THERMODYNAMIC CONSTRAINTS ON KINETIC PROOFREADING IN BIOSYNTHETIC PATHWAYS

MÅNS EHRENBERG, *Department of Medical Biophysics, Karolinska Institutet, S-104 01 Stockholm 60, Sweden*
CLAS BLOMBERG, *Research Group for Theoretical Biophysics, Royal Institute of Technology, S-100 44 Stockholm, Sweden*

ABSTRACT We develop a quantitative theory of kinetic proofreading with an arbitrary number of checking steps after the hydrolysis of a nucleoside triphosphate. In particular, we investigate the relationship between the minimum dissipation of free energy required for a given error frequency in such systems. Several conclusions can be drawn from the present treatment: first, the ultimate accuracy of error correcting selective pathways is set by the displacement from equilibrium of the nucleoside triphosphates. Second, it is advantageous to achieve a desired accuracy at a small energy dissipation with several checking steps rather than a single one. This could explain antinomies in the amino acylation reaction as well as in mRNA translation, where small structural differences lead to large differences in flow rates between right and wrong substrates. Third, all checking steps should contribute equally to the accuracy, which implies a specific and symmetrical set of rate constants for the checking events on the enzyme.

INTRODUCTION

Gene replication, translation of the messenger-RNA on the ribosomes, and the amino acylation reaction have low error rates. For DNA replication the probability of an error ranges between 10^{-8} and 10^{-10} (Fowler et al., 1974). The probability of a mistake in the amino acylation of transfer-RNA or by a misreading of the codon triplet is $\sim 3 \times 10^{-4}$ or smaller (Loftfield, 1963; Loftfield and Vanderjagt, 1972).

In the acylation reaction several amino acids have competitors with small structural differences. Pauling (1958) argued that valine and isoleucine are so similar that it would be difficult for the synthetase to achieve an error frequency below a few percent.

Experiments on tRNAs with complementary or mismatching anticodon loops (Eisinger and Gross, 1975; Grosjean et al., 1978) show that as a rule the specificity of triplet-triplet binding free energy in such systems is not consistent with the error rate of ribosomal translation (Grosjean et al., 1978). These examples point towards a possible dilemma in selective biosynthetic pathways: the intrinsic structural selectivity d, where (c.f. Hopfield, 1974; Fersht, 1977*a*) $d = e^{\Delta G_{max}/RT}$ may not be large enough to account for the low error frequencies which have been observed in vivo. There are two different ways to resolve the problem. In the first we may argue that existing theoretical considerations and that experiments performed in artificial systems underestimate the free energy difference ΔG_{max} that exist in real systems (c.f. Kurland, et al. 1975). The other way the dilemma can be resolved (and this will be of primary interest here) is by the operation of error correcting devices.

When no error correction exists the ratio between the flow of correct (J^c) and incorrect

(J^w) product formation always fulfills the inequality

$$J^c/J^w \leq e^{\Delta G_{max}/RT} = d,$$

provided that activation free energies are also included in ΔG_{max} (c.f. Fersht, 1977*a* and below).

This limit can be exceeded by kinetic proofreading. General properties connected with such mechanisms were clarified in two theoretical works (Hopfield, 1974; Ninio, 1975). First, a discard branch is necessary in the enzymatic pathway from substrate S to product P:

$$\begin{array}{ccccc} E + S + A & \rightleftharpoons & E_1 & \rightleftharpoons & E + P + M \\ & & \downarrow\uparrow & & \\ & & E + S + M & & \end{array}$$

Second, the discard step must be thermodynamically driven out from the main pathway by its coupling to the hydrolysis of a nucleoside triphosphate A. The driving force is provided by the displacement from equilibrium of A with its hydrolytic products M as demonstrated by (Kurland, 1978):

$$\frac{[A]}{[M]} = K_{AM}\gamma,$$

$$\gamma > 1.$$

Several DNA polymerases have an exonuclease activity (Goulian et al., 1968; Brutlag and Kornberg, 1972). The suggestion (Kornberg, 1969) that this property of the polymerases is used as an error correcting device in gene replication is now widely accepted.

Experiments by Norris and Berg (1964) and Baldwin and Berg (1966) have shown that an erroneous activation of an amino acid in the acylation reaction can be subsequently corrected by hydrolysis of the AA-AMP complex.

Direct evidence for free energy driven editing mechanisms has recently been obtained with quenched stopped flow systems in several cases in the amino acylation reaction (e.g., Fersht and Kaethner, 1976; Fersht, 1977*b*).

The existence of kinetic proofreading in at least two selective biosynthetic pathways may therefore be regarded as experimentally well established. In this work theoretical aspects of error correcting mechanisms are considered. We start with the assumption that ΔG_{max} for certain combinations of competing substrates corresponds to a d value that is smaller than the required discrimination D of the system:

$$\frac{J^c}{J^w} = D > d.$$

We formulate the kinetic proofreading schemes so that the law of detailed balance is made explicit. In this way the absolute limits of accuracy as well as the necessary free energy dissipation of such systems can be determined. Finally, the optimization of these parameters, along with its consequences for the kinetic properties of the enzyme, are discussed. An alternative treatment with complementary information to this work is presented elsewhere.[1]

[1]Blomberg, C., and M. Ehrenberg. 1980. Manuscript submitted for publication.

REDUCTION OF NONBRANCHED STEADY STATE DIAGRAMS

Consider a nonbranched multistep enzymatic pathway leading from substrate S to the formation of product P:

$$E + S + A \underset{k_{10}}{\overset{k_{01}}{\rightleftharpoons}} E_1 \underset{k_{21}}{\overset{k_{12}}{\rightleftharpoons}} E_2 \underset{k_{32}}{\overset{k_{23}}{\rightleftharpoons}} \cdots \underset{k_{n,n-1}}{\overset{k_{n-1,n}}{\rightleftharpoons}} E_n \underset{k_{n+1,n}}{\overset{k_{n,n+1}}{\rightleftharpoons}} E + P + M. \tag{1}$$

The enzyme couples the reaction S → P with the degradation A → M of a cosubstrate. Define equilibrium constants for all steps in scheme 1

$$K_i = \frac{k_{i-1,i}}{k_{i,i-1}}; i = 1, 2 \ldots n + 1. \tag{2}$$

In equilibrium the concentrations of P and S are connected by

$$\left\{\frac{[P]}{[S]}\right\}_{eq} = K_{PS} = e^{(\mu_{0S} - \mu_{0P})/RT}. \tag{3}$$

μ_{0S} and μ_{0P} are standard free energies for S and P, respectively. Similarly, for A and M

$$\left\{\frac{[A]}{[M]}\right\}_{eq} = K_{AM} = e^{(\mu_{0M} - \mu_{0A})/RT}. \tag{4}$$

A state E_i is in equilibrium connected with the initial state E + S + A by products of the equilibrium constants K_i:

$$\left\{\frac{[E_i]}{[E][S][A]}\right\}_{eq} = C_i = \prod_{j=1}^{i} K_j; i = 1, 2 \ldots n. \tag{5}$$

We can also define

$$C_{n+1} = \frac{K_{PS}}{K_{AM}}.$$

A steady state flow over the enzyme with net formation of product will be established if the concentrations of P and M are shifted below equilibrium with S and A:

$$[P][M] = \frac{K_{PS}}{K_{AM}} \cdot [S][A] \cdot 1/\beta, \tag{6}$$

$$\beta > 1.$$

β is composed of the displacement of P from equilibrium with S and the displacement of A from equilibrium with M:

$$[P] = K_{PS} \cdot [S]\delta_P, \tag{7}$$

$$[A] = K_{AM} \cdot [M]\gamma, \tag{8}$$

so that

$$1/\beta = \delta_P/\gamma. \tag{9}$$

A large value of γ may be used to concentrate product molecules in relation to substrate, i.e., $\delta_P \gg 1$ can be associated with net formation of P provided that $\beta > 1$. The dissipation of free energy per mole of product is

$$W_{diss} = RT[\log(\gamma) - \log(\delta_P)] = RT \log(\beta), \quad (10)$$
$$W_{diss} > 0.$$

Define displacements from equilibrium δ_i for all states of scheme 1 according to

$$[E_i] = [E][S][A]C_i\delta_i;\ i = 1, 2 \ldots n, \quad (11)$$
$$1/\beta = \delta_{n+1}.$$

The flows over the individual steps are now

$$J_0 = J = [E][S][A]R_0(1 - \delta_1),$$
$$J_i = J = [E][S][A]R_i(\delta_i - \delta_{i+1});\ i = 1, 2 \ldots n, \quad (12)$$

where

$$R_0 = k_{01}$$
$$R_i = C_i k_{i,i+1}. \quad (13)$$

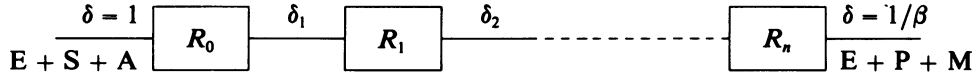
From Eq. 12 it follows that scheme 1, with respect to its input-output properties, may be reduced (c.f. Ninio, 1975):

$$J = [E][S][A]R(1 - 1/\beta),$$

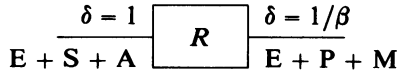
where

$$\frac{1}{R} = \sum_{i=0}^{n} \frac{1}{R_i}. \quad (14)$$

It is convenient to use the shorthand notation



for scheme 1 which illustrates the flow equations (12). Via the transformation in Eq. 14 this can now be written



SELECTIVITY IN NONBRANCHED KINETIC SCHEMES

The enzyme E discriminates between two substrates S^c and S^w by creating different rate constants R_i for the species. The flow of cognate substrate is

$$J^c = [E][S^c][A]R^c(1 - 1/\beta^c),$$

and of noncognate

$$J^{w} = [E][S^{w}][A]R^{w}(1-1/\beta^{w})$$

where, according to Eq. 14,

$$\frac{1}{R^{c}} = \Sigma \frac{1}{R_i^{c}},$$

$$\frac{1}{R^{w}} = \Sigma \frac{1}{R_i^{w}}.$$

Every step of the enzyme will give its contribution to the overall selectivity. Define d_i, the intrinsic discrimination of step i, according to

$$d_i = R_i^{c}/R_i^{w} = \frac{C_i^{c} k_{i,i+1}^{c}}{C_i^{w} k_{i,i+1}^{w}} = e^{\Delta G_i/RT}, \tag{15}$$

where

$$\Delta G_i = (\mu_{0i}^{w} - \mu_{0i}^{c}) + (\mu_{0i,i+1}^{*w} - \mu_{0i,i+1}^{*c}). \tag{16}$$

μ_{0i}^{c} and μ_{0i}^{w} are standard free energies for the state E_i for right and wrong substrates, respectively. $\mu_{0i,i+1}^{*c}$ and $\mu_{0i,i+1}^{*w}$ denote the activation free energies related to a transition from E_i to E_{i+1} (Fig. 1).

If there exists a limitation in the enzyme structure with respect to how much it can separate between two similar substrates (c.f. Introduction), then

$$\Delta G_i \leq \Delta G_{max},$$

or, equivalently,

$$d_i \leq d = e^{\Delta G_{max}/RT}. \tag{17}$$

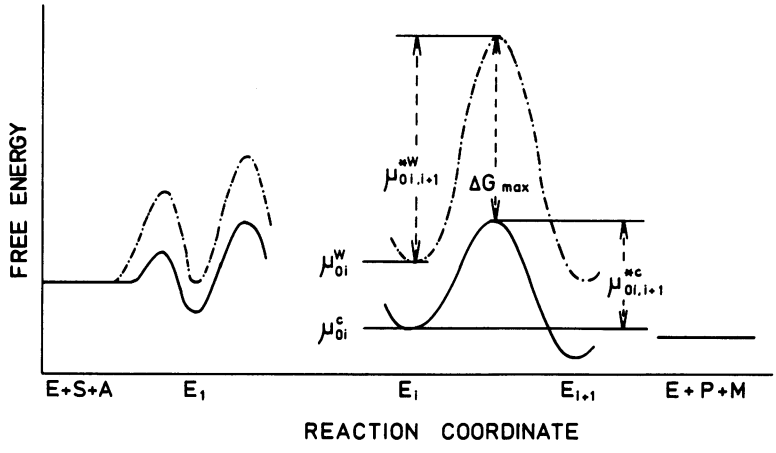


FIGURE 1 Illustration of the accuracy determining parameter ΔG_{max} in steady state flows. Free energy contour for cognate substrates (——); noncongate substrates (—·—).

The overall discrimination of the enzyme is

$$\frac{R^c}{R^w} = \frac{\sum_{i=0}^{n} \frac{d_i}{R_i}}{\sum_{i=0}^{n} \frac{1}{R_i}}, \tag{18}$$

where for convenience the index c has been deleted on R_i^c. The ratio R^c/R^w increases with increasing d_i irrespective of the set $\{R_i\}^n$. Therefore the overall selectivity is limited by the maximum selectivity of a single step:

$$R^c/R^w \leq d.$$

The maximum value of R^c/R^w is reached if every step i has the selectivity d or if the most accurate step is much slower than all the others. This proves formally that without an energy driven error correction mechanism there is no way to enhance the selectivity above d by introducing several reaction steps. For a somewhat less general but similar statement, see Ninio (1975).

SELECTIVITY IN BRANCHED KINETIC SCHEMES WITHOUT COSUBSTRATE DISPLACEMENT

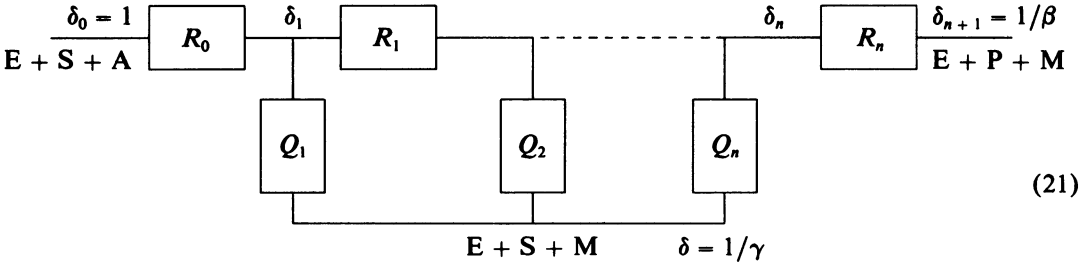
Consider a branched kinetic scheme of the type:

$$\begin{array}{c} \mathrm{E} + \mathrm{S} + \mathrm{A} \underset{k_{10}}{\overset{k_{01}}{\rightleftharpoons}} \mathrm{E}_1 \underset{k_{21}}{\overset{k_{12}}{\rightleftharpoons}} \mathrm{E}_2 \underset{k_{32}}{\overset{k_{23}}{\rightleftharpoons}} \cdots \underset{k_{n,n-1}}{\overset{k_{n-1,n}}{\rightleftharpoons}} \mathrm{E}_n \underset{k_{n+1,n}}{\overset{k_{n,n+1}}{\rightleftharpoons}} \mathrm{E} + \mathrm{P} + \mathrm{M}. \\ q_{-1} \Big\| \; q_1 q_{-2} \Big\| \; q_2 \qquad\qquad q_{-n} \Big\| \; q_n \\ \mathrm{E} + \mathrm{S} + \mathrm{M} \quad \mathrm{E} + \mathrm{S} + \mathrm{M} \qquad \mathrm{E} + \mathrm{S} + \mathrm{M} \end{array} \tag{19}$$

The law of detailed balance (e.g., de Groot and Mazur, 1969) implies that

$$q_i = q_{-i} \cdot C_i \cdot K_{\mathrm{AM}}; \; i = 1, 2 \ldots n,. \tag{20}$$

These relations can be used to eliminate the set $\{q_i\}_{i=1}^n$. In shorthand notation scheme 19 may be written



(21)

where

$$Q_i = q_{-i} \cdot C_i. \tag{22}$$

If $\gamma = 1$ in scheme 21, i.e., if the cosubstrate A is in equilibrium with its degradation products, M, then the overall selectivity can never exceed d. To understand this, note that with $\gamma = 1$ schemes of this type may also be reduced. For cognate substrates:

$$\mathrm{E} + \mathrm{S^c} + \mathrm{A} \xrightarrow{\delta = 1} \boxed{R^c} \xrightarrow{\delta = 1/\beta_c} \mathrm{E} + \mathrm{P^c} + \mathrm{M}$$

and for noncognate

$$\mathrm{E} + \mathrm{S^w} + \mathrm{A} \xrightarrow{\delta = 1} \boxed{R^w} \xrightarrow{\delta = 1/\beta_w} \mathrm{E} + \mathrm{P^w} + \mathrm{M}$$

The contracted schemes R^c and R^w may be obtained by iteration

$$R^c = Y_n^c,$$

$$Y_i^c = R_i - \frac{(R_i)^2}{Y_{i-1} + R_i + Q_i},$$

$$Y_0^c = R_0$$

and

$$R^w = Y_n^w,$$

$$Y_i^w = (R_i/d_i) - \frac{(R_i/d_i)^2}{Y_{i-1}^w + R_i/d_i + Q_i/f_i}$$

$$Y_0^w = R_0/d_0.$$

According to our assumptions, $d_i \leq d$, and $f_i \leq d$. For an arbitrary set of parameters $\{R_i, Q_i\}_1^n$ the ratio R^c/R^w increases with increasing d_i and f_i. The same maximum d is obtained when $d_i = f_i = d$; $i = 1, 2 \ldots n$;. Therefore $R^c/R^w \leq d$ in this case also.

STEADY STATE EQUATIONS IN KINETIC PROOFREADING

In this section the basic definitions and equations for a system using proofreading to enhance selectivity will be given. We assume that the concentrations of correct and incorrect substrates are the same:

$$[\mathrm{S^c}] = [\mathrm{S^w}] = [\mathrm{S}]. \tag{23}$$

Two types of product molecules are formed in the process, correct ones $\mathrm{P^c}$ and wrong ones $\mathrm{P^w}$. The cosubstrate A is displaced from equilibrium with M according to Eq. 8 and γ is now larger than one.

The standard free energies of right and wrong substrates and products are considered identical:

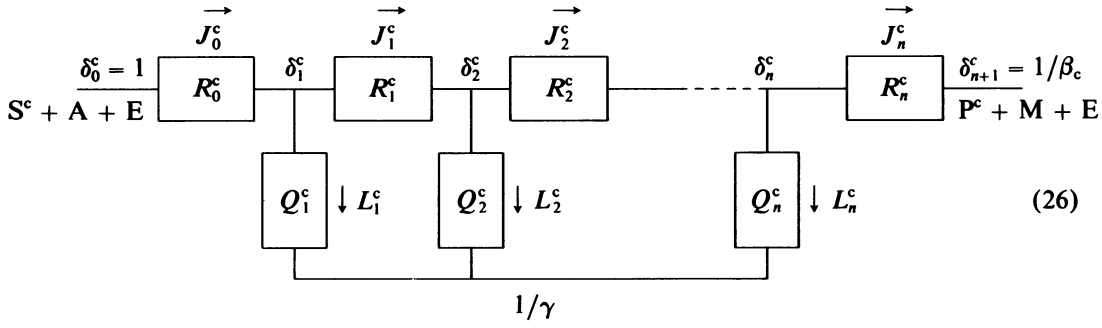
$$\begin{aligned} \mu_{0s}^c &= \mu_{0s}^w, \\ \mu_{0p}^c &= \mu_{0p}^w, \end{aligned} \tag{24}$$

so that

$$[P^c] = K_{PS}\delta_p^c[S],$$
$$[P^w] = K_{PS}\delta_p^w[S]. \tag{25}$$

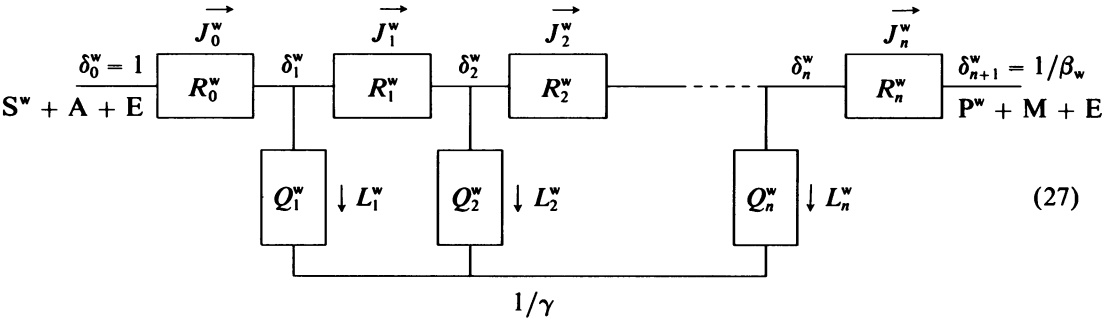
The correct product is displaced from equilibrium with correct substrate by the factor δ_p^c and the noncorrect product by the factor δ_p^w.

Restrictions 23 and 24, which have been introduced to simplify the derivations, do not limit the generality of the treatment, since they can easily be relieved in the final expressions of the theory (c.f. Eq. A24). The enzyme has one set of rate constants for the correct molecules and another for the wrong ones. Using the generalized rate constants R_i and Q_i defined in Eqs. 13 and 22 we may write schematically for correct molecules:



(26)

and correspondingly for the wrong ones:



(27)

The flow J_i is the number of moles passing from state i to $i + 1$ in the enzyme-substrate complex per unit time. L_i is the loss flow through checking step i. The displacements $1/\beta_c$ and $1/\beta_w$ are defined from

$$1/\beta_c = \frac{[P^c][M]}{[S^c][A]} \cdot \frac{K_{AM}}{K_{PS}^c} = \frac{\delta_p^c}{\gamma},$$
$$1/\beta_w = \frac{[P^w][M]}{[S^w][A]} \cdot \frac{K_{AM}}{K_{PS}^w} = \frac{\delta_p^w}{\gamma}. \tag{28}$$

For a positive net flow from substrate to product the second law of thermodynamics states that (c.f. Eq. 10) $1/\beta_{c,w} < 1$.

The enzyme makes a twofold use of the cosubstrate A: three free energy drop $RT \log(\gamma)$ available from hydrolysis of A is as before used to shift the reaction to the product side. In addition, the displacement γ now drives the exit flows L_i outwards.

The steady state treatment requires for its consistency a net exit flow from the product pool. We shall assume that the outflow is proportional to the concentration of product molecules:

$$\begin{aligned} J_n^c &= k_E \cdot [P^c], \\ J_n^w &= k_E \cdot [P^w]. \end{aligned} \tag{29}$$

The discrimination ratio D we define as the product flow of correct molecules divided by the product flow of incorrect ones:

$$D = J_n^c / J_n^w. \tag{30}$$

It is clear from Eqs. 25 and 28, and the boundary conditions of Eq. 29 that D can be expressed according to:

$$D = \frac{[P^c]}{[P^w]} = \frac{\delta_p^c}{\delta_p^w} = \frac{\beta_w}{\beta_c}. \tag{31}$$

This ratio is related to the probability of an error P_E in a location in a polymer by $P_E = 1/(D + 1)$. Before developing the analysis further we shall relate the formalism to real cases. In the amino acylation reaction, E is an aminoacyl-tRNA synthetase, for which S^c is a correct amino acid *and* a correct tRNA. S^w represents an incorrect amino acid *or* an incorrect tRNA. P^c is an aminoacyl-tRNA molecule where the amino acid correctly matches its tRNA. The cosubstrate A is ATP and M is AMP + PP_i. The in vivo level of γ can be estimated to be $\sim 10^{10}$. For translation of mRNA, the substrate S^c is an aminoacyl-tRNA interacting with a matching codon and forming an additional peptide bond releasing a free tRNA molecule. Correspondingly, S^w represents an aminoacyl-tRNA with a mismatching codon-anticodon interaction. The cosubstrate A here is GTP, and M is GDP + phosphate. The in vivo level of γ can be estimated to be $\sim 10^7$. Finally, in the case of DNA replication, the substrate molecules are the four deoxyribonucleoside triphosphates, and the products are inserted nucleoside monophosphates in the DNA chain. The cosubstrate A is in this case identical with the substrate, and proofreading is made possible by the release of pyrophosphate in the reaction. The excess level γ should in this case be defined as

$$\frac{[d\text{XTP}]}{[d\text{XMP}][\text{PP}_i]} = K_{AM} \cdot \gamma.$$

Schemes 26 and 27 are directly applicable to the growth of a polymer, where the parameter k_E in Eq. 29 is related to the rate of chain elongation. The steady state conditions may be applied either to a growing system where the full products, proteins, or DNA are diluted by cell division or, as a very close approximation, to cases where proteins are created and subsequently degraded, with their amino acids returning to the amino acid pool. The

parameters β_c and β_w, which define the boundary conditions for schemes 26 and 27, determine the concentration of polymers in relation to monomers in the system, given the standard free energies of substrates and products.

The dissipation of free energy over the enzyme originates from two different flows. The first (I) is the hydrolysis of cosubstrate molecules A over the checking steps without product formation. The second (II) is the dilution of the products P + M below equilibrium with the substrates S + A. Denoting the dissipation of free energy per mole of formed product by W_{diss} we have (c.f. Eq. 10):

$$\frac{W_{diss}}{RT} = \left\{\frac{[(J_0^c - J_n^c) + (J_0^w - J_n^w)] \log(\gamma)}{J_n^c + J_n^w}\right\}_I + \left\{\frac{J_n^c}{J_n^c + J_n^w} \log(\beta^c) + \frac{J_n^w}{J_n^c + J_n^w} \log(\beta^w)\right\}_{II}. \quad (32)$$

Ths discrimination ratio D is determined by the requirements of low error levels in gene replication and protein synthesis. In our treatment we shall establish how the enzyme can use a given intrinsic discrimination d to achieve the required D with as small losses of free energy as possible. In mathematical terms: minimize W_{diss} with respect to all rate constants in schemes 26 and 27 and with the subsidiary condition

$$D = J_n^c / J_n^w.$$

The minimization may also include the external parameters γ and β_c. A limitation in ΔG_{max} may be expressed as two inequalities:

$$\begin{aligned} 1/d \le R_i^c / R_c^w \le d, \\ 1/d \le Q_i^c / Q_i^w \le d. \end{aligned} \quad (33)$$

The first inequality states that the cognate molecules may not go more than a factor d faster in the forward direction, and the second that the noncognate molecules may not go more than a factor d faster over the exit step. The most advantageous case, given ΔG_{max} is obtained when

$$\begin{aligned} R_i^c = R_i,\ R_i^w = R_i/d;\ i = 0,1,2 \ldots n; \\ Q_i^c = Q_i,\ Q_i^w = dQ_i;\ i = 1,2 \ldots n;. \end{aligned} \quad (34)$$

The definitions of R_i and Q_i in Eqs. 13 and 22 show that these two conditions can be realized simultaneously if all states E_i have the same standard free energy for correct and incorrect substrates. Furthermore, two different selection mechanisms are required: one for the forward step and one for the vertical editing step.

Such mechanisms are favored by present knowledge: the exonuclease activity of the DNA polymerase is separated from its polymerase activity (Kornberg, 1969; Jovin et al., 1969). Also, the deacylating activity of aminoacyl synthetases is related to a specific site (Fersht and Dingwall, 1979). The double advantage for cognate molecules in Eq. 34 will lead to a very efficient mechanism for proofreading with almost negligible losses of nucleoside triphosphates associated with cognate product formation.

The inequalities in expression 33 have been made symmetrical with respect to the advantage for the cognate in the forward direction and its disadvantage to pass through the

vertical proofreading steps. Special constructions, implying an asymmetry in the two cases like the "double sieve mechanism" (Fersht, 1977*a*), can, after a minor modification, be included in the theoretical framework developed here. Alternatives to the choice (Eq. 34) are discussed elsewhere (see footnote 1).

Given Eq. 34, the flows in schemes 26 and 27 are determined by the $2n + 1$ rate constants R_i and Q_i and the boundary conditions

$$\delta_0^c = \delta_0^w = 1; \delta_{n+1}^c = 1/\beta_c, \delta_{n+1}^w = 1/\beta_w.$$

We have (c.f. Eq. 12)

$$J_i^c = [E][S][A]R_i(\delta_i^c - \delta_{i+1}^c); J_i^w = [E][S][A]\frac{R_i}{d}(\delta_i^w - \delta_{i+1}^w). \tag{35}$$

The flows over the discard steps are

$$L_i^c = [E][S][A]Q_i(\delta_i^c - 1/\gamma); L_i^w = [E][S][A]Q_i d(\delta_i^w - 1/\gamma). \tag{36}$$

Conservation of mass implies

$$L_i^c + J_i^c = J_{i-1}^c; L_i^w + J_i^w = J_{i-1}^w. \tag{37}$$

At the boundaries:

$$J_0^c = [E][S][A]R_0(1 - \delta_1^c), J_0^w = [E][S][A]\frac{R_0}{d}(1 - \delta_1^w) \tag{38}$$

and

$$J_n^c = [E][S][A]R_n(\delta_n^c - 1/\beta_c), J_n^w = [E][S][A]\frac{R_n}{d}(\delta_n^w - 1/\beta_w). \tag{39}$$

Relationships between the exit rate constant from the product pool k_E and the displacements $1/\beta_c$ and $1/\beta_w$ can be obtained from Eqs. 28, 29, and 39:

$$\frac{[E][A]}{K_{PS}}R_n(\delta_n^c - 1/\beta_c) = k_E \cdot \gamma/\beta_c; \frac{[E][A]}{K_{PS}} \cdot \frac{R_n}{d}(\delta_n^w - 1/\beta_w) = k_E \cdot \gamma/\beta_w. \tag{40}$$

It follows from Eqs. 36 and 37 that the displacements and the flows are connected by the relations

$$\delta_i^c = \frac{J_{i-1}^c - J_i^c}{[E][S][A]Q_i} + 1/\gamma, \delta_i^w = \frac{J_{i-1}^w - J_i^w}{[E][S][A]Q_i d} + 1/\gamma \tag{41}$$

and that the recursive relations

$$J_i^c = R_i\left(\frac{J_{i-1}^c - J_i^c}{Q_i} - \frac{J_i^c - J_{i+1}^c}{Q_{i+1}}\right), J_i^w = \frac{R_i}{d^2}\left(\frac{J_{i-1}^w - J_i^w}{Q_i} - \frac{J_i^w - J_{i-1}^w}{Q_{i+1}}\right) \tag{42}$$

are valid. It is possible to solve Eqs. 42 by iterative formulae or to assume special cases for the rate constants where analytical expressions for J_n^c and J_n^w exist. However, in a search for optimum properties of the mechanism, a different approach is more suitable.

SYMMETRY PROPERTIES OF THE PROOFREADING DIAGRAM

In this section we develop the necessary mathematical tools to determine the extremum properties of proofreading diagrams 26 and 27. A detailed derivation is given in the Appendix.

To make the problem tractable we need a new set of independent variables to replace the generalized rate constants R_i and Q_i defined in Eqs. 13 and 22. Define α_i and ϕ_i from

$$J_i^c = J_0^c \prod_{j=1}^{i} (1 - \alpha_j),\ J_i^w = J_0^w \prod_{j=1}^{i} (1 - \phi_j);\ i = 1, 2, \ldots n;. \tag{43}$$

The parameters α_i and ϕ_i, which constitute the new variable set, describe the fractions of molecules lost over the checking step i:

$$L_i^c = \alpha_i \cdot J_i^c,\ L_i^w = \phi_i \cdot J_i^w. \tag{44}$$

From Eqs. 31 and 43 we obtain:

$$D = \frac{J_n^c}{J_n^w} = \frac{J_0^c}{J_0^w} \cdot \frac{P_\alpha}{P_\phi}, \tag{45}$$

where

$$P_\alpha = \prod_{i=1}^{n} (1 - \alpha_i),\ P_\phi = \prod_{i=1}^{n} (1 - \phi_i). \tag{46}$$

In the Appendix we show how the two sets of variables $\{R_i/R_0, Q_i/R_0\}_1^n$ and $\{\alpha_i, \phi_i\}_1^n$ are connected. We further demonstrate that the discrimination ratio D is expressed symmetrically by the variables $\{\alpha_i, \phi_i\}_1^n$:

$$D = \frac{J_n^c}{J_n^w} = d\,\frac{\left(1 - \frac{1}{\gamma}\right)\frac{1}{P_\Psi P_\phi} + \frac{1}{\gamma} - \frac{1}{\beta_c}}{\left(1 - \frac{1}{\gamma}\right)\frac{1}{P_\Psi P_\alpha} + \frac{1}{\gamma} - \frac{1}{\beta_w}}, \tag{47}$$

where

$$P_\Psi = \prod_{i=1}^{n} \Psi_i \tag{48}$$

and

$$\Psi_i = \frac{\alpha_i d^2 - \phi_i}{d^2\alpha_i(1 - \phi_i) - \phi_i(1 - \alpha_i)}. \tag{49}$$

Expression 47 is valid for arbitrary boundary conditions. In what follows we restrict the treatment to boundary conditions according to Eq. 29.

By putting $\beta_c = \beta$ for convenience and using Eq. 31, the discrimination ratio D may be written

$$D = d \cdot \frac{\left(1 - \frac{1}{\gamma}\right)\frac{1}{P_\phi P_\Psi} + \frac{1}{\gamma} - \frac{d-1}{d} \cdot \frac{1}{\beta}}{\left(1 - \frac{1}{\gamma}\right)\frac{1}{P_\alpha P_\Psi} + \frac{1}{\gamma}}. \tag{50}$$

The dissipation of free energy per mole of total product according to Eq. 32 can now be written

$$\frac{W_{diss}}{RT} = \left[\frac{D}{D+1}\left(\frac{1}{P_\alpha} - 1\right) + \frac{1}{D+1}\left(\frac{1}{P_\phi} - 1\right)\right] \log(\gamma) + \log(\beta) + \frac{1}{D+1}\log(D). \tag{51}$$

Both the discrimination ratio D and W_{diss} depend only on the full products P_α, P_ϕ, and P_ψ. This immediately indicates the equivalence of all steps in the mechanism independent of the values of β and γ. Therefore, an unconstrained maximization of D as well as a minimization of W_{diss} with Eq. 50 as a subsidiary condition are connected with the relations

$$\begin{aligned} \alpha_i &= \alpha, \\ \phi_i &= \phi. \end{aligned} \quad ;i = 1, 2, \ldots n;. \tag{52}$$

This means that the loss fraction in each proofreading step should be the same both for cognate and noncognate molecules. It also follows from Eq. 52 that (see Appendix)

$$\begin{aligned} R_i/R_0 &= r_1(r)^{i-1};\ i = 1, 2, \ldots n-1; \\ Q_i/R_0 &= q_1(r)^{i-1};\ i = 1, 2, \ldots n;. \end{aligned} \tag{53}$$

Therefore, if the mechanism minimizes its dissipative losses of free energy, the rate constants R_i and Q_i vary with the same factor r from one step to the next, except at the boundaries. It is important to recognize that the equipartitioning of flows in Eq. 52 is associated not only with a set of rate constants according to Eq. 53, but also with a particular choice of β and γ. If these boundary conditions are changed, then Eq. 52 is no longer valid. In fact, setting up difference equations for the flow factors α_i and ϕ_i for a mechanism where the rate constants obey Eq. 53 leads to two solutions for each (c.f. Eqs. A30 and A31):

$$\begin{aligned} 1 - \alpha &= \frac{1 + r + q}{2} (\pm) \frac{1}{2}\sqrt{(1 + r + q)^2 - 4r}, \\ 1 - \phi &= \frac{1 + r + d^2 q}{2} (\pm) \frac{1}{2}\sqrt{(1 + r + d^2 q)^2 - 4r}, \end{aligned} \tag{54}$$

where

$$q = (q_1/r_1)r.$$

In optimum cases the mechanism has been adjusted so that only solutions corresponding to positive values of α and ϕ are retained. When the boundary conditions are changed the other two solutions contribute.

It is instructive to formulate D directly in terms of the rate constants. This can be done simply when Eq. 52 is valid. We have (c.f. Eqs. A29 and A30)

$$D = d\,\frac{(1 - 1/\gamma)\,\dfrac{P_\alpha}{P_r} + \dfrac{1}{\gamma} - \dfrac{d-1}{d}\cdot\dfrac{1}{\beta}}{(1 - 1/\gamma)\,\dfrac{P_\phi}{P_r} + \dfrac{1}{\gamma}} \tag{55}$$

where

$$P_r = r^n \tag{56}$$

and where α and ϕ are chosen positive in Eq. 54. Two limiting cases can be used to clarify the properties of this expression. First, when $r \ll 1$, both $1 - \alpha$ and $1 - \phi$ are small. It follows directly from Eq. 51 that in this case there are large excess flows of nucleoside triphosphates. A first order approximation in r yields:

$$1 - \alpha = r/(1 + q),\ 1 - \phi = r/(1 + d^2q)$$

and this gives

$$D = d\,\frac{(1 - 1/\gamma)\left(\dfrac{1}{1+q}\right)^n + \dfrac{1}{\gamma} - \dfrac{d-1}{d}\cdot\dfrac{1}{\beta}}{(1 - 1/\gamma)\left(\dfrac{1}{1+d^2q}\right)^n + \dfrac{1}{\gamma}}. \tag{57}$$

Second, when r is large the flow factors are approximately given by

$$1 - \alpha = \frac{r}{r+q},\ 1 - \phi = \frac{r}{r+d^2q}$$

and therefore

$$D = d\,\frac{(1 - 1/\gamma)\left(\dfrac{1}{r+q}\right)^n + \dfrac{1}{\gamma} - \dfrac{d-1}{d}\cdot\dfrac{1}{\beta}}{(1 - 1/\gamma)\left(\dfrac{1}{r+d^2q}\right)^n + \dfrac{1}{\gamma}}. \tag{58}$$

In this case the losses of cosubstrate molecules over the discard steps are small. However, since now $(1/r) \ll 1$, a large value of D necessarily requires large values of both β and γ. Eq. 58 is therefore also connected with a large energy dissipation.

A proofreading mechanism that uses many steps will have better properties with respect to its energy consumption than one that has one or a few. The case where the number of steps goes to infinity is therefore of considerable interest, since this always corresponds to the best solution of the editing problem with respect to its energy dissipation. We can visualize such a mechanism as a diffusion-like relative motion of the substrate in the enzyme-substrate complex from the degrading step of the cosubstrate to the release of product. The vertical outgoing steps in diagrams 26 and 27 correspond in this limit to an outflow of products per "length unit" of the enzyme.

The substitutions

$$R_i \to R_0 r_1 n(1 + r/n)^i,$$

$$Q_i \to \frac{R_0 q_1}{n}(1 + r/n)^i$$

lead straightforwardly to

$$\lim_{n\to\infty} D = d\frac{(1 - 1/\gamma)\,e^{-\alpha - r} - \frac{d-1}{d}\cdot\frac{1}{\beta} + \frac{1}{\gamma}}{(1 - 1/\gamma)\,e^{-\phi - r} + \frac{1}{\gamma}},$$

where α and ϕ are given by

$$\alpha = -r/2 + \sqrt{q + r^2/4},$$
$$\phi = -r/2 + \sqrt{qd^2 + r^2/4}.$$

The dissipation of free energy is in this limit given by

$$\lim_{n\to\infty} \frac{W_{diss}}{RT} = \left[\frac{D}{D+1}(e^{\alpha} - 1) + \frac{1}{D+1}(e^{\phi} - 1)\right]\log(\gamma) + \log(\beta) + \frac{1}{D+1}\log(D).$$

ABSOLUTE LIMITS OF KINETIC PROOFREADING

In this section we shall determine unconstrained upper limits for the discrimination ratio D, given the intrinsic discrimination d and the external parameters γ and β. In the literature one usually ascribes to proofreading the capacity of squaring the intrinsic discrimination of the first step. From an advantage d the mechanism may by editing obtain the advantage d^2. Since we have assumed here that the free energy difference ΔG_{max} may be used in two different ways for forward and outward discrimination according to Eq. 34 we would obtain d^3 instead. For a mechanism with n steps where the discrimination increases by the factor d^2 in each, the corresponding limit would be given by

$$D_{max} \leq d(d)^{2n}. \tag{59}$$

However, when $(d)^{2n} > \gamma$ there exists a much lower upper bound (Appendix):

$$D_{max} < d\gamma\left(1 - \frac{d-1}{d}\frac{1}{\beta}\right). \tag{60}$$

The corresponding limit without error correction, obtainable by an elementary consideration, is given by:

$$D_{max} = d\left(1 - \frac{d-1}{d}\frac{1}{\beta}\right). \tag{61}$$

The term $(d - 1)/d \cdot 1/\beta$ describes the influence of the backreaction from product to substrate. The factor γ in expression 60 is a direct and unavoidable consequence of the law of detailed balance: when the concentration of noncognate substrate molecules in a state on the

enzyme is displaced more than a factor $1/\gamma$ below its equilibrium value, the cosubstrate hydrolysis will be insufficient to drive the discard flow outwards. Instead, an inflow will appear and no further selective advantage is obtained. In the Appendix we outline how to derive exact expressions for D_{max}. Since the unconstrained maximum is always connected with the condition $r = 0$, D_{max} may be obtained from Eq. 57 by taking the derivative of D with respect to q. When $1/\beta = 0$, which corresponds to negligible backflow, an expression for D_{max} in closed form can be derived (Eq. A33). The highest possible accuracy for a given d and a given displacement γ is obtained in the limit of infinitely many steps in the mechanism:

$$\lim_{\substack{\beta\to\infty \\ n\to\infty}} D_{max} = \frac{[(\gamma - 1)(d^2 - 1)]^{(d^2-1)/d^2}}{d} \tag{62}$$

When the discrimination ratio is maximized with no restrictions, the extreme value is always connected with an infinite dissipation of free energy per mole of product because the reciprocal yield:

$$\frac{D}{D+1}\left(\frac{1}{P_\alpha} - 1\right) + \frac{1}{D+1}\left(\frac{1}{P_\phi} - 1\right)$$

in Eq. 51 tends to infinity.

We may therefore conclude that no enzyme is constructed to achieve the highest possible accuracy without taking the free energy dissipation into account, since this would turn it into a nucleoside triphosphatase with negligible product formation. We have explicitly calculated elsewhere (see footnote 1) how the excess losses of nucleoside triphosphates vary when ultimate limits as in Eqs. A33 and 62 are approached.

RELATIONSHIP BETWEEN FREE ENERGY DISSIPATION AND ACCURACY IN OPTIMUM CASES

In this work, we consequently consider the total dissipation of free energy W_{diss} over the mechanism as the relevant measure of its efficiency. The boundary conditions at the product side, summarized by the displacement β, and the boundary condition at the proofreading exit pathways, expressed by the displacement γ, may either be varied and thus included in the minimization of W_{diss} or kept constant at values chosen *a priori*. In the first case the underlying assumption is that both γ and β are determined by the requirements of the particular proofreading mechanism under investigation. This will give the absolute minimum energy dissipation for a given error and a given intrinsic selectivity d. In the second case γ or β are determined by other demands of the cell, external to the proofreading process, and the enzyme structure has been adjusted to these levels.

The power of the multistep mechanism is illustrated by an example chosen to be close to the conditions of the amino acylation reaction. The intrinsic, structurally determined selectivity d is 50. An error probability of 1/50,000, which is a thousandfold decrease, is required. The ATP displacement γ is fixed at 10^{10} and the optimum proofreading process is determined with the parameter β as a free variable. In Fig. 2 *A* the dissipation of free energy, and in Fig. 2 *B* the excess loss of ATP defined as $(J_0^c - J_n^c)/J_0^c$, are shown as functions of the number of steps in the process. In this particular case a one-step mechanism is close to its ultimate limit and

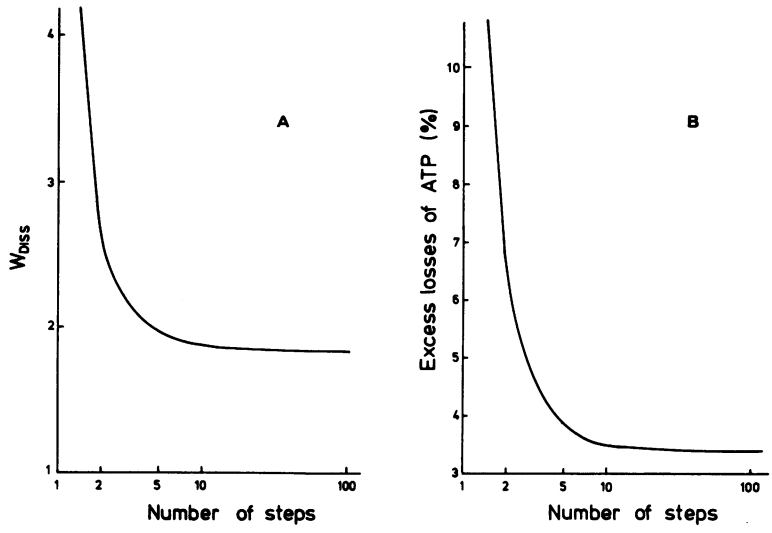


FIGURE 2 (*A*) The total dissipation of free energy W_{diss} as a function of the number of steps n in the proofreading mechanism. $d = 50$, $D = 50{,}000$, $\gamma = 10^{10}$. (*B*) The triphosphate losses related to discarding of cognate molecules as a function of the number of steps for the same mechanism.

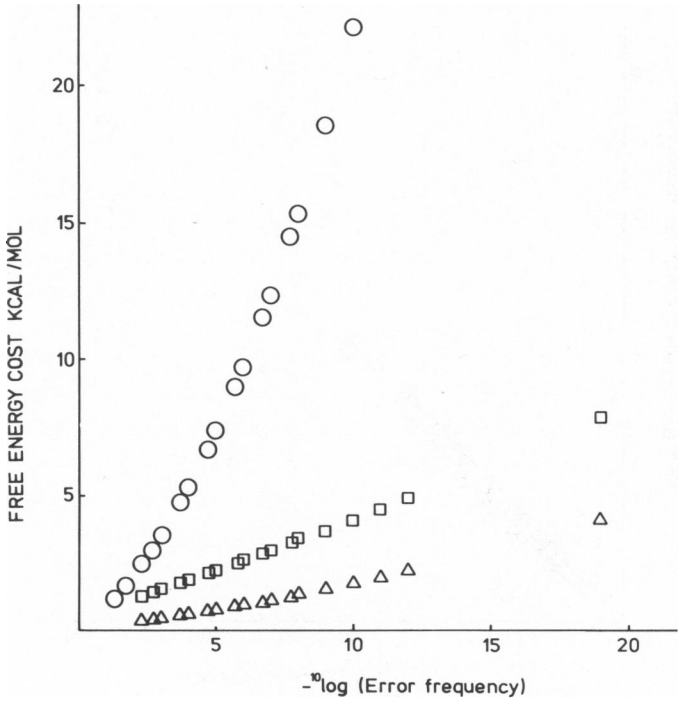


FIGURE 3 The total dissipation of free energy for a continuous mechanism ($n = \infty$) as a function of the error frequency. $d = 100$, $[S^c]/[S^w] = 1$ (△); $d = 100$, $[S^c]/[S^w] = 0.05$ (□); $d = 10$, $[S^c]/[S^w] = 1$ (○).

therefore dissipates much energy because of a high level of ATP hydrolysis. At 10 steps the dissipation of free energy is close to its minimum level, and there is only a 3% chance for a correctly activated amino acid to be discarded. When $d \cdot d^{2n} \gg D$, an n-step process behaves very much like a continuous ($n \rightarrow \infty$) process. In this case the most dramatic change is therefore seen when the number of steps increases from one to five.

We have investigated two continuous ($n \rightarrow \infty$) processes with respect to how the energy dissipation varies as a function of the error rate (Fig. 3). Both γ and β were used as independent variables and were thus determined by the requirement of minimum dissipation of free energy. In the first, $d = 100$ and in the second, $d = 10$. The curves are approximately linear for $d = 100$ but display a distinct upward curvature for $d = 10$. Clearly, an error frequency of 10^{-10} can be obtained without serious energy losses when $d = 100$ but not when $d = 10$.

In Fig. 4, the displacement γ of the nucleoside triphosphates is shown as a function of the error frequency. There is a remarkable linearity over 20 orders of magnitude:

$$\gamma = f(P_E) \approx 1/P_E = D.$$

This functional relationship is furthermore approximately invariant with respect to the parameter d.

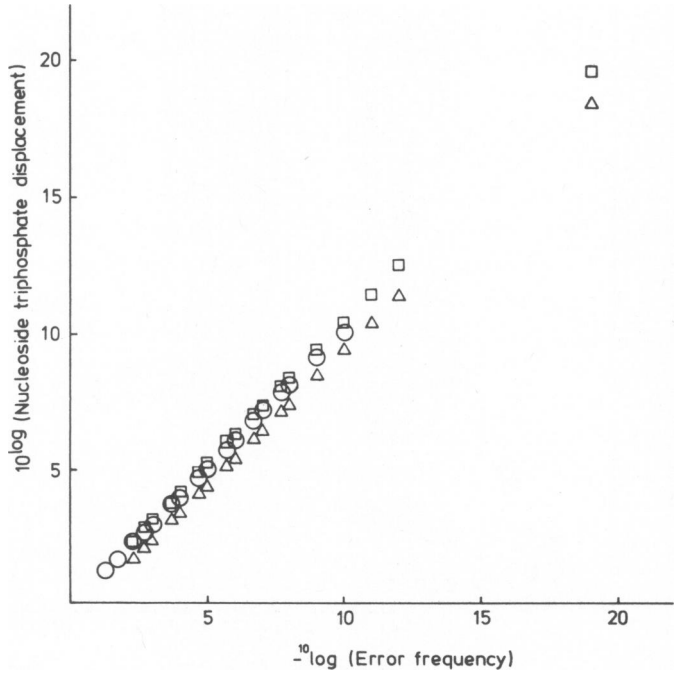


FIGURE 4 The nucleoside triphosphate displacement γ determined by the requirements of optimal proofreading, as a function of the error level. The same mechanisms and symbols as in Fig. 3.

CONCLUSIONS

In our study we have calculated the dissipative losses of a proofreading system from the displacements from equilibrium of the products of the reaction. This is in contrast to the usual identification of the costs of selection with the excess number of nucleoside triphosphates hydrolyzed in proofreading systems (Galas and Branscomb, 1978; Savageau and Freter, 1979). The latter measures only the fluxes in the editing branches and must be combined with the molar free energy loss to give a true measure of the dissipative cost of the editing function. Accordingly, the dissipative losses of selection are determined by both the total fluxes of all products and the displacements from equilibrium of all products.

In addition, we have exploited a novel formulation of branched kinetic systems to obtain the limits of accuracy in such systems. In particular, when the displacements from equilibrium of the selected product as well as of the cosubstrate (e.g., ATP) are fixed, the accuracy of the system cannot exceed the term:

$$D < d\gamma\left(1 - \frac{d-1}{d} \cdot \frac{\delta_P}{\gamma}\right).$$

The necessary relationship between accuracy of selection and dissipative losses can also be used to formulate an optimized kinetic scheme in which a given accuracy is to be obtained at a minimum dissipative loss. Such an optimization is realized by introducing multiple editing steps after the hydrolysis of the cosubstrate. In addition, the minimum entropy production is observed when the system is arranged symmetrically with the flows partitioned so that a constant fraction is discarded at each branch point. This result is independent of the absolute magnitude of the displacements at either the exit or discard branches. To the extent that naturally occurring selection systems have evolved to minimize free energy losses, these principles of optimal function should be reflected in the kinetic structure of enzyme systems.

That such principles may indeed be relevant is suggested by recent work of Fersht (1977b). His results indicate in at least one case that multiple editing steps are employed by an enzyme that amino-acylates tRNA. Furthermore, this selection system operates with a virtually negligible excess hydrolysis of the cosubstrate ATP (Mulvey and Fersht, 1977).

The advantages of symmetrically arranged multistage editing schemes have been recognized by Freter and Savageau.[2] However, there are important differences between their treatment and ours. Since they make no explicit use of the displacements from equilibrium of the products of the reaction, the necessary relationship between the accuracy of a selection and its energetic cost is missing in this treatment.

The variation of their formal parameters does not take into account the law of detailed balance. Consequently the displacement-determined upper limits on the accuracy in multistep editing processes are inaccessible. Furthermore, their operationally defined selectivity parameters do not correspond to the intrinsic structural selectivities of the mechanism except in the special case of complete irreversibility in all steps. Under these limiting conditions, their treatment and ours converge.

---

[2]Freter, R., and M. Savageau. Manuscript submitted for publication.

APPENDIX

The investigation of how a proofreading mechanism should be constructed to achieve a given accuracy at minimum dissipation of free energy is facilitated by a change of variables from the rate constants $\{R_i/R_0, Q_i/R_0\}_1^n$ to a new set of variables $\{\alpha_i, \phi_i\}_1^n$, which is directly related to the division of flows over the checking steps. Define α_i and ϕ_i from

$$J_i^c = J_0^c \prod_{j-1}^{i} (1 - \alpha_j), J_i^w = J_0^w \prod_{j-1}^{i} (1 - \phi_j). \tag{A1}$$

To make the notation more suitable express the rate constants as

$$\frac{R_i^c}{R_0} = \prod_{j-1}^{i} r_j, \frac{R_i^w}{R_0} = \frac{1}{d} \prod_{j-1}^{i} r_j, \tag{A2}$$

and

$$\begin{aligned} \frac{Q_i^c}{R_0} &= \prod_{j-1}^{i-1} r_j \cdot q_i, \\ \frac{Q_i^w}{R_0} &= \frac{1}{d} \prod_{j-1}^{i-1} r_j \cdot q_i d^2. \end{aligned} \tag{A3}$$

Eq. 41 and definition A1 connect the displacements $\delta_i^{c,w}$ with the flow factors α_i and ϕ_i:

$$\begin{aligned} \delta_i^c &= \frac{J_0^c}{R_0[E][S][A]} \prod_{j-1}^{i-1} \frac{1 - \alpha_j}{r_j} \cdot \frac{\alpha_i}{q_i} + \frac{1}{\gamma}, \\ \delta_i^w &= \frac{J_0^w}{R_0/d[E][S][A]} \prod_{j-1}^{i-1} \frac{1 - \phi_j}{r_j} \cdot \frac{\phi_i}{q_i d^2} + \frac{1}{\gamma}, \end{aligned} \tag{A4}$$

The boundary conditions in Eq. 38 at the substrate side of the enzyme together with the definitions in Eq. A3 give

$$\begin{aligned} J_0^c &= [E][S][A] R_0 (1 - 1/\gamma) \frac{q_1}{\alpha_1 + q_1}, \\ J_0^w &= [E][S][A] \frac{R_0}{d} (1 - 1/\gamma) \frac{q_1}{\frac{\phi_1}{d^2} + q_1} \end{aligned} \tag{A5}$$

At the product side the boundary conditions in Eq. 39 give

$$\begin{aligned} J_n^c \left(1 - \frac{\alpha_n}{1 - \alpha_n} \cdot \frac{r_n}{q_n}\right) &= [E][S][A] R_0 \prod_{j-1}^{n} r_j \left(\frac{1}{\gamma} - \frac{1}{\beta_c}\right), \\ J_n^w \left(1 - \frac{\phi_n}{1 - \phi_n} \cdot \frac{r_n}{q_n d^2}\right) &= [E][S][A] \frac{R_0}{d} \prod_{j-1}^{n} r_j (1 - 1/\beta_w). \end{aligned} \tag{A6}$$

Define two products related to the division of flows over all checking steps:

$$P_\alpha = \prod_{j=1}^{n} (1 - \alpha_j), P_\phi = \prod_{j=1}^{n} (1 - \phi_j). \tag{A7}$$

A combination of the boundary conditions at the entry (Eq. A5) and exit (Eq. A6) side of the enzyme leads to

$$\begin{aligned} P_r\left(\frac{1}{\gamma} - \frac{1}{\beta_c}\right) &= \left(1 - \frac{1}{\gamma}\right)\frac{q_1}{\alpha_1 + q_1} P_\alpha \left(1 - \frac{\alpha_n}{1 - \alpha_n} \cdot \frac{r_n}{q_n}\right), \\ P_r\left(\frac{1}{\gamma} - \frac{1}{\beta_w}\right) &= \left(1 - \frac{1}{\gamma}\right)\frac{q_1 P_\phi}{\frac{\phi_1}{d^2} + q_1}\left(1 - \frac{\phi_n/d^2}{1 - \phi_n} \cdot \frac{r_n}{q_n}\right), \end{aligned} \tag{A8}$$

where

$$P_r = \prod_{j=1}^{n} r_j. \tag{A9}$$

The recursive relations in Eq. 42 lead to difference equations for the new variables α_i and ϕ_i as defined in Eq. A1:

$$\begin{aligned} 1 - \alpha_i &= \frac{r_i}{q_i}\alpha_i - \frac{(1 - \alpha_i)\alpha_{i+1}}{q_{i+1}}; i = 1, 2, \ldots n - 1; \\ d^2(1 - \phi_i) &= \frac{r_i}{q_i}\phi_i - \frac{1}{q_{i+1}}(1 - \phi_i)\phi_{i+1}. \end{aligned} \tag{A10}$$

These can be used to express the variables r_i/q_i and q_{i+1} which are related to the rate constants and are defined in Eqs. A2 and A3 in terms of the flow factors:

$$\frac{r_i}{q_i} = \frac{(1 - \alpha_i)(1 - \phi_i)(d^2\alpha_{i+1} - \phi_{i+1})}{(1 - \alpha_i)\phi_i\alpha_{i+1} - (1 - \phi_i)\alpha_i\phi_{i+1}}; i = 1, 2, \ldots n - 1; \tag{A11}$$

$$\frac{1}{q_{i+1}} = \frac{\alpha_i \cdot d^2(1 - \phi_i) - \phi_i(1 - \alpha_i)}{(1 - \alpha_i)\phi_i \cdot \alpha_{i+1} - (1 - \phi_i)\alpha_i \cdot \phi_{i+1}}; i = 1, 2, \ldots n - 1;. \tag{A12}$$

Given $\{\alpha_i, \phi_i\}_1^n$, Eq. A11 and A12 uniquely determine the parameters r_i; $i = 2, 3, \ldots n - 1$; q_i; $i = 2, 3, \ldots n$; as well as the ratio r_1/q_1. Two parameters, e.g., q_1 and r_n/q_n are not explicitly expressed by the flow factors. The next step in the derivation will therefore be to complete the relationship between $\{\alpha_i, \phi_i\}_1^n$ and $\{R_i/R_0, Q_i/R_0\}_1^n$.

From Eqs. A11 and A12 we define variables u_i according to

$$u_i = \frac{r_i}{q_i} q_{i+1} = \frac{(1 - \alpha_i)(1 - \phi_i)(d^2\alpha_{i+1} - \phi_{i+1})}{\alpha_i d^2(1 - \phi_i) - \phi_i(1 - \alpha_i)}.$$

We may write

$$\prod_{i=1}^{n-1} u_i = \frac{1}{q_1} P_r \frac{q_n}{r_n} = \frac{P_\alpha}{1 - \alpha_n} \cdot \frac{P_\phi}{1 - \phi_n} \cdot \frac{\alpha_n - \frac{\phi_n}{d^2}}{\alpha_1 - \frac{\phi_1}{d^2}} \prod_{i=1}^{n-1} \Psi_i, \tag{A13}$$

where

$$\Psi_i = \frac{\alpha_i d^2 - \phi_i}{d^2\alpha_i(1-\phi_i) - \phi_i(1-\alpha_i)}; i = 1, 2 \ldots n;. \tag{A14}$$

Use Eq. A13 to obtain

$$P_r = q_1 \frac{r_n}{q_n} \cdot \frac{P_\alpha P_\phi}{(1-\alpha_n)(1-\phi_n)} \cdot \frac{\left(\alpha_n - \frac{\phi_n}{d^2}\right)}{\left(\alpha_1 - \frac{\phi_1}{d^2}\right)} \cdot P_{\Psi/\Psi_n} \tag{A15}$$

where

$$P_\Psi = \prod_{i=1}^{n} \Psi_i. \tag{A16}$$

There are now two equations, the boundary conditions in Eq. A8, which can be used to make the determination of the rate constants from the flow factors α_i and ϕ_i complete. If the expression A15 for P_r is used in the two equations in A8 and, furthermore, r_n/q_n is eliminated, we obtain

$$q_1 = \frac{(1-1/\gamma)\left(\alpha_1 - \frac{\phi_1}{d^2}\right) - P_\Psi[(1/\gamma - 1/\beta_w)(\phi_1/d^2)P_\alpha - (1/\gamma - 1/\beta_c)\alpha_1 P_\phi]}{D_N} \tag{A17}$$

where

$$D_N = P_\Psi\left[\left(\frac{1}{\gamma} - \frac{1}{\beta_w}\right)P_\alpha - \left(\frac{1}{\gamma} - \frac{1}{\beta_c}\right)P_\phi\right]. \tag{A18}$$

From this it follows that

$$q_1 + \alpha_1 = \frac{\left(\alpha_1 - \frac{\phi_1}{d^2}\right)[1 - 1/\gamma + P_\Psi P_\alpha(1/\gamma - 1/\beta_w)]}{D_N} \tag{A19}$$

and also

$$q_1 + \frac{\phi_1}{d^2} = \frac{[\alpha_1 - (\phi_1/d^2)][1 - 1/\gamma + P_\Psi P_\phi(1/\gamma - 1/\beta_c)]}{D_N}. \tag{A20}$$

Similarly r_n/q_n can be expressed as a function of $\{\alpha_i, \phi_i\}_1^n$. Therefore, we have shown that there exists a one to one correspondence between the variables r_i and q_i and the flow factors α_i and ϕ_i:

$$\{\alpha_i, \phi_i\}_{i-1}^n \Longleftrightarrow \{r_i, q_i\}_{i-1}^n,$$

or, as is evident from Eqs. A2 and A3, equivalently

$$\{\alpha_i, \phi_i\}_{i-1}^n \Longleftrightarrow \left\{\frac{R_i}{R_0}, \frac{Q_i}{R_0}\right\}_1^n.$$

We have furthermore explicitly shown how to relate the parameter set $\{\alpha_i, \phi_i\}_1^n$ to the rate constants of the mechanism. Therefore, when searching for optimum properties of the proofreading mechanism we

may first determine the set $\{\alpha_i, \phi_i\}_1^n$ and from this derive $\{r_i, q_i\}_1^n$. All rate constants except R_0, which determines the absolute time scale of the process, then follow. The discrimination ratio D can now, by using Eqs. A5 for J_0 together with Eqs. A19 and A20, be formulated simply:

$$D = \frac{J_n^c}{J_n^w} = d \frac{(1 - 1/\gamma)\dfrac{1}{P_\phi P_\psi} + \dfrac{1}{\gamma} - \dfrac{1}{\beta_c}}{(1 - 1/\gamma)\dfrac{1}{P_\alpha P_\psi} + \dfrac{1}{\gamma} - \dfrac{1}{\beta_w}}. \tag{A21}$$

P_α and P_ϕ are defined in Eq. A7 and P_ψ in Eq. A16. With our particular choice of boundary conditions according to Eqs. 29 we can use Eq. 31 to obtain $\beta_c = \beta_w/D$. Putting $\beta_c = \beta$ for convenience we obtain a simplified expression for D:

$$D = d \frac{(1 - 1/\gamma)\dfrac{1}{P_\phi P_\psi} + \dfrac{1}{\gamma} - \dfrac{d-1}{d} \cdot \dfrac{1}{\beta}}{(1 - 1/\gamma)\dfrac{1}{P_\alpha P_\psi} + \dfrac{1}{\gamma}}. \tag{A22}$$

The dissipation of free energy per mole of product according to Eq. (32) can now be written

$$\frac{W_{diss}}{RT} = \left[\frac{D}{D+1}\left(\frac{1}{P_\alpha} - 1\right) + \frac{1}{D+1}\left(\frac{1}{P_\phi} - 1\right)\right] \log(\gamma) + \log(\beta) + \frac{1}{D+1}\log(D). \tag{A23}$$

At this point it is suitable to relax the restrictions $[S^c] = [S^w]$ and $K_{PS}^c = K_{PS}^w$ in Eqs. 23 and 24. If we keep the definition of D as the ratio J_n^c/J_n^w or, equivalently, $[P^c]/[P^w]$, then Eq. 31 must be modified according to

$$D = \frac{J_n^c}{J_n^w} = \frac{[S^c]}{[S^w]} \cdot \frac{K_{PS}^c}{K_{PS}^w} \cdot \frac{\beta^w}{\beta^c}.$$

Therefore

$$\beta^w = D \cdot \beta \cdot \frac{K_{PS}^w}{K_{PS}^c} \cdot \frac{[S^w]}{[S^c]}.$$

The more general expression

$$D = d \frac{[S^c]}{[S^w]} \cdot \frac{(1 - 1/\gamma)\dfrac{1}{P_\phi P_\psi} + \dfrac{1}{\gamma} - \dfrac{1}{d}\left(d - \dfrac{K_{PS}^c}{K_{PS}^w}\right)\dfrac{1}{\beta}}{(1 - 1/\gamma)\dfrac{1}{P_\alpha P_\psi} + \dfrac{1}{\gamma}} \tag{A24}$$

now replaces Eq. A22, but the expression for the dissipation of free energy per mole of product remains the same.

Both the discrimination ratio D in Eq. A22 and W_{diss} in Eq. A23 are invariant with respect to any interchange of pairs

$$\alpha_i, \phi_i \rightleftharpoons \alpha_j, \phi_j.$$

Minimum values of W_{diss}, calculated with all rate constants of the enzyme as independent variables, are

for a given D always obtained with the constraint that all flow factors α_i and ϕ_i are identical. In short, minimum energy dissipation implies that:

$$\left.\begin{aligned}\alpha_i &= \alpha = \text{constant},\\ \phi_i &= \phi = \text{constant.}\end{aligned}\right\}; i = 1, 2, \ldots n;.$$

From Eqs. A11 and A12 it now follows that any proofreading mechanism that is efficient in this sense will have rate constants R_i and Q_i which, except at the boundaries, change with the same factor as we move along the steps of the enzyme

$$\begin{aligned}R_i &= R_0 r_1 (r)^{i-1};\ i = 1, 2 \ldots n-1,\\ Q_i &= R_0 q_1 (r)^{i-1};\ i = 1, 2, \ldots n;\end{aligned} \tag{A25}$$

where

$$\frac{r_1}{q_1} = \frac{r}{q}.$$

It is clarifying to relate D in Eq A22 directly to the rate constants of the mechanism. In optimum cases this can be done simply. Define therefore variables ρ_i and σ_i from the relations (c.f. Eqs. A10)

$$\begin{aligned}\sigma_i(1-\alpha_i) &= \rho_i\alpha_i - \alpha_i(1-\alpha_i),\\ d^2\sigma_i(1-\phi_i) &= \rho_i\phi_i - (1-\phi_i)\phi_i.\end{aligned} \tag{A26}$$

It follows directly that

$$\begin{aligned}\rho_i &= \frac{(1-\alpha_i)(1-\phi_i)(d^2\alpha_i - \phi_i)}{d^2\alpha_i(1-\phi_i) - \phi_i(1-\alpha_i)},\\ \sigma_i &= \frac{\alpha_i\phi_i(\phi_i - \alpha_i)}{d^2\alpha_i(1-\phi_i) - \phi_i(1-\alpha_i)};\ i = 1, 2, \ldots n;.\end{aligned} \tag{A27}$$

With

$$\mathrm{P}_\rho = \sum_{i=1}^{n} \rho_i, \tag{A28}$$

the discrimination ratio can now be written

$$D = d\,\frac{(1 - 1/\gamma)\mathrm{P}_\alpha/\mathrm{P}_\rho - \dfrac{d-1}{d}\cdot\dfrac{1}{\beta} + \dfrac{1}{\gamma}}{(1 - 1/\gamma)\mathrm{P}_\phi/\mathrm{P}_\rho + \dfrac{1}{\gamma}}. \tag{A29}$$

Expression A29 is completely general. However, in optimum cases, the variables ρ_i and σ_i can be directly identified with the rate constants of the enzyme according to

$$\left.\begin{aligned}\rho_i &= \rho = r,\\ \sigma_i &= \sigma = q.\end{aligned}\right\}; i = 1, 2, \ldots n; \tag{A30}$$

It also follows from Eqs. A26 that

$$1 - \alpha = \frac{1 + \rho + \sigma}{2} (\pm) \frac{1}{2} \sqrt{(1 + \rho + \sigma)^2 - 4\rho},$$
$$1 - \phi = \frac{1 + \rho + d^2\sigma}{2} (\pm) \frac{1}{2} \sqrt{(1 + \rho + d^2\sigma) - 4\rho}. \tag{A31}$$

The maximum value of D with no restriction on energy dissipation, given d, β and γ, can now be obtained by using the relations in A30 and taking the derivatives of D in Eq. A29 with respect to ρ and σ. We note that the inequalities

$$0 < \frac{1 - \alpha}{\rho} < 1, \quad 0 < \frac{1 - \phi}{\rho} < 1$$

follow directly from Eqs. A31. If we put $(1 - \alpha)/\rho = 1$ and $(1 - \phi)/\rho = 0$ in Eq. A29 a simple expression for an upper bound of the maximum of D is obtained.

$$D_{\max} < d\gamma\left(1 - \frac{d - 1}{d} \cdot \frac{1}{\beta}\right). \tag{A32}$$

The more detailed analysis shows that $D_{\max}$ is always characterized by the condition $\rho = r = 0$. The true maximum can therefore be obtained from Eq. 57 by taking the derivative of D with respect to q. When $\beta = 0$, i.e., when the concentration of product molecules is negligible, expressions for $D_{\max}$ in closed form can easily be obtained.

In the discrete case with n steps

$$\lim_{\beta \to \infty} D_{\max} = (\gamma - 1) \frac{d^2 - 1}{d} \cdot \frac{1}{\left\{1 + \frac{[(\gamma - 1)(d^2 - 1)]^{1/(n+1)} - 1}{d^2}\right\}^{n+1}}. \tag{A33}$$

In the limit $\gamma \to \infty$ Eq. A33 is reduced to expression 59.
In the continuous case, when $n \to \infty$, we obtain

$$\lim_{\substack{\beta \to \infty \\ n \to \infty}} D_{\max} = \frac{(\gamma - 1)(d^2 - 1)^{(d^2-1)/d^2}}{d}. \tag{A34}$$

In all cases where we give numerical examples on the relationship between dissipation of free energy and accuracy in the main text we have chosen to include the displacement β as a free variable. The variable β has been eliminated using the subsidiary condition (Eq. A29) according to

$$\beta = \frac{(d - 1)\gamma P_\rho}{(dP_\alpha - DP_\rho)(\gamma - 1) - P_\rho(D - d)}.$$

W_{diss} was subsequently minimized with respect to the two (or three) variables ρ, σ, (and γ).

We thank Prof. Ch. Kurland (Uppsala) for keen interest and penetrating discussions, Mrs. Róza Jakabffy for excellent technical assistance, and Mrs. Karin Dahlberg for typing the manuscript. We are grateful to Anders Ehrenberg, Inga Fischer-Hjalmars, Sven Forsblom, Bo Forslind, and Lennart Nilsson for valuable suggestions regarding the text. We also thank Prof. M. Savageau (Michigan) for providing us with preprints.

This work was supported by grants from the Swedish Natural Science Research Council and the Swedish Cancer Society.

Received for publication 30 October 1979 and in revised form 20 March 1980.

REFERENCES

BALDWIN, A. N., and P. BERG. 1966. Transfer ribonucleic acid-induced hydrolysis of valyladenylate bound to isoleucyl ribonucleic acid synthetase. *J. Biol. Chem.* **241**:839.

BRUTLAG, D., and A. KORNBERG. 1972. Enzymatic synthesis of deoxyribonucleic acid. XXXVI. A proofreading function for the 3'–5' exonuclease activity in deoxyribonucleic acid polymerases. *J. Biol. Chem.* **247**:241.

DEGROOT, S. R., and P. MAZUR. 1969. Non Equilibrium Thermodynamics. North Holland Publishing Company, Amsterdam.

EISINGER, J., and N. GROSS. 1975. Conformers, dimers and anticodon complexes of $tRNA^{Glu}_{2}$. *Biochemistry.* **14**:4031.

FERSHT, A. R. 1977*a*. Enzyme Structure and Mechanism. W. H. Freeman & Company, Publishers. San Francisco. 283.

FERSHT, A. R. 1977*b*. Editing mechanisms in protein synthesis. Rejection of valine by the isoleucyl-tRNA synthestase. *Biochemistry.* **16**:1025.

FERSHT, A. R., and C. DINGWALL. 1979. Establishing the misacylation deacylation of the tRNA pathway for the editing mechanism of prokaryotic and eukaryotic valyl-tRNA synthetases. *Biochemistry.* **18**:1238.

FERSHT, A. R., and M. KAETHNER. 1976. Enzyme hyperspecificity. Rejection of threonine by the valyl-tRNA synthetase by misacylation and hydrolytic editing. *Biochemistry.* **15**:3342.

FOWLER, R. G., G. E. DEGNEN, and E. C. COX. 1974. Mutational specificity of a conditional *Escherichia coli* mutator, mutD5. *Mol. Gen. Genet.* **133**:179.

GALAS, D., and E. BRANSCOMB. 1978. Enzymatic determinants of DNA polymerase accuracy. Theory of coliphage T4 polymerase mechanisms. *J. Mol. Biol.* **124**:653–687.

GOULIAN, M., Z. J. LUCAS, and A. KORNBERG. 1968. Enzymatic synthesis of deoxyribonucleic acid. XXV. Purification and properties of deoxyribonucleic acid polymerase induced by infection with phage T4. *J. Biol. Chem.* **243**:627.

GROSJEAN, H. J., S. DEHENAU, and D. CROTHERS. 1978. On the physical basis for ambiguity in genetic coding interactions. *Proc. Natl. Acad. Sci. U.S.A.* **75**:610.

HOPFIELD, J. J. 1974. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc. Natl. Acad. Sci. U.S.A.* **71**:4135.

JOVIN, T. M., P. T. ENGLUND, and A. KORNBERG. 1969. Enzymatic synthesis of deoxyribonucleic acid. XXXVII. Chemical modifications of deoxyribonucleic acid polymerase. *J. Biol. Chem.* **244**:3009–3018.

KORNBERG, A. 1969. Active center of DNA polymerase. *Science (Wash. D.C.).* **163**:1410–1418.

KURLAND, CH. 1978. The role of guanine nucleotides in protein biosynthesis. *Biophys. J.* **22**:373.

KURLAND, CH., R. RIGLER, M. EHRENBERG, and C. BLOMBERG. 1975. Allosteric mechanism for codon-dependent tRNA selection on ribosomes. *Proc. Natl. Acad. Sci. U.S.A.* **72**:4248.

LOFTFIELD, R. B. 1963. The frequency of errors in protein biosynthesis. *Biochem. J.* **89**:92.

LOFTFIELD, R. B., and D. VANDERJAGT. 1972. The frequency of errors in protein biosynthesis. *Biochem. J.* **128**:1353.

MULVEY, R., and A. FERSHT. 1977. Editing mechanisms in aminoacylation of tRNA: ATP consumption and the binding of aminoacyl-tRNA by elongation factor Tu. *Biochemistry.* **16**:4731.

NINIO, J. 1975. Kinetic amplification of enzyme discrimination. *Biochimie.* **57**:587.

NORRIS, A., and P. BERG. 1964. Mechanism of aminoacyl RNA synthesis: studies with isolated aminoacyl adenylate complexes of isoleucyl RNA synthetase. *Biochemistry.* **52**:330.

PAULING, L. 1958. *In* Festschrift Arthur Stoll. Birkhauser, A. G. Basel. 597–602.

SAVAGEAU, M., and R. FRETER. 1979. Energy cost of proofreading to increase fidelity of tRNA aminoacylation. *Biochemistry.* **18**:3486.