Energy Cost of Proofreading to Increase Fidelity of Transfer Ribonucleic Acid Aminoacylation†

Michael A. Savageau* and Rolf R. Freter

ABSTRACT: The paradox of relatively error free function in biological systems composed of relatively error prone components has recently come under intensive investigation. In the case of tRNA aminoacylation, aminoacyl-tRNA synthetases were discovered to have a separate function that allows misacylated molecules to be hydrolyzed more rapidly than correctly acylated molecules. This additional function of the synthetases provides a proofreading or verification mechanism that is believed to improve significantly the overall accuracy of tRNA aminoacylation. In this paper we provide an explicit relationship between the accuracy achieved by proofreading and the energy cost. Experimental data available in the literature are examined in light of this relationship. The following are the principal conclusions from our study: (1) high-accuracy proofreading of tRNA aminoacylation has a high energy cost, as much as 100 times greater than indications from early experimental work; (2) the minimum net error derived in previous theoretical studies is never actually reached; (3) mechanisms in which misacylation and subsequent proofreading occur on the surface of the same synthetase molecule achieve a much higher accuracy than mechanisms in which these functions occur on the surface of different synthetase molecules.

The biosynthesis of proteins occurs with remarkable fidelity. For example, Loftfield and his colleagues showed that isoleucine was accidentally replaced by valine in fewer than 3 of 10^4 positions during the biosynthesis of chicken ovalbumin (Loftfield, 1963) or rabbit hemoglobin (Loftfield & Vanderjagt, 1972). However, Pauling (1958) had previously suggested on theoretical grounds that the ability to discriminate between isoleucine and valine should be no better than about 20-fold. Thus, it appears that the accuracy of the overall process of protein synthesis is significantly greater than that of its individual component processes, e.g., aminoacylation of tRNA.

The first hint at a resolution of this paradox in the case of aminoacylation of tRNA occurred in the early 1970's when it was discovered that aminoacyl-tRNA synthetases possess a separate function that allows them to hydrolyze the aminoacyl-tRNA, thereby releasing free amino acid and free tRNA (Yarus, 1972a; Schreier & Schimmel, 1972). Since incorrect amino acid–tRNA pairs were split much more rapidly than the correct amino acid–tRNA pair, it was clear that this hydrolytic function of the synthetases provided a proofreading or verification mechanism that was capable of improving the accuracy of tRNA aminoacylation (Yarus, 1972a; Schreier & Schimmel, 1972). It was also clear that such proofreading had an energy cost. This work has been confirmed and greatly extended in several laboratories by studies designed to reveal the chemical basis [e.g., see von der Haar & Cramer (1976) and Alford & Hecht (1978)] and detailed kinetics [e.g., see Fersht & Jakes (1975) and Fersht & Kaethner (1976)] of proofreading mechanisms. The questions remain—how much can the accuracy be improved by this mechanism and at what cost?

Independently, Hopfield (1974) and Ninio (1975) provided a partial answer to the first question. They showed that the *maximum* discrimination in tRNA aminoacylation obtained with such proofreading is equal to the product of two factors: (1) the factor by which synthetases are able to discriminate between correct and incorrect substrates for the initial ligase reaction and (2) the factor by which synthetases are able to discriminate between incorrect and correct substrates in the proofreading reaction. Thus, a 100:1 discrimination in the initial ligase reaction and a 100:1 discrimination in the proofreading reaction theoretically could produce an overall discrimination of 10 000:1 for tRNA aminoacylation.

The second question, that of the energy cost for this increased fidelity, has remained unanswered. Although energy expenditure is an essential element in all the arguments for proofreading, the precise relation between energy cost and accuracy achieved has remained undetermined. In the following pages we provide a theory of proofreading that explicitly yields the energy costs. These results are compared with experimental evidence regarding the proofreading of misacylated molecules with tRNA errors and misacylated molecules with amino acid errors.

Theory

Single Type of Error and Proofreading. The essence of our theory can be illustrated with a simple model that includes a single type of error and proofreading. More complex cases, or generalizations, will be discussed later. Consider two aminoacyl-tRNA synthetases and their cognate tRNAs and amino acids and assume that the network of reactions is symmetrical and operating in a normal steady state.

Figure 1 is a schematic representation of the case in which errors in tRNA recognition occur. The synthetases recognize their cognate amino acids with perfect fidelity. Following activation, the (AA-AMP)–synthetase complex can aminoacylate either the cognate or the noncognate tRNA. The aminoacyl-tRNAs, while still bound to synthetase molecules, are either released for incorporation into protein or hydrolyzed to yield free amino acid and free tRNA. Half of the total network defined by the two tRNA species is shown; the other half is identical by symmetry and need not be considered here. Only the key intermediates of interest in this network are represented in Figure 1. Because the network is operating in a steady state, the number of intermediate steps between these key intermediates is irrelevant [see Savageau (1976), chapters

† From the Department of Microbiology, The University of Michigan, Ann Arbor, Michigan 48109. *Received January 11, 1979.* This work was supported in part by a grant to M.A.S. from the National Science Foundation (PCM 7807635).

0006-2960/79/0418-3486$01.00/0

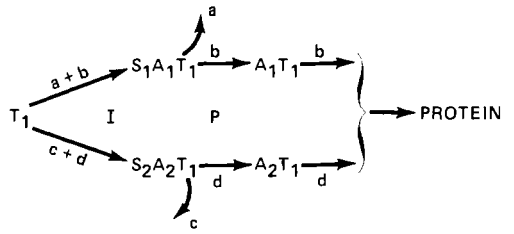

FIGURE 1: Schematic model representing error in tRNA recognition and subsequent proofreading by aminoacyl-tRNA synthetases. The symbols are as follows: T_1, tRNA; A_1, cognate amino acid; S_1, cognate synthetase; A_2, noncognate amino acid; S_2, noncognate synthetase. The fundamental fluxes are represented by a, b, c, and d. I is the initial discrimination ratio, and P is the proofreading discrimination ratio. See the text for further discussion.

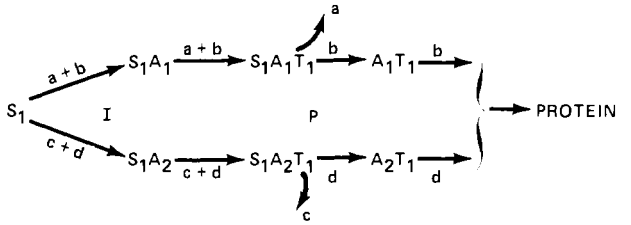

FIGURE 2: Schematic model representing error in amino acid recognition and subsequent proofreading by aminoacyl-tRNA synthetases. The symbols are as follows: S_1, synthetase; A_1, cognate amino acid; T_1, cognate tRNA; A_2, noncognate amino acid. The fundamental fluxes are represented by a, b, c, and d. I is the initial discrimination ratio, and P is the proofreading discrimination ratio. See the text for further discussion.

3 and 5 and references cited therein]. The arrows represent the resultant flux in pathways that are generally reversible, with the exception of the arrows representing the hydrolytic proofreading reactions. These latter reactions, which release an ATP equivalent of energy, are highly exergonic and essentially irreversible.

Figure 2 is a schematic representation of the case in which errors in amino acid recognition occur. Following activation of either the cognate or the noncognate amino acid, the (AA-AMP)–synthetase complexes recognize the cognate tRNA with perfect fidelity. The aminoacyl-tRNA molecules, while bound to the synthetase, are either released for incorporation into protein or hydrolyzed to yield free amino acid and free tRNA. This model is otherwise like that in Figure 1. The theory that follows applies equally well to either model. We proceed by defining four elemental equations. The solution of these equations yields an explicit relationship between the net error and the energy cost of proofreading.

Elemental Equations. There are four independent fluxes that define the state of these networks: a, b, c, and d (see Figures 1 and 2). These in turn may be used to define other parameters of interest.

(*1*) *Initial Discrimination Ratio.* In Figure 1 the initial discrimination ratio I is defined for a given tRNA species as the ratio of input flux of tRNA involved in correct ligation to that involved in incorrect ligation:

$$I = (a + b)/(c + d) \tag{1}$$

In Figure 2 the initial discrimination ratio, defined for a given tRNA species as the ratio of input flux of correct amino acid to that of incorrect amino acid, is also given by eq 1. This definition applies in vivo whatever the detailed mechanism of discrimination may be.

(*2*) *Proofreading Discrimination Ratio.* The ratio of incorrect flux recycled to that continuing on, c/d, divided by the ratio of correct flux recycled to that continuing on, a/b, is defined as the proofreading discrimination ratio P for a given tRNA species:

$$P = bc/(ad) \tag{2}$$

This definition applies regardless of the detailed mechanism and whether amino acid errors or tRNA errors are being proofread.

(*3*) *Net Flux.* The flux of amino acids through the network and into protein per tRNA species is defined as F:

$$F = b + d \tag{3}$$

(*4*) *Recycling.* The flux of amino acids recycled as a result of proofreading per tRNA species is defined as R:

$$R = a + c \tag{4}$$

Net Error. The net error E is expressed as a probability or fraction and defined as the flux of incorrect amino acids divided by the net flux of amino acids through the network and into protein for a given tRNA species:

$$E = d/(b + d) \tag{5}$$

Net error can be expressed explicitly in terms of the parameters I, P, F, and R by solving the elemental eq 1–4 for the independent fluxes and substituting the resultant values for b and d into eq 5. The result is

$$E = [[(P - 1) - (1 + C) - I(1 + PC)] + [[(P - 1) - (1 + C) - I(1 + PC)]^2 + 4(P - 1)(I + 1)(1 + C)]^{1/2}]/[2(P - 1)(I + 1)] \tag{6}$$

In this expression, the parameters R and F occur only in the ratio R/F, which, according to the above definitions, is the recycling/net flux ratio for a given tRNA species. Thus, we have redefined this ratio as the new parameter C, which is the cost of proofreading expressed as moles of ATP per mole of amino acid inserted into protein for a given tRNA species. It must be emphasized that the cost of proofreading, according to this definition, includes the hydrolysis of *correct* as well as incorrect amino acid–tRNA pairs ($R = a + c$) and normalization by the flux of *incorrect* as well as correct amino acids into protein ($F = b + d$). Certain features of the relationship in eq 6 are particularly significant and will be identified both algebraically and graphically in Figure 3.

(*1*) *Low-Cost Asymptote.* In the limit as $C \rightarrow 0$ in eq 6, the net error approaches the asymptotic value

$$E_0 = 1/(I + 1) \tag{7}$$

which is the value without proofreading. [This is also the net error when there is indiscriminate proofreading ($P = 1$), regardless of the energy expended (C).]

(*2*) *High-Cost Asymptote.* In the limit as $C \rightarrow \infty$ in eq 6, the net error approaches the asymptotic value

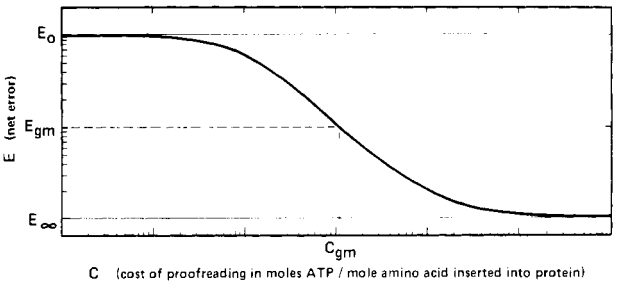

FIGURE 3: Cost–accuracy curve for proofreading. E_0 is the low-cost error asymptote; E_∞ is the high-cost error asymptote; E_{gm} is the geometric mean error; C_{gm} is the corresponding energy cost. The scales are logarithmic. See the text for further discussion.

$$E_\infty = 1/(PI + 1) \quad (8)$$

which is the minimum net error that can be realized. Thus, even in the case of indiscriminate substrate recognition ($I = 1$), fidelity could be improved solely by proofreading if sufficient energy were to be expended.

(3) Cost Corresponding to the Geometric Mean Error. The geometric mean of the two net error asymptotes is given by

$$E_{gm} = \sqrt{E_0 E_\infty} = (I + 1)^{-1/2}(PI + 1)^{-1/2} \quad (9)$$

The cost of proofreading to achieve this degree of fidelity can be obtained from eq 6 and 9:

$$C_{gm} = \left(\frac{P - 1}{PI + 1}\right) + \left(\frac{I + 1}{PI + 1}\right)^{1/2} \quad (10)$$

This is also the cost of proofreading to achieve 50% (on a log scale) of the reduction in net error from E_0 to E_∞.

For convenience, proofreading systems that achieve less than 25% (on a log scale) of the reduction in net error from E_0 to E_∞ will be called low cost–low accuracy systems; those that achieve between 25 and 75% of this reduction will be called intermediate cost–intermediate accuracy systems; and those that achieve greater than 75% of this reduction will be called high cost–high accuracy proofreading systems.

The graphical representation of eq 6 and the significant features described above are given in Figure 3.

The relationship in eq 6 provides a general constraint among the various quantities that must be satisfied in vivo for any branched network of the form considered, independently of the detailed mechanisms that determine the discrimination at each branch point. However, since the quantities in eq 6 are not easily measured in vivo, the comparisons in the next section involve extrapolation from measurements made in vitro.

Generalizations. The results given in this section represent the essential features of our theory. In addition, several generalizations have been considered, four of which will be discussed briefly.

(1) If more than one stage of (hydrolytic) proofreading is involved, the same theory can be applied to each stage in succession. The ratio of the "correct" to "incorrect" flux emerging from any given stage of proofreading can be regarded as if it were an "initial discrimination ratio" preceding the next stage of proofreading. The accuracy will improve with each additional stage of proofreading, and the cost will equal the sum of the individual costs for each stage of proofreading. The cost of proofreading can be distributed among two or more stages to yield a system that is more accurate than one in which the same total amount of energy is expended in a single stage of proofreading.

(2) For cases in which there are multiple types of amino acid errors or multiple types of tRNA errors, the parameters c and d in the preceding analysis can be replaced by $\sum_{i=1}^{n} c_i$ and $\sum_{i=1}^{n} d_i$, where n is the number of types, and exactly the same results are obtained.

(3) For an asymmetrical network (i.e., one in which the I_1, P_1, E_1, and C_1 associated with tRNA species 1 are not all equal to the corresponding I_2, P_2, E_2, and C_2 associated with tRNA species 2), decreasing the net error with respect to one of the two different tRNA species by increasing C_1 results in a coincident *increase* in the net error with respect to the other tRNA species, E_2.

(4) If errors in amino acid recognition and errors in tRNA recognition occur in the same system, then the errors are nearly additive. That is, the net error for the system with errors only in amino acid recognition can be added to the net error for the system with errors only in tRNA recognition, and the resulting error will be nearly the same as the net error for the system with both types of recognition errors. There is a slight correction for the case in which an error in amino acid recognition and an error in tRNA recognition result in a correctly aminoacylated tRNA.

The results in these last two cases may not be intuitively clear, but a more detailed elucidation is beyond the scope of this article.

Cost–Accuracy Performance of Various Proofreading Systems

In this section we use data available in the literature to examine the performance of various proofreading systems in light of the theory developed in the preceding section.

Phenylalanyl-tRNA Synthetase and Isoleucyl-tRNA^{Phe}. This system in *Escherichia coli* was among the first for which experimental data became available and a model for proofreading tRNA errors was proposed (Yarus, 1972a). In this model, the synthetase cognate with the tRNA portion of the incorrect aminoacyl-tRNA was assigned the proofreading role. Thus, it was assumed that the incorrect aminoacyl-tRNA is formed on the amino acid's cognate synthetase, released, and then with a certain probability hydrolyzed by the tRNA's cognate synthetase. Although this model does not correspond to the schematic model in Figure 1, it can be shown that the theory in the previous section applies in either case.

From the work of Yarus, one can obtain estimates of the relevant parameters. In vitro, the relative affinity of tRNA^{Ile} and tRNA^{Phe} for isoleucyl-tRNA synthetase is greater than or equal to 2.4×10^4, and isoleucyl-tRNA synthetase has a maximal velocity for the formation of isoleucyl-tRNA^{Ile} about 350 times that for the formation of isoleucyl-tRNA^{Phe} (Yarus, 1972b). If it is assumed that in vivo tRNA^{Ile} and tRNA^{Phe} are present at the same concentration and compete for isoleucyl-tRNA synthetase, then the relative rates of isoleucyl-tRNA^{Ile} and isoleucyl-tRNA^{Phe} synthesis are given by $(V_{max}/K_M)_{Ile}/(V_{max}/K_M)_{Phe}$. Substitution of the parameter values determined in vitro by Yarus (1972b) into this expression yields a value of 8.5×10^6. When this value is corrected for the fact that phenylalanyl-tRNA^{Phe} is synthesized at approximately 0.71 times the rate of isoleucyl-tRNA^{Ile} (Roberts et al., 1963), an estimate of 6.0×10^6 is obtained for the initial discrimination ratio I. Phenylalanyl-tRNA synthetase hydrolyzes isoleucyl-tRNA^{Phe} approximately 200 times faster than phenylalanyl-tRNA^{Phe} (Yarus, 1972a) (i.e., $c/d = 200a/b$ when the different sizes of pools for the species to be proofread are taken into account). Thus, the value of P, the proofreading discrimination ratio, is approximately 200. Furthermore, phenylalanyl-tRNA^{Phe} is hydrolyzed about 200 times slower than the rate of phenylalanyl-tRNA^{Phe} synthesis (i.e., $a/b \simeq 5 \times 10^{-3}$) since isoleucyl-tRNA^{Phe} is hydrolyzed at about the same rate as phenylalanyl-tRNA^{Phe} synthesis (Yarus, 1972a).

From the values in the previous paragraph, the proofreading cost C and the corresponding error E for this system are calculated to be about 0.0050 (mol of ATP per mol of amino acid inserted into protein) and 8.4×10^{-8}, respectively. This "operating point" for the phenylalanine–isoleucine system is plotted in Figure 4 on the theoretical "cost–accuracy" curve for this system (curve a). These results suggest that this is a low cost–low accuracy system. However, because the initial discrimination ratio is large, the net error is not inconsistent with the overall accuracy of protein synthesis in vivo.

Isoleucyl-tRNA Synthetase and Isoleucyl-tRNA^{Phe}. Experimental data for this system in *E. coli* are provided by the work of Yarus (1972b) and more recently by Yamane &

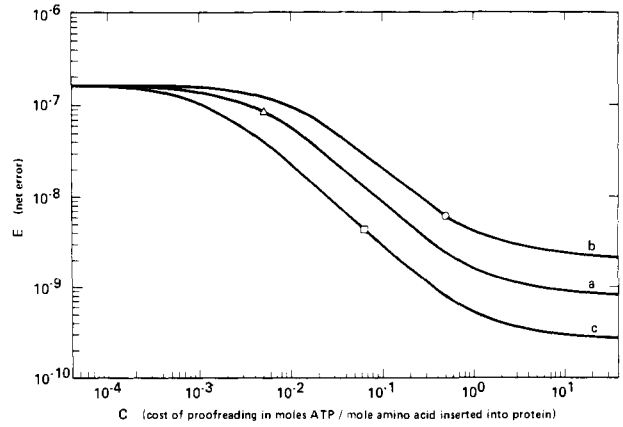

FIGURE 4: Cost–accuracy curves and operating points calculated for various proofreading systems. (a) Phenylalanyl-tRNA synthetase proofreading isoleucyl-tRNA^{Phe} according to the data of Yarus (1972a) (△). (b) Isoleucyl-tRNA synthetase proofreading isoleucyl-tRNA^{Phe} according to the data of Yamane & Hopfield (1977) (○). (c) Isoleucyl-tRNA synthetase proofreading isoleucyl-tRNA^{Phe} according to the data of Mulvey & Fersht (1977) (□). The scales are logarithmic. See the text for further discussion.

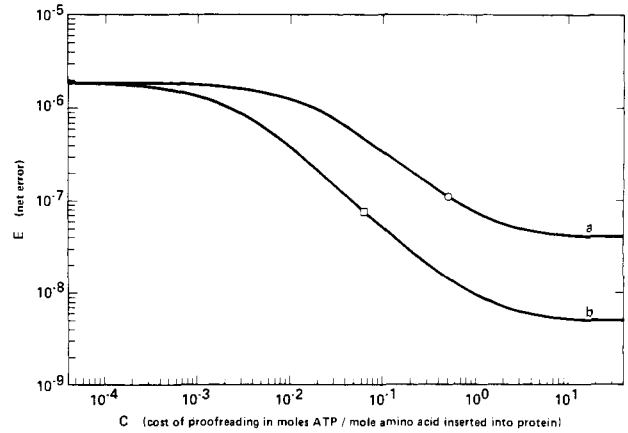

FIGURE 5: Cost–accuracy curves and operating points calculated for vairous proofreading systems. (a) Isoleucyl-tRNA synthetase proofreading isoleucyl-tRNA^{fMet} according to the data of Yamane & Hopfield (1977) (○). (b) Isoleucyl-tRNA synthetase proofreading isoleucyl-tRNA^{fMet} according to the data of Mulvey & Fersht (1977) (□). The scales are logarithmic. See the text for further discussion.

Hopfield (1977) and Mulvey & Fersht (1977). The proofreading in this case is according to the model in Figure 1. The initial discrimination ratio I is 6.0×10^6 according to the data of Yarus (1972b) and the arguments given in the previous subsection.

According to Hopfield and his colleagues, the c/d ratio is 39 (Yamane & Hopfield, 1977) whereas the a/b ratio is 0.50 (Hopfield et al., 1976). Thus, the proofreading discrimination ratio P has a value of 78. From these values a proofreading cost C of about 0.50 (mol of ATP per mol of amino acid inserted into protein) and a net error of 6.3×10^{-9} can be calculated. Thus, C is about fivefold more than C_{gm}, which has a value of 0.11 for this case. The net error E is about threefold greater than the minimum net error E_∞, which in this instance is 2.1×10^{-9}. The operating point is plotted in Figure 4 on the theoretical cost–accuracy curve for this case (curve b). These results suggest that this is a high cost–high accuracy system with a net error that is consistent with the overall accuracy of protein synthesis. The proofreading cost for this case is 100 times greater than that for the system in the previous subsection, assuming that these systems function independently. The resulting net error, however, is 13 times less.

According to Mulvey & Fersht (1977), the a/b ratio for this system is 0.064. If the corresponding c/d ratio has the value 39 determined by Yamane & Hopfield (1977), then the proofreading discrimination ratio P has a value of 610. A proofreading cost C of 0.064 (mol of ATP per mol of amino acid inserted into protein) and a net error of 4.4×10^{-9} can be calculated from the above data. In this case C is about 60% greater than the corresponding C_{gm}, which has a value of 0.040. The net error E is about 16-fold greater than the minimum net error E_∞, which is 2.7×10^{-10}. The operating point is plotted in Figure 4 on the theoretical cost–accuracy curve for this case (curve c). The system in this case appears to be an intermediate cost–intermediate accuracy system. The proofreading cost is about 13-fold higher than the corresponding estimate calculated with the data of Yarus (1972b) and about eightfold lower than the corresponding estimate calculated with the a/b ratio of Hopfield et al. (1976). The net error calculated with the a/b ratio of Mulvey & Fersht (1977) is about 30% less than that calculated with the a/b ratio of Hopfield et al. (1976).

Isoleucyl-tRNA Synthetase and Isoleucyl-tRNA^{fMet}. Data for this system in *E. coli* are provided by Mertes et al. (1972), Yamane & Hopfield (1977), and Mulvey & Fersht (1977). Again, the proofreading is according to the model in Figure 1; the analysis is slightly different from that in the previous case.

In vitro, the relative affinity of tRNA^{Ile} and tRNA^{fMet} for isoleucyl-tRNA synthetase is greater than or equal to 3.3×10^3, and isoleucyl-tRNA synthetase has a maximal velocity for the formation of isoleucyl-tRNA^{Ile} about 1900 times that for the formation of isoleucyl-tRNA^{fMet} (Mertes et al., 1972). From an average molecular weight of approximately 50 000 for the proteins of *E. coli* (O'Farrell, 1975) and the amino acid composition of *E. coli* (Roberts et al., 1963), it can be calculated that the rate of synthesis for fMet-tRNA^{fMet} in vivo is about 0.086 times that for isoleucyl-tRNA^{Ile}. With these values and the assumptions stated in the preceding subsections, the initial discrimination ratio I is estimated to be 5.4×10^5.

Yamane & Hopfield (1977) have determined values of 24 for the c/d ratio and 0.50 for the a/b ratio. Their proofreading discrimination ratio P is therefore 48. With these parameter values and the value of I given in the previous paragraph, the proofreading cost C is calculated to be 0.50 (mol of ATP per mol of amino acid inserted into protein), about fourfold greater than C_{gm} for this system. The corresponding net error E is 1.1×10^{-7}, or about threefold higher than the minimum net error $E_\infty = 3.9 \times 10^{-8}$. The relative position of the operating point on the theoretical cost–accuracy curve for this system is shown in Figure 5 (curve a).

The a/b ratio for this same system has a value of 0.064 according to Mulvey & Fersht (1977). Again, if the corresponding c/d ratio has the value 24 determined by Yamane & Hopfield (1977), then the proofreading ratio P has a value of 380. A proofreading cost $C = 0.064$ (mol of ATP per mol of amino acid inserted into protein) and a net error $E = 7.9 \times 10^{-8}$ can be calculated from the above data. This value of C is approximately 25% greater than the corresponding C_{gm}, which has a value of 0.051. The net error E is about 16-fold greater than the minimum net error E_∞, which is 4.9×10^{-9}. The operating point is plotted in Figure 5 on the corresponding cost–accuracy curve (curve b). Again, the cost calculated with the a/b ratio of Mulvey & Fersht (1977) is about eightfold lower than that calculated with the a/b ratio of Hopfield et al. (1976), and the corresponding net error is about 30% less.

Isoleucyl-tRNA Synthetase and Valyl-tRNA^{Ile}. The initial

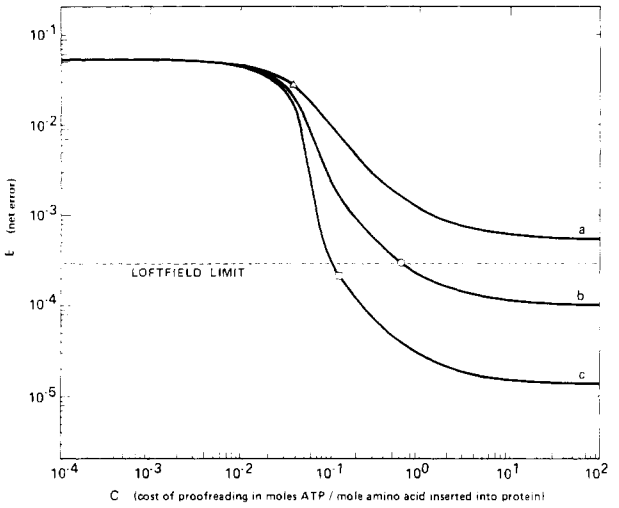

FIGURE 6: Cost–accuracy curves and operating points calculated for various proofreading systems. (a) Isoleucyl-tRNA synthetase proofreading valyl-tRNA^{Ile} according to the data of Schreier & Schimmel (1972) (△). (b) Isoleucyl-tRNA synthetase proofreading valyl-tRNA^{Ile} according to the data of Hopfield et al. (1976) (○). (c) Isoleucyl-tRNA synthetase proofreading valyl-tRNA^{Ile} according to the data of Mulvey & Fersht (1977) (□). The scales are logarithmic. See the text for further discussion.

discrimination ratio I for this system, which is represented in Figure 2, can be estimated from the kinetic studies of Baldwin & Berg (1966) and Loftfield & Eigner (1966). These authors found that, in *E. coli*, isoleucyl-tRNA synthetase has an affinity for valine that is 1/100 that for the correct substrate isoleucine, whereas the maximal velocities appear to be approximately the same for both substrates. If it is assumed that isoleucine and valine compete for isoleucyl-tRNA synthetase and that the ratio of their concentrations in vivo is 0.18 (Raunio & Rosenqvist, 1970), then the initial discrimination ratio I can be estimated with the formula $(V_{max}/K_M)_{Ile}$·[Ile]/[$(V_{max}/K_m)_{Val}$[Val]] to be 18.

Early experiments by Schimmel and his colleagues were designed to examine the hydrolysis of preformed isoleucyl-tRNA^{Ile} and valyl-tRNA^{Ile} by isoleucyl-tRNA synthetase.[1] From the data of Eldred & Schimmel (1972), it can be estimated that isoleucyl-tRNA synthetase hydrolyzes valyl-tRNA^{Ile} approximately 100 times faster than isoleucyl-tRNA^{Ile} (i.e., $c/d = 100a/b$ when the differences in pool size for the species to be proofread are taken into account). Thus, the value of P, the proofreading discrimination ratio, is approximately 100. Furthermore, Schreier & Schimmel (1972) found that isoleucyl-tRNA synthetase hydrolyzes isoleucyl-tRNA^{Ile} about 100 times slower than it synthesizes isoleucyl-tRNA^{Ile} (i.e., $a/b \simeq 10^{-2}$). From these values and the value of I in the previous paragraph, the proofreading cost C and the net error E for this system are calculated to be 0.037 (mol of ATP per mol of amino acid inserted into protein) and 2.7×10^{-2}, respectively. The operating point is plotted in Figure 6 on the corresponding cost–accuracy curve for this system (curve a). The cost is about fivefold less than C_{gm}, and the net error is reduced only twofold by proofreading. This is a low cost–low accuracy system that fails to reduce the net error below the Loftfield limit[2] of 3×10^{-4}.

Hopfield et al. (1976) have examined this same system but under conditions in which synthesis and proofreading of an aminoacyl-tRNA molecule can occur on the same synthetase molecule. They obtained an a/b ratio of 0.50 and a c/d ratio of 270, which yield a proofreading discrimination ratio P of 540. From these values and the value given above for I, the proofreading cost C is calculated to be 0.58 (mol of ATP per mol of amino acid inserted into protein). This is about sixfold greater than $C_{gm} = 0.10$. The corresponding net error E is 3.4×10^{-4}, which is about threefold higher than the minimum net error, $E_\infty = 1.0 \times 10^{-4}$, for this system. The operating point is plotted in Figure 6 on the theoretical cost–accuracy curve for the system (curve b). These results suggest that this case represents a high cost–high accuracy proofreading system. The resulting net error is about that associated with the overall accuracy of protein synthesis.

Mulvey & Fersht (1977) also have examined this system under conditions similar to those used by Hopfield et al. (1976), but the value they obtained for the a/b ratio was less, 0.064 vs. 0.50. If the corresponding c/d ratio has the value 270 determined by Hopfield et al. (1976), then the proofreading discrimination ratio P would have a value of 4200. A proofreading cost C of about 0.12 (mol of ATP per mol of amino acid inserted into protein) and a net error E of 2.2×10^{-4} can be calculated as before but with the a/b value determined by Mulvey & Fersht (1977). In this case C is about 69% greater than C_{gm}, which has a value of 0.071. The net error E is about 17-fold greater than the minimum net error E_∞, which is 1.3×10^{-5}. The operating point for this case is plotted in Figure 6 on the corresponding cost–accuracy curve (curve c). The proofreading system in this case is intermediate cost–intermediate accuracy. However, because of the large proofreading discrimination ratio P, the resulting net error is just below the Loftfield limit.

Valyl-tRNA Synthetase and Threonyl-tRNA^{Val}. This, the last system we shall examine, is also represented by the model in Figure 2. The K_M of valyl-tRNA synthetase for threonine is about 100-fold greater than that for its true substrate, valine, whereas its maximal velocity with threonine as substrate is about 30% of that with valine as substrate (Bergmann et al., 1961). By assuming that the concentration of valine is about 40 times the concentration of threonine in the cell (Raunio & Rosenqvist, 1970) and that these two amino acids are competing for valyl-tRNA synthetase, an initial discrimination ratio I = 13 000 can be calculated.

Yamane & Hopfield (1977) have examined this system and obtained estimates of 0.40 for the a/b ratio and 400 for the c/d ratio. These values, which yield a proofreading discrimination ratio P = 1000, together with the value for I given in the previous paragraph can be used to calculate the proofreading cost C. In this case the cost, 0.40 (mol of ATP per mol of amino acid inserted into protein), is 13 times greater than $C_{gm} = 0.032$. The corresponding net error $E = 2.7 \times 10^{-7}$, and the minimum net error $E_\infty = 7.7 \times 10^{-8}$. Thus, there is an approximate threefold difference as in the previous three examples that involved the a/b ratio determined by Hopfield et al. (1976). The operating point and theoretical cost–accuracy curve for this proofreading system are represented in Figure 7 (curve a). These results suggest that this is a high

[1] In this case the synthesis and proofreading of an aminoacyl-tRNA molecule occur on the surface of different synthetase molecules. The schematic model of the implied proofreading system would be slightly different from that represented in Figure 2, but it can be shown that the theory in the previous section applies in either case.

[2] The error limits determined originally by Loftfield and his colleagues were for valine incorrectly substituted for isoleucine in chicken ovalbumin (Loftfield, 1963) and rabbit hemoglobin (Loftfield & Vanderjagt, 1972). However, Edelmann & Gallant (1977) recently have obtained approximately the same error limit for cysteine incorrectly substituted for arginine in *E. coli* flagellin.

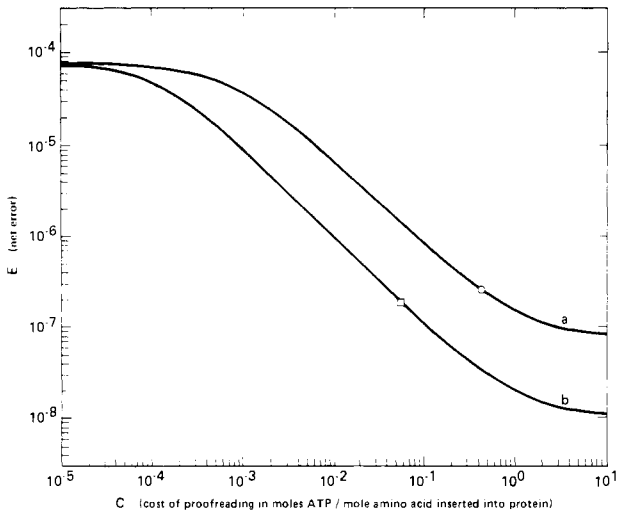

FIGURE 7: Cost–accuracy curves and operating points calculated for various proofreading systems. (a) Valyl-tRNA synthetase proofreading threonyl-tRNA^{Val} according to the data of Yamane & Hopfield (1977) (○). (b) Valyl-tRNA synthetase proofreading threonyl-tRNA^{Val} according to the data of Mulvey & Fersht (1977) (□). The scales are logarithmic. See the text for further discussion.

cost–high accuracy proofreading system.

Mulvey & Fersht (1977) also have measured the stoichiometry of ATP hydrolysis and valyl-tRNA^{Val} synthesis by valyl-tRNA synthetase and determined the a/b ratio to be 0.053. By use of the other parameter values given in the previous paragraph, a proofreading cost $C = 0.053$ (mol of ATP per mol of amino acid inserted into protein) and a net error $E = 2.0 \times 10^{-7}$ can be calculated. In this case, C is about fivefold greater than C_{gm}, which is 0.012, and E is about 20-fold greater than the minimum net error E_∞, which is 1.0×10^{-8}. The operating point and theoretical cost–accuracy curve for this proofreading system are given in Figure 7 (curve b). This system, with the a/b ratio of Mulvey & Fersht (1977), appears to be an intermediate cost–intermediate accuracy system.

Discussion

The first major conclusion to be drawn from this study is that high-accuracy proofreading of tRNA aminoacylation has a high cost in terms of energy consumed in proofreading. The cost is about 100 times greater than the early indications from the work of Yarus (1972a) and Schreier & Schimmel (1972) and greater than most people had first expected [e.g., see Hopfield (1974)]. The cost of proofreading to approach the minimum error is consistent with the high energy expenditure obtained in recent experiments (Hopfield et al., 1976; Yamane & Hopfield, 1977). Yamane & Hopfield (1977) have speculated that the cost of such "surveillance functions" might account for the large fraction of energy generated by the cell that is currently unaccounted for by known biosynthetic processes (Forrest & Walker, 1971; Hempfling & Mainzer, 1975; Lagunas, 1976). According to calculations presented elsewhere (Savageau & Freter, 1979), the cost of proofreading all types of tRNA aminoacylations would be approximately 4% of the currently unaccounted for energy.

The method used by Hopfield et al. (1976) for determining the a/b ratio gives a value that is an overestimate (Mulvey & Fersht, 1977) and should be corrected by 50%. The corresponding net error E would then be reduced by 17% (Savageau & Freter, 1979). However, a more serious problem is the choice of appropriate conditions in vitro to represent the situation in vivo. This is particularly difficult because values for the a/b ratio appear to be highly dependent upon assay conditions; the data of Mulvey & Fersht (1977), which appear to be most accurate, yield values that cover almost the entire range from the low estimates of Schreier & Schimmel (1972) to the high estimates of Hopfield et al. (1976).

For calculations involving the data of Mulvey & Fersht (1977), we have used their average values for the a/b ratios. Since these authors did not report independent measurements for the corresponding c/d ratios, we assumed the values[3] reported by Hopfield and his colleagues. The result in each case is a net error similar to that estimated with the a/b ratio of Hopfield and his colleagues, but the proofreading cost is five- to eightfold less. This fidelity is possible without the accompanying cost because the proofreading discrimination ratio P has increased, which is a necessary consequence of our assumption concerning c/d ratios. In the case of isoleucyl-tRNA synthetase editing valyl-tRNA^{Ile}, the proofreading discrimination ratio increases to a value of 4200. This value appears to be unreasonably high because it represents the ability of the synthetase to discriminate between two molecules that differ only by a methyl group (see the beginning of the text). By contrast, the proofreading discrimination ratio obtained solely from the data of Hopfield et al. (1976) has a value of 540.

An alternative but equally justifiable assumption, which has been used elsewhere (Savageau & Freter, 1979), is that the proofreading discrimination ratio P is the same for the assay conditions of Mulvey & Fersht (1977) and those of Hopfield et al. (1976). In other words, the a/b ratio and the corresponding c/d ratio are equally affected by the differences in assay conditions. The values for the c/d ratios corresponding to the a/b ratios determined by Mulvey & Fersht (1977) follow accordingly. The result in each case is a cost–accuracy curve that is identical with that determined from the data of Hopfield and colleagues (see Figures 4–7). However, since the proofreading costs calculated from the data of Mulvey & Fersht (1977) are five- to eightfold less, the corresponding net errors are about four- to fivefold greater (see Table I). In the case of isoleucyl-tRNA synthetase proofreading valyl-tRNA^{Ile} (Figure 6), the net error is considerably greater than the Loftfield limit.

Another alternative is that the error following a single stage of proofreading is, in fact, higher than the Loftfield limit. If this view is correct, then there must be one or more subsequent stages of proofreading that eventually reduce the error to a value less than 3×10^{-4} for stable, finished protein. This possibility is treated in detail elsewhere (Freter and Savageau, unpublished results).

The second major conclusion of this study is that the minimum net error for tRNA aminoacylation, which was first derived by Hopfield (1974) and Ninio (1975), is never achieved. This statement implies that the minimum net error is reached only asymptotically as the cost of proofreading becomes infinite. More importantly, it implies that no system whose errors are critical would operate on or near the high-cost asymptote, since a lower cost of proofreading could always be selected without an appreciable increase in net error. As we have argued elsewhere (Savageau & Freter, 1979), the operating point for such a system will generally lie along the inclined portion of the cost–accuracy curve for the system. This conclusion is further supported by our analysis of the available experimental data for tRNA aminoacylation networks (see Figures 4–7).

[3] The possibility that these experimental values are overestimates will be considered elsewhere along with our treatment of multistage proofreading (Freter and Savageau, unpublished results).

Table I: Cost of Proofreading and the Accuracy Achieved in Various Proofreading Systems[a]

proof-reading synth-etase	proof-read amino-acyl-tRNA	ref	C[b]	E	E_∞
		tRNA Errors			
Phe	Ile-tRNA^{Phe}	*c*	0.0050	8.4×10^{-8}	8.3×10^{-10}
Ile	Ile-tRNA^{Phe}	*d*	0.50	6.3×10^{-9}	2.1×10^{-9}
		e	0.064	4.4×10^{-9}	2.7×10^{-10}
		f	0.064	3.0×10^{-8}	2.1×10^{-9}
Ile	Ile-tRNA$_f^{Met}$	*d*	0.50	1.1×10^{-7}	3.9×10^{-8}
		e	0.064	7.9×10^{-8}	4.9×10^{-9}
		f	0.064	4.9×10^{-7}	3.9×10^{-8}
		Amino Acid Errors			
Ile	Val-tRNA^{Ile}	*g*	0.037	2.7×10^{-2}	5.6×10^{-4}
		d	0.58	3.1×10^{-4}	1.0×10^{-4}
		e	0.12	2.2×10^{-4}	1.3×10^{-5}
		f	0.12	1.7×10^{-3}	1.0×10^{-4}
Val	Thr-tRNA^{Val}	*d*	0.40	2.7×10^{-7}	7.7×10^{-8}
		e	0.053	2.0×10^{-7}	1.0×10^{-8}
		f	0.053	1.5×10^{-6}	7.7×10^{-8}

[a] See Figures 4–7 and the text for further discussion. [b] Cost of proofreading in mol of ATP per mol of amino acid inserted into protein. [c] Calculated from the experimental data of Yarus (1972a,b). [d] Calculated from the experimental data of Hopfield et al. (1976) and Yamane & Hopfield (1977). [e] Calculated from the experimental data of Mulvey & Fersht (1977) by assuming the c/d ratios determined from the data of Hopfield et al. (1976) and Yamane & Hopfield (1977). [f] Calculated from the experimental data of Mulvey & Fersht (1977) by assuming the P ratios determined from the data of Hopfield et al. (1976) and Yamane & Hopfield (1977). [g] Calculated from the experimental data of Schreier & Schimmel (1972) and Eldred & Schimmel (1972).

Our results also bear on the relevance of different mechanisms for proofreading. Those that involve aminoacyl-tRNA synthesis with subsequent proofreading on the surface of the same synthetase molecule are relatively efficient, whereas mechanisms that require aminoacyl-tRNA molecules synthesized on one synthetase molecule to diffuse to a second synthetase molecule for proofreading are relatively inefficient. The proofreading discrimination ratio may be identical for these two types of mechanisms, yet the former achieve much lower net errors because they are able to proofread at much higher rates. This is clearly seen in Figure 6 by comparing the results of experiments that allow the first type of mechanism to operate [e.g., Hopfield et al. (1976) and Mulvey & Fersht (1977)] with the results of experiments that allow the second type of mechanism to operate [e.g., Schreier & Schimmel (1972)]. The proofreading rates measured for the same aminoacyl-tRNA and the same synthetase differ by a factor of 10–50. The inefficiency results at least in part from the inevitable dilution of effective substrate concentration for the proofreading reaction that occurs in the second type of mechanism. Almost all of the early assays for proofreading were performed in the second manner involving hydrolysis of preformed aminoacyl-tRNA by synthetase. Hence, one expects these estimated rates to be low. This explanation accounts for the low accuracy–low cost proofreading of isoleucyl-tRNA^{Phe} by phenylalanyl-tRNA synthetase (see Figure 4a). It also may account in part for reports that some misacylated tRNAs are not proofread (Schreier & Schimmel, 1972; Ebel et al., 1973). Furthermore, if efficient proofreading must occur on the surface of the same synthetase involved in the misacylation of tRNA, then early reports of indiscriminate proofreading of correct aminoacyl-tRNAs by noncognate synthetases (Bonnet et al., 1972; Ebel et al., 1973) may not be physiologically relevant. [See also the kinetic arguments of Mulvey & Fersht (1977) and the discussion of chemical mechanisms by Hecht (1977).]

Finally, it should be noted that our theory of proofreading is applicable to more than just tRNA aminoacylation. As has been noted already by many others, similar types of proofreading appear to occur along with many important cellular functions including DNA replication, RNA transcription, and protein translation. All such proofreading processes can be represented by branching networks of the kind we have analyzed and, in principle, similar results may be obtained.

Added in Proof

Recent calculations (Fersht, 1979) show that the data of Yamane & Hopfield (1977) for hydrolytic proofreading of tRNA^{fMet} and tRNA^{Phe} errors by isoleucyl-tRNA synthetase are undoubtedly artifactual. In these systems there is now no evidence for proofreading. This conclusion in no way affects the validity of the theory we have presented and is consistent with our own calculations showing that the high initial discrimination ratios for these systems ensure sufficient accuracy even without proofreading. However, this conclusion does emphasize the point that a complete set of reliable data with which to apply our theory still has not been reported for any system.

Acknowledgments

We thank Michael Yarus, who introduced us to this problem and who contributed useful comments and suggestions during the preparation of the manuscript, and Ann Marie Kotre, who critically read and considerably improved the manuscript.

References

Alford, B., & Hecht, S. M. (1978) *J. Biol. Chem. 253*, 4844–4850.
Baldwin, A. N., & Berg, P. (1966) *J. Biol. Chem. 241*, 839–845.
Bergmann, F. H., Berg, P., & Dieckmann, M. (1961) *J. Biol. Chem. 236*, 1735–1740.
Bonnet, J., Giege, R., & Ebel, J. P. (1972) *FEBS Lett. 27*, 139–144.
Ebel, J. P., Giege, R., Bonnet, J., Kern, D., Befort, N., Bollack, C., Fasiolo, F., Gangloff, J., & Dirheimer, G. (1973) *Biochimie 55*, 547–557.
Edelmann, P., & Gallant, J. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 3396–3398.
Eldred, E. W., & Schimmel, P. R. (1972) *J. Biol. Chem. 247*, 2961–2964.
Fersht, A. (1979) in *Transfer RNA* (Schimmel, P., Söll, D., & Abelson, J., Eds.) Part 1, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (in press).
Fersht, A. R., & Jakes, R. (1975) *Biochemistry 14*, 3350–3356.
Fersht, A. R., & Kaethner, M. M. (1976) *Biochemistry 15*, 3342–3346.
Forrest, W. W., & Walker, D. J. (1971) *Adv. Microb. Physiol. 5*, 213–274.
Hecht, S. M. (1977) *Tetrahedron 33*, 1671–1696.
Hempfling, W. P., & Mainzer, S. E. (1975) *J. Bacteriol. 123*, 1076–1087.
Hopfield, J. J. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 4135–4139.
Hopfield, J. J., Yamane, T., Yue, V., & Coutts, S. M. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 1164–1168.
Lagunas, R. (1976) *Biochim. Biophys. Acta 440*, 661–674.
Loftfield, R. B. (1963) *Biochem. J. 89*, 82–92.

Loftfield, R. B., & Eigner, E. A. (1966) *Biochim. Biophys. Acta 130*, 426–448.
Loftfield, R. B., & Vanderjagt, D. (1972) *Biochem. J. 128*, 1353–1356.
Mertes, M., Peters, M. A., Mahoney, W., & Yarus, M. (1972) *J. Mol. Biol. 71*, 671–685.
Mulvey, R. S., & Fersht, A. R. (1977) *Biochemistry 16*, 4731–4737.
Ninio, J. (1975) *Biochimie 57*, 587–595.
O'Farrell, P. H. (1975) *J. Biol. Chem. 250*, 4007–4021.
Pauling, L. (1958) in *Festschrift Arthur Stoll* pp 597–602, Birkhaeuser, Basel, Switzerland.
Raunio, R., & Rosenqvist, H. (1970) *Acta Chem. Scand. 24*, 2737–2744.
Roberts, R. B., Abelson, P. H., Cowie, D. B., Bolton, E. T., & Britten, R. J. (1963) *Studies of Biosynthesis in Escherichia coli*, Publication No. 607, Carnegie Institute of Washington, Washington, D.C.
Savageau, M. A. (1976) *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology*, Addison-Wesley, Reading, MA.
Savageau, M. A., & Freter, R. R. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76* (in press).
Schreier, A. A., & Schimmel, P. R. (1972) *Biochemistry 11*, 1582–1589.
von der Haar, F., & Cramer, F. (1976) *Biochemistry 15*, 4131–4138.
Yamane, T., & Hopfield, J. J. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 2246–2250.
Yarus, M. (1972a) *Proc. Natl. Acad. Sci. U.S.A. 69*, 1915–1919.
Yarus, M. (1972b) *Biochemistry 11*, 2352–2361.

Ambiguity and Transcriptional Errors as a Result of Methylation of N-1 of Purines and N-3 of Pyrimidines†

M. Kröger‡ and B. Singer*

ABSTRACT: Poly(A), poly(C), and poly(U) containing about 10% modified nucleoside were used as templates in transcription experiments using DNA-dependent RNA polymerase in the presence of Mn^{2+}. The presence of 3-methylcytidine, 3-methyluridine, 1-methyladenosine, and N^6-methyladenosine did not prevent transcription but decreased the rate. Polymers containing xanthosine were transcribed with the same rate as homopolymers, but those with 11% 1-methylguanosine were virtually inactive as templates. Nearest-neighbor analysis of products of transcription of various copolymers showed that, in the presence of all four nucleoside triphosphates, 3-methylcytidine was able to direct AMP, CMP, and UMP equally well into the complementary strand. 1-Methyladenosine also directed incorporation of AMP, GMP, CMP, and UMP but with a preference for AMP and UMP. 3-Methyluridine directed AMP and UMP incorporation. 3-Methyluridine-directed CMP incorporation was low but could be verified when ATP and UTP were absent. Polymers containing N^6-methyladenosine, xanthosine, or 5-fluorouridine directed incorporation of the expected complementary nucleotide only. The data presented indicate that, under competitive conditions (all four NTPs present), nucleosides modified on the N-3 of Urd or Cyd and the N-1 of Ado have little or no specificity in transcription, thus behaving ambiguously. Although incorporation of GMP could only be found for 1-methyladenosine due to technical problems, it is likely that the N-3 methylated pyrimidines also direct GMP. It is postulated that no specific hydrogen bonds are formed between the modified bases in the template and the nucleoside triphosphate, but instead other factors such as stacking forces and the RNA polymerase itself direct incorporation of two, three, or four nucleotides. In templates, 3-methylcytidine, 3-methyluridine, and 1-methyladenosine are by this criterion mutagenic.

The general understanding regarding the mechanism of point mutation is that a base is modified in such a way that it behaves in transcription or translation as if it were another base. This assumes that the proper number and location of potential hydrogen bonds are required for a base to become incorporated. However, in vitro testing of the template activity of modified bases in polynucleotides has led to results which cannot be interpreted as resulting from normal hydrogen bonding (Topal & Fresco, 1976; Singer et al., 1978). Certain alkylated bases have already been studied in terms of their base-pairing abilities, but there is no uniform theory to fit the reported properties: 3-MeC is reported to be a mutagenic base, acting like A, U, or G in transcription (Ludlum, 1970a; Singer & Fraenkel-Conrat, 1970). Three *O*-alkyl bases have been studied in the same way more recently. O^6-MeG acts like A or U (Gerchman & Ludlum, 1973; Mehta & Ludlum, 1978), while O^2- and O^4-alkyl U act like C or G in transcription (Singer et al., 1978) but like U in tRNA binding (Singer et al., 1979). The presence of 3-MeU was considered to be a lethal event (Szer & Shugar, 1961; Grunberger et al., 1968), but in doublets (pGp3MeU) it was found that 3-MeU could function as C in tRNA binding (Singer et al., 1979).

Thrierr & Leng (1971) found that 15% 1-MeA in poly(A) appears to allow helix formation with poly(U), but the 1-MeA prevented binding of DNA-dependent RNA polymerase to such polymers. However, these polymers were prepared by methylation of poly(A) and not by de novo polymerization. Pochon & Michelson (1967) reported 1-MeG as being incapable of forming a base pair, both on experimental and theoretical grounds.

Methylation of the N-1 of purines and the N-3 of pyrimidines should have the same effect in preventing transcription if only base pairing were involved. The fact that 3-MeC is accepted as being mutagenic while 3-MeU is generally

† From the Department of Molecular Biology and Virus Laboratory, University of California, Berkeley, California 94720. *Received March 22, 1979.* This investigation was supported by Grant CA12316 from the National Cancer Institute.
‡ M. K. is a Fellow of the Deutsche Forschungsgemeinschaft.

0006-2960/79/0418-3493$01.00/0