J. theor. Biol. (1987) **125**, 1–14

Mini Review

Theoretical Modelling of Protein Synthesis

GUNNAR VON HEIJNE, CLAS BLOMBERG AND HANS LILJENSTRÖM

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

This article provides an overview of the use of mathematical and computer modelling in furthering the understanding of protein synthesis. In particular, we discuss issues such as the nature of the rate limiting step(s), error rates, tRNA-codon adaptation, codon bias, attenuation control, and problems of selection and error corrections, focussing on their theoretical treatment.

Introduction

Protein synthesis is a process of sufficient complexity to make interpretation of experimental data difficult and unguided intuition a rather risky guide. Theoretical modelling, either in the form of mathematical derivation or computer simulation, is necessary in order to gain a full understanding of the workings and evolution of the mechanisms involved, and has been successfully used on many occasions. In this review, we will try to provide a reasonably complete survey of the various models that have been constructed so far and the main results they have provided.

When a mature mRNA is delivered to the protein-synthesizing machinery, it brings about an ordered succession of chemical reactions, leading ultimately to the production of a characteristic amount of protein per unit time. In eukaryotes, initiation requires the binding of an activated 40 S ribosomal subunit (including initiation factors and Met-tRNA) to an initiation region on the mRNA (possibly also complexed with one or more initiation factor), and subsequent addition of a 60 S ribosomal subunit. At each step in the ensuing translation process, the ribosome searches for a cognate EF-1 (EfTu in prokaryotes) bound aminoacyl-tRNA (aa-tRNA) to match the codon in the acceptor (A) site, then catalyzes transpeptidation of the nascent chain to the A-site, and finally moves the whole peptidyl-tRNA back into the peptidyl-tRNA (P) site in an EF-G dependent translocation step. During translation, both misincorporation of non-cognate aa-tRNAs and frameshifts due to codon–anticodon "slippage" are possible, and must be kept at sufficiently low levels. Finally, when the ribosome reaches a termination codon, termination and release of the finished protein is induced by a release factor, and the ribosome is dissociated into subunits by a dissociation factor.

From the point of view of a tRNA molecule, one cycle is composed of aminoacylation by the appropriate aminoacyl-tRNA synthetase, binding to EF-1 (EfTu in procaryotes), capture by a ribosomal A-site, translocation into the P-site, and release in free form as the nascent chain is transferred to the next incoming tRNA.

0022-5193/87/050001+14 $03.00/0

Obviously in such a complex series of events, the physiologically relevant characteristics of the process (e.g., the total rate of production of completed chains and the error frequency) depend in intricate ways on the primary events and the characteristics of the molecules involved. This is amply illustrated by the theoretical works described below: different steps become rate-limiting under different conditions; molecular details such as the mRNA secondary structure can influence the process in various ways; and the optimization of the overall design of the machinery in terms of, e.g., proofreading efficiency or production rate is by no means trivial.

We will concentrate on five main areas where modelling has been particularly important: (i) "experimental simulations" with the aim of facilitating the interpretation of *in vitro* experiments and identifying the rate-limiting steps; (ii) adaptation between tRNA levels and codon frequencies and the effects of biased utilization of codons; (iii) influences from mRNA secondary structure; (iv) attenuation control; and (v) problems related to selection and error correction, including proofreading, error propagation and accuracy economy.

1. Experimental Simulations

The first serious attempts to model protein synthesis were made by MacDonald *et al.* (1968) and MacDonald & Gibbs (1969). These authors treated the movement of ribosomes along the mRNA as a random walk, subject to the restriction that the ribosomes may interfere with one another through steric hindrance. This interdependence introduces serious mathematical difficulties, and it was only possible to obtain (numerical) solutions for the steady-state case. Nevertheless, three types of steady-state solutions were shown to be possible: a low-density solution with few ribosomes on the mRNA when $k_i < k_t < k_e/(1+\sqrt{S})$ (k_i, k_e, k_t refer to the rate constants for initiation, elongation and termination, and S is the number of codons covered by a ribosome) corresponding to the total production rate (number of completed chains per unit time) being limited by the rate of initiation; a high-density solution when $k_t < k_i < k_e/(1+\sqrt{S})$ corresponding to termination being rate-limiting; and a "traffic jam" solution with a transition from low to high density of ribosomes at some internal point on the mRNA when $k_i = k_t < k_e/(1+\sqrt{S})$. These solutions were later studied by Hiernaux (1974), who demonstrated that the position of the transition on the mRNA in the last case is ambiguous and depends on the initial conditions chosen. At any rate, these early works clearly showed that the analytical approach was far too difficult to be practical other than on a qualitative scale, and subsequent modelling has either been very restricted in scope, or relied on computer simulation.

Basically, all computer simulations of protein synthesis keep track of all ribosomes on a number (100, say) of mRNAs. Ribosomes are moved one at a time, provided that they are not obstructed by the preceding ribosome, according to whether a random number between zero and one is less than a pre-set step-probability or not. The step-probability may or may not be the same for all elongation steps, and the initiation and termination probabilities per unit time can be varied to simulate various limitations in the availability of activated 40 S subunits, initiation-inhibitor concentrations, or the presence of competing mRNAs. After a sufficient number of

ribosomes to reach steady state have been run through the system, one collects statistics on experimentally measurable quantities such as polysome-sizes, transit-times, rates of production, and distribution of ribosomes along the mRNAs. One can thus make sure that the theoretical curves are obtained under conditions that are similar to one's experimental set-up, and that they incorporate the molecular details one happens to be interested in.

Works by Gordon (1969) and Vassart *et al.* (1971) are two early examples of the simulation approach. Gordon studied the progression of ribosomes on a single mRNA and calculated ribosome fluxes and polysome size-distributions for various ratios of initiation to elongation rate. Vassart *et al.* extended this to allow for ribosome recycling, and could show that total protein synthesis increases up to a saturating value as the initiation rate constant is increased (at high k_i values the actual number of initiations per unit time is limited by how fast the preceding ribosome moves out of the initiation region, i.e. by k_e); that interference between ribosomes (slow ribosomes delay fast ones following but not *vice versa*) leads to a decrease in the mean rate of elongation as the rate of initiation (and degree of crowding) increases; that a reasonable increase or decrease of k_e has effects on the polysome pattern but does *not* appreciably affect total synthesis (a point stressed already by Itano (1966)); and that the average ribosome density does indeed behave as in the models by MacDonald *et al.* discussed above.

However, the most systematic, complete, and cited simulation study designed to reproduce experimental *in vitro* results is that by Bergmann & Lodish (1979). Although in many respects similar to the earlier ones, one now finds a much stronger insistence on the physiologically relevant ranges of parameter variation, and a specific application to the case of translational control of protein synthesis through competition between mRNAs with intrinsic differences in initiation rates under conditions of varying availability of activated 40 S subunits. The idea is that mRNAs with small k_i produce as much protein as those with larger k_i when the 40 S concentration is saturating, but that they are at a growing disadvantage when this concentration (or the concentration of some "discriminatory initiation factor") drops. This phenomenon had earlier been discussed in a more qualitative way by Lodish (1974), Ninio (1975*a*) and Lodish & Froshauer (1977). Again, Bergmann & Lodish stress that initiation is the major rate-limiting step under most physiological conditions, and that the mean elongation rate plays no important role for the total rate of production. Further developments along these lines are provided by Godefroy-Colburn & Thach (1981)—discussing the "discriminatory factor" concept as well as improving slightly on the simulation model—and by Chu & Rhoads (1980)—who apply the model to initiation inhibition by cap analogs.

The increase in error frequencies under amino acid starvation has also been analysed with simulations of this type. Harley *et al.* (1981) took advantage of the fact that severe starvation for one or more amino acids most probably leads to queueing of ribosomes all the way back to the initiation site; thus the production rate will be limited by the time it takes a ribosome to read through the largest cluster of "hungry" codons in the mRNA. This fact made it possible to extrapolate the error frequency observed under starvation back to unstarved conditions, and further

suggested that, on average, short mRNAs should be preferentially translated under severe starvation since they are less likely to contain long "hungry" codon-clusters.

Menninger (1983) provided a variation on this theme by including the possibility of "ribosome editing", i.e. the hypothesis that the probability of premature termination is high when, after an erroneous aa-tRNA has been incorporated, an incorrect peptidyl-tRNA is transferred to the P-site. In this model, the production of complete protein is very sensitive to the probability of erring, and thus to amino acid starvation, and significant reductions of the mean polysome size are expected since many ribosomes will terminate translation on one of the "hungry" codons. The effect of amino acid starvation on the error rate is discussed more fully in section 9.

In conclusion, methods to simulate *in vitro* experiments, in particular when dealing with problems of initiation of protein synthesis and amino acid starvation, are now well developed and have served their purpose of identifying the parameters that are most critical for the rate of protein production and polysome size.

2. tRNA-codon Adaptation

In most of the models discussed above, it was assumed that the elongation rate is the same for all codons. This is not strictly true. Experimentally, nascent chains of intermediate lengths corresponding to pausing at rare codons are sometimes seen (Varenne *et al.*, 1984; Misra & Reeves, 1985). Since there is a striking correlation between tRNA levels and codon frequencies in many cells (Garel, 1974; Ikemura, 1981*a*,*b*), the rate of elongation is most likely modulated by tRNA availability.

The observed tRNA-codon adaptation (frequently used codons correspond to abundant tRNAs) makes one suspect that there exists some kind of underlying optimization principle. Many physiologically important quantities such as mean elongation rate, mean error frequency, and energy spent on proofreading (see section 7) depend on the mean number of aa-tRNAs that a ribosome has to sort through before it finds a cognate one. This number is given by

$$\langle M \rangle = \sum x_i / y_i \tag{1}$$

where x_i and y_i denote the frequency of codons corresponding to amino acid i and the corresponding relative tRNA concentration, respectively. Minimizing $\langle M \rangle$ with respect to the y_i's y yields $y_i \sim \sqrt{x_i}$ as the optimal choice, whereas the experimental correlation seems to be more linear. Optimal correlations of this kind have been repeatedly discussed in the literature (Pontier, 1970; von Heijne & Blomberg, 1979; Tsukamoto, 1979; Hearon & Tsukamoto, 1980; Chavancy & Garel, 1981).

In these treatments, only the aa-tRNA levels are considered. In reality, however, the cell can probably only directly manipulate the *total* tRNA concentrations, and the effects of a change in one of these will be distributed among all the tRNA pools (free, synthetase-bound, aminoacylated, A-site and P-site bound tRNA; see Gouy & Grantham, 1980). When the whole tRNA cycle is taken into account and the mean elongation rate is maximized with respect to the total tRNA concentrations, a much more complex situation appears (Liljenström *et al.*, 1985). The optimal correlation is now seen to depend both on the tRNA/ribosome and synthetase/ribo-

some ratios, and at high growth rates (corresponding to small tRNA/ribosome ratios) the optimal correlation tends to be linear rather than square-root. Thus, it seems that the translation system is tuned to be maximally efficient at high rather than low growth rates.

Apart from any tRNA-codon adaptations, codon usage is also strikingly non-random in most genes (Grantham *et al.*, 1981; Gouy & Gautier, 1982). As a rule, highly expressed genes tend to use only a subset of all codons (corresponding to the most abundant tRNAs), whereas codon usage in weakly expressed genes is less constrained. There are many, not necessarily mutually exclusive, explanations for these observations; again, maximization of the mean rate of elongation is one.

A low frequency of rare codons obviously gives a low value for $\langle M \rangle$ (eqn (1)) and a high elongation rate. In view of the results discussed in section 1, however, it is not correct to equate elongation rate with production rate: a high elongation rate in the initiation region ensures that the next ribosome can bind to the mRNA with minimum delay caused by the preceding one and thus gives a higher total productivity, whereas a high elongation rate in the *internal* parts of the mRNA affects the productivity only indirectly, through a slight increase in the number of free ribosomes available for initiation. Rare codons in weakly expressed mRNAs thus have a negligible effect on the total production rate of the particular mRNA as well as of the whole cell, and can be allowed to accumulate (Liljenström & von Heijne, 1987). This point is worth stressing, since it has been repeatedly overlooked in the literature.

3. mRNA Secondary Structure

A high percentage of all nucleotides in an mRNA is expected to be part of base-paired secondary structures such as hairpin loops (Fitch, 1974). It is conceivable that such loops may interfere with ribosome movement, and it has been suggested that particular hairpins may influence the error frequency by their effects on translation (Shpaer, 1985).

This kind of translational modulation has also been analysed with the aid of computer simulation along the lines discussed in section 1, but with an additional routine to determine whether a potential hairpin loop about to be translated is prevented from refolding by the preceding ribosome (von Heijne *et al.*, 1977, 1978). The main result from these studies was that an increased ribosome density on the mRNA does *not* in this case necessarily lead to a reduction of the mean transit time; this is a qualitatively different behaviour to that expected from the simpler models. The reason is that more ribosomes on the mRNA means less possibilities for refolding of hairpins, and thus less obstruction of ribosome movement. The expected effect is not very dramatic, however, and has not been experimentally verified.

4. Attenuation Control

In bacteria, a number of amino acid biosynthetic operons are controlled by an attenuator mechanism that directly determines whether or not RNA polymerase

transcribes through an early potential termination-site. A ribosome translating a short "leader peptide"-region in the early parts of the transcript is thought to either stall at a codon corresponding to a rare aa-tRNA, thus interfering with the formation of a particular hairpin loop in the mRNA and preventing transcription termination, or else to translate the whole leader peptide and allow termination (Yanofsky, 1981).

This control mechanism integrates a number of the themes discussed above: overall kinetics of protein synthesis, modulation of the elongation rate through aa-tRNA availability, and ribosomes influencing mRNA folding. It therefore comes as no great surprise that it has also been the subject of theoretical modelling. Manabe (1981) derived an expression for the probability that a ribosome will be stalled at one of the critical codons when the polymerase reaches the termination site, and constructed theoretical response-curves for the attenuator under various degrees of amino acid starvation. This was extended by von Heijne (1982), who simulated the effects of "mutations" in the critical codons and discussed the design principles of the attenuator region.

More recently, Suzuki *et al.* (1986) incorporated an experimentally detected *transcriptional* pausing into the model, and showed that this pause, which leads to better synchronization between ribosome and polymerase movement (Fisher *et al.*, 1985), makes the control mechanism much more responsive to changes in aa-tRNA concentrations and less sensitive to stochastic variations in the timing of ribosome-binding to the transcript.

5. Selection

In the remaining sections, certain theoretical aspects on the accuracy of the ribosome selection will be considered. The selection depends on the triplet anti-codon–codon binding, and as there are tRNAs with similar anticodon triplets that also can bind to the codon, errors can always occur when incorrect tRNAs are accepted. There are measurements of the error frequencies *in vivo* that are relevant to the discussion, see e.g. Edelman & Gallant (1977). Commonly, these show values of the order 10^{-3}–10^{-4}. Probably, error frequencies vary among the amino acids, and the situation is not quite clear, see Kurland & Ehrenberg (1984).

One feature that appeared puzzling when first found in experiments is that the accuracy is influenced by structural changes in the tRNA or the ribosome at places rather far from the relevant codon and anticodon bases. (Experiments on this often consider bacterial cells and their response to the drug streptomycin—which is known to bind at the ribosome, to interfere with the selection, and then lead to a higher error frequency (Gorini, 1971).) These results can, however, easily be explained as the selection accuracy is always influenced by general, unspecific features of the path from substrate to product. This was discussed by Schwartz & Lysikov (1974) and Ninio (1974) and in many later works. They also discussed the related feature that mutations, e.g. in streptomycin-immune bacteria, could lead to an improved accuracy, which is not favoured by wild-type bacteria. It was pointed out that an improved selection would lead to a prolonged process time for the product formation, which clearly is a disadvantage. This principle is developed further in later works.

It is easy to see the general arguments in this question. The simplest selection process where a substrate S (tRNA) is selected by an enzyme (ribosome) complex to form a product (polypeptide) can be written as

$$E+S \leftrightarrows (ES) \rightarrow (EP) \rightarrow E+P.$$

In most types of biochemical selections, the specificity arises from the enzyme-bound state. Correct substrates are bound tighter, and therefore the dissociation rates from the enzyme bound states are faster for incorrect substrates. The best accuracy is obtained if the product formation step at the enzyme, from (ES) to (EP), is much slower than the dissociations. The first two states, E+S and (ES) are then almost equilibrated. The (ES) concentration is given by the free energy of formation of this state, ΔG, and is proportional to $\exp(-\Delta G/RT)$. The quotient between equilibrium constants for this state of correct and incorrect substrates is a measure of the intrinsic selection capacity.

To reach the highest accuracy in this manner, the intermediate step must be slow, and thus also the product formation. For a cell, it is advantageous not to achieve the maximum accuracy, but to work with an acceptable accuracy and a satisfactory product formation rate.

Another feature of the selection process is due to the fact that there is no difference in the free energy of formation of peptide bonds of correct and incorrect amino acids. Accordingly, if the process proceeds close to equilibrium, no selection would be obtained. To achieve a high accuracy, there must be a considerable dissipation (see e.g. Blomberg *et al.*, 1981). As the free energy sources may be precious for a cell and some economy is needed, this is also a disadvantage.

Thus, the highest levels of accuracy are not favourable and should be avoided from the time and/or dissipation aspects. This is developed further in a number of works, but, before going on to that, we consider the question of low accuracy.

6. Error Propagation

Many erroneous polypeptides have a reduced activity or are inactive, which means that they are rather harmless and essentially represent a waste of resources. However, in certain cases they can be quite harmful. In particular this is so for proteins that partake in the selection processes with the function of regulating the accuracy. If such a protein is erroneous, it may not function properly in the selection, and can give rise to a larger number of errors. In this way, successive generations of selection proteins may contain an increasing number of errors, which might eventually yield a complete breakdown of the selection process: an error catastrophe. This problem was first discussed by Orgel (1963, 1970). This error propagation has been studied by mathematical models which show that an error catastrophe is not inevitable, and one may expect that a stable situation is reached. The models usually predict some sort of threshold effect: if the error at some stage is too large, it will increase until the system breaks down (Hoffman, 1974; Kirkwood & Holliday, 1975).

The situation has been studied experimentally with streptomycin-influenced bacteria in which initial errors are varied to increased levels. An error propagation to

stable levels is clearly seen, and this has been analyzed by modelling (Gallant & Prothero, 1980; Liljenström *et al.*, 1985). Many other features of error catastrophe are less clear at present and thus it is uncertain if, as proposed, this error propagation and eventual catastrophe is related to the ageing of eukaryote cells (see e.g. Kirkwood *et al.*, 1984).

One important result of these experiments is that cells still appear to function with rather high error levels. Experiments with streptomycin-induced and mutant bacteria show that these levels can be increased up to about 25 times without yielding any catastrophe (see Kurland & Ehrenberg, 1984). This suggests that normal error levels are much lower than those necessary for proper functioning. In the past, it has been suggested for theoretical reasons that error levels about 10^{-3}–10^{-4} are necessary for maintaining a proper function of a cell. This does not seem to be the case, and actual levels are probably set by the requirements for efficient functioning, which we will return to. It has been an important question (and this to some extent remains) as to how a primitive cell at an early stage of the origin of life could attain a sufficiently high accuracy. If the requirements of accuracy are not as high as has been suggested, the arguments on this point should in some way be re-evaluated.

7. Proofreading

It was apparent at an early stage that the high accuracy levels found for systems *in vivo* are difficult to interpret from simple selection schemes, such as that in section 5, and estimates of the binding energies of anticodon–codon triplets. The latter may be obtained from theoretical calculations or from measurements. Such estimates have provided intrinsic discrimination ratios of the order 100 or so, one to two orders of magnitude lower than what is really accomplished, which suggests that there must be some kind of accuracy improvement. Concepts of that kind certainly existed in many places in the early 1970s, but it was Hopfield (1974) and Ninio (1975*b*) who clearly discussed that possibility with its requirements. From Hopfield, the error correction has been called *proofreading*. (One also finds the term "editing"). In principle, it means that the selection scheme is branched with further possibilities of rejecting substrates from the pathway to product

$$\begin{array}{c} E+S \rightleftarrows (ES)_1 \rightarrow (ES)_2 \rightarrow (EP) \rightarrow E+P \\ \qquad\qquad\quad \downarrow \\ \qquad\qquad\quad (E+S)' \end{array}$$

Selection occurs at two stages during substrate processing: in the initial step and in the branch step. The probability that an incorrect substrate is accepted is significantly decreased. In principle, there can be an arbitrary number of branches out from the pathway with further selection steps and a still higher accuracy. What was emphasized in the works of Hopfield and Ninio is that the branches must be driven by a decrease in free energy in the branch: there must be a net flow of substrates out through the branch. This is what is meant by $(E+S)'$, which is not the same as the initial state, but some component—the substrate, the enzyme, or a cosubstrate—must be degraded. Usually this is accomplished by hydrolysis of a nucleoside

triphosphate, and this seems to be the role of the GTP that accompanies the acylated tRNA and elongation factor Tu at the ribosome (in bacteria).

It is relevant to seek the minimum dissipation that is necessary to attain a certain accuracy level, and this is considered in a number of works. The first treatment is due to Bennet (1979). A more complete discussion was made by Ehrenberg & Blomberg (1980), and Blomberg & Ehrenberg (1981). In these works, the rate constants along the branched pathway were varied to yield the optimum dissipation-accuracy relations. In the works by Ehrenberg & Blomberg, a formalism was developed that was suitable for treating this optimization problem in the branched reaction network. This is described in most detail in Blomberg (1983). In these works, all steps along the branched scheme are treated as reversible to account for the free energy changes, and an arbitrary number of branches is possible. The concentrations of different states along the pathway are replaced by displacements from equilibrium with the initial states (displacement × equilibrium value = actual concentration), and instead of varying the rate constants, certain quotients of flows of correct and incorrect substrates are introduced which lead to simpler and more symmetric quantities of selection. Note that in models of this kind, one aims to study optimized processes, which means that intrinsic rate constants are given by their optimal values, and there is no arbitrariness in their choice.

To get the minimum dissipation consistent with a certain accuracy level, the displacements from equilibrium of the product and the degradation product (at the branch) are also varied. In actual cases these probably have assigned values, fixed by other constraints. They are probably very small in the ribosome selection, see Blomberg (1985), which also means that the steps of the selection process are essentially irreversible. The main quantity that describes the excess dissipation in such cases is the excess ratio: (flow out of the branches)/(total incoming flow). This was considered by Blomberg & Ehrenberg (1981) (which in fact was a first version of the formalism of the branched network with reversible steps). Savageau & Freter (1979), and Freter & Savageau (1980) have considered selection processes with a similar aim in a formalism which included only flow quantities (and thus not concentrations/displacements from equilibrium). This gives results equivalent to other formalisms for a process with purely irreversible steps (and a large dissipation), but its significance for a process with reversible steps is questionable. For a full account of this problem, see Blomberg (1983*a*).

The time delay in proofreading is quite significant, and may be more relevant to actual processes than the dissipation. This aspect was treated by Blomberg (1977), and later by using similar formalism to the dissipation losses (Blomberg, 1985). There are some notable differences between the models of dissipation and time delay. In principle, the free energies are more directly accounted for than the "time cost". For instance, saturation effects of the ribosome (or a selection enzyme) are not appearing in the quotients that determine the free energy quantities. As for the time effect, saturation features are relevant, and there may also be constraints, not really known, on the rates along the pathway. Still, one can treat a relatively general scheme, and reach relevant conclusions. For a non-saturated ribosome, the largest difference to the dissipation case is the treatment of the initial selection step

$(E+S \rightarrow (ES)_1)$: to obtain minimum dissipation, this shall be equilibrated, and will appear different to proofreading steps. For the time delay model, the initial step and the branches are equivalent. Indeed, this seems to be experimentally verified in some cases (Kurland & Ehrenberg, 1984). On the other hand, for a saturated ribosome, the initial state is always close to equilibrium. Then the dissipation loss and the time delay are both determined by the loss at the branches, and are described by similar quantities.

A further complication related to what has just been said is that the ribosome does not proceed uniformly along the mRNA (sections 1 and 3) and thus certain features (in particular saturation) depend on the site along the chain. For instance, time delay may be more crucial for the first amino acids of the protein chain. The consequences of this have not yet been considered in models of selection.

8. Accuracy Economy

One can go further with the consideration of the accuracy cost. High accuracy requires a cost in terms of free energy or time as just discussed, but low accuracy means that there is a significant cost due to the erroneous proteins which represent a waste of resources. It should be reasonable that a cell optimizes the cost per functionable protein. By taking into account features of individual amino acids, such a requirement can provide optimum error levels of the individual amino acids of the protein as well as optimum features of the selection process. This has been treated by Blomberg (1983*b*, 1985). Ehrenberg & Kurland (1984) discuss this in a slightly different way using somewhat different expressions which are equivalent to those of Blomberg.

An extended analysis of this kind can yield more precise predictions about the selection features which can be compared to actual situations. It is possible to investigate different cell strategies to get an understanding of the actual performance of a cell. The optimization means that actual error levels are determined by a cost requirement, and not by any low accuracy constraint required to produce a functioning system. As already stated, cells can still work with a considerable decrease in accuracy, but they are then unfit to compete with wild type cells. (Note that the protein synthesis process is one of the main processes of a cell, and one of the most time and free energy consuming.)

Concentrations of the different substrates are to be considered in this framework, which generalizes the aspects of tRNA adaptation in section 2. Of course, a substrate in larger concentrations is selected more accurately than one with lower concentrations. Clearly, there are mutual asymmetric selections, where one substrate occurs more frequently than another in the product, or is more difficult to distinguish from the other than vice versa, or may be more relevant for the function of the protein. In such cases, the cost optimization involves the establishment of the monomer concentration levels and the design of the individual processes, possibly modifications of the monomers to cope with an efficient mutual selection. These aspects are considered in a recent work (Blomberg, 1986).

Ehrenberg & Kurland (1984) have considered the design of the selection mechanism and the relevant concentrations using the requirement that the selection time shall be minimized for a growing cell culture, which introduces further new aspects. They also discuss different growth conditions, whether there is an abundant supply of food and free energy or a poor medium, where the cell also must use a considerable free energy to produce, for instance, amino acids. It is suggested that the cell may have evolved different types of strategies for these cases.

9. Selection Under Starvation Conditions

Situations where there is a severe depletion of one or more amino acids have been mentioned in section 1. That situation is of course quite relevant to the selection problem. Usually, one would expect that a depleted amino acid would be related to a high error frequency—other, more frequent, monomers could take its place. However, this is not really the case. The rate will be reduced, but the accuracy is not much affected. The effect is known to be regulated by the nucleoside-four-phosphate ppGpp, which is synthesized in starvation conditions. Also, amino acids of normal concentrations are not incorporated with an enhanced accuracy.

A couple of different models have been suggested to account for this effect. They are all supported by experiment, but at present no model can account for all the known features in a satisfactory way. One model, suggested by O'Ferrel (1978) and developed in later works (Rojas *et al.*, 1984), is that ppGpp binds efficiently to elongation factor Tu, and reduces the possibilities of amino acyl tRNA binding to that protein. This means that the tRNA-EfTu binding becomes a rate-limiting step in the tRNA-cycle, and most tRNA-molecules will be in the free, amino-acylated state. Then, the concentrations of these would not reflect the concentrations of amino acids. This model explains much of the features of the effect, including the fact that amino acids in normal concentrations do not show improved accuracy. However, it is difficult to conceive that this model can actually work with given concentrations and known coupling constants. Another possibility, first suggested by Kurland & Ehrenberg (1984) and later taken up by Dix & Thompson (1986) is that ppGpp enters the proofreading scheme and makes proofreading more efficient. (The way this actually occurs in the proofreading scheme is that the back step $(EP) \rightarrow (ES)_2$ is made faster by an increased concentration of ppGpp, and thus facilitates rejection.) Even if this explains certain facts, it should mean that proofreading for all substrates, and also those of amino acids in normal concentrations, would be increased. This does not seem to be the case. Ninio (1986), motivated by this fact, suggested a model where ribosomes, with some time delay, can be transformed to a new state which yields an enhanced accuracy. This state would be reached only for rare substrates when there will be time for the ribosome to transform. This model is at present speculative, and it is not obvious how ppGpp enters the mechanism. Certainly, accuracy enhancement of this kind is not found in cells which lack the gene for ppGpp synthesis.

10. Drop-off, Translocation

It is known that peptidyl-tRNA can drop off the ribosome before protein synthesis is complete, which fact was considered in the simulation work by Menninger mentioned in section 1. An idea of Menninger was that this could serve as a proofreading step as any incorrect peptidyl-tRNA would have a high probability of dropping off the ribosome. Ehrenberg & Kurland (1986) have considered this in a recent work. Indeed, this mechanism could be used for accuracy enhancement, but it is not particularly efficient and can only in special situations yield a marginal improvement in accuracy. This is because a great improvement would necessarily mean that many correct but incomplete polypeptides drop off, with a corresponding energy waste. Also, many errors do not influence the performance of the protein. Ehrenberg & Kurland argue that this effect cannot be used as an efficient means of correcting missence errors. It is found that this is effective only if most errors lead to a loss of function and if the intrinsic selection potential of the drop-off step can provide a better selection than is accomplished in previous steps.

On the other hand, they argue that this can be used for error correction by reducing frameshift errors, i.e. errors that ensue when a tRNA does not read the three consecutive bases of the cognate codon but misses one or reads one too many. This evidently makes all the later amino acids wrong. Although rather costly, the drop-off could be a possibility for error-correction in this case, and this also gives an explanation for the GTP-hydrolysis that accompanies the translocation at the ribosome: this is required to drive the drop-off as a proofreading step.

Conclusion

We have considered a number of models which have been used in describing different features of protein synthesis. These models usually take up single aspects and consider the general simulation of ribosome movement, features that influence the elongation and ribosome movement which include the mRNA secondary structure and tRNA availability, and finally the problems of describing the accuracy, which in some way seems to be related to some economy efficiency. Rather few works take up the connections between these aspects, and this would be important in accomplishing a more complete picture of the protein synthesis. In fact, important points in one aspect may be ignored in other studies. So, as discussed here, the most realistic simulations of ribosome movement seem to indicate that variations of the elongation rates of amino acids in the interior part of the polypeptide chains do not really affect the total process time. But many works have still strongly emphasized the importance of a rapid elongation rate. Thus, it will be important to investigate further in the future how to fit these pieces together in a more general view.

REFERENCES

BENNET, CH. H. (1979). *Biosystems* **11**, 85.
BERGMANN, J. E. & LODISH, H. F. (1979). *J. biol. Chem.* **254**, 11927.

BLOMBERG, C. (1977). *J. theor. Biol.* **66,** 307.
BLOMBERG, C. (1983*a*). *Quart. Rev. Biophys.* **16,** 415.
BLOMBERG, C. (1983*b*). *Int. J. quant. Chem.* **23,** 687.
BLOMBERG, C. (1985). *J. theor. Biol.* **115,** 241.
BLOMBERG, C. (1986). *Eur. J. Biophys.* (submitted.)
BLOMBERG, C. & EHRENBERG, M. (1981). *J. theor. Biol.* **88,** 631.
BLOMBERG, C., EHRENBERG, M. & KURLAND, C. G. (1981). *Quart. Rev. Biophys.* **13,** 231.
BLOMBERG, C., LILJENSTRÖM, H. & JOHANSSON, J. (1985). *J. theor. Biol.* **114,** 407.
CHAVANCY, G. & GAREL, J. P. (1981). *Biochimie* **63,** 187.
CHU, L-Y. & ROADS, R. E. (1980). *Biochemistry* **19,** 184.
DIX, D. B. & THOMPSON, R. C. (1986). *Proc. natn. Acad. Sci. U.S.A.* **83,** 2027.
EDELMAN, P. & GALLANT, J. A. (1977). *Cell* **10,** 131.
EHRENBERG, M. & BLOMBERG, C. (1980). *Biophys. J.* **31,** 333.
EHRENBERG, M. & KURLAND, C. G. (1984). *Quart. Rev. Biophys.* **17,** 45.
EHRENBERG, M. & KURLAND, C. G. (1986). *Quart. Rev. Biophys.* (in press).
FISHER, R. F., DAS, A., KOLTER, R., WINKLER, M. E. & YANOFSKY, C. (1985). *J. mol. Biol.* **182,** 397.
FITCH, W. M. (1974). *J. mol. Evol.* **3,** 279.
FRETER, R. R. & SAVAGEAU, M. H. (1980). *J. theor. Biol.* **85,** 99.
GALLANT, J. A. & PROTHERO, J. (1980). *J. theor. Biol.* **97,** 257.
GAREL, J. P. (1974). *J. theor. Biol.* **43,** 211.
GORINI, L. (1971). *Nature New Biol.* **234,** 261.
GRANTHAM, R., GAUTIER, C., GOUY, M. & MERCIER, R. (1981). *Nucleic Acids Res.* **9,** r43.
GODEFROY-COLBURN, T. & THACH, R. E. (1981). *J. biol. Chem.* **256,** 11762.
GORDON, R. (1969). *J. theor. Biol.* **22,** 515.
GOUY, M. & GRANTHAM, R. (1980). *FEBS Lett.* **115,** 151.
GOUY, M. & GAUTIER, C. (1982). *Nucleic Acids Res.* **10,** 7055.
HARLEY, C. B., POLLARD, J. W., STANNERS, C. P. & GOLDSTEIN, S. (1981). *J. biol. Chem.* **256,** 10786.
HEARON, J. Z. & TSUKAMOTO, Y. (1980). *J. theor. Biol.* **83,** 525.
HIERNAUX, J. (1974). *Biophys. Chem.* **2,** 70.
HOFFMAN, G. W. (1974). *J. mol. Biol.* **86,** 349.
HOPFIELD, J. J. (1974). *Proc. natn. Acad. Sci. U.S.A.* **71,** 4135.
IKEMURA, T. (1981*a*). *J. mol. Biol.* **146,** 1.
IKEMURA, T. (1981*b*). *J. mol. Biol.* **151,** 389.
ITANO, H. A. (1966). *J. cell. Physiol.* **67,** Sup. 1, 65.
KIRKWOOD, T. B. L. & HOLLIDAY, R. (1975). *J. mol. Biol.* **97,** 257.
KIRKWOOD, T. B. L., HOLLIDAY, R. & ROSENBERGER, R. F. (1984). *Int. Rev. Cytol.* **92,** 93.
LILJENSTRÖM, H., VON HEIJNE, G., BLOMBERG, C. & JOHANSSON, J. (1985). *Eur. Biophys. J.* **12,** 115.
LILJENSTRÖM, H. & VON HEIJNE, G. (1987), *J. theor. Biol.* **124,** 43.
LODISH, H. F. (1974). *Nature* **251,** 385.
LODISH, H. F. & FROSHAUER, S. (1977). *J. biol. Chem.* **252,** 8804.
MACDONALD, C. T., GIBBS, J. H. & PIPKIN, A. C. (1968). *Biopolymers* **6,** 1.
MACDONALD, C. T. & GIBBS, J. H. (1969). *Biopolymers* **7,** 707.
MANABE, T. (1981). *J. theor. Biol.* **91,** 527.
MENNINGER, J. R. (1983). *J. mol. Biol.* **171,** 383.
MISRA, R. & REEVES, P. (1985). *Eur. J. Biochem.* **152,** 151.
NINIO, J. (1974). *J. mol. Biol.* **84,** 297.
NINIO, J. (1975*a*). *Nature* **255,** 429.
NINIO, J. (1975*b*). *Biochimie* **57,** 587.
NINIO, J. (1986). *FEBS Lett.* **196,** 1.
O'FERREL, P. H. (1978). *Cell* **14,** 545.
ORGEL, L. E. (1963). *Proc. natn. Acad. Sci. U.S.A.* **49,** 517.
ORGEL, L. E. (1970). *Proc. natn. Acad. Sci. U.S.A.* **67,** 1476.
PONTIER, J. (1970). *Bull. Math. Biophys.* **32,** 83.
ROJAS, A.-M., EHRENBERG, M., ANDERSSON, S. & KURLAND, C. G. (1984). *Mol. Gen. Genet.* **197,** 36.
SAVAGEAU, M. A. & FRETER, R. R. (1979). *Proc. natn. Acad. Sci. U.S.A.* **76,** 1902.
SCHWARTZ, V. S. & LYSIKOV, B. N. (1974). *Dokl. Akad. Nauk. SSSR* **217,** 1446.
SHPAER, E. G. (1985). *Nucleic Acids Res.* **13,** 275.
SUZUKI, H., KUNISAWA, T. & OTSUKA, J. (1986). *Biophys. J.* **49,** 425.
TSUKAMOTO, Y. (1979). *J. theor. Biol.* **78,** 451.
VARENNE, S., BUC, J., LLOUBES, R. & LAZDUNSKI, C. (1984). *J. mol. Biol.* **180,** 549.

VASSART, G., DUMONT, J. E. & CANTRAINE, F. R. L. (1971). *Biochim. biophys. Acta* **247**, 471.
VON HEIJNE, G. (1982). *J. theor. Biol.* **97**, 227.
VON HEIJNE, G., NILSSON, L. & BLOMBERG, C. (1977). *J. theor. Biol.* **68**, 321.
VON HEIJNE, G., NILSSON, L. & BLOMBERG, C. (1978). *Eur. J. Biochem.* **92**, 397.
VON HEIJNE, G. & BLOMBERG, C. (1979). *J. theor. Biol.* **78**, 113.
YANOFSKY, C. (1981). *Nature* **289**, 751.