Integrated Function of a Kinetic Proofreading Mechanism: Double-Stage Proofreading by Isoleucyl-tRNA Synthetase†

Masahiro Okamoto and Michael A. Savageau*

Department of Microbiology and Immunology, The University of Michigan, Ann Arbor, Michigan 48109

Received April 26, 1985; Revised Manuscript Received December 12, 1985

ABSTRACT: Experimental measurements for isoleucyl-tRNA synthetase proofreading valyl-tRNA^{Ile} in *Escherichia coli* previously have been incorporated into the conventional Michaelis–Menten model for this system. This model was augmented to include two stages of proofreading—the aminoacyl adenylate and aminoacyl-tRNA stages—and used to predict the values of four additional rate constants that have been determined experimentally. The results suggest (1) that two stages of conventional kinetic proofreading with binding sites designed for isoleucine (the "correct" substrate) are inconsistent with the experimental data, (2) that a double-stage mechanism in which one stage (the "double-sieve") involves a binding site designed for valine (the "incorrect" substrate) and the other involves a binding site designed for isoleucine is consistent with all the experimental data, and (3) that the experimental data are not sufficiently accurate to distinguish the stage at which the double-sieve mechanism operates in vivo. Furthermore, analysis of the model suggests that four parameters have the most questionable values and that experimental refinement of their estimates will be needed to determine which of the two stages involves the double-sieve mechanism.

The fundamental character of kinetic proofreading mechanisms was revealed in the pioneering studies of Hopfield (1974) and Ninio (1975), and they noted that an expenditure of energy was necessary in order to improve the accuracy of biological recognition. A general theory has been developed that yields the energy cost to achieve a given degree of accuracy and shows how this cost–accuracy relationship is determined by the factors that influence the design of such mechanisms (Savageau & Freter, 1979; Freter & Savageau, 1980). These factors include the total energy cost for proofreading, the initial discrimination between correct and incorrect substrate molecules, the number of proofreading stages, the proofreading discrimination at each stage, and the distribution of proofreading effort among the stages for which there is an optimum distribution.

This general theory has been applied to the specific case of multistage Michaelis–Menten mechanisms and used to explore the influence of specific constraints upon the optimum design of such mechanisms (Savageau & Lapointe, 1981). These methods recently have been used to obtain the first complete characterization of a specific proofreading mechanism—isoleucyl-tRNA synthetase proofreading valyl-tRNA^{Ile} in *Escherichia coli* (Okamato & Savageau, 1984a). The results provide an excellent illustration of the value of integrating fragmentary data into a model of the intact system. (1) Such integration provides a rigorous test for consistency of the individual measurements. In this case the experimental data were found to be internally inconsistent. (2) Such integration predicts which experimental data are most suspect. In this case, one of the three most questionable measurements was found upon reexamination to be in error by 10–15-fold. Correction of this value produced a self-consistent set of data. (3) The integrated analysis provides predictions for various parameter values. In many cases, these provide estimates for parameter values that are difficult to determine directly or that have yet to be measured experimentally. (4) A sensitivity analysis provides an indication of the relative importance of various parameter values and, hence, an indication of where future experimental effort might be focused most profitably.

Having completely characterized the parameter values of the system, we were in a position to examine the dynamic behavior of the intact system by numerical solution of the system's kinetic equations. By these methods we were able to separate the effects of substrate competition and reaction velocity upon the accuracy of the system. We also were able to demonstrate the temporal development of these effects in response to a sudden change in competition or reaction velocity and to assess the stability of the system (Okamoto & Savageau, 1984b).

The proofreading discrimination ratio for this system was estimated from experimental data to be $P = 1100$. This value is an order of magnitude higher than one would expect for the rejection of valine by a binding site designed for isoleucine (Pauling, 1958; Loftfield & Eigner, 1966; Fersht, 1977a; Freist & Cramer, 1983). Fersht (1977a) has provided one possible explanation in terms of a "double-sieve" mechanism. The first level of discrimination occurs during substrate recognition through rejection of valine by an "isoleucine site". The second level occurs through rejection of isoleucyl adenylate by a separate "valine site" designed for hydrolysis of valyl adenylate. This latter mechanism can be expected to achieve P values approaching 10^5 (Fersht, 1979). Another possible explanation would be the existence of two stages of proofreading, each designed with a conventional isoleucine site having a P value of about 100 (Freter & Savageau, 1980). It is difficult to distinguish these two hypothesis on the basis of the individual experimental measurements that have been made upon this system.

In this paper we shall combine the existing experimental data with the accepted Michaelis–Menten model for isoleucyl-tRNA synthetase (Yarus, 1969; Eldred & Schimmel, 1972; Holler & Calvin, 1972; Mulvey & Fersht, 1977) to determine which of the above two hypotheses is best able to account for all the data. The results show clearly that two equally selective stages of conventional proofreading cannot account for the integrated set of experimental data. The results also are clearly in agreement with a double-sieve mechanism.

† This work was supported in part by a grant to M.A.S. from the National Institutes of Health (GM-30054).

0006-2960/86/0425-1969$01.50/0

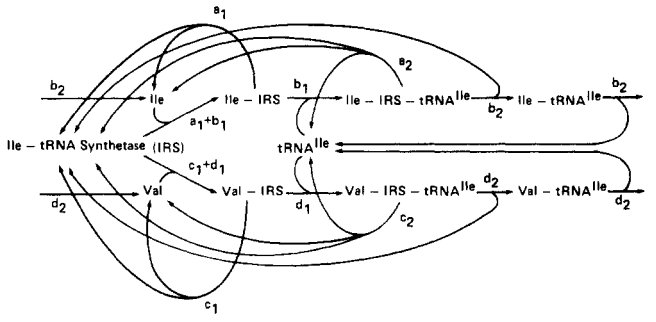

FIGURE 1: Isoleucyl-tRNA synthetase proofreading valyl-tRNA^{Ile}. Branching diagram that characterizes the pattern of macroscopic fluxes in a double-stage proofreading mechanism.

However, whether the double sieve occurs at the first stage, as suggested by Fersht (1977b), or at the second stage is still in question. We have identified those kinetic parameters that are most critical in determining the stage at which the double-sieve mechanism acts, but the experimental values of these parameters are currently too uncertain to answer the question.

MODEL

Figure 1 illustrates the general branching structure of a double-stage proofreading mechanism. There are eight independent fluxes in this model: a_1, a_2, b_1, b_2, c_1, c_2, d_1, and d_2. a_1 and a_2 are net fluxes representing hydrolysis of the correct aminoacyl adenylate and the correct aminoacyl-tRNA, respectively. b_1 and b_2 are net fluxes representing conversion of the correct aminoacyl adenylate to the correctly acylated tRNA and the release of correctly charged tRNA, respectively. c_1 and c_2 are net fluxes representing hydrolysis of the incorrect aminoacyl adenylate and the incorrect aminoacyl-tRNA, respectively. d_1 and d_2 are net fluxes representing the conversion of the incorrect aminoacyl adenylate to the incorrectly acylated tRNA and the release of incorrectly charged tRNA, respectively.

Macroscopic Parameters. In this case, the macroscopic behavior can be represented by a set of 14 macroscopic parameters (I_1, I_2, P_1, P_2, F_1, F_2, R_1, R_2, E_1, E_2, C_{12}, C_{22}, C_{11}, and C_2) defined in terms of these net fluxes (Freter & Savageau, 1980): The input discrimination ratio for the first stage of proofreading is

$$I_1 = (a_1 + b_1)/(c_1 + d_1) = (a_1 + a_2 + b_2)/(c_1 + c_2 + d_2) \quad (1)$$

and for the second stage is

$$I_2 = b_1/d_1 = (a_2 + b_2)/(c_2 + d_2) \quad (2)$$

The proofreading discrimination ratio for the first stage of proofreading is

$$P_1 = (b_1c_1)/(a_1d_1) = [(a_2 + b_2)c_1]/[a_1(c_2 + d_2)] \quad (3)$$

and for the second stage is

$$P_2 = (b_2c_2)/(a_2d_2) \quad (4)$$

The net forward flux from the first stage of proofreading is

$$F_1 = b_1 + d_1 = a_2 + b_2 + c_2 + d_2 \quad (5)$$

and from the second stage is

$$F_2 = b_2 + d_2 \quad (6)$$

The net recycling flux from the first stage of proofreading is

$$R_1 = a_1 + c_1 \quad (7)$$

and from the second stage is

$$R_2 = a_2 + c_2 \quad (8)$$

The net error after the first stage of proofreading is

$$E_1 = d_1/(b_1 + d_1) = (c_2 + d_2)/(a_2 + b_2 + c_2 + d_2) \quad (9)$$

and after the second stage is

$$E_2 = d_2/(b_2 + d_2) \quad (10)$$

The cost of proofreading at the first of two stages is

$$C_{12} = R_1/F_2 = (a_1 + c_1)/(b_2 + d_2) \quad (11)$$

and at the second of two stages is

$$C_{22} = R_2/F_2 = (a_2 + c_2)/(b_2 + d_2) \quad (12)$$

The apparent cost of proofreading at the first of two stages is

$$C_{11} = R_1/F_1 = (a_1 + c_1)/(b_1 + d_1) = (a_1 + c_1)/(a_2 + b_2 + c_2 + d_2) \quad (13)$$

and total cost of double-stage proofreading is

$$C_2 = C_{12} + C_{22} = (a_1 + a_2 + c_1 + c_2)/(b_2 + d_2) \quad (14)$$

or

$$C_2 = (C_{11} + 1)(C_{22} + 1) - 1 \quad (15)$$

Cost–Accuracy Relationship. Freter and Savageau (1980) derived the specific relationship between improved accuracy due to proofreading (E_1, E_2) and the associated energy cost (C_{11}, C_{12}, C_{22}, C_2) for a double-stage proofreading mechanism. The apparent cost for the first of two stages of proofreading, C_{11}, and the cost of the second stage of proofreading, C_{22}, can be represented as

$$C_{11} = [1 - (I_1 + 1)E_1][1 + (P_1 - 1)E_1]/[(I_1P_1 + 1)E_1 - 1] \quad (16)$$

$$C_{22} = [1 - (1/E_1)E_2] \times [1 + (P_2 - 1)E_2]/[[[(1 - E_1)/E_1]P_2 + 1]E_2 - 1] \quad (17)$$

Then, by the use of eq 15, one can write the total cost of double-stage proofreading as

$$C_2 = (I_1 + 1)(P_1 - 1)(P_2 - 1)(1 - E_1)(1 - E_2)E_1E_2/[[(I_1P_1 + 1)E_1 - 1][(1 - E_1)P_2E_2 - (1 - E_2)E_1]] - 1 \quad (18)$$

Optimal Distribution of Proofreading Effort. For given values of the parameters I_1, P_1, P_2, and E_2, the cost of proofreading C_2 is determined by the intermediate error E_1 (eq 18). Different values of E_1 correspond to different distributions of proofreading effort among the two stages. As all the proofreading effort becomes focused upon the first stage, E_1 approaches E_2; as it becomes focused upon the second stage, E_1 approaches $1/(I_1 + 1)$ (or I_2 approaches I_1). The total cost of proofreading C_2 varies with the distribution of proofreading effort (E_1) and reaches a minimum with the optimal distribution (Freter & Savageau, 1980). The value of E_1 that minimizes the total cost is obtained by differentiating C_2 with respect to E_1 and setting the derivative equal to zero:

$$E_{1(\min)} = ([(I_1P_1)(1 - E_2)/(P_2E_2)]^{1/2} + 1)^{-1} \quad (19)$$

Substituting this value into eq 18 yields the minimum value of the total cost C_2:

$$C_{2(\min)} = [(I_1 + 1)(P_1 - 1)(P_2 - 1) \times (1 - E_2)E_2/[(I_1P_1P_2E_2)^{1/2} - (1 - E_2)^{1/2}]^2] - 1 \quad (20)$$

Thus, the number of parameters in the cost–accuracy relationship is reduced by one for a mechanisms optimally designed with respect to the distribution of proofreading effort.

Michaelis–Menten Mechanism. Figure 2 shows the specific Michaelis–Menten mechanism corresponding to the general

model in Figure 1. This mechanism is assumed to operate within *E. coli* cells growing exponentially at 37 °C in a glucose minimal medium with an average doubling time of 47 min. Except for proofreading at the aminoacyl adenylate stage, this model is identical with that considered in our previous paper (Okamoto & Savageau, 1984a).

Fundamental Equations

The kinetic equations that govern the system represented in Figure 2 are given under steady-state conditions:

$$\dot{X}_1 = (k_{-3} + k'_1)X_5 + (k_{-4} + k'_2)X_6 + (k_7 + k_1)X_7 + (k_8 + k_2)X_8 - k_3X_1X_3 - k_4X_1X_4 - k_{-7}X_1X_9 - k_{-8}X_1X_{10} = 0 \quad (21)$$

$$\dot{X}_2 = (k_{-5} + k_1)X_7 + (k_{-6} + k_2)X_8 + k_9X_9 + k_{10}X_{10} - k_5X_2X_5 - k_6X_2X_6 = 0 \quad (22)$$

$$\dot{X}_3 = f_1 + (k_{-3} + k'_1)X_5 + k_1X_7 - k_3X_1X_3 = 0 \quad (23)$$

$$\dot{X}_4 = f_2 + (k_{-4} + k'_2)X_6 + k_2X_8 - k_4X_1X_4 = 0 \quad (24)$$

$$\dot{X}_5 = k_3X_1X_3 + k_{-5}X_7 - (k_{-3} + k'_1)X_5 - k_5X_2X_5 = 0 \quad (25)$$

$$\dot{X}_6 = k_4X_1X_4 + k_{-6}X_8 - (k_{-4} + k'_2)X_6 - k_6X_2X_6 = 0 \quad (26)$$

$$\dot{X}_7 = k_5X_2X_5 + k_{-7}X_1X_9 - (k_{-5} + k_7 + k_1)X_7 = 0 \quad (27)$$

$$\dot{X}_8 = k_6X_2X_6 + k_{-8}X_1X_{10} - (k_{-6} + k_8 + k_2)X_8 = 0 \quad (28)$$

$$\dot{X}_9 = k_7X_7 - k_{-7}X_1X_9 - k_9X_9 = 0 \quad (29)$$

$$\dot{X}_{10} = k_8X_8 - k_{-8}X_1X_{10} - k_{10}X_{10} = 0 \quad (30)$$

Equations 21 and 22 are not independent, e.g.

$$\dot{X}_1 = -(\dot{X}_5 + \dot{X}_6 + \dot{X}_7 + \dot{X}_8)$$

$$\dot{X}_2 = -(\dot{X}_7 + \dot{X}_8 + \dot{X}_9 + \dot{X}_{10})$$

or alternatively

$$X_S = X_1 + X_5 + X_6 + X_7 + X_8 \quad (31)$$

$$X_T = X_2 + X_7 + X_8 + X_9 + X_{10} \quad (32)$$

where X_S and X_T are the total isoleucyl-tRNA synthetase and total $tRNA^{Ile}$ concentrations, respectively. We also shall use the following physical and algebraic constraints.

Equilibrium constants:

$$K_5 = k_5/k_{-5} \quad (33)$$

$$K_7 = k_7/k_{-7} \quad (34)$$

Aminoacylation ratio:

$$A = (X_7 + X_9)/X_2 \quad (35)$$

Symmetry coefficients:

$$S_{57} = k_{-5}/k_7 \quad (36)$$

$$S_{68} = k_{-6}/k_8 \quad (37)$$

where S_{57} and S_{68} are symmetry coefficients with respect to the release of uncharged (k_{-5} or k_{-6}) and charged (k_7 or k_8) tRNA that are initially taken as unity.

Discrimination factors:

$$D_{34} = K_3/K_4 \quad (38)$$

$$D_{56} = K_5/K_6 \quad (39)$$

D_{34} is a discrimination factor between the corresponding cognate ($K_3 = k_3/k_{-3}$) and noncognate ($K_4 = k_4/k_{-4}$) steps that is taken as 180 (Loftfield & Eigner, 1966; Freist & Cramer, 1983); D_{56} is a discrimination factor between the corresponding cognate ($K_5 = k_5/k_{-5}$) and noncognate ($K_6 = k_6/k_{-6}$) steps that should be greater than 1 for effective proofreading and to which we have assigned a value of 2.4.

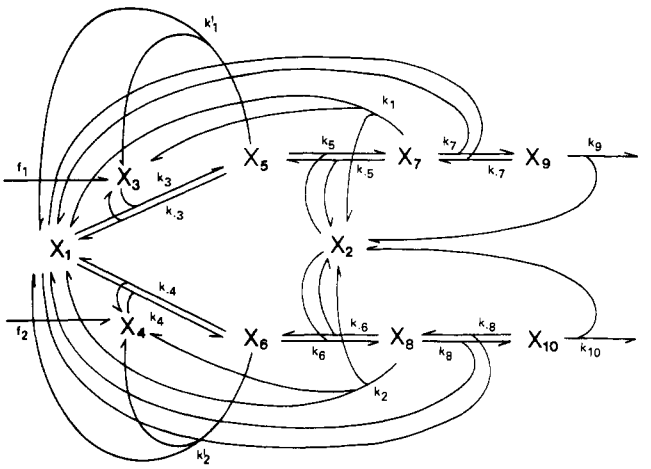

FIGURE 2: Isoleucyl-tRNA synthetase proofreading valyl-$tRNA^{Ile}$. Michaelis–Menten model corresponding to the branched diagram in Figure 1. Free isoleucyl-tRNA synthetase, X_1; free $tRNA^{Ile}$, X_2; free isoleucine, X_3; free valine, X_4; enzyme-bound isoleucyl adenylate, X_5; enzyme-bound valyl adenylate, X_6; enzyme-bound isoleucyl-$tRNA^{Ile}$, X_7; enzyme-bound valyl-$tRNA^{Ile}$, X_8; free isoleucyl-$tRNA^{Ile}$, X_9; free valyl-$tRNA^{Ile}$, X_{10}.

Table I: Independent Parameters Determining the Allowable Range of Values for the Concentration Variables of the Model in Figure 2[a]

parameter	value	parameter	value
I_1	36.0	S_{57}	1.0
C_2	0.093	k_3	4.4×10^6 M^{-1} s^{-1}
E_2	4.14×10^{-4}	k_{-3}	30.0 s^{-1}
X_S	5.0 μM	K_5 $(=k_5/k_{-5})$	2.86×10^6 M^{-1}
X_T	5.0 μM	K_7 $(=k_7/k_{-7})$	7.1×10^{-9} M
A	4.0	X_3	550 μM
f_1	20.0 μM s^{-1}	k_7	20 s^{-1}

[a] The experimental determination of these parameters is described in Okamoto and Savageau (1984a).

Since codon–anticodon recognition is unaffected by the amino acid residue bound to tRNA, we shall assume

$$k_9 = k_{10} \quad (40)$$

At steady-state, the final net error E_2 can be represented as

$$E_2 = f_2/(f_1 + f_2) \quad (41)$$

and the macroscopic factors E_1, I_1, P_1, and P_2 can be represented as

$$E_1 = (k_6X_2X_6 - k_{-6}X_8)/(k_5X_2X_5 - k_{-5}X_7 + k_6X_2X_6 - k_{-6}X_8) \quad (42)$$

$$I_1 = (k_3X_1X_3 - k_{-3}X_5)/(k_4X_1X_4 - k_{-4}X_6) \quad (43)$$

$$P_1 = [k'_2X_6(k_5X_2X_5 - k_{-5}X_7)]/[k'_1X_5(k_6X_2X_6 - k_{-6}X_8)] \quad (44)$$

$$P_2 = [k_2X_8(k_7X_7 - k_{-7}X_1X_9)]/[k_1X_7(k_8X_8 - k_{-8}X_1X_{10})] \quad (45)$$

Experimental Data

The experimentally determined values given in our previous paper (Okamoto & Savageau, 1984a) are listed in Table I. These values apply equally well to the double-stage model in Figure 2, because the same values would have been determined experimentally regardless of whether the actual mechanism involves single- or double-stage proofreading. For most of these parameters, this is obvious; the case of C_2 requires some additional explanation. In estimates of the c/d and a/b ratios

for single-stage proofreading, Hopfield et al. (1976) and Mulvey & Fersht (1977) measured moles of ATP hydrolyzed per mole of aminoacyl-tRNA^{Ile} formed. If the underlying mechanism in fact involves two stages of proofreading, then these measurements would correspond to $(c_1 + c_2)/d_2$ and $(a_1 + a_2)/b_2$. From eq 10 and 14, one can show that $C_2 = (1 - E_2)(a_1 + a_2)/b_2 + E_2(c_1 + c_2)/d_2$, which is identical with C in the previous paper, because $E_2 = E$, $(a_1 + a_2)/b_2 = a/b$, and $(c_1 + c_2)/d_2 = c/d$.

Although the total cost of proofreading remains unchanged, the distribution of costs between the two stages is undetermined at this point. The value of one of the proofreading discrimination factors, P_1 or P_2, also is undetermined. These two additional degrees of freedom arise because we have introduced two additional rate constants (k'_1 and k'_2) without additional experimental measurements or conditions to constrain their values. Fersht (1977b) has estimated the values of the rate constants k_1, k'_1, k_2, and k'_2 for the isoleucyl-tRNA synthetase (IRS) system of *E. coli* K12. He obtained values of $k_1 = 0.014$ s^{-1}, $k'_1 = 1.7 \times 10^{-3}$ s^{-1}, and $k_2 = 10$ s^{-1} at 25 °C and pH 7.78. He also measured the hydrolysis rate of IRS-Val-AMP; however, he could not estimate precisely the corresponding value of k'_2 because 80% of the Val-AMP bound to enzyme was lost before the first measurement at 1 min. By assuming a Q_{10} of 2 for these reactions, we estimate that at 37 °C the values of k_1, k'_1, k_2, and k'_2 are

$$k_{1(\text{expt})} = 0.028 \text{ s}^{-1} \qquad k'_{1(\text{expt})} = 0.0034 \text{ s}^{-1}$$
$$k_{2(\text{expt})} = 20 \text{ s}^{-1} \qquad k'_{2(\text{expt})} \geq 0.36 \text{ s}^{-1} \tag{46}$$

The use of two of these values in an attempt to eliminate the two degrees of freedom mentioned above leads to inconsistencies with the data in Table I or with the remainder of Fersht's data.

Rather than arbitrarily distribute the error among these parameters to obtain consistency, we shall use the model and the data in Table I to provide the best fit to the data for the parameters k_1, k'_1, and k_2. (The data for k'_2 will not be used for this purpose because of its high degree of uncertainty.) As noted above, there are two degrees of freedom that may be used to obtain this fit. These correspond to the values of the two macroscopic parameters E_1 and P_1. For each choice of values for E_1 and P_1, the experimental data in Table I can be used, together with the constraints imposed by the equations governing the double-stage mechanism (eq 16, 17, and 23–45), to estimate the remaining rate constants and concentrations in steady state. As in the previous paper, we proceed by first obtaining conditions that ensure nonnegative concentration values; then, these conditions are refined so as to ensure nonnegative rate constants as well.

Results

Optimal Distribution of Proofreading Effort. One degree of freedom can be removed by assuming that proofreading effort is optimally distributed between the two stages. This condition leads to eq 20, which with the values of I_1, E_2, and C_2 given in Table I can be solved numerically to yield a family of $P_1 - P_2$ values. This is shown in Figure 3. The points marked A, B, C, and E represent the extreme conditions (maximum or minimum P_1 or P_2). The dotted line represents the condition $P_1 = P_2$.

All the points on this curve represent systems with their proofreading optimally distributed between the two stages. But do these points represent physically realizable systems? As in our previous paper (Okamoto & Savegeau, 1984a), we have used a computer program to determine whether a given set

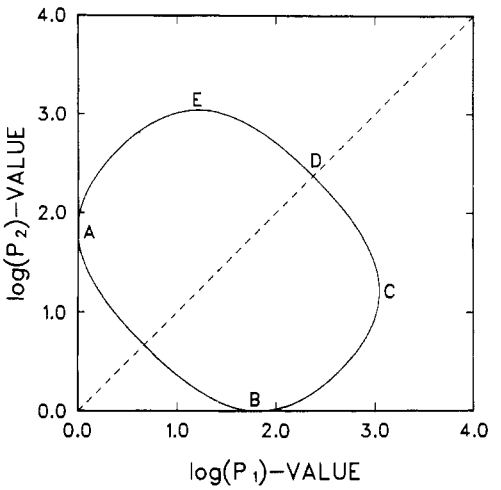

FIGURE 3: Family of $P_1 - P_2$ values. These values are all consistent with the experimental data in Table I and with the assumption that proofreading effort is optimally distributed among the two stages. See text for discussion.

Table II: Allowable Range of Values for the Rate Constant k_9 Corresponding to Different $P_1 - P_2$ Values in Figure 3

P_1 value	P_2 value	allowable value for k_9 (s^{-1})	P_1 value	P_2 value	allowable value for k_9 (s^{-1})
1.0	67.1	none[a]	100.0	508	32.8–33.5
	67.1	none		1.04	none
5.0	753	none	200.0	279	32.8–33.5
	4.33	none		1.38	none
10.0	1030	none	500.0	102	32.8–33.5
	2.34	none		2.93	none
16.4	1100	32.9–33.3	1000.0	29.9	32.7–33.6
	1.60	none		9.06	none
40.0	896	32.8–33.4	1100.0	16.41	32.6–33.7
	1.07	none		16.38	none

[a] Systems with no allowable value for the rate constant k_9 are physically unrealizable.

of experimental data (Table I plus P values) is consistent with the requirement for nonnegative concentration values. The results, shown in Table II, indicate that systems represented along arc CDE in Figure 3 correspond to physically realizable systems; those along arc EABC do not. Furthermore, the value of the rate constant k_9 must be between 32.91 and 33.3 s^{-1} for physically realizable systems. Thus, with respect to the condition that all concentration values be nonnegative, all systems represented along arc CDE are consistent with the experimental data in Table I.

In addition to the parameters listed in Table I, there are three parameters (S_{68}, D_{34}, D_{56}) that are involved in determining whether values for the rate constants will be nonnegative. It should be emphasized that these three parameters have nothing to do with the conditions for nonnegative concentration values (or the results in Table II). When the concentration variables are positive, the following equations are necessary and sufficient conditions for nonnegative rate constants:

$$D_{34} < K_3X_1X_4/X_6 \tag{47}$$

$$S_{68}K_5X_2X_6f_2/[X_8[E_1(C_{22} + 1)(f_1 + f_2) + S_{68}f_2]] < D_{56} < K_5X_2X_6/X_8 \tag{48}$$

The concentration values determined in the previous paragraph and parameter values from Table I can be substituted into eq 47 and 48, and one finds a threshold value of 379 for D_{34} and bounds of 1.2 and 2.5 for D_{56}. The values for these two parameters that we have selected under Fundamental Equations

Table III: Error between Predicted and Experimental Values of the Rate Constants k_1, k'_1, and k_2 for Different $P_1 - P_2$ Values in Figure 3

P_1 value	P_2 value	ϵ value[a]	k_1 (s^{-1})	k'_1 (s^{-1})	k_2 (s^{-1})	k'_2 (s^{-1})[b]
16.4[e]	1100	14400	0.378	1.8×10^{-7}	60.0	7.3×10^{-6}
16.5[f]	1100	17.4	0.376	3.4×10^{-3}	59.1	0.140
20.0	1088	37.7	0.342	7.9×10^{-2}	41.9	3.31
50.0	803	96.2	0.245	0.292	13.9	13.5
100	508	118	0.211	0.366	6.74	17.8
236[d]	236	138	0.187	0.420	2.56	20.5
500	102	167	0.163	0.472	0.895	21.2
1000	29.9	366	8.7×10^{-2}	0.640	0.115	19.5
1100[c]	16.4	170000	2.0×10^{-4}	0.832	1.2×10^{-4}	17.5

[a] Total error; $\epsilon = \sum e_i$, where $e_i = k_i/k_{i(\text{expt})}$ when $k_i > k_{i(\text{expt})}$, $e_i = k_{i(\text{expt})}/k_i$ when $k_i < k_{i(\text{expt})}$, and k_i represents k_1, k'_1, or k_2. The values of $k_{i(\text{expt})}$ are given in eq 46. [b] Although values of k'_2 are predicted by the model, this parameter is not used in calculating the error because of the large uncertainty in the experimental estimate of $k'_{2(\text{expt})}$. See text for discussion. [c] Point C. [d] Point D. [e] Point E. [f] The case with minimum error.

Table IV: Error between Predicted and Experimental Values of the Rate Constants k_1, k'_1, and k_2 for Different Values of $E_1 - P_1$[g]

E_1 value	P_1 value	P_2 value	min ϵ value[e]	k_1 (s^{-1})	k'_1 (s^{-1})	k_2 (s^{-1})	k'_2 (s^{-1})[f]
2.70×10^{-2}	1.000	1930	117	0.211	0.367	25.4	0.649
2.69×10^{-2a}	16.5	1100	17.4	0.376	3.4×10^{-3}	59.1	0.140
2.23×10^{-2b}	900.0	904	16.9	0.377	3.3×10^{-3}	48.5	6.16
9.19×10^{-3c}	8.0×10^3	358	15.5	0.376	3.4×10^{-3}	19.2	22.9
3.38×10^{-3d}	2.9×10^4	121	17.5	0.376	3.4×10^{-3}	6.50	30.1
1.02×10^{-3}	1.1×10^5	25.4	29.1	0.377	3.3×10^{-3}	1.37	33.1
4.14×10^{-4}	1.0×10^6	1.01	382	0.378	9.0×10^{-4}	5.5×10^{-2}	33.9

[a] Case a. [b] Case b. [c] Case c. [d] Case d. [e] Minimum total error for a given E_1 value, calculated as in Table III. [f] See footnote *b* in Table III. [g] Optimal distribution of proofreading effort among the two stages is not assumed in this case. See text for discussion.

fall within these limits and, therefore, are consistent with the requirement for nonnegative values of rate constants.

Each choice of P_1 value along arc CDE leads to a different set of predicted values for the rate constants and concentrations of the system. In particular, the values of the parameters k_1, k'_1, k_2, and k'_2 are predicted. Thus, we can now ask the question, which value of P_1 leads to the best fit between the predicted values for these parameters and the corresponding experimental data of Fersht (1977b)? The answer is provided in Table III. A value of $P_1 = 16.5$ (and, thus, $P_2 = 1100$) produces the least error between the experimental and predicted values of the parameters k_1, k'_1, and k_2.

This result does not correspond to double-stage proofreading with catalytic sites designed for isoleucine ($P_1 \doteq P_2 \doteq 100$). The large value of $P_2 = 1100$ is clearly consistent with a double-sieve mechanism as Fersht has proposed. However, the double-sieve, according to this calculation, would operate at the second (aminoacyl-tRNA) stage, rather than the first (aminoacyl adenylate) stage as suggested by Fersht (1977b).

Nonoptimal Distribution of Proofreading Effort. Although it seems reasonable to assume that multiple-stage proofreading mechanisms would be selected for optimal distribution of proofreading effort (minimum cost for a given degree of accuracy), we can test this assumption. If eq 18 rather than eq 20 is used and E_1 and P_1 values are varied to produce physically realizable predictions, then again one can determine the best fit to the experimental data of Fersht (1977b). If a significantly better fit to the data were to be obtained, then one might conclude that the previous assumption of optimality was invalid.

The results of such a two-dimensional search of $E_1 - P_1$ values is shown in Table IV. For each choice of E_1 value, one can find the value of P_1 that minimizes the error between predicted and experimental values of k_1, k'_1, and k_2. By examining the range of possible E_1 values, one can determine the lowest of the minima. This situation (case c) corresponds to values of $E_1 = 9.19 \times 10^{-3}$ and $P_1 = 8000$. The corresponding value of P_2, as determined from eq 18, is 358. The predicted values of k_1, k'_1, k_2, and k'_2 also are shown in Table IV.

This result is again inconsistent with two stages of proofreading with catalytic sites based on the structure of isoleucine ($P_1 \doteq P_2 \doteq 100$). Although a mechanism with equal P values (case b) is superior to the mechanism with an optimum distribution of proofreading effort found in the previous section (Tables III and IV, case a), the P values are much greater than 100 ($P_1 \doteq P_2 \doteq 900$). Case c, which produces the best fit to the experimental data, indicates a double-sieve mechanism, and in this case the double sieve appears to operate at the first stage as Fersht (1977b) has suggested.

Conventional Proofreading at the Second Stage. The results in the previous section are consistent with Fersht's proposal of a double-sieve mechanism and his suggestion that it operates at the first stage of proofreading. However, the P_2 value of 358 is relatively high for conventional proofreading at the second stage. Discrimination based on a site designed for isoleucine would be expected to have a value no larger than about 20–200 (Pauling, 1958; Loftfield & Eigner, 1966; Fersht, 1977a; Freist & Cramer, 1983).

If we require P_2 values of about 100, then the error increases to a value (case d) comparable to that found in the case with optimal distribution of proofreading effort (case a). The predicted values of k_1, k'_1, k_2, and k'_2 also are shown in Table IV. This result is consistent with (1) the proposal of a double-sieve mechanism, (2) its operation at the first stage of proofreading, and (3) a second (or "mopping up") stage employing a conventional mechanism based on the structure of isoleucine.

Discussion

In our previous paper we have shown that experimental data for proofreading of valyl-$tRNA^{Ile}$ by isoleucyl-tRNA synthetase is consistent with a single stage of proofreading (Okamoto & Savageau, 1984a). However, the proofreading discrimination has a rather high value of 1100, much greater than the limit of about 20–200 expected for a binding site designed for isoleucine (Pauling, 1958; Loftfield & Eigner, 1966; Fersht, 1977a; Freist & Cramer, 1983).

Fersht (1977a) has provided an explanation for these high P values by postulating that the proofreading mechanism is

actually designed to select for and hydrolyze the valine derivative, while ignoring the corresponding isoleucine derivative. This mechanism, which he called a double sieve, would permit P values up to about 10^5 (Fersht, 1979).

An examination of factors influencing the design of multiple-stage proofreading mechanisms had shown that high proofreading discrimination (P values) and multiple stages with lower P values were alternative means to improving accuracy with a given expenditure of energy for proofreading (Freter & Savageau, 1980). In this paper we have asked whether double-stage proofreading with lower P values might account for the experimental data regarding proofreading of valine errors by isoleucyl-tRNA synthetase of *E. coli*.

In our first attempt to fit the experimental data, we assumed that proofreading effort in the double-stage mechanism would be optimally distributed among the two stages. In this case, the best fit to the data was provided by a quasi-single-stage mechanism in which essentially all of the proofreading occurred at the second of the two stages. The predicted proofreading discrimination factors were $P_1 = 16.5$ and $P_2 = 1100$. This result clearly argues against two stages, each of comparable significance and each involving a binding site designed for isoleucine. The large value of P_2 is consistent with Fersht's double-sieve proposal. However, the location of the binding site designed for valine at the aminoacyl-tRNA rather than the aminoacyl adenylate stage is contrary to the suggestion in Fersht (1977b).

In view of this inconsistency, we were led to question the assumption that proofreading effort might be optimally distributed. By relaxing this assumption and fitting the model to the experimental data, we found that the best fit was provided by a double-stage mechanism in which the first stage was predominant ($P_1 = 8000$ and $P_2 = 358$). The result again clearly argues against two comparable stages involving isoleucine sites. In fact, the case with equal P values (case b in Table IV) requires that these proofreading discrimination factors have values of about 900, well in excess of that expected for an isoleucine site. The high P_1 value of 8000 indicates a site designed for the valine derivative, and in this case, it is located at the aminoacyl adenylate level as suggested by Fersht (1977b). Still this result is not entirely satisfactory because the second stage has a proofreading discrimination factor of 358, which is significantly higher than that of the conventional isoleucine site.

If we require that the second stage have a proofreading discrimination about equal to 100, then the best fit to the experimental data occurs with $P_1 = 29\,000$ and $P_2 = 121$. Such a mechanism would be consistent with both a valine site (double sieve) operating at the aminoacyl adenylate stage and an isoleucine site (conventional kinetic proofreading) operating at the aminoacyl-tRNA stage. However, in this case the error among the experimental data of Fersht (1977b) is as great as that in our first approach described above.

Attempts to distinguish between these two alternatives, each with the same total error, according to dynamic criteria were unsuccessful. The methods used were those previously described in Okamoto and Savageau (1984b). No dynamic advantages could be consistently associated with one or the other of these alternatives (data not shown).

From all these results we conclude (1) that a conventional double-stage mechanism with binding sites designed for isoleucine is inconsistent with the experimental data, (2) that a double-sieve mechanism with a binding site designed for valine is consistent with the experimental data, and (3) that the experimental data are not sufficiently accurate to determine whether the double sieve is located at the aminoacyl adenylate or aminoacyl-tRNA stage.

Errors in the estimates of k_1, k'_1, and k_2 are likely to be large. This is suggested by the experimental data themselves (Fersht, 1977b). In all the cases that we have examined, the value of k_1 is 13.4-fold greater than Fersht's experimental estimate. Attempts to reduce this error by varying the values of other parameters led to even greater errors in the other parameters. For example, the predicted value of the parameter k_1 can be made to agree with the experimental estimate, but discrepancies between the predicted and experimental values of k'_1 are then greater than 227-fold (results not shown). Thus, an even larger error in a parameter whose value is among the most accurate of those determined by Fersht (1977b) would be required to lower the error in k_1. These results suggest that the value of k_1 itself is questionable. This is in agreement with an earlier observation by Schreier and Schimmel (1972) that k_1 is remarkably influenced by temperature, pH, and Mg^{2+} concentration. An estimate of Q_{10} from their data is 7.4. If this value rather than our assumed value of $Q_{10} = 2$ were used to adjust Fersht's estimate for k_1, our results would be largely unaffected except that the error between predicted and experimental values for k_1 would be reduced from 13.4- to 3.6-fold.

These results and the results of our previous study suggest that the values of the parameters k_1, K_5, K_7, and X_3 must be estimated more precisely under conditions corresponding to those that exist in vivo if we are to definitively decide the stage at which the double-sieve mechanism normally operates—aminoacyl adenylate or aminoacyl-tRNA hydrolysis. There are other parameters that might have some affect on these results because their values involve considerable experimental uncertainty; e.g., their values were measured under conditions that do not correspond to those that exist in vivo. However, analysis shows that the sensitivities of these parameters are generally so low that an inordinately large experimental error (>100-fold in many cases) would be required for them to have a significant affect (Okamoto & Savageau, 1984a; results not shown).

Following the submission of this paper, we became aware of a recent paper by Lin, Baltzinger, and Remy (Lin et al., 1984) in which they have rigorously demonstrated a two-stage proofreading mechanism for rejection of tyrosine by the phenylalanyl-tRNA synthetase of yeast. They showed that previously measured rate constants for transfer of amino acid to tRNA are actually apparent rate constants that are the sum of the rate constant for transfer and the rate constant for hydrolysis of the enzyme-bound aminoacyl adenylate. Thus, previous estimates of the transfer rate are considerably in error, and they suggest that this is true of the threonine-valyl-tRNA synthetase and valine-isoleucyl-tRNA synthetase systems as well. These are the same rate constants that our approach indicates are most suspect in the case of isoleucyl-tRNA synthetase.

Registry No. Isoleucyl-tRNA synthetase, 9030-96-0.

REFERENCES

Eldred, E. W., & Schimmel, P. R. (1972) *Biochemistry 11*, 17–23.

Fersht, A. R. (1977a) *Enzyme Structure and Mechanism*, W. H. Freeman, San Francisco.

Fersht, A. R. (1977b) *Biochemistry 16*, 1025–1030.

Fersht, A. R. (1979) in *Transfer RNA: Structure Properties and Recognition* (Schimmel, P. R., Soll, D., & Abelson, J. N., Eds.) pp 247–254, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Freist, W., & Cramer, F. (1983) *Eur. J. Biochem. 131*, 65–80.
Freter, R. R., & Savageau, M. A. (1980) *J. Theor. Biol. 85*, 99–123.
Holler, E., & Calvin, M. (1972) *Biochemistry 11*, 3741–3752.
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.
Lin, S. X., Baltzinger, M., & Remy, P. (1984) *Biochemistry 23*, 4109–4116.
Loftfield, R. B., & Eigner, E. A. (1966) *Biochim. Biophys. Acta 130*, 426–448.
Mulvey, R. S., & Fersht, A. R. (1977) *Biochemistry 16*, 4731–4737.
Ninio, J. (1975) *Biochimie 57*, 587–595.
Okamoto, M., & Savageau, M. A. (1984a) *Biochemistry 23*, 1701–1709.
Okamoto, M., & Savageau, M. A. (1984b) *Biochemistry 23*, 1710–1715.
Pauling, L. (1958) in *Festschrift Arthur Stoll*, pp 597–602, Birkhaeuser, Basel, Switzerland.
Savageau, M. A., & Freter, R. R. (1979) *Biochemistry 18*, 3486–3493.
Savageau, M. A., & Lapointe, D. S. (1981) *J. Theor. Biol. 93*, 157–177.
Schreier, A. A., & Schimmel, P. R. (1972) *Biochemistry 11*, 1582–1589.
Yarus, M. (1969) *J. Mol. Biol. 42*, 171–189.

Cadmium Binding and Metal Cluster Formation in Metallothionein: A Differential Modification Study†

Werner R. Bernhard, Milan Vašák, and Jeremias H. R. Kägi*
Biochemisches Institut der Universität Zürich, CH-8057 Zürich, Switzerland
Received September 11, 1985

ABSTRACT: Mammalian metallothioneins (MT) contain 20 Cys in a total of 61 amino acid residues and bind 7 Cd and/or Zn ions. The metal is localized in two clusters made up of three and four metal–thiolate complexes in the NH_2- and COOH-terminal half of the chain, respectively [Otvos, J. D., & Armitage, I. M. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 7094–7098]. The formation of these oligonuclear complexes designated as Cd_4 and Cd_3 clusters has now been monitored in MT reconstituted with varying amounts of Cd by using differential chemical modification of Cys with [^{14}C]iodoacetamide. At ratios below 2–3 mol of Cd/mol of MT bound, no differential protection of Cys by the metal, and hence no preferred binding, is detectable. At Cd-to-protein ratios between 3 and 5 mol of Cd/mol of MT, the modification profiles reveal preferred and cooperative binding in the COOH-terminal half of the chain, indicating formation of the Cd_4 cluster. At still higher ratios, formation of the Cd_3 cluster is initiated in the NH_2-terminal section of the polypeptide chain. Comparison of the differential modification data of Cd_6–MT and Cd_7–MT suggests that the last Cd to be bound is coordinated to Cys ligands located mainly between positions 20 and 30 of the sequence. The extent of labeling of the different Cys in Cd_7–MT indicates that the ligands of the Cd_3 cluster are 3 times as accessible to iodoacetamide than those of the Cd_4 cluster, suggesting a greater thermodynamic or kinetic stability of the latter.

Metallothioneins (MT)[1] are widely occurring proteins characterized by a low molecular weight (6500–6800) and an extremely high metal and sulfur content. All known mammalian forms contain 20 Cys in a total of 61 amino acids. All Cys are conserved in the sequence and are serving as ligands for binding a total of seven bivalent d^{10} metal ions (Nordberg & Kojima, 1979). The synthesis of MTs can be induced by different metal ions, by glucocorticoid hormones, and by a variety of stress conditions. Zn and Cd are the major natural metallic constituents of mammalian MTs. The currently accepted model suggests that the seven diamagnetic metal ions are partitioned into two separate metal–thiolate clusters containing three and four metal ions (Otvos & Armitage, 1980) and located in domains formed by the NH_2- and COOH-terminal halves of the polypeptide chain, respectively (Winge & Miklossy, 1982).

Several lines of evidence indicate that binding of Cd to these clusters is not uniform. Thus, the observation that digestion of partially metal-depleted MT with subtilisin results in a residual fragment composed of the COOH-terminal half of the chain and containing the complete Cd_4 cluster signifies indirectly that the metal is lost preferentially from the Cd_3 cluster (Winge & Miklossy, 1982). Similarly, potentiometric titration studies (Kägi & Vašák, 1983; Avdeef et al., 1985) and the demonstration that at neutral pH EDTA removes a single Cd ion from the protein (Nicholson et al., 1986) strongly indicate the existence of at least two classes of Cd–thiolate complexes differing in thermodynamic stability.

† This research was supported by the Schweizerischer Nationalfonds (Grant 3.207-82).

[1] Abbreviations: MT, metallothionein; apo-MT, apometallothionein; Cd-MT, cadmium metallothionein; EDTA, ethylenediaminetetraacetic acid; TFA, trifluoroacetic acid; DTE, dithioerythritol; DTNB, 5,5′-dithiobis(2-nitrobenzoic acid); Tris, tris(hydroxymethyl)aminomethane; HPLC, high-performance liquid chromatography; CAM-Cys, (carboxamidomethyl)cysteine; NEM, *N*-ethylmaleimide.

0006-2960/86/0425-1975$01.50/0