J. theor. Biol. (1987) **124**, 43-55

Translation Rate Modification By Preferential Codon Usage: Intragenic Position Effects

HANS LILJENSTRÖM AND GUNNAR VON HEIJNE

Research Group for Theoretical Biophysics, Department of Theoretical Physics, Royal Institute of Technology, S-100 44 Stockholm, Sweden

(*Received* 7 *December* 1985, *and in revised form* 24 *July* 1986)

We present a model for calculating the protein production rate as a function of the translation rate. The model takes into account that the elongation rate along an mRNA molecule is non-uniform as a result of different tRNA availabilities for different codons. Initiation of ribosomes on an mRNA is normally the rate-limiting step in the translation process, and blocking of the initiation site can be avoided if the codons closest to this site allow fast translation by the ribosome. Hence, different selective forces may act on the choice of synonymous codons in the initiation region than elsewhere on a given mRNA. We show that the elongation rate along the whole mRNA influences the production rate of abundant proteins, whereas only the elongation rate in the initiation region is of importance for the production rate of rare proteins. We also present an analysis of the codon distribution along known mRNAs coding for abundant and rare proteins.

1. Introduction

A great deal of attention has been paid lately to the question of preferential codon usage. It is now a well established fact that different proteins are coded for by messenger RNAs using different sets of codons. Highly expressed genes preferentially use a subset of all codons, whereas weakly expressed genes tend to use the whole set (Grantham *et al.*, 1981; Gouy & Gautier, 1982; Konigsberg & Godson, 1983; Blake & Hinds, 1984). The codons used in the highly expressed genes correspond to the most abundant isoacceptor tRNAs, thus asserting the quickest possible translation of their messengers. For weakly expressed genes there is even some evidence of preferential usage of codons with an inverse correlation to tRNA abundance. One reason for this may be that there is a modulation of gene expression at the translational level, in addition to regulation at the transcriptional level (through the amount of messengers), as has been proposed by e.g. Fiers & Grosjean (1979).

In this work we discuss codon usage with respect to the efficiency of protein synthesis. In so doing, we focus on the mean time spent by a ribosome at each codon and the effect of the codons' position in the mRNA, rather than looking at aspects concerning, e.g., nucleotide content (Bibb *et al.*, 1984, Perrin, 1984), codon-anticodon binding strength (Grosjean & Fiers, 1982), mRNA secondary structure (Shpaer, 1985), different mutation rates (Golding & Strobeck, 1982), or contextual constraints (Lipman & Wilbur, 1983; Yarus & Folley, 1985). The idea of different elongation rates for different codons rests on the notion of codon adaptation, which

0022-5193/87/010043 + 13$03.00/0

has been discussed thoroughly in previous works (Garel, 1974; von Heijne & Blomberg, 1979; Ikemura, 1981*a b*). Presumably, the more tRNA there is of a certain kind the shorter is the waiting time at the corresponding codon. The optimal correlation between tRNA levels and codon frequencies is dependent on several factors in the tRNA cycle, such as the relative degree of saturation of the ribosomes and of the synthetases (Liljenström *et al.*, 1985). Experimentally, it has been observed that the translation rate is non-uniform, most likely as a result of variations in tRNA availability (Varene *et al.*, 1984; Pedersen, 1984).

Several models of protein synthesis have been developed in order to describe various aspects of the process (e.g. MacDonald *et al.*, 1968; Gordon, 1969; Bergmann & Lodish, 1979). Most of these models deal with the problem of how fast ribosomes can move across the messenger. The different steps in translation (i.e. initiation, elongation and termination), as well as problems like queuing, have been extensively studied.

Although it has been recognized by some authors (e.g. Bergmann & Lodish, 1979) that the initiation step is rate limiting in normal situations (with no queuing), it is surprising that the consequences of this fact do not seem to have been fully realized. We now present a simple model of the translation system that is based on this fact, and deal with some of its consequences.

First, it has to be appreciated that it is only the frequency of initiations that determines the rate of protein production (as long as queueing of ribosomes does not block the initiation region). That is to say, in steady-state, for each initiation there will also be a termination and consequently a completion of one polypeptide. This suggests two things: (1) the amount of free ribosomes (and initiation factors) is crucial for the overall rate of production, and (2) there may be different selective forces affecting the codon composition in the region close to the initiation site than elsewhere. These assumptions form the basis of our study.

In the model presented below, each mRNA is regarded as consisting of two regions, an initiation region and a major region. The initiation region corresponds to the number of codons a ribosome has to translate in order to leave place for a new ribosome to initiate (approximately equal to the number of codons that are covered by a ribosome). By studying the effect of decreasing or increasing the elongation rate in the two regions, we show that, for abundant proteins, the elongation rate in both the initiation region and the major region is important for the production of the corresponding protein as well as for the total protein production of the whole cell. For rare proteins, however, only the elongation rate of the initiation region is found to be of major importance.

We have also made an analysis of the codon bias along known mRNAs coding for rare and abundant proteins compared to randomized controls. The results show that there is a progressive reduction in the mean elongation rate along mRNAs coding for rare proteins.

2. Theory

To illustrate the effects of codon composition in various parts of the mRNA, we use a simple model in which we regard the translation system as consisting of ribosomes, messenger RNA and acylated transfer RNA (aa-tRNA). All other molecules or molecular complexes (such as initiation factors, elongation factors, synthetases etc.) are regarded as being present in excess and thus not limiting. We can reduce the problem even further by letting the aa-tRNA molecules enter only indirectly, through the time spent by a ribosome at the various codons in the messengers (the step time). Thus, we are left with ribosomes and messengers as the basic molecules.

The ribosomes can be divided into two fractions, free and bound. The free ribosomes (actually consisting of dissociated small and large subunits, bound to the appropriate initiation factors) here represent the activated complexes that can attach to a messenger. The bound ribosomes are those translating a messenger. Free ribosomes can attach to the initiation site of an mRNA molecule if there is not already a ribosome blocking this site. In steady-state it is the number of initiations per unit time that determines the total protein production rate. Hence, the more free ribosomes, the more initiations per unit time, and the more products made. The problem is thus to find an expression for the time interval between two initiations, and an expression for the amount of free ribosomes available.

If we denote the total number of ribosomes in a cell by R_{tot} and the total number of mRNA molecules by m, we have

$$R_{tot} = mR_b + R_f, \tag{1}$$

where R_b denotes the (average) number of bound ribosomes per messenger (i.e. the polysome size), and R_f denotes the number of free ribosomes.

The time interval between two initiations on a given mRNA is the sum of the time it takes for a free ribosome to bind to the messenger and the time it takes for the ribosome to move away from the initiation region. With the rate constant for the first process denoted by k_I, we calculate the time for that process as

$$t_{I1} = (k_I \cdot R_f)^{-1} \tag{2}$$

and the time for the latter process as

$$t_{I2} = \sum_{j=1}^{L} t_j \tag{3}$$

where t_j is the step time at codon j (counting from the initiation site). This time also includes the translocation and transpeptidation times, which we assume to be constant and negligible for this problem. The upper summation limit, L, corresponds to the number of codons that constitute the initiation region, i.e. the number of codons the ribosome has to translate in order to leave room for a new initiation (this number is generally estimated to be somewhere between 10 and 20). The total time for translating the whole mRNA is

$$t_S = \sum_{j=1}^{S} t_j = \sum_{j=1}^{S} 1/v_j \tag{4}$$

where S is the total number of codons in the coding sequence, and the elongation rate $v_j = 1/t_j$.

Now, if the protein production rate (i.e. the number of proteins made per unit time) is to be, say, increased, the initiation time, $t_I = t_{I1} + t_{I2}$, has to be shortened. This may be accomplished either by increasing the number of free ribosomes, or by increasing the elongation rates, v_j, for the first L codons. (A third possibility would be to vary the ribosomal binding strength, i.e. to change the rate constant k_I, but since that is apparently not dependent on codon usage, we will not treat this possibility here). The number of free ribosomes increases the faster the ribosomes traverse the messengers and leave the bound state, and therefore a higher mean elongation rate for the whole mRNA leads to an increase in R_f. Whereas changing the initiation time of a certain messenger only has a major effect for the production rate of that particular protein, an increase in the total mean elongation rate for that protein favours the production of *all* proteins in the cell through the increase of free ribosomes. These simple observations suggest that there may be different selective forces acting on the initiation region and on the rest of the messenger (major region). This could presumably result in different mean elongation rates for the two regions.

To demonstrate these effects quantitatively, let v_I denote the mean elongation rate in the initiation region, and v the mean rate in the major region. Asssuming $1/t_I < v/L$ (i.e. no queueing), we then have

$$t_{I2} = \sum_{j=1}^{L} v_j^{-1} = L/v_I \quad \text{and} \quad t_S = \sum_{j=1}^{S} v_j^{-1} = L/v_I + (S-L)/v. \qquad (5)$$

The average number of ribosomes on the mRNA (i.e. the average polysome size) can now be calculated as the transition time divided by the initiation time

$$\begin{aligned} R_b = t_S/t_I &= [L/v_I + (S-L)/v]/[L/v_I + 1/k_I R_f] \\ &= [1 + (v_I/v)(S-L)/L]/[1 + v_I/k_I R_f L] \end{aligned} \qquad (6)$$

where R_f is given by (1).

If we look at the whole cell and assume that all messengers have the same mean initiation time, t_I, the total amount of protein made in the cell per unit time is

$$P = m/t_I = [v_I k_I m R_f]/[v_I + k_I R_f L]. \qquad (7)$$

R_f is calculated from equations (1) and (6)

$$R_f = A/2 + (A^2/4 + R_{tot} v_I/k_I L)^{1/2} \qquad (8)$$

where

$$A = R_{tot} - v_I/k_I L - m(1 + v_I(S-L)/vL). \qquad (9)$$

The simplest case where we can study the effect of preferential codon usage, is to make a "model cell" with only two mRNAs, one abundant and one rare. We can then study the effect of changing the elongation rate in either the initiation region or in the major region of the messengers.

Consider a system where one particular mRNA (which we call α) constitutes a fraction f of all mRNAs in the cell, i.e. $m_\alpha = f \cdot m$. The other mRNA (β) then constitute $(1-f)$ of all mRNAs, i.e. $m_\beta=(1-f)m$. Accordingly, we assign the rates and rate constants $v_{I\alpha}$, v_α, $k_{I\alpha}$, and $v_{I\beta}$, v_β, $k_{I\beta}$ to the α and β species, respectively. Thus, for the calculation of the amount of free ribosomes, R_f, we use the expressions for the average polysome sizes for each species (from eqn (6))

$$R_{b\alpha}=[1+(v_{I\alpha}/v_\alpha)(S-L)/L]/[1+v_{I\alpha}/k_{I\alpha}R_fL] \tag{10}$$

$$R_{b\beta}=[1+(v_{I\beta}/v_\beta)(S-L)/L]/[1+v_{I\beta}/k_{I\beta}R_fL] \tag{11}$$

which yield a cubic equation for R_f

$$R_f^3+(\gamma_\alpha+\gamma_\beta+m_\alpha\delta_\alpha+m_\beta\delta_\beta-R_{tot})R_f^2$$
$$+(\gamma_\alpha\gamma_\beta+m_\alpha\delta_\alpha\gamma_\beta+m_\beta\delta_\beta\gamma_\alpha-(\gamma_\alpha+\gamma_\beta)R_{tot})R_f-\gamma_\alpha\gamma_\beta R_{tot}=0 \tag{12}$$

where

$$\gamma_\alpha=v_{I\alpha}/k_{I\alpha}L \tag{13}$$

$$\gamma_\beta=v_{I\beta}/k_{I\beta}L \tag{14}$$

$$\delta_\alpha=1+(v_{I\alpha}/v_\alpha)(S-L)/L \tag{15}$$

$$\delta_\beta=1+(v_{I\beta}/v_\beta)(S-L)/L. \tag{16}$$

Equation (12) is solved in an ordinary manner, and the protein production rates for each mRNA species is calculated from

$$P_\alpha=[v_{I\alpha}k_{I\alpha}m_\alpha R_f]/[v_{I\alpha}+k_{I\alpha}R_fL] \tag{17}$$

$$P_\beta=[v_{I\beta}k_{I\beta}m_\beta R_f]/[v_{I\beta}+k_{I\beta}R_fL]. \tag{18}$$

These expressions allow one to calculate the dependence of the protein production rate on codon usage in different parts of the messenger, provided that the step time for each codon is known. The model allows for queuing of ribosomes in the major region as long as this does not disturb initiation (i.e. as long as $1/t_I<v/L$).

3. Results

(A) MODEL CALCULATIONS

Let us first look at what happens to the total protein production rate as well as to the individual production rates of the two mRNAs when we change, in turn, the mean elongation rates $v_{I\alpha}$, $v_{I\beta}$, v_α and v_β (with all other parameters kept constant). For clarity, we may choose an extreme case with $f=0{\cdot}99$ (i.e. 99% of the mRNA codes for protein α, and 1% codes for protein β). The reference parameters chosen for this model system are of the same magnitudes as in an average *E. coli* cell: total number of ribosomes, $R_{tot}=10\,000$; total number of mRNAs, $m=800$; number of codons in each mRNA, $S=300$; number of codons in the initiation region, $L=15$; and mean elongation rates $v_{I\alpha}=v_{I\beta}=v_\alpha=v_\beta=20$ amino acids/second. The initiation rate constants are chosen so that about 85% of the ribosomes are bound, i.e. $k_{I\alpha}=k_{I\beta}=0{\cdot}001$ (Gouy & Grantham, 1980; Ingraham *et al.*, 1983).

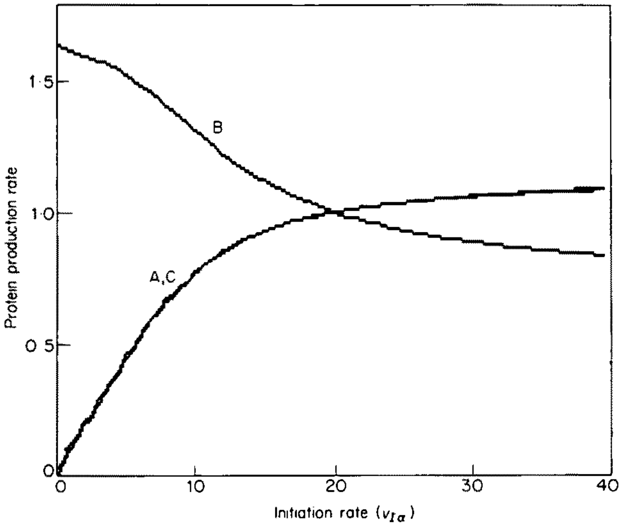

FIG. 1. Protein production rate dependence on the elongation rate in the initiation region of an *abundant* mRNA. Curve A shows the production rate, P_α, of an abundant protein, whose mRNA constitutes 99% of the total mRNA in the cell. It is given as a function of the mean elongation rate, $v_{I\alpha}$, for the first 15 codons in this mRNA. Curve B shows how the production rate, P_β, of a rare protein (whose mRNA constitutes 1% of the total mRNAs) is affected by a variation in $v_{I\alpha}$. Curve C, which almost coincides with curve A, gives the total protein production rate in the cell ($P = P_\alpha + P_\beta$) as a function of $v_{I\alpha}$. All protein production rates are normalized, so that the values of P_α, P_β, and P are *one* when all rates are the same ($=20$). (P_α and P_β are given by eqns (17) and (18).)

(i) In Fig. 1 we have plotted the normalized (see legend) protein production rates P_α, P_β, and P ($=P_\alpha + P_\beta$) versus $v_{I\alpha}$, with $f = 0{\cdot}99$. P_α and P increase and P_β decreases as $v_{I\alpha}$ increases from 0 to 40 (i.e. from zero to twice the reference value). P_β decreases because more ribosomes become bound to the α-messengers, leaving fewer free ribosomes for initiation on the β-messengers.

(ii) Increasing $v_{I\beta}$, on the other hand, has a major effect only on P_β, since the β-messengers bind a negligible part of the free ribosomes, and the protein production for the α-messengers remain almost the same. This case is illustrated in Fig. 2.

(iii) If v_α is increased, both protein production rates P_α and P_β increase, since the pool of free ribosomes grows due to the faster translation of ribosomes on the α-messengers (Fig. 3.)

(iv) Consequently, increasing v_β only marginally affects the protein production rate of both messengers. (The protein production rate is only indirectly dependent on v_α and v_β, through the amount of free ribosomes, R_f).

These results show that it may be meaningful for the cell to optimize the elongation rate along the whole messenger for abundant proteins, whereas the production rate of rare proteins can be regulated primarily through the elongation rate of the initation region. (Often, regulation is accomplished through variations in k_I, but since this is not dependent on codon composition we do not treat this possibility here.)

Also, since one single mutation may change the step time for a particular codon by as much as a factor of ~10, due to differences in the concentrations of iso-acceptor

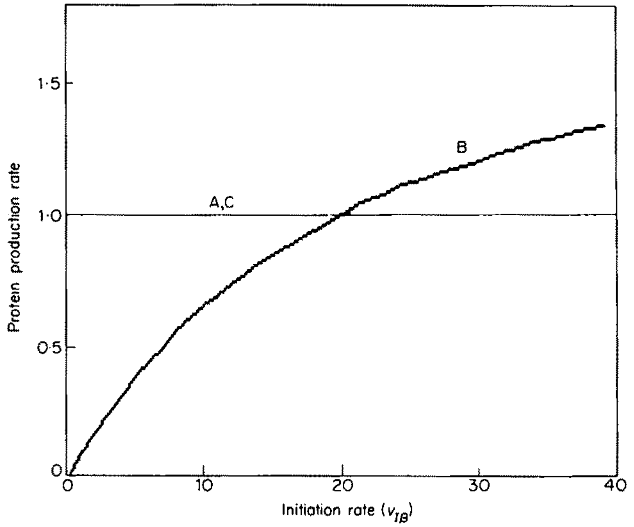

FIG. 2. Protein production rate dependence on the elongation rate in the initiation region of a *rare* mRNA. Curve B shows the (normalized) production rate of the rare protein, P_β, as a function of the mean elongation rate, $v_{I\beta}$, for the first 15 codons of the rare mRNA. Curves A and C, which coincide, show the (normalized) production rates P_α and P, respectively, as a function of $v_{I\beta}$.

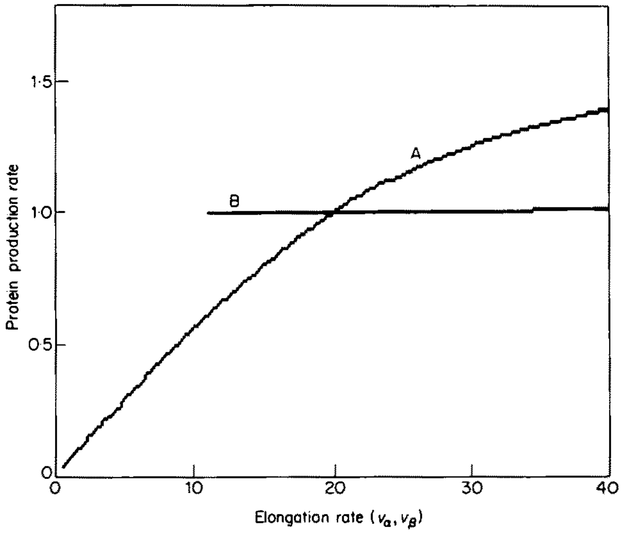

FIG. 3. Protein production rate dependence on the elongation rate in the major region (see text) of a *rare* and an *abundant* mRNA. Curve A shows the (normalized) production rates of both the rare *and* abundant proteins (P_α, P_β, and P) as a function of the mean elongation rate, v_α, in the major region (codons 16 to 300) of the *abundant* mRNA. Curve B shows the same production rates as a function of the mean elongation rate, v_β, in the major region of the *rare* mRNA. This latter curve starts at $v_\beta = 12$, since queueing of ribosomes occur at the rare mRNA below this value. (All parameters are the same as in Fig. 1.)

tRNAs (Varenne *et al.*, 1984), and since the initiation region is only some 15 codons long, a mutation that changes the elongation rate of a codon in this region will have a substantial effect on v_l, whereas a similar mutation in the major region (typically a couple of hundred codons long) will only affect v marginally. Thus, v_l should be quite sensitive even to single point mutations.

(B) CODON USAGE IN DIFFERENTIALLY EXPRESSED MRNAS

In order to check whether codon usage is different in the initiation region and the major region of mRNAs, we have analyzed sequences coding for abundant *E. coli* proteins (12 proteins, $\gtrsim$ 10 000 molecules per cell) and for rare ones (13 proteins, $\lesssim$ 100 molecules per cell). In the former group we find e.g. ribosomal proteins; in the latter group we find, e.g., repressors—see Table 1.

TABLE 1

E. Coli mRNA sequences for rare and abundant proteins

Rare		Abundant	
Gene	No. Codons	Gene	No. Codons
ampC	378	lpp	79
araC	317	ompA	347
dnaG	580	recA	354
ecoRI	278	rplA	235
m.ecoRI	327	rplJ	166
eltA	254	rplK	143
galR	343	rplL	122
lacI	360	rpsG	81
lexA	202	rpsI	88
lysA	312	rpsL	125
trpR	108	rpsU	72
tn3 repressor	186	tufB	395
tn9CAT	220		

Sequences are from Genbank (version 32·0) and EMBL (version 5·0).

In the analysis we use values for the step times as derived from data on codon frequencies and tRNA abundance by Varenne *et al.* (1984). We calculate the mean step time in a window of fixed length, which is allowed to slide along the sequence, moving one codon at a time. The mean elongation rate for the codons in the window is the reciprocal value of the mean step time. This procedure is repeated for each of the selected sequences. Average translation-rate profiles for the mRNAs coding for rare and abundant proteins, respectively, are obtained by adding the values for the sequences in either group at each window position and dividing by the number of sequences in the group.

The window length ought to coincide with the number of codons covered by the ribosome. This number should be between 10 and 20; we use 15 throughout. However, the results of the sequence analysis do not change appreciably if the window length is varied between 10 and 20.

Analyzing the first 270 codons (only three abundant proteins and eight rare proteins are longer) for each protein in the two groups gave the following results:

(1) Both the abundant *and* the rare proteins are translated with a rate higher than that obtained for a random choice of synonymous codons (weighted by the average amino acid usage in *E. coli* (Blake & Hinds, 1984)), see Fig. 4. That is, even the mRNAs coding for rare proteins have a selective codon usage, though much weaker than is the case for the mRNAs coding for abundant proteins. Individual mRNAs coding for rare proteins may of course have elongation rate values that lie below the "random choice" value at certain window positions, the lowest observed value being 1·1 on our scale. The mean elongation rate for the abundant proteins, is, on the average, 1·64 times higher than for the rare proteins.

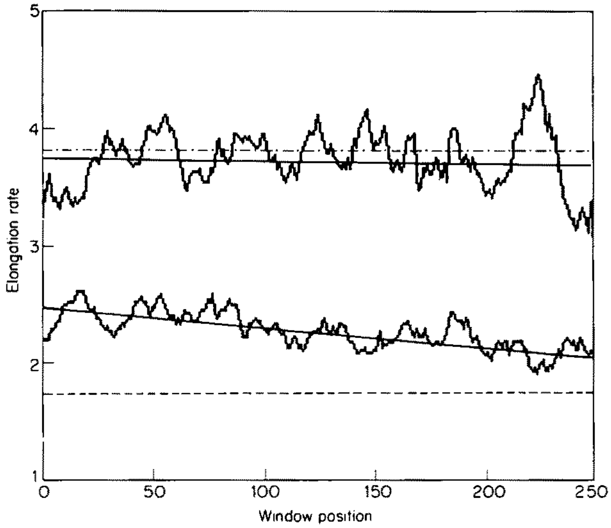

FIG. 4. Elongation rate variation along mRNAs coding for rare and abundant proteins. The upper curve is the average for 12 abundant protein mRNAs and the lower curve is the average for 13 rare protein mRNAs (see table 1). The value of the elongation rate at each window position is the reciprocal of the mean step time for 15 contiguous codons, as calculated from data on codon frequencies and tRNA abundance. A least square approximation to each curve is also indicated. The dot-dashed line (- · - · -) is obtained when optimal codons are selected for an average *E. coli* amino acid composition. The dashed line (- - - - -) is obtained for a random choice of synonymous codons (weighted by the average amino acid usage in *E. coli*).

(2) The abundant proteins are translated at a rate close to the maximum rate possible in *E. coli* (indicated by the dot-dashed line in Fig. 4) when the average amino acid usage is taken into account. In several regions the mean rate is even higher than the rate obtained when the optimal synonymous codons are selected, reflecting the use of particularly abundant amino acids coded by "fast" codons in these regions.

(3) The elongation rate is highly variable along the mRNAs for both the rare and the abundant proteins. The difference in elongation rate for two adjacent codons

may often be 10-fold in the rare protein mRNAs, whereas this factor is normally less than five, and typically less than three in the abundant protein mRNAs. (Since the latter avoid rare codons, this reduced variability comes as no surprise.)

(4) There is a striking feature of the mRNAs coding for rare proteins that is not seen in the mRNAs coding for abundant proteins: the elongation rate tends to be higher in the beginning than later in the mRNA sequences (11 out of 13 cases). There is an 18% drop in the mean elongation rate (average of 13 sequences) over the first 270 codons for the rare protein mRNAs, whereas the least square approximation for the abundant protein mRNAs, averaged over 12 individual mRNA sequences, coincides almost exactly with the mean elongation rate for these sequences, see Fig. 4.

To assess the significance of this feature we generated 100 sets of randomized sequences, with 13 sequences in each set, and each sequence having the same codon composition as the corresponding mRNA sequence coding for one of the rare proteins but with the codons randomly shuffled. None of the 100 randomized sets had such a large slope as the set of rare protein mRNAs (±18%; the largest slope found for any of the randomized sets being 13%). Thus, the general trend seems to be significant.

An analysis of the heights of the individual peaks relative to the linear approximation (Fig. 4) and the corresponding peak-distribution for the 100 randomized sets did not reveal any unexpectedly large departures, either for rare or abundant mRNAs. Using spectral analysis we also looked for significant periodicities in the observed elongation rate pattern, as compared with the randomized sequences, but none was found.

(C) PERTINENT EXPERIMENTAL RESULTS

There is as yet little direct experimental evidence for the production rate of a given protein being modulated by preferential codon usage. However, although not a direct effect of selective codon usage, a stable hairpin loop in the very beginning of an mRNA has been shown to interfere with initiation and thus to reduce the protein production rate of yeast cytochrome c (Baim *et al.*, 1985). A mutation in codons 6 and 7 of the yeast CYC1 mRNA apparently produces a stable hairpin structure that decreases the production of the corresponding protein to 20% of the normal level. It is clear that initiation is affected, since the average polysome size drops from 4–7 ribosomes/mRNA to 1–4 ribosomes/mRNA. A similar hairpin loop in the major region of the mRNA would presumably not result in a corresponding reduction in protein synthesis. To the extent that this hairpin loop affects initiation by slowing down ribosome movement, its effect on the protein production rate should be similar to that of one or more hungry codons at the same position on the mRNA.

The effects of codon usage *per se* have so far only been anayzed in positions away from the initiation region. Chloramphenicol acetyltransferase (CAT), which has 25% rare codons and yet can be expressed at high levels, is a good case in point. It has been shown that codon usage can affect the production rate of this protein, but only at very high levels of expression (Robinson *et al.*, 1984). Four common codons coding for arginine (CGT) and four rare codons (AGG) coding for the same amino acid were inserted into the EcoRI site in the CAT genes of two different plasmids. A difference in gene expression could be seen only at very high transcription rates. The EcoRI site of this gene is at nucleotide 213 (codon 71) from the start of the coding sequence and thus not in the initiation region (Alton & Vapnek, 1979). According to our model, using different synonymous codons in this area of the gene (major region) will not affect the protein production rate until there is a very high amount of messengers for this protein, or if queueing back to the initiation region occurs. A corresponding substitution of synonymous codons in the initiation region would presumably have a greater effect on the expression of the gene.

Further, ribosome pausing has been reported to occur at a rate Arg codon (AGA) at codon position 402 of the tolC mRNA (Misra & Reeves, 1985). By increasing the amount of the corresponding tRNA, this pause can be removed. However, there is no apparent increase in the total amount of tolC made (although no quantification is given, this seems to be the case from an examination of the gel photographs), which is in keeping with the ideas presented above. Here too, it would be interesting to look at the effects of ribosome pausing very early in the sequence.

It has also been suggested that the use of rare codons in the dnaG gene causes a specific drop in the protein production rate (this gene is flanked by the highly expressed genes rpoD and rpsU on the same mRNA (Konigsberg & Godson, 1983)). The elongation rate pattern of the dnaG gene does not, however, support this idea. On the contrary, the calculated elongation rate is exceptionally high in the beginning of the coding sequence (not shown); thus, the low dnaG protein production rate is most likely *only* a consequence of an exceptionally poor ribosome binding-site.

4. Discussion

In this paper we have pointed to the crucial role of the initiation region in determining the rate of protein synthesis. In this context, we have looked at the possible modulating effects of preferential codon usage in different regions of the mRNA, and have shown that it makes sense for the cell to optimize the elongation rate along the whole mRNA for abundant proteins. This can be done by selecting the synonymous codons that match the most abundant isoacceptor tRNAs. Messengers coding for rare proteins, however, use the whole set of codons and in some cases preferentially use codons corresponding to rare tRNAs. It has been suggested repeatedly that this may have a regulatory purpose. However, for rare proteins only the initiation region (the first 10–20 codons) should be of major importance for regulating the protein production rate, the following 10–20 codons having a weaker effect (Bergmann & Lodish, 1979).

In the sequence analysis of mRNAs coding for rare proteins, it was found that the overall pattern is a relatively high rate of elongation in the beginning of the sequences and a subsequent linear decrease of the rate towards the end. A more detailed analysis of the mean elongation rate-profile for this limited set of mRNAs does not, at present, reveal any further significant features. As there is no abrupt change in the elongation rate in the region between codons 10 and 20, there does not seem to be a general tendency among rare mRNAs to maximize their production rate by specifically enhancing the rate of ribosome movement in the initiation region. Beyond this, the reason(s) for the observed linear drop in the mean elongation rate is not obvious from the considerations in this paper. However, a recent study of simulated mRNA sequences, with more detailed elongation rate profiles, shows that a linear decrease in elongation rate can be more profitable than a sudden drop right after the initiation region. This is simply because the queuing back to the initiation region may be avoided that way; the further away from this region "slow" codons occur, the lesser risk for interfering with initiation (work in preparation).

Clearly, non-uniform translation rates as a result of biased codon usage may have biological consequences aside from its effect on the protein production rate. Thus, since translation and transcription appear to be coupled (Varenne *et al.*, 1984; Fisher *et al.*, 1985; Bonekamp *et al.*, 1985), the variation in translation rate may not, in itself, be of primary importance, but simply an adaptation to a non-uniform transcription rate. Even if transcription proceeds at a uniform rate, the variation in translation rate may just be a consequence of a general slow-down to avoid collision between ribosomes and polymerases. Codon composition may also influence processes such as proofreading (Kurland & Ehrenberg, 1984). In these cases, there would be selective pressures not only on the initiation region, but also on the rest of the coding sequence. Our results cannot at this point distinguish between these possibilities.

To test the importance of the ideas put forth in this work, it would be interesting to see what effects a synonymous codon substitution (e.g. the Arg codon CGU for AGA) in the very beginning of a coding sequence would have on the expressivity of the gene. As discussed above, it has already been shown that the introduction of an exceptionally stable hairpin loop in this region drastically reduces the protein production rate (Baim *et al.*, 1985).

We wish to thank Dr C. Blomberg for critically reading the manuscript. This work was supported in part by a grant from the Swedish Natural Sciences Research Council (GvH).

REFERENCES

ALTON, N. P. & VAPNEK, D. (1979). *Nature* **282,** 864.

BAIM, S. B., PIETRAS, D. F., EUSTICE, D. C. & SHERMAN, F. (1985). *Mol. cell. Biol.* **5,** 1839.

BERGMAN, J. E. & LODISH, H. F. (1979). *J. biol. Chem.* **254,** 11927.

BIBB, M. J., FINDLAY, P. R. & JOHNSON, M. W. (1984). *Gene* **30,** 157.

BLAKE, R. D. & HINDS, P. W. (1984). *J. Biomol. Struct. Dyn.* **2,** 593.

BONEKAMP, F., ANDERSEN, H. D., CHRISTENSEN, T. & JENSEN, K. F. (1985). *Nucleic Acids Res.* **13,** 4113.

FIERS, W. & GROSJEAN, H. (1979). *Nature* **277,** 328.

FISHER, R. F., DAS, A., KOLTER, R., WINKLER, M. E. & YANOFSKY, C. (1985). *J. mol. Biol.* **182,** 397.

GAREL, J. P. (1974). *J. theor. Biol.* **43,** 211.

GOLDING, G. B. & STROBECK, C. (1982). *J. mol. Biol.* **18,** 379.

GORDON, R. (1969). *J. theor. Biol.* **22,** 515.

GOUY, M. & GAUTIER, C. (1982). *Nucleic Acids Res.* **10,** 7055.

GOUY, M. & GRANTHAM, R. (1980). *FEBS Lett.* **115,** 151.

GRANTHAM, R., GAUTIER, C., GOUY, M., JACOBZONE, M. & MERCIER, R. (1981). *Nucleic Acids Res.* **9,** r43.

GROSJEAN, H. & FIERS, W. (1982). *Gene* **18**, 199.

VON HEIJNE, G. & BLOMBERG, C. (1979). *J. theor. Biol.* **78**, 113.

IKEMURA, T. (1981*a*). *J. mol. Biol.* **146**, 1.

IKEMURA, T. (1981*b*). *J. mol. Biol.* **151**, 389.

INGRAHAM, J. L., MAALOE, O. & NEIDHARDT, F. C. (1983). *Growth of the Bacterial Cell.* Sunderland, Mass.: Sinauer Associates, Inc.

KONIGSBERG, W. & GODSON, G. N. (1983). *Proc natn. Acad. Sci. U.S.A.* **80**, 687.

KURLAND, C. G. & EHRENBERG, M. (1984). *Prog. Mol. Biol. Nucl. Acids Res.* **31**, 191.

LILJENSTRÖM, H., VON HEIJNE, G., BLOMBERG, C. & JOHANSSON, J. (1985). *Eur. Biophys. J.* **12**, 115.

LIPMAN, D. J. & WILBUR, W. J. (1983). *J. Mol. Biol.* **163**, 363.

MACDONALD, C. T., GIBBS, J. H. & PIPKIN, A. C. (1968). *Biopolymers* **6**, 1.

MISRA, R. & REEVES, P. (1985). *Eur. J. Biochem.* **152**, 151.

PEDERSEN, S. (1984). *EMBO J.* **3**, 2895.

PERRIN, P. (1984). *Nucleic Acids Res.* **12**, 5515.

ROBINSON, M., LILLEY, R., EMTAGE, J. S., YARRANTON, G., STEPHENS, P., MILLICAN, A., EATON, M. & HUMPHREYS, G. (1984). *Nucleic Acids Res.* **12**, 6663.

SHPAER, E. G. (1985). *Nucleic Acids Res.* **13**, 275.

VARENNE, S., BUC, J., LLOUBES, R. & LAZDUNSKI, C. (1984). *J. mol. Biol.* **180**, 549.

YARUS, M. & FOLLEY, L. S. (1985). *J. mol. Biol.* **182**, 529.