 2002). Translation initiation efficiency is in particular influenced by the codon following the initiation codon and adenine is abundant in highly expressed genes (Stenstrom et al., 2001).

As reviewed by Sprengart and Porter (1997), translation initiation is affected by several factors. A consensus Shine-Dalgarno (SD) sequence and the proper spacing and sequence before the initiation codon are certainly beneficial. A downstream box (DB) may also be helpful (Etchegaray and Inouye, 1999), however, an even more important factor may be possible mRNA secondary structures that block ribosome binding (De Smit and van Duin, 1990). A single base change that affected secondary structure stability near the SD region caused a 500-fold change in the expression of the coat protein of RNA bacteriophage MS2. These secondary structures can possibly be disrupted by RNA

helicases such as the DEAD protein of E. coli. Iost and Dreyfus (1994) showed that overexpression of the DEAD protein stimulated β-galactosidase expression 30-fold from the T7 promoter, but not from the lac promoter. Without the DEAD protein overexpression, β-galactosidase production from the T7 promoter was tenfold lower than from the lac promoter even though transcription was tenfold faster. Further experiments suggested that the DEAD protein stabilized the transcript, apparently independent of ribosome coverage. Yet, the E. coli DEAD-box proteins share important homologies with proteins having demonstrated RNA helicase activities (Iost et al., 1999, Linder and Daugeron, 2000). It is still tempting to recommend testing the overexpression of these proteins for genes with suspected problematic mRNA secondary structure. An alternative approach is to modify the 5′ coding sequence (without changing the amino acid sequence) to discourage predicted secondary structure.

Stop codon

In bacterial genome, UAA is the most commonly used termination codon. UGA is used more than UAG, although it appears to cause more reading frame errors (Table 5). An error in reading termination codon when an aminoacyl-tRNA (suppressor) improperly responds to it results in the continuation of protein synthesis until another termination codon is encountered in the phase reading frame. This readthrough leads to the synthesis of a longer protein, with additional C-terminal amino acid residues. Amber UAG suppressors tend to be relatively efficient, usually in the range of 10-50%, depending on the system. Opal stop codons (UGA) are generally less efficient as translational terminators than are ochre stop codons (UAA) (Scolnick et al., 1968). Ochre (UAA) suppressor is always much less efficient, usually with activities below 10%. UGA is the least efficient of the termination codons in its natural function and it is misread by tRNATrp as frequently as 1-3% in wild-type situations. Change of UGA into UAA terminators in human IFN-α2b leads to doubling the protein expression and production (Sanchez et al., 1998). Transcription terminators stabilize the mRNA by forming a stem loop at the 3' end (Newbury et al., 1987). Translation termination is preferably mediated by the stop codon UAA in Escherichia coli. Increased efficiency of translation termination is achieved by insertion of consecutive stop codons or the prolonged UAAU stop codon (Poole et al., 1995).

Frameshifting with altered translational apparatus

General Basic

Translation errors during protein synthesis are a well-documented phenomenon (Parker, 1989). They can have several different effects including frameshifts (Spanjaard et al., 1990), premature truncation, low expression (Brinkman et al., 1989; Rosenberg et al., 1993) and misincorporation (Seetharam et al., 1988; Schimmel, 1989). The majority of previously reported translation errors have fallen into the first three categories. This has led to the general view that the codon usage in recombinant proteins is important primarily for determining expression levels, particularly in the first ten to 20 codons (Parker, 1989; Spanjaard et al., 1990; Rosenberg et al., 1993). If expression levels are sufficient and the protein shows the correct apparent molecular mass on an SDS-PAGE gel, translation integrity is often assumed to be intact. This assumption is not always valid as previously demonstrated by (Calderone et al., 1996).

Table (3): Low-usage codons in four different species (E.coli (E.C.), Saccharomyces cervisia (s.c.), Drosophila melanogaster (d.m.), Primates (p.). (Zhang et al., 1991; Wada et al., 1992; Forman et al., 1998).

Amino acid	e.c.	s.c.	d.m.	p.
Arg	AGG	AGG		
Arg	AGA		AGA	
Ile	AUA		AUA	
Leu	CUA			
Arg	CGA	CGA	CGA	CGA
Arg	CGG	CGG	CGG	CGG
Pro	CCC			
Ser	UCG			UCG
Arg		CGC		CGC
Pro		CCG		CCG
Leu		CUC		
Ala		GCG		GCG
Thr		ACG		ACG
Leu		ACG	UUA	
Gly			GGG	
Ser			AGU	
Cys			UGU	
Arg				CGU

We shall define codon usage as the number of times a codon is translated per unit time. The most common way of measuring codon usage is by summing the number of times codons appear in the reading frames of genome. Second way of estimating codon usage is to examine the distribution of usage in different genes.

In Human Diseases

Fidelity of transcription depends on proofreading by a 3'-5' exonuclease activity found in a number of eukaryotic and prokaryotic RNA polymerases (Thomas et al., 1998). The Escherichia coli RNA polymerase exhibits error rates ranging from 1 to 10 misincorporations per 105 synthesized nucleotides (for review, see Libby and Gallant, 1991). Aberrant RNAs also include the presence of altered nucleotides in the mRNA caused either by direct alteration of the mRNA or by misincorporation of altered mutagenic nucleotides, such as 8-oxo-GTP, by the RNA polymerase (Taddei et al., 1997). The presence of 8-oxo-G in cellular RNA was shown to be restricted to vulnerable neurons in dementia such as Alzheimer's disease (AD) (Nunomura et al., 1999) or Parkinson's disease (Zhang et al., 1999) and to be correlated with the synthesis of altered proteins (Dukan et al., 2000). The rate of transcriptional frameshifting is unknown (in human diseases), but it is proposed that it leads to some neurodegenerative diseases (van Leeuwen et al., 1998b). In patients with AD, the accumulation of aberrant protein species shifted to the +1 frame with no related mutation in the DNA blueprint has been reported. The appearance of those proteins was shown to be caused by a molecular misreading of a coding GA repeat resulting in the deletion of a dinucleotide GA in the mRNA (Hol et al., 1998; van Leeuwen et al., 1998a, b).

In *Escherchia coli*

Translational errors have been characterized in E. coli and measured in vivo and/or in vitro (for review, see Kurland, 1992). These include missense errors resulting in the substitution of one amino acid for another, or incorporation of extra amino acids caused by

readthrough of termination codons. Overall, translational errors occur at a frequency of 10-3 to 10-4 (Loftfield and Vanderjagt, 1972; Edelmann and Gallant, 1977; Ellis and Gallant, 1982; Parker, 1989) and are limited by the ability of the ribosome to discriminate between correct and incorrect tRNAs entering the A-site (Thompson, 1988; Nierhaus, 1990; Czworkowski and Moore, 1996). When induced by the addition of some antibiotics (e.g., streptomycin), translational misreading leads to increased concentrations of oxidized proteins (Dukan et al., 2000), which were found to accumulate in aging cells (Stadtman, 1992). Other translational errors include premature termination (Menninger, 1977; Kurland and Gallant, 1996) and translational frameshifting. The latter generally gives rise to a shorter peptide because of the subsequent encounter of stop codons in the shifted frame. Spontaneous frameshifts resulting from translational errors occur at a frequency of <10-5 per codon (Kurland, 1992), but were shown to increase in E. coli cells entering stationary phase (Barak et al., 1996).

Framshift mechanism

The frameshift occurs by realignment of the peptidyl-tRNA from the codon UGA immediately onto the overlapping UUU triplet (Craigen et al., 1985). The frameshift mechanism is depicted in Figure 2. The four frameshift enhancing features are numbered in the figure and described here. First, frameshifting is facilitated by slow translation of the UGA codon (Craigen and Caskey, 1986; Adamski et al., 1993), which is consistent with the autoregulatory function of the frameshift. Frameshifting also occurs with codons substituted for the UGA (Weiss et al., 1987; Curran and Yarus, 1988), and frameshift frequency is inversely related to the rate of aa-tRNA selection at those codons (Sipley and Goldman, 1993). Second, the 34 end of 16 S rRNA (the anti-Shine–Dalgarno sequence, Shine and Dalgarno, 1974) base pairs with a run of purines upstream of the slip site to enhance frameshifting (Weiss et al., 1988). The mechanism by which this interaction stimulates frameshifting is not known, but it is worth noting that a Shine-Dalgarno-like interaction also stimulates leftward (–1) frameshifting in E. coli dnaX (Larsen et al., 1994), and that frameshift direction and efficiency may be related to the size of the spacing between that interaction and the P site of t-RNA (Larsen et al., 1995). Third, a G:U wobble base pair in the pre-shift codon : anticodon complex is associated with high frequency frameshifting, as if this weak pair facilitates slippage from the initial frame (Curran, 1993). Weak pairing in the initial frame has also been shown to contribute to high frequency frameshifting at other programmed frameshift sites (Tsuchihashi and Brown, 1992; Weiss et al., 1989). Fourth, the realigned complex includes stable base pairs (Weiss et al., 1987; Curran, 1993). This feature is also observed at virtually all other programmed frameshift sites.

Recently, Fu and Parker (1994) showed that a specific UUU UAC site near the 3'end of *E. coli argI* undergoes one nucleotide rightward frameshift with a frequency of several percent. Derivatives that preserve the UUU Ynn theme are frameshift prone, but changes of either the phenylalanine codon to UUC or the 34 neighbor to adenosine inhibit frameshifting. Thus, the sequence requirements for frameshifting at this site are at least superficially similar to those found for the extensively characterized RF2 programmed frameshift slippage site (the third and fourth features, Fig. 2). However, unlike the RF2 slippage site, the context surrounding the *argI* slippage site does not include any other element, such as a Shine-Dalgarno sequence, known to facilitate frameshifting at

programmed frameshift sites. Furthermore, this frameshift has no apparent cellular utility; it does not generate a useful polypeptide, and it has no apparent regulatory function. It may, therefore, be a high frequency 'error.' In both contexts, high frequency frameshifting requires that the rephased tRNA:

Table (4): E.coli usage-codons, (Fuglsang, 2003; http://www.novagen.com/protocols).

	Escherichia coli			
Amini acid	**codons**	**%**	**f.genes[1]**	**f.classII[2]**
	UUA	11	0.131	0.055
	UUG	11	0.129	0.034
	CUU	10	0.104	0.056
Leu	CUC	10	0.104	0.080
	CUA	03	0.37	0.008
	CUG	55	0.496	0.767
	AUU	47	0.507	0.335
Ile	AUC	46	0.420	0.659
	AUA	07	0.073	0.006
	CCU	15	0.159	0.112
Pro	CCC	10	0.124	0.016
	CCA	19	0.191	0.153
	CCG	56	0.525	0.719
	CGU	43	0.378	0.643
	CGC	37	0.398	0.330
Arg	CGA	05	0.065	0.011
	CGG	08	0.098	0.008
	AGA	04	0.039	0.006
	AGG	02	0.022	0.003
	GGU	38	0.337	0.508
Gly	GGC	40	0.403	0.428
	GGA	09	0.109	0.020
	GGG	13	0.151	0.044

1. fraction in all genes, 2. fraction in class II, % indicates of used synonymous codon. Codon usage is expressed as the fraction of all possible codons for a given amino acid. "All genes" is the fraction represented in all 4290 coding sequences in the *E.coli* genome. "Class II" is the fraction represented in 195 genes highly and continuously expressed during exponential growth.

message complex be stable, but the partially mismatched complex at UUC CGG is relatively frameshift-prone only in the RF2 context. Together with previous work, these data suggest that the Shine-Dalgarno may be able to help hold the shifted phase following tRNA : message realignment. As previously shown many RF2 alleles frameshift with considerable frequencies despite forming moderately mismatched complexes in the rightward frame (Curran, 1993). The intermediate frameshift frequency for RF2-UUC CGG, which has a single mismatch in the rightward frame (an A:C in the middle position), is consistent with that earlier work. In contrast, the NSD alleles, that have UUC as the phenylalanine codon, essentially do not frameshift. The simplest explanation is that those alleles, which can form neither a perfectly realigned tRNA : message complex nor a Shine-Dalgarno interaction, cannot stabilize the rephased complex.

Examples of codon misreadings in both the second and third bases of the codon-anticodon pairs have been reported. A second base misread of the rare arginine codon AGA being substituted with AAA lysine was observed in insulin-like growth factor-1 (Steetharam *et al.*, 1988) and an HIV-1 gp160 fragment (Calderone *et al.*, 1996). Substitution

of CAA/G glutamine for CAU histidine in granulocyte colony stimulating factor is an instance of mistranslation through a third base wobble. A synthetic test protein designed to magnify translational problems using consecutive AGG–AGG arginine codons resulted in an exceedingly high frequency of frameshifting in the +1 direction (Spanjaard *et al.*, 1988). Kane *et al.* (1992) found an interesting two amino acids in-frame ''hop'' at a single disfavored AGA codon.

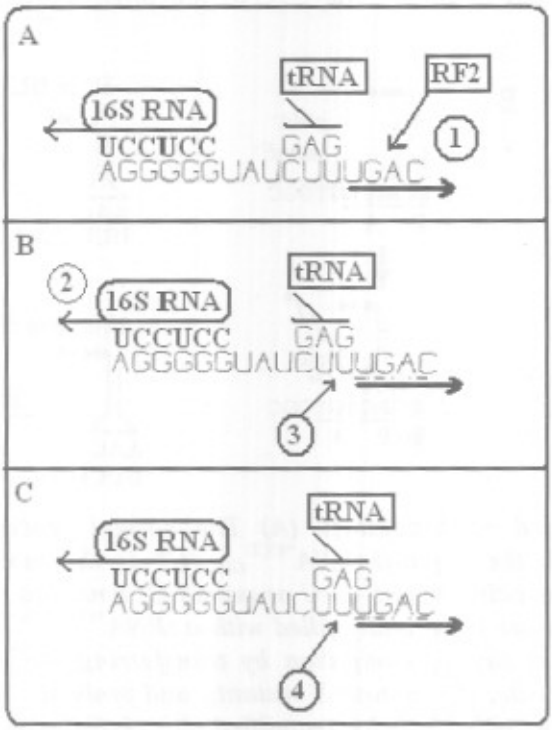

Fig. (2): Model of the RF2 programmed frameshift mechanism. The mechanism is outlined illustrating the four features that contribute to high frequency frameshifting. These features are marked by circled numbers and are described below and in Introduction. (A) Prior to the shift, the ribosomal complex contains peptidyl-tRNALeu in the P site base paired to the CUU triplet. The UGA codon (underlined) may either be translated normally, or a frameshift may occur. The first shifty feature is slow translation of the UGA (or of other codons substituted for it), which allows time for frameshifting. (B) The probability of a frameshift is increased if the 34 end of 16 S RNA (the anti-Shine-Dalgarno) base pairs with the message (feature 2). Realignment is also facilitated by a G:U wobble base pair between the peptidyl-tRNA and the message (feature 3). (C) The fourth shifty feature is stable base pairing between the tRNA and message in the shifted frame (Schwartz and Curran, 1997).

The RF2 programmed frameshift site (Fig. 3) has been extensively analyzed genetically and biochemically. The RF2 protein terminates translation at UGA and UAA codons (Scolnick *et al.*, 1968). *prfB* contains a UGA near its 54 end, and RF2

terminates synthesis of nascent RF2 polypeptide in a concentration-dependent manner (Adamski *et al.*, 1993; Craigen and Caskey, 1986). To synthesize RF2, ribosomes bypass the UGA termination codon by a rightward (+1) frameshift. In conclusion, this ribosomal frameshift results from the coincidence of three events: (1) decreased codon–anticodon affinity at the P-site, which is caused by tRNA hypomodification in *mnmE*- and *gidA*- strains; (2) a repetitive mRNA sequence predisposing to slippage; and (3) increased translational pausing attributable to the presence of a rare codon at the A-site.

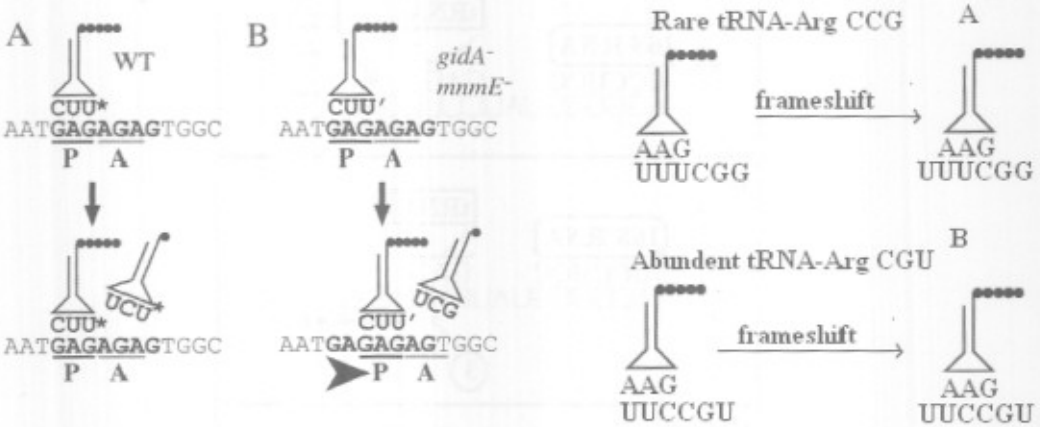

Fig.(3): Model for the observed +2 frameshift. (A) The normal process of translation in NECB1 lacZ(+GA) begins with the peptidyl-tRNA$^{*UCU}_{Glu}$ (U* is the modified uridine at the wobble position, which could be either mnm5-s2-U or mnm5-U) bound to the mRNA in the P-site of the ribosome. The empty A-site is normally filled with a tRNA$^{*UCU\ Glu}$ and the translational complex can proceed to the next step of elongation by transferring the peptide and translocating the accepted tRNA in the P-site. (B) In mnmE mutants, and probably in gidA mutants, the uridine at the wobble position of the tRNA^{Glu} is hypomodified (U': s2-U or some other hypomodified form of the mnm5-s2-U). This hypomodified nucleotide is less efficient at pairing with a guanine in the wobble position. The presence of a rare AGA codon at the A-site results in a pause during the translation process. Therefore, the conjunction of a less efficient pairing between the tRNA and the mRNA at the P-site and the presence of a rare codon at the A-site renders the ribosome prone to shifting. This translation complex moves to (or possibly scans for) the next GAG cognate codon of the tRNA in the P-site, which is easy on this repetitive sequence and then resumes translation, but in a different frame. Therefore, the conjunction of these three events - the presence of the hypomodified tRNAGlu UUC in the P-site, the repetitive sequence of the mRNA and the rare AGA codon in the A-site allows the ribosome to shift to another frame, as described in Figure 2. Therefore on the lacZ(+GA) mRNA, the ribosome slips over the 2 bases (GA) toward the 3' end on the mRNA, producing a +2 frameshift that restores the normal reading frame and subsequently the deactivate the protein function (Bregeon et al., 2001).

Extension of the dual error model may explain tRNA hopping

Most of the described nonprogrammed ribosomal frameshifts are thought to slip in either +1 or -1 frame (Farabaugh and Björk, 1999). The +1/-1 dual error model proposed previously (Qian *et al.*, 1998) suggests several explanations for nonprogrammed +1 and -1 frameshifts that could occur, especially at CCCN and GGGN sites (N, any nucleotide).

This model assumes that slipping of the ribosome would happen at a precise stage after the three nucleotide translocation. At this step, the ribosome is prone to slip on the mRNA, in particular when a near-cognate tRNA or a hypomodified tRNA is present in the P-site. In this case, the shift is attributable to a weaker codon–anticodon pairing at the wobble position. Even if a normal cognate tRNA is in the P-site, the ribosome could slip when a translational pause is induced because of the presence of a rare codon in the A-site.

Table (5): Missing codon pairs in the Escherichia coli genome (Boycheva et al., 2003).

Amino acid: stop codon	codon pairs
Phe:stop	UUU:UAG
Ser:stop	UCU:UAG
Ser:stop	UCC:UAG
Tyr:stop	UAC:UAG
Cys:stop	UGC:UAG
Leu:stop	CUC:UAG
Leu:stop	CUA:UAG
Pro:stop	CCU:UAG
Pro:stop	CCC:UAG
Pro:stop	CCA:UAG
His:stop	CAU:UAG
Thr:stop	ACU:UAG
Thr:stop	*ACC*:UAG
Thr:stop	ACG:UAG
Val:stop	GUC:UAG
Gly:stop	GGC:UAG
Thr:stop	ACU:UAG
Thr:Arg	ACU:AGA
Pro:Arg	CCU:AGG

Underlined codon pairs are missing in the subset of 2656 protein coding sequences and the non-underlined pairs are missing in the entire *E.coli* genome. Rare codons are bolded, frequently used codons are given in italics and the rest of the codons are moderately used.

An extended version of the +1/-1 dual error model could also explain the tRNA hopping described by Weiss *et al.* (1987). The mutation they studied is a leaky -1 frameshift allele *trpE91* in *Salmonella* (Riyasaty and Atkins, 1968). The low-level translational frameshifting that occurs at this site was studied by transferring the region surrounding the mutation into a *lacZ* expression vector (Weiss *et al.*, 1987). The sequence of the expressed peptide revealed that the frameshift occurred by a translational hopping, creating a +2 rather than a -1 frameshift, via a tRNAVal hopping by +2. Two other sequences were identified on which the ribosome is prone to slip (or hop) over 5 or 6 nucleotides. Protein sequencing confirmed that tRNAAsn hops 5 nucleotides between cognate codons and that tRNALeu hops over an in-frame stop codon between near-cognate codons. The presence of an in-frame stop codon for both +2 and +6 hopping apparently creates a pausing site during translation, a favorable condition for the frameshifting (Farabaugh, 1996). The +5 event, however, occurred at an in-frame UCA codon that is not predicted to cause a translational pause. The strain used by Weiss and colleagues (1988, 1989) were find that the CSH26 with a deletion of the *pro-lac* region, where the *strC* and *trmB* genes have been mapped, respectively, by P1 co-transduction (Roberts and Reeve, 1970) and conjugation

(Marinus *et al.*, 1975), which may have an influence on this +5 frameshift. The *strC* gene is known to be involved in streptomycin resistance and to modify the ribosome structure (Roberts and Reeve, 1970). The TrmB protein is a tRNA methyltransferase, which catalyzes the formation of methyl-7-guanosine (m7G) in some tRNAs, including the tRNAAsn hopping over 5 nucleotides. Whether this modification is important to stabilize the translation complex remains to be tested. All observations of ribosome frameshifting with hypomodified tRNA or in presence of a pausing site during translation are, therefore, in favor of the extended version of the dual error model described above.

Though ribosomal frameshifting and mistranslation events associated with the rarest codons AGG and AGA have been well-documented (Seetharam *et al.*, 1988; Calderone *et al.*, 1996; Curran, 1993; Spanjaard and van Duin, 1988). The composition of these Class II genes is likely to better reflect the resident tRNA populations needed during high level target protein synthesis, and in this context, CGG becomes the third rarest arginine codon. CGG occurs at a frequency of 0.54% in all *E. coli* open reading frames (Nakamura *et al.*, 2000). The frequency and severity of CGG-associated mistranslational events described in the current report are remarkable. Without supplemental enhancement of arginyl $tRNA^{CGG}$, fully 78% of the naturally encoded protein was frameshifted in the +1 direction at the second and third CGG codons in the triplet cluster. This exceeds the paradigmatic example of Spanjaard, who observed a 50% frameshift in the +1 direction when a consecutive AGG–AGG codon pair was artificially inserted into a model system to study processivity errors (Spanjard and van Duin, 1988).

Missense substitutions of CAG glutamine residues for CGG encoded arginine were also found at an exceedingly high frequency. Mass spectrometry of the intact frameshifted molecule demonstrated that a weighted average of 2.1, out of the 9 remaining CGG codons not directly involved in the frameshift phenomena, were misincorporated with glutamine. Detectable populations of up to four misincorporations were evident representing 11% of the total. Relative to previous reports, this rate of missense error introduction of 23% at CGG arginines is surprisingly high. The majority of reports on amino acid missense substitution cite frequencies of 1–2% (Scorer and Carrier, 1991; Santos and Tuite, 1993). Exceptions include reports by Calderone *et al.* (1996) who determined by using genetic codes optimization and/or mass spectrometry analysis that lysine for arginine misincorporation at three AGA sites at very high frequencies of 36–42% and Parker (1989) who reported translational error frequencies of about 10%. However, in the study by Calderone *et al.* (1996), production levels approached 100 mg/L, or a specific yield approximately 25 times that achieved in the current report. It should be stressed that in the present case, expression was evident only by Western immunoblot analysis of unfractionated whole cell lysates and represents less than 1% of total cellular protein.

The mechanism for CAG glutamine substitution for CGG arginine is the result of a second base codon misreading in the codon–anticodon pair. This is yet another example of what has been described as the ''hungry codon syndrome'' (Kurland and Gallant, 1996). The overutilization of CGG arginine in the heterologous gene placed extraordinary physiological demands upon the host cell protein synthetic apparatus. The hypothesis is that the ribosomes stall at the CGG positions because there is a low level of cognate arginyl-tRNA. In addition, there are only three glutamine

residues (positions 101, 247, and 257) in the cloned HSV-2 protease domain and 2 of them are after the tandem CGG codons. Since compositionally glutamine represents about 4% of *E. coli* proteins, this low level of glutamine codons (3 out of 258 residues, about 1%) suggests that the relative concentration of glutaminyl tRNA was high. Presumably, this high level of glutaminyl-tRNA coupled with stalled ribosomes at the CGG codons likely facilitated the misincorporation event (McNulty *et al.*, 2003).

Conclusions

Although many alternative organisms and expression systems are now being used for recombinant protein production, exciting progress continues to be made with *E. coli*. Overexpression of target genes in *E. coli* is often the method of choice, because of extensive knowledge of its genome, availability of versatile vector systems, and host strains, the ease of use, low costs, and high expression levels, exceeding often more than 30% of total cellular protein. However, despite the many advantages of *E. coli*, high-level expression is not routinely achieved. Among the many reported causes preventing efficient heterologous protein production in *E. coli* are biased codon usage, gene product toxicity, solubility, mRNA secondary structure, and mRNA stability. In addition, rare codon gene expression can lead to translational errors as a result of ribosomal stalling at a position requiring incorporation of amino acids coupled to minor tRNAs, or even at sites requiring major tRNAs, but which are depleted because of overutilization of a particular amino acid. The mistranslational events related to rare tRNAs are observed as codon misreadings and as processing errors and they manifest themselves as amino acid substitutions or frameshift events. Specifically, the rare arginine (AGG, AGA, CGG, and CGA), leucine (CUA), isoleucine (AUA), and proline codons (CCC) often lead to frameshift errors and ultimately to undesired products. These are the all encountered solved problems for *E.coli* protein expression, so it is prefer to take the advantages of *E.coli* expression machinery after optimization of the gene interest.

REFERENCES

Adamski, F., Donly, B.C. and Tate, W.P. (1993). Competition between frameshifting, termination and suppression at the frameshift site in the Escherichia coli release factor-2 mRNA. Nucl. Acids Res. 21:5074-8.

Baca, A.M. and Hol, W.G., (2000). Overcoming codon bias: a method for high-level overexpression of Plasmodium and other AT-rich parasite genes in Escherichia coli. Int. J. Parasitol. 30: 113-18.

Barak, Z., Gallant, J., Lindsley, D., Kwieciszewki, B. and Heidel, D. (1996). Enhanced ribosome frameshifting in stationary phase cells. J. Mol. Biol. 263: 140-48.

Boycheva, S., Chkodrov, G. and Ivanov, I. (2003). Codon pairs in the genome of Escherichia coli. Bioinformatics 19: 987-98.

Brégeon, D., Colot V., Radman, M. and Taddei, F. (2001). Ribosomal frameshift translational misreading: a tRNA modification counteracts a +2. Genes & Dev. 15: 2295-06.

Brinkman, U., Mattes, R.E. and Buckel, P. (1989). High-level expression of recombinant gene in Escherichia coli is dependent on the availability of the dnaY gene product. Gene 85: 109-14.

Calderone, T.L., Stevens, R.D. and Oas, T.G. (1996). High-level misincorporation of lysine for arginine at AGA codons in a fusion protein expressed n Escherichia coli. J.Mol.Biol., 262: 407-12.

Craigen, WJ. and Caskey, CT. (1986). Expression of peptide chain release factor 2 requires high-efficiency frameshift. Nature. 322: 273-5.

Craigen, WJ., Cook, RG., Tate, WP. and Caskey, CT. (1985). Bacterial peptide chain release factors: conserved primary structure and possible frameshift regulation of release factor 2. Proc Natl Acad Sci USA. 82: 3616-20.

Curran, JF. (1993). Analysis of effects of tRNA:message stability on frameshift frequency at the Escherichia coli RF2 programmed frameshift site. Nucl. Acids Res. 21:1837-43.

Curran, J.F. and Yarus, M. (1988). Use of tRNA suppressors to probe regulation of Escherichia coli release factor 2. J Mol Biol. 203: 75-83.

Czworkowski, J. and Moore, P.B. (1996). The elongation phase of protein synthesis. Prog. Nucleic Acid Res. Mol. Biol. 54: 293-32.

De Smit, MH. And van Duin, J. (1990). Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. Proc Natl Acad Sci USA 87:7668-72.

Dieci, G., Bottarelli, L., Ballabeni, A. and Ottonello, S. (2000). tRNAassisted overproduction of eukaryotic ribosomal proteins. Protein Expr. Purif. 18: 346-54.

Dukan, S., Farewell, A., Ballesteros, M., Taddei, F., Radman, M. and Nystrom, T. (2000). Protein oxidation in response to increased transcriptional or translational errors. Proc. Natl. Acad. Sci. USA. 97: 5746-49.

Edelmann, P. and Gallant, J. (1977). Mistranslation in *E. coli*. Cell 10: 131-37.

Ellis, N. and Gallant, J. (1982). An estimate of the global error frequency in translation Mol. Gen. Genet. 188: 169-72.

Etchegaray J-P. and Inouye M (1999). Translational enhancement by an element downstream of the initiation codon in Escherichia coli. J Biol Chem. 274:10079-85.

Farabaugh, P.J. (1996). Programmed translational frameshifting. Microbiol. Rev. 60: 103-34.

Farabaugh, P.J. and Björk, G.R. (1999). How translational accuracy influences reading frame maintenance. EMBO J.18: 1427-34.

Forman, M.D., Stack, R.F., Masters, P.S., Hauer, C.R. and Baxter, S.M. (1998). High level, context dependent misincorporation of lysine for arginine in *Saccharomyces cerevisiae* a1 homeodomain expressed in *Escherichia coli*. Protein Sci., 7: 500-03.

Fu, C. and Parker, J. (1994). A ribosomal frameshifting error during translation of the argI mRNA of *Escherichia coli.* Mol Gen Genet. 243: 434-41.

Fuglsang, A. (2003). Codon optimizer: a freeware tool for codon optimization. Protein Expression and Purification 31: 247-49.

Hol, E.M., Neubauer, A., de Kleijn, D.P., Sluijs, J.A., Ramdjielal, R.D. et al., (1998). Dinucleotide deletions in neuronal transcripts: A novel type of mutation in non-familial Alzheimer's disease and Down syndrome patients. Prog. Brain Res. 117: 379-95.

Iost, I. and Dreyfus, M. (1994). mRNAs can be stabilized by DEAD-box proteins. Nature 372:193-96.

Iost, I., Dreyfus, M. And Linder, P. (1999). Ded1p, a DEAD-box protein required for translation initiation in Saccharomyces cerevisiae, is an RNA helicase. J Biol Chem. 274: 17677-83.

Jeong, W. and Shin, H.-C. (1998). Supply of the argU gene product allows high-level expression of recombinant human interferon-

α2a in Escherichia coli. Biotechnol.letter 20, 19-22.

Kane, J.F. (1995). Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. Curr. Opin. Biotechnol. 6, 494–500.

Kane, J.F., Violand, B.N., Curran, D.F., Staten, N.R., Duffin, K.L. and Bogosian, G. (1992). Novel in-frame two codon translational hop during synthesis of bovine placental lactogen in a recombinant strain of *Escherichia coli*. Nucleic Acids Res. 20, 6707–12.

Kim, R., Sandler, S.J., Goldman, S., Yokota, H., Clark, A.J. and Kim, S.H. (1998). Overexpression of archaeal proteins in *Escherichia coli*. Biotechnol. Lett. 20, 207–10.

Kurland, C. and Gallant, J. (1996). Errors of heterologous protein expression. Curr. Opin. Biotechnol. 7, 489-93.

Kurland, C.G. (1992). Translational accuracy and the fitness of bacteria. Annu. Rev. Genet. 26: 29-50.

Larsen, B., Peden, J., Matsufuji, S., Matsufuji, T., Brady, K. Et al. (1995). Upstream stimulators for recoding. Biochem Cell Biol. 73:1123-9.

Larsen, B., Wills, NM., Gesteland, RF. And Atkins, JF. (1994). rRNA-mRNA base pairing stimulates a programmed -1 ribosomal frameshift. J Bacteriol. 176: 6842-51.

Laursen, B.S., Steffensen, S., Hedegaard, J., Moreno, J.M., Mortensen, K.K. and Sperling-Petersen, H.U. (2002). Structural requirements of the mRNA for intracistronic translation initiation of the enterobacterial *infB* gene. Genes Cells. 7, 901-10.

Libby, R.T. and Gallant, J.A. (1991). The role of RNA polymerase in transcriptional fidelity. Mol. Microbiol. 5: 999-04.

Linder, P. and Daugeron, M-C. (2000). Are DEAD-box proteins becoming respectable helicases? Nat Struct Biol., 7:97-99.

Loftfield, R.B. and Vanderjagt, D. (1972). The frequency of errors in protein biosynthesis. Biochem. J. 128: 1353-56.

Marinus, M.G., Morris, N.R., Soll, D. and Kwong, T.C. (1975). Isolation and partial characterization of three *Escherichia coli* mutants with altered transfer ribonucleic acid methylases. J. Bacteriol. 122: 257-65.

McNulty, D.E., Claffee, B.A., Huddleston, M.J. and Kane J.F. (2003). Mistranslational errors associated with the rare arginine codon CGG in *Escherichia coli* Protein Expr. Purif. 27: 365-74.

Menninger, J.R. (1977). Ribosome editing and the error catastrophe hypothesis of cellular aging. Mech. Ageing. Dev. 6: 131-42.

Miller, J.H. (1972). Experiment in molecular genetics. Cold Spring Harbor Laboratory Press, Cold spring Harbor, N.Y.

Nakamura,Y., Gojobori, T. and Ikemura, T. (2000). Codon usage tabulated from the international DNA sequence databases: status for the year 2000. Nucl. Acids Res. 28: 292.

Newbury, S.F., Smith, N.H., Robinson, E.C., Hiles, I.D. and Higgins, C.F. (1987). Stabilization of translationally active mRNA by prokaryotic REP sequences. Cell 48, 297-10.

Nierhaus, K.H. (1990). The allosteric three-site model for the ribosomal elongation cycle: Features and future. Biochemistry 29: 4997-08.

Nunomura, A., Perry, G., Pappolla, M.A., Wade, R., Hirai, K., Chiba, S. and Smith, M.A. (1999). RNA oxidation is a prominent feature of vulnerable neurons in Alzheimer's disease. J. Neurosci. 19: 1959-64.

Parker J. (1989). Errors and alternative sin reading the universal genetics code. Microbiol. Rev.,53: 273-298.

Poole, E.S., Brown, C.M. and Tate, W.P. (1995). The identity of the base following the stop codon determines the efficiency of in vivo translational termination in *Escherichia coli*. EMBO J. 14: 151-58.

Qian, Q., Li, J.N., Zhao, H., Hagervall, T.G., Farabaugh, P.J. and Björk, G.R. (1998). A new model for phenotypic suppression of frameshift mutations by mutant tRNAs. Mol. Cell 1: 471-82.

Ringquist, S., Shinedling, S., Barrick, D., Green, L., Binkley, J., Stormo, G.D. and Gold, L. (1992). Translation initiation in *Escherichia coli*: sequences within the ribosome-binding site. Mol. Microbiol., 6: 1219-29.

Riyasaty, S. and Atkins, J.F. (1968). External suppression of a frameshift mutant in *Salmonella*. J. Mol. Biol., 34: 541-57.

Roberts, L.M. and Reeve, E.C. (1970). Two mutations giving low level streptomycin resistance in *Escherichia coli K 12*. Genetic Res., 16: 359-65.

Rosenberg, A.H., Goldman, E., Dunn, J.J., Studier, F.W. and Zubay, G. (1993). Effects of consecutive AGC codons on translation in Escherichia coli, demonstrated with a versatile codon test system. J.Bacteriol., 175: 716-22.

Sanchez J.C., Padron G., Santana H. and Herrera L. (1998). Elimination of an $_{HU}$IFN 2b readthrough specifies produced in Escherichia coli, by replacing its natural translation stop signal. J.biotechnol., 63: 179-86.

Santos, MAS. and Tuite, M.F. (1993). New insights into mRNA decoding-implications for heterologous protein synthesis, Trends Biotechnol. 11: 500-5.

Schimmel, P. (1989). Hazards of deducing enzyme structure-activity relationships on the basis of chemical applications of molecular biology. Acc.Chem.Res., 22: 232-233.

Schwartz, R. and Curran, J.F. (1997). Analyses of frameshifting at UUU-pyrimidine sites. Nucl. Acids Res., 25: 2005-11.

Shine, J. and Dalgarno, L. (1974). The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc Natl Acad Sci USA 1974 71:1342-6.

Scolnick, E., Tompkins, R., Caskey, T. and Nirenberg, M. (1968). Release factors differing in specificity for terminator codons. Proc Natl Acad Sci. USA. 61:768-74.

Scorer, CA. M.J. Carrier, R.F. (1991). Rosenberger, Amino acid misincorporation during high-level expression of mouse epidermal growth factor in Escherichia coli, Nucl. Acids Res., 19: 3511-16.

Seetharam, R., Heeren, R.A.,Wong, E.Y., Braford, S.R., Klein, B.K. et al. (1988). Mistranslation in IGF-1 during over-expression of the protein in *Escherichia coli* using a synthetic gene containing low frequency codons. Biochem. Biophys. Res. Commun., 155, 518–23.

Sipley J. and Goldman E. (1993). Increased ribosomal accuracy increases a programmed translational frameshift in Escherichia coli. Proc Natl Acad Sci. USA. 90: 2315-9.

Sørensen, H.P., Laursen, B.S., Mortensen, K.K. and Sperling-Petersen, H.U. (2002). Bacterial translation initiation–mechanism and regulation. Recent Res. Dev. Biophys. Biochem., 2: 243-70.

Sørensen, H.P., Sperling-Petersen, H.U. and Mortensen, K.K. (2003a). A favorable solubility partner for the recombinant expression of streptavidin. Protein Expr. Purif., 32, 252-59.

Sørensen, H.P., Sperling-Petersen, H.U. and Mortensen, K.K. (2003b). Dialysis

strategies for protein refolding. Preparative streptavidin production. Protein Expr. Purif., 32, 252-59.

Spanjaard, S.A. and Van Duin, J. (1988). Translation of the sequence AGG-AGG yields 50% ribosomal frameshift, Proc. Natl. Acad. Sci. USA 85: 7967-71.

Spanjaard, R.A., Chen, K., Walker, J.R. and Van Duin, J. (1990). Frameshift suppression at tandem AGA and AGC codons by cloned tRNA genes: assigning a codon to argU tRNA and T4 tRNAarg. Nucl.Acids Res.,18: 5031-36.

Sprengart, ML. and Porter, AG. (1997). Functional importance of RNA interactions in selection of translation initiation codons. Mol Microbiol., 24: 19-28.

Stadtman, E.R. (1992). Protein oxidation and aging. Science 257: 1220-24.

Stenstrom, C., Jin, H., Major, L., Tate, W. and Isaksson, L.A. (2001). Codon bias at the 3'-side of the initiation codon is correlated with translation initiation efficiency in *Escherichia coli*. Gene 263: 273-84.

Taddei, F., Hayakawa, H., Bouton, M.F., Cirinesi, A.M., Matic, I., Sekiguchi, M. and Radman, M. (1997). Counteraction by MutT protein of transcriptional errors caused by oxidative damage. Science 278: 128-30.

Thomas, M.J., Platas, A.A. and Hawley, D.K. (1998). Transcriptional fidelity and proofreading by RNA polymerase II. Cell 93: 627-37.

Thompson, R.C. (1988). EFTu provides an internal kinetic standard for translational accuracy. Trends Biochem. Sci., 13: 91-93.

Tsuchihashi, Z. and Brown, P.O. (1992). Sequence requirements for efficient translational frameshifting in the Escherichia coli dnaX gene and the role of an unstable interaction between tRNA(Lys) and an AAG lysine codon. Genes Dev., 6: 511-9.

Valente, C.A., Prazeres, D.M.F., Cabral, J.M.S. and Monteriro, G.A. (2004). Translation feature of human alpha 2b interferon production Escherichia coli. Appied Envirmnetal Microbiol., 70: 5033-36.

van Leeuwen, F.W., Burbach, J.P., and Hol, E.M. (1998a). Mutations in RNA: A first example of molecular misreading in Alzheimer's disease. Trends Neurosci., 21: 331-5.

van Leeuwen, F.W., de Kleijn, D.P., van den Hurk, H.H., Neubauer, A. et al. (1998b). Frameshift mutants of beta amyloid precursor protein and ubiquitin-B in Alzheimer's and Down patients. Science 279: 242-7.

Wada, K.-N., Wada, Y., Ishibashi, F., Gojobori, T and Ikemura, T. (1992). Codon usage tabulated from the GenBank genetic sequence data. Nucl. Acids Res., 20, Supplement: 2111-118.

Weiss, RB., Dunn, DM., Atkins, JF. and Gesteland, RF. (1987). Slippery runs, shifty stops, backward steps, and forward hops: -2, -1, +1, +2, +5, and +6 ribosomal frameshifting. Cold Spring Harb Symp Quant Biol., 52: 687-93.

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

Weiss, RB., Dunn, DM., Shuh, M., Atkins, JF. and Gesteland RF. (1989). *E. coli* ribosomes re-phase on retroviral frameshift signals at rates ranging from 2 to 50 percent. New Biologest. 2:159-69.

Yarian, C., Marszalek, M., Sochacka, E., Malkiewicz, A., Guenther, R., Miskiewicz, A. and Agris, P.F. (2000). Modified nucleoside dependent Watson-Crick and

wobble codon binding by tRNALysUUU species. Biochemistry 39: 13390-95.

Zhang, J., Perry, G., Smith, M.A., Robertson, D., Olson, S.J.,Graham, D.G., and Montine, T.J. (1999). Parkinson's disease is associated with oxidative damage to cytoplasmic DNA and RNA in *substantia nigra* neurons. Am. J. Pathol., 154: 1423-29.

Zhang, S., Zubay, G. and Goldman E. (1991). Low-usage codons in Escherichia coli, yeast, fruit fly, and primates. Gene 105: 61-72.

المستخلص العربى

التتابع الجينى الأمثل للحصول على افضل بروتين: متطلبات أساسية

الراشدى مصطفى رضوان

قسم بحوث البروتين، معهد بحوث الهندسة الوراثية والتكنولوجيا الحيوية

مدينة مبارك للابحاث العلمية والتطبيقات التكنولوجية- برج العرب 21934- الاسكندرية، مصر.

يظل إنتاج البروتين المهندس وراثيا بكميات كبيرة مطلبا مهما للاكتشافات الدوائية وتطوير الصناعات، وكذلك لأبحاث التركيب البلوري ثلاثي الأبعاد. وتمثل الخلايا البكتيرية المعروفة "اشرشيا كولاى" أحد أهم الكائنات المستخدمة لهذا الغرض لما لها من مميزات فريده مثل: سرعة النمو وسهولة التعامل معها ومعرفة الجينوم الخاص بها، حيث يمكن الحصول على بروتينات من هذه الخلايا تصل الى 30% من محتواها البروتينى فى الحاله العاديه وحتى 65% بعد حثها كيميائيا. وبالرغم من هذه المميزات الفريدة إلا انه يمكن ان تتعرض هذه الكائنات لمخاطر عديدة نتيجه لإدخال جينات ذات شفرات وراثية غريبة عليها، أو غير مصممة بشكل جيد لتتوافق مع جينومها. وهذا يؤدى الى أخطاء عديدة فى البروتين المنتج منها. من المخاطر التى تتعرض لها البكتيريا: إدخال جينات بها نسب مرتفعه او مجموعات من الكودونات (الشفرات) الغريبة عليها، وفى حالة حدوث ذلك فان المنتج البروتينى، أيما كان تجريبى او علاجى، فانه يحتوى على تغيير فى متواليتة ومن ثم التركيب الفراغي ثلاثى الابعاد والذى يؤثر بدوره على وظيفته.

يجب اخذ الاحتياطات الاتية أثناء تصميم الجين المراد كلونتة وتحويله لبروتين: 1- يجب تجنب احتواء الجين على شفرات وراثية نادرة فى الخلايا البكتيرية، 2- يجب تجنب المحتوى العالى (80%) او المحتوى الأدنى (اقل من 30%) من القواعد GC فى تركيبة الجين، 3- يجب الابتعاد عن (حزم المتواليات المتجمعة TATA- وحزم كاى- التكرار الداخلي لمستقبلات الريبوزوم- متواليات القواعد GC-rich or AT-rich- تكرار المتواليا ت- والتراكيب الثانية للحامض النووىRNA). تجنب استخدام شفرلت إيقاف (arg/UAG, arg/UGA) وخاصة المسبوقة بشفرة الحمض الامينى آرجينين، 4- استخدام شفره ايقاف UAA لما لها من مميزات.