Protection of an Unstable Reaction Intermediate Examined with Linear Free Energy Relationships in Tyrosyl-tRNA Synthetase†

Tim N. C. Wells‡ and Alan R. Fersht*

Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, U.K.

Received February 6, 1989; Revised Manuscript Received June 14, 1989

ABSTRACT: Linear free energy relationships (LFERs) are powerful tools in the search to understand the relationship between molecular structure and activity. They frequently link the changes in the rate constants for a reaction to changes in the equilibrium constant caused by alterations in structure. In physical-organic chemistry, these have been interpreted to give information on the structure of the transition state. Similar phenomena have been observed for reactions catalyzed by a series of engineered mutants of tyrosyl-tRNA synthetase from *Bacillus stearothermophilus*. LFERs are applied in this study to probe how the enzyme minimizes its side reactions. A linear free energy relationship is shown between the binding of the unstable enzyme–tyrosyl adenylate complex and its rate constant of hydrolysis. However, mutations of a key residue, His48, show significant deviation from the relationship, implying a role for the side chain in protection of the complex from hydroxide attack. A second linear free energy relationship is shown linking the rate and equilibrium constants for tyrosyl adenylate binding to the enzyme. Four distinct classes of mutation are discussed in the context of this relationship. The data from all but one of these groups of mutations conform well to a linear free energy relationship between the dissociation rate and dissociation equilibrium constants for the enzyme–tyrosyl adenylate complex with slope $\beta = 1.01 \pm 0.08$. The specificity of binding of tyrosyl adenylate is determined solely by its dissociation rate constant of the intermediate, and the mutations have relatively little effect on the enzyme–tyrosyl adenylate association rate. A detailed discussion of the methods of analyzing the statistical significance of linear free energy relationships and a guideline for the identification of outliers are given.

Understanding how enzymes achieve their remarkable catalytic power demands that the answers to two basic questions are understood. The first is how the enzyme optimizes the rate of the desired chemical reaction. Part of the answer to this involves tuning the internal thermodynamics of enzyme–reactant complexes [see, for example, Fersht (1974) and Albery and Knowles (1976)]. The limit to this rate enhancement has been theoretically defined and termed catalytic perfection (Albery & Knowles, 1976). The second question is how to minimize the rate at which unwanted side reactions are catalyzed. By analogy with synthetic organic chemistry, this process can be termed maximization of catalytic yield. In discussions of the optimization of enzyme catalysis this second factor is often overlooked.

Tyrosyl-tRNA synthetase from *Bacillus stearothermophilus* provides a system for investigating one aspect of the way in which enzymes enhance catalytic yield—the sequestering of a reactive intermediate within the active site cleft. The production of tyrosyl-tRNA proceeds via an enzyme–tyrosyl adenylate intermediate (E·Tyr-AMP):[1]

$$E + Tyr + Mg{\cdot}ATP = E{\cdot}Tyr\text{-}AMP + Mg{\cdot}PP \quad (1)$$

Tyrosyl adenylate is a mixed phosphoanhydride and as such is very susceptible to nucleophilic attack. In aqueous solution at pH 7.0, 16.7 kcal·mol^{-1} of free energy is released when it is hydrolyzed, the highest reported phosphotransfer potential for a biological molecule (Wells et al., 1986). To overcome the unfavorable equilibrium for its production, tyrosyl adenylate must be bound more tightly within the active site of the enzyme than the initial substrates [p 313 of Jencks (1969)]. A systematic protein engineering study of some of the strengths of the hydrogen bonds between the enzyme active site and the tyrosyl adenylate intermediate identified residues that contribute more than 3 kcal·mol^{-1} to the differential stabilization energy (Wells & Fersht, 1986).

As well as being thermodynamically unstable, tyrosyl adenylate is kinetically reactive and acts as a nonspecific tyrosylating agent (Wells et al., 1986). Similar results have been reported for the aspartyl- (Kern et al., 1985; Mejdoub et al., 1987) and phenylalanyl- (Rapaport et al., 1985) tRNA synthetases. As well as nucleophilic attack on tyrosyl adenylate bound within the active site, hydrolysis of tyrosyl adenylate occurs in free solution. This provides a futile cycle for ATP hydrolysis, which would increase the ratio of molecules of ATP consumed per molecule of tRNA charged in vivo. Tyrosyl-tRNA synthetase reduces the significance of the latter side reaction by a factor of 1500. In solution the hydrolysis rate for tyrosyl adenylate is 7.6×10^{-2} s^{-1} (Fersht, 1975; Mulvey & Fersht, 1977) whereas within the active site of the *B. stearothermophillus* enzyme the value is reduced to 5.14×10^{-5} s^{-1} (Wells et al., 1986).

The studies reported here investigate two of the possible ways in which the enzyme operates to reduce this futile cycle: by preventing the tyrosyl adenylate from being hydrolyzed while bound within the active site and by preventing its release into free solution. The approach used here to investigate these two mechanisms is to measure the rate constants for dissociation and hydrolysis of the enzyme–tyrosyl adenylate complex and the equilibrium constant for dissociation. By use of standard thermodynamic formulas, the rate and equilibrium

† This work was funded by the Medical Research Council of the U.K.

‡ Current address: Smith Kline & French Research Limited, The Frythe, Welwyn, AL6 9AR, U.K.

[1] Abbreviations: Tyr-AMP, tyrosyl 5′-adenylate; E·Tyr-AMP, enzyme–tyrosyl adenylate complex; pp, pyrophosphate.

0006-2960/89/0428-9201$01.50/0 © 1989 American Chemical Society

constants can be converted into the free energy of binding of tyrosyl adenylate and the free energy of the transition states for binding and hydrolysis. (The transition state for binding is defined as the point of highest free energy along the enzyme–tyrosyl adenylate binding pathway.) These free energies can then be regressed to determine whether a significant linear free energy relationship exists. Such an LFER indicates general trends in the relationship between structure and function. The slope β of the plot gives information about the energetics of spectroscopically inaccessible transition states. Furthermore, points deviating from the line indicate mutant proteins where either the structure of the protein has been grossly distorted or else a residue critical in stabilizing the transition state for the reaction has been modified.

Two LFERs are constructed: one compares the strength of binding of the E·Tyr-AMP complex with its rate of hydrolysis, and the second links the logarithm of the E·Tyr-AMP dissociation equilibrium and rate constants. The second LFER is of a different type from those found in physical-organic chemistry or from those reported previously for an enzyme system (Fersht et al., 1986, 1987). In this case, the event studied does not involve a chemical reaction taking place, merely a binding process followed by an enzyme–substrate isomerization. The mutant enzymes which have been included in this LFER analysis fall into four distinct categories: (i) mutations that according to the X-ray crystallographic structure will remove interactions between enzyme and the tyrosyl adenylate; (ii) mutations that remove interactions between the enzyme and the transition state for phosphotransfer, indicated by model-building studies (Leatherbarrow et al., 1985); (iii) mutations that alter residues on a highly mobile loop, which has recently been implicated in binding the transition state (Fersht et al., 1988); (iv) mutation Gly → Ala 47 (A. Vrielink, T. N. C. Wells, M. Panico, P. Carter, and D. M. Blow, unpublished results), which is designed to distort the binding of the tyrosyl adenylate by altering van der Waals contacts within the active site. The existence of a linear free energy relationship for these mutants can be used to describe a general two-step model for the binding of tyrosyl adenylate. This model is compared with the data for the binding of the alkylamino adenylate inhibitor tyrosinyl 5′-adenylate to the *Escherichia coli* tyrosyl-tRNA synthetase. The comparison assesses the validity of such an analogue as a mimic of the reaction intermediate.

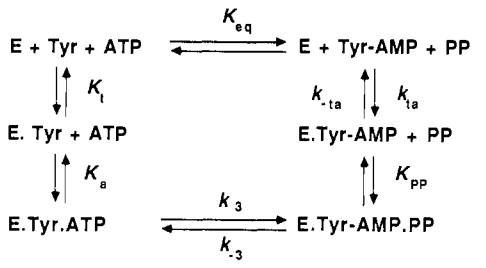
Materials and Methods

All reagents were obtained from Sigma, except for radiochemical which was from Amersham International. Reactions were carried out at 25 °C in 144 mM Tris-HCl buffer, pH 7.8, containing 10 mM $MgCl_2$. Enzymes were purified from overproducing clones in the bacteriophage M13 as described previously (Wells & Fersht, 1986), by use of the clones constructed by oligonucleotide site-directed mutagenesis in earlier studies.

Dissociation Rate Constant (k_{ta}; Scheme I). The rate constant for dissociation of [^{14}C]tyrosyl adenylate from tyrosyl-tRNA synthetase was measured by forming an enzyme–tyrosyl adenylate complex (200 nM) in the presence of 10 mM Mg·ATP and stoichiometric quantities of [^{14}C]tyrosine and enzyme. The radiolabeled tyrosyl adenylate was then "chased" off the active site. At the start of the experiment a large excess of "cold" tyrosine was added to the reaction mixture to prevent the rebinding of labeled tyrosyl adenylate to the enzyme, effectively chasing it from the active site.

Dissociation Equilibrium Constant ($K_{ta} = k_{-ta}/k_{ta}$). The dissociation constant for tyrosyl adenylate from the unligated

Scheme I



enzyme may be calculated from data obtained in constructing the reaction free energy profiles and the associated experiments (Wells & Fersht, 1986). Using a combination of equilibrium dialysis and pre-steady-state methods, it is possible to measure the equilibrium constant ($K_f = k_3K_{pp}/k_{-3}K_tK_{a'}$) (Scheme I) for the formation of an enzyme–tyrosyl adenylate complex and free pyrophosphate from enzyme, ATP, and tyrosine. For a limited number of mutant enzymes [those which lie within an "observation window" (Wells et al., 1986)], it is possible to measure the equilibrium constant K_{eq} for the formation of tyrosyl adenylate from ATP and tyrosine.

The value of this constant is independent of the enzyme used. Comparison of the results from studies of wild-type and three mutant enzymes gives an average value of 3.5×10^{-7} for K_{eq}. From thermodynamic considerations (the Haldane relationship), it can be seen that

$$K_{ta} = K_{eq}/K_f \quad (2)$$

This equation gives values of K_{ta} for the mutants which lie outside of the observation window.

pH Dependence of Enzyme–Tyrosyl Adenylate Hydrolysis. Enzyme-bound tyrosyl adenylate was prepared in situ (Wells & Fersht, 1986) except that the buffer was 10 mM Tris-HCl, pH 7.78. The excess ligands were removed on a Sephadex G-25 column. The complex was then diluted into a variety of buffers and incubated at 25 °C. The buffers used were as follows: 126 mM Bis-Tris-HCl at pH 6.0; 111 mM Tris-HCl, pH 7.16; 144 mM Tris-HCl, pH 7.78; 323 mM Tris-HCl, pH 8.54; 108 mM glycine, pH 9.0. All these buffers have constant ionic strength of $I = 0.1$M. To investigate the effect of the concentration of Tris-HCl buffer, two other buffer strengths, 646 mM Tris-HCl, pH 8.58 ($I = 0.2$ M), and 162 mM Tris-HCl, pH 8.52 ($I = 0.05$ M), were used. (The small shifts in the pH are caused by the effects of ionic strength on the pK_a of the buffer.)

The concentration of E·Tyr-AMP complexes in the reaction mixtures was always above 200 nM to ensure that virtually all the adenylate was bound to the enzyme (Wells et al., 1986) and thus that an insignificant proportion of the observed rate is due to hydrolysis in solution. A further control for this was to repeat the hydrolysis experiment with an equal concentration of the unligated enzyme being added. If a significant part of the adenylate is free in solution, then the addition of more free enzyme should sequester a fraction of that free adenylate and so reduce the observed rate of decay of E·Tyr-AMP complexes. In no circumstances was any reduction in rate observed. Samples were taken at defined time intervals and filtered through 2.5 cm diameter, 0.22 μm pore size nitrocellulose filters (Sartorius). The filters were washed with 5 mL of ice-cold buffer (144 mM Tris-HCl, pH 7.8), dried, and scintillation counted. Equivalent samples were quenched into trichloroacetic acid and filtered through glass fiber filters (Whatman GFC) in order to monitor the rate at which covalent [^{14}C]tyrosine–enzyme complexes were being formed.

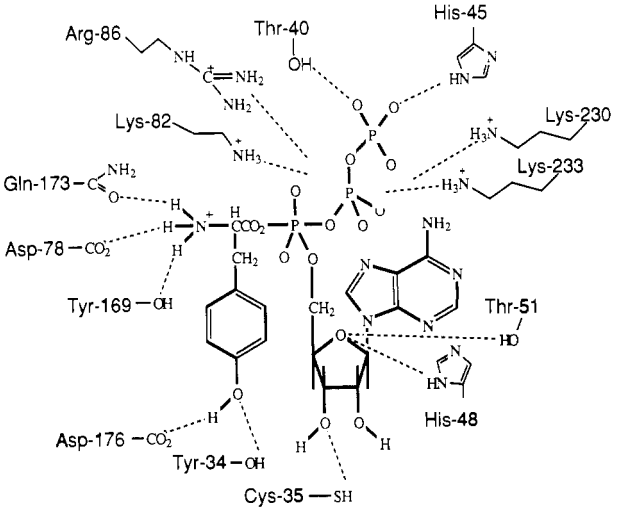


FIGURE 1: Model of the active site of tyrosyl-tRNA synthetase from *B. stearothermophilus*.

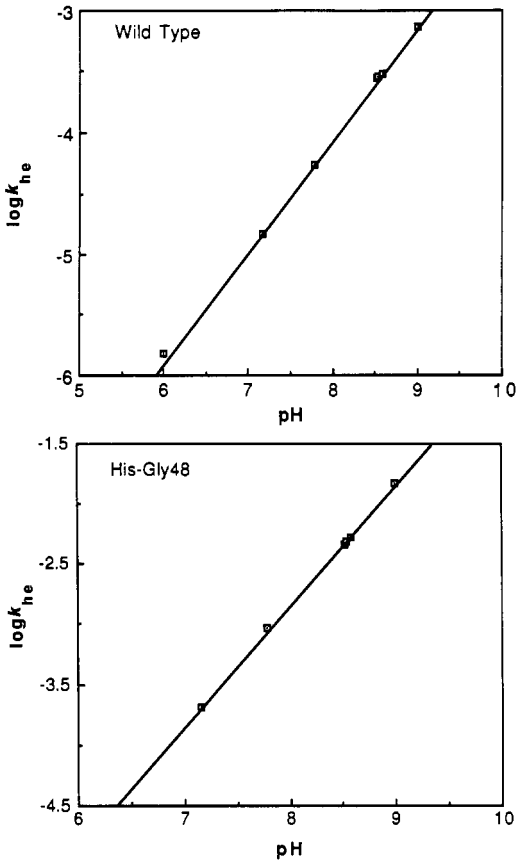


FIGURE 2: Dependence of the rate of hydrolysis of E·Tyr-AMP complexes on pH at 25 °C and pH 7.78. Reaction conditions described in the text. The three points at pH ~8 are for variation of buffer concentration (see text).

In no case was the level of [^{14}C]tyrosine covalently bound to the enzyme significantly above background levels.

Results

Decay of E·[^{14}C]Tyr-AMP Complex Is Caused by Attack of Hydroxide Ion. The effect of pH on the observed rate of hydrolysis of the enzyme-bound tyrosyl adenylate was studied for the wild-type enzyme and one mutant, His → Gly 48. It is seen in Figure 2 that in both cases the plot of the logarithm of the observed decay rate against pH is linear with a slope of unity. This confirms that the reaction is an attack of a single

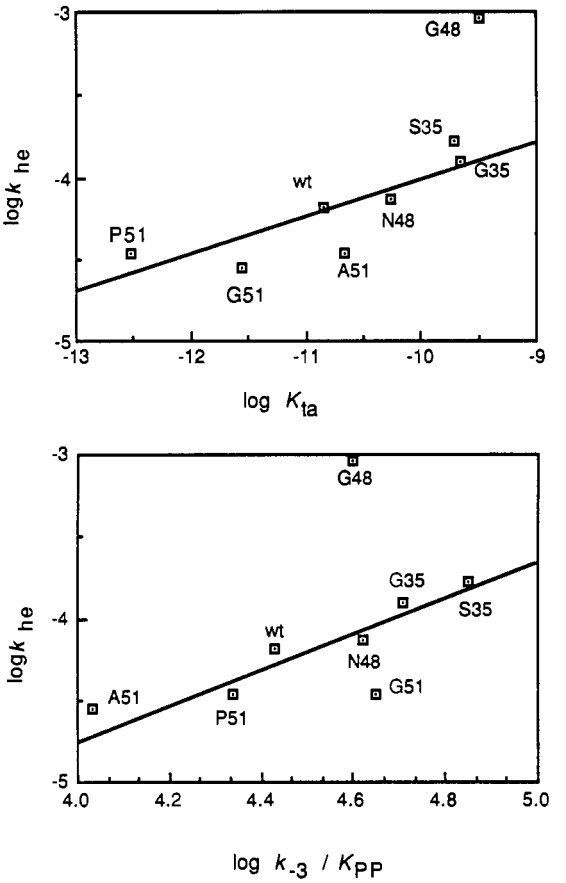


FIGURE 3: LFER for hydrolysis of E·Tyr-AMP complexes against dissociation constant and pyrophosphorolysis rate, respectively. All reactions carried out in 144 mM Tris-HCl, pH 7.78 (I = 0.1) at 25 °C.

hydroxide ion on the enzyme-bound intermediate and follows the pseudo-first-order rate law:

$$k_{he} = k_{bimolecular}[OH^-] \quad (3)$$

It is also seen in Figure 2 that the experiments to alter the buffer strength resulted in only small shifts in the overall decay rate. These shifts are consistent with the small perturbation in pH caused by the change in buffer strength and confirm there is no catalysis from the buffer. Qualitative observations on the other mutants in this study produced similar conclusions (TNCW and ARF, unpublished observations).

Linear Free Energy Relationship for E·Tyr-AMP Hydrolysis. In Figure 3 the data are plotted for log k_{he} (the rate constant for hydrolysis of enzyme-bound tyrosyl adenylate) against log K_{ta} (the dissociation constant for tyrosyl adenylate from its complex with the enzyme).

The hydrolysis rate of the E·Tyr-AMP complex was measured for the wild-type enzyme and seven mutants (Table I). These mutants involve changes of either Cys35, His48, or Thr51 (Figure 1), residues that have been shown to interact directly with the enzyme-bound tyrosyl adenylate (Rubin & Blow, 1981). Assuming an error variance ratio, $\lambda = 1$ (see Appendix). The data can be fitted to a linear functional relationship with slope $\beta = 0.36 \pm 0.12$ (correction coefficient $r = 0.77$).

An initial interpretation is that there is a link between the dissociation constant of the enzyme–tyrosyl adenylate complex and the rate constant for its hydrolysis. However, the LFER is not such a good fit as those previously published. One explanation of the bad fit is that Gly48 is an outlier, an idea backed up by the abnormal temperature dependence of hy-

Table I: Comparison of the Hydrolysis Rate of E·Tyr-AMP Complexes with the Dissociation Constant of E·Tyr-AMP Complex and Pyrophosphorolysis Rate[a]

enzyme	hydrolysis, log k_{he}[b]	tyrosyl adenylate binding		pyrophosphorolysis	
		log k_{ta}[c]	Δ[d]	log k_{-3}[c]	Δ[d]
wild type	−4.19	−10.85	+0.05	4.43	+0.08
Cys → Gly35	−3.91	−9.65	+0.03	4.71	+0.06
Cys → Ser35	−3.78	−9.71	+0.17	4.85	+0.04
His → Gly48	−3.04	−9.49	+0.84	4.60	+0.72
His → Asn48	−4.13	−10.26	−0.03	4.62	−0.02
Thr → Ala51	−4.56	−11.55	−0.28	4.03	+0.13
Thr → Gly51	−4.47	−1.67	−0.14	4.65	−0.27
Thr → Pro51	−4.47	−12.53	+0.18	4.34	−0.04

[a] Experiments carried out at pH 7.78 and 25 °C. [b] k_{he} is written as a unimolecular rate constant (eq 3), whereas k_3/k_{pp} is bimolecular. [c] Data calculated from Wells and Fersht (1986), Ho and Fersht (1986), and Fersht et al. (1987). [d] Minimum distance of point from maximum-likelihood LFER for Figure 3 (top) in the case of tyrosyl adenylate binding and Figure 3 (bottom) in the case of pyrophosphorolysis.

drolysis of Gly48 (Wells & Fersht, 1987). To test this, the best-fit line was calculated for the data set without Gly48 (see Appendix) and the minimum deviation from the line, Δ, calculated (Table I). The best-fit line has slope $\beta = 0.25 \pm 0.07$ (with an improved correlation coefficient of $r = 0.83$). The standard error of deviation from this best-fit line is 0.16. This means that Gly48 lies 5.3 standard errors away, giving a probability $p = 0.001$ that His → Gly48 does not deviate from the calculated LFER (Figure 3).

Before this can be accepted as the working hypothesis, two alternative explanations have to be examined. First, there is the possibility that some other point is the outlier. From Figure 3, the obvious candidate for an alternative outlier is Pro51. The analysis was therefore repeated, taking the data set with Pro51 missing. In this case a steeper line ($\beta = 0.56 \pm 0.16$) is obtained, with a standard error of deviation of 0.26 unit. Thr → Pro51 deviates from the line by 0.70 unit, and the probability that it does not deviate from the line is ~0.05. In other words, in a normal population one point in 20 would be this far off the line. Classification of Thr → Pro51 as an outlier is, therefore, not justified.

The second alternative hypothesis is that His48 is not an outlier but that the plot curves, the classic problem when an outlier is at the end of a series. To gain more insight into this problem, a second LFER was plotted (Figure 3). Here hydrolysis of the enzyme–tyrosyl adenylate complex is compared with pyrophosphorolysis, in other words the attack of magnesium pyrophosphate on E·Tyr-AMP to produce ATP and tyrosine. This rate constant is labeled k_{-3}/K_{pp} in keeping with earlier papers (Wells & Fersht, 1986). The choice of log (k_{-3}/K_{pp}) as one axis may seem somewhat strange initially, but there are good reasons for it. First, it has been shown (Wells, 1987) that, for the side chains in question, there is a significant correlation between log (k_{-3}/K_{pp}) and log K_{ta}, even when the mutual dependence of parameters has been allowed for. Some of the physicochemical components which go to make up log K_{ta} are thus retained in log (k_{-3}/K_{pp}). Outliers from the previous plot (Figure 3) should, therefore, remain outliers, provided that the slope of the new plot is significantly different from zero. It should be stressed that although this type of transformation is useful in outlier detection, interpretation of the resultant β value would be highly speculative. The second reason for plotting log (k_{-3}/K_{pp}) is that it is a much cleaner experimental parameter—the standard error on the data is only 58% of that for log K_{ta} [see Wells and Fersht (1986)]—making trends more apparent. Once again, the data show a linear relationship. Here, $\beta = 0.89 \pm 0.10$, i.e., 95%

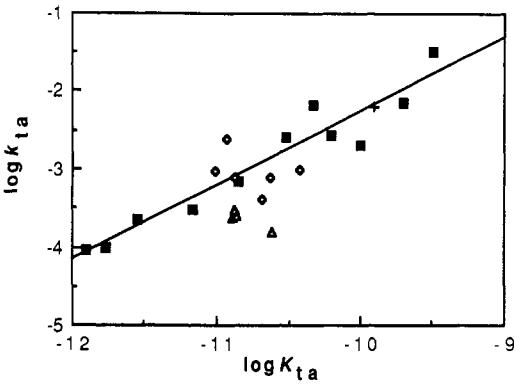


FIGURE 4: Linear free energy relationship between the dissociation rate constant and the dissociation equilibrium constant for tyrosyl adenylate and the tyrosyl-tRNA synthetase: (■) Mutations involving direct interaction in the E·Tyr-AMP complex. (◇) Mutations to residues implied in transition state binding from model building studies (Leatherbarrow et al., 1985). (△) Mutations to residues implied in highly mobile transition state binding loop (Fersht et al., 1987b). (+) Mutation (Gly → Ala47) constructed to perturb adenine conformation in E·Tyr-AMP (A. Vrielink, T. N. C. Wells M. Panico, P. Carter, and D. M. Blow, unpublished results).

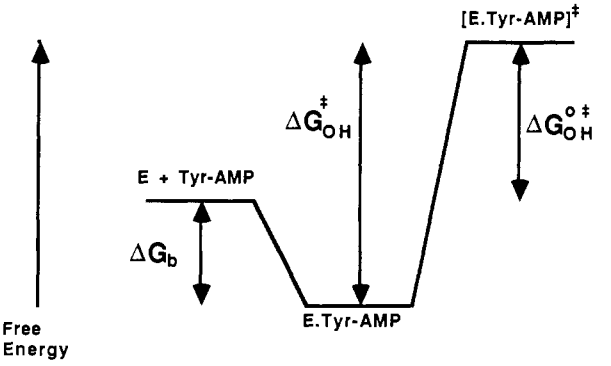


FIGURE 5: Free energy changes on binding and hydrolysis in the E·Tyr-AMP complex.

confidence limits at $\beta = 0.64$–1.30 (see eq A3). As expected, the data show a slightly smaller variation about the best-fit straight line. In this case, however, the mutant His → Gly48 clearly deviates from the line. The standard error of deviation from the best-fit line for this LFER is 0.12 kcal·mol^{-1}. His → Gly48 lies 0.72 kcal·mol^{-1} away from the LFER. By use of Student's t test, this gives a probability $p < 0.001$ that His → Gly48 is not significantly different from the other proteins in the study, confirming that the mutant His → Gly48 has an abnormally high rate hydrolysis of the E·Tyr-AMP complex.

Linear Free Energy Relationship for Tyrosyl Adenylate Binding. (i) Mutations in the Tyrosyl Adenylate Binding Pocket. The majority of the mutant proteins selected for this study are those which have modifications to hydrogen-bonding interactions in the region of the protein involved in tyrosyl adenylate binding. (The *E. coli* enzyme is included to enable comparisons with earlier studies of tyrosinyl 5′-adenylate binding to be made.)

There is a remarkably high correlation between the rate and equilibrium constants for tyrosyl adenylate dissociation from tyrosyl-tRNA synthetases over a range of 3 orders of magnitude, from His → Gly48, which contributes +1.85 kcal·mol^{-1} to tyrosyl adenylate binding, to Thr → Cys51, which contributes −1.44 kcal·mol^{-1}.

The data are plotted in Figure 4 and fit well to a line with slope $\beta = 1.01 \pm 0.08$. This is a good standard deviation, compared to those typically published for reactions from

Table II: Rate and Equilibrium Constants for Tyrosyl Adenylate Dissociation from Tyrosyl-tRNA Synthetase

enzyme	K_{ta} (M)[a]	k_{-ta} (s^{-1})	k_{ta} (s^{-1} M^{-1})	Δ[d]
wild type	1.41×10^{-11}	6.89×10^{-4}	4.89×10^{7}	+0.03
E. coli[b]	6.22×10^{-11}	2.77×10^{-3}	4.45×10^{7}	+0.06
(i) Mutations in the Tyrosyl Adenylate Binding Pocket				
Tyr → Phe34	6.29×10^{-11}	2.11×10^{-3}	3.35×10^{7}	+0.21
Cys → Ser35	2.01×10^{-10}	7.08×10^{-3}	3.52×10^{7}	+0.20
His → Gly48	3.22×10^{-10}	2.54×10^{-2}	7.88×10^{7}	−0.28
His → Asn48	4.63×10^{-11}	6.60×10^{-3}	1.43×10^{8}	−0.23
Thr → Ala51	2.82×10^{-12}	2.23×10^{-4}	7.91×10^{7}	−0.07
Thr → Cys51	1.25×10^{-12}	9.14×10^{-5}	7.37×10^{7}	−0.04
Thr → Gly51	3.00×10^{-11}	2.65×10^{-3}	8.86×10^{7}	+0.08
Arg → Ala86	6.71×10^{-12}	3.01×10^{-4}	4.48×10^{7}	+0.11
Arg → Gln86	1.70×10^{-12}	9.5×10^{-5}	5.58×10^{7}	−0.04
(ii) Mutations Implied in Transition State Binding from Model Building Studies (Leatherbarrow et al., 1985)				
Thr → Ala40	9.53×10^{-12}	9.5×10^{-4}	1.00×10^{8}	−0.13
His → Gly45	3.81×10^{-11}	9.8×10^{-4}	2.57×10^{7}	+0.28
His → Ala45	2.31×10^{-11}	7.9×10^{-4}	3.42×10^{7}	+0.20
His → Gln45	1.17×10^{-11}	2.4×10^{-3}	2.05×10^{8}	−0.35
His → Asn 45	1.34×10^{-11}	7.7×10^{-4}	5.75×10^{7}	+0.03
HG45TA40[c]	2.06×10^{-11}	4.0×10^{-4}	1.94×10^{7}	+0.38
(iii) Mutations Implicated in Transition State Stabilization from the Kinetics of E·Tyr-AMP Production (Fersht et al., 1987b)				
Lys → Ala230	1.38×10^{-11}	2.50×10^{-4}	1.81×10^{7}	+0.38
Lys → Asn230	1.34×10^{-11}	2.98×10^{-4}	2.23×10^{7}	+0.33
Lys → Ala233	1.26×10^{-11}	2.40×10^{-4}	1.90×10^{7}	+0.38
Lys → Asn233	2.42×10^{-11}	1.57×10^{-4}	6.48×10^{6}	+0.69
(iv) Mutation Designed To Produce Small Structural Changes within the Active Site				
Gly → Ala47	1.25×10^{-10}	6.45×10^{-3}	5.15×10^{7}	+0.07

[a] The dissociation constant for the E·Tyr-AMP complex is defined as $K_{ta} = k_{-ta}/k_{ta}$. [b] All mutations are made in the full-length tyrosyl-tRNA synthetase from *B. stearothermophilus*. The *E. coli* data are included to allow comparison with the tyrosinyl 5′-adenylate data (Grosse et al., 1979). [c] The double mutant His → Gly45, Thr → Ala40 (Leatherbarrow et al., 1985). [d] Minimum distance of point from maximum-likelihood LFER for data some type i mutations.

physical-organic chemistry. However, once again, there is more scatter than in previously published LFER plots from the enzymology of the tyrosyl-tRNA synthetase. One reason for the increased scatter appearing in the plot is the cumulative errors involved in calculating the free energy of binding of tyrosyl adenylate (Wells & Fersht, 1986). Despite the accuracy of the individual experimental data used, results from three processes have to be combined in the calculation, and so the variances from these three steps are combined. The data (Table II) show that there are no outliers within this group (see Appendix for details of calculation). The standard error of these points is 0.150 in the direction perpendicular to the best fit line.

(ii) Mutations Not in the Tyrosyl Adenylate Pocket but Which from Model Building Studies Affect Binding to the Transition State. Model building and subsequent mutagenesis studies show that some side chains which do not interact with the adenylate are involved in catalysis by transition state stabilization (Leatherbarrow et al., 1985). Mutation at these positions slows down the rate of formation of tyrosyl adenylate from ATP and tyrosine by as much as 3×10^5 in the case of the double mutant (Thr → Ala40; His → Gly45). However, these mutations appear to make no significant effect on either the dissociation rate or the binding constant for tyrosyl adenylate. When the data are overlapped with the data from section i, there is no significant change in the slope with $\beta = 1.03 \pm 0.10$. Similarly, it can be seen from the displacement distances that none of these points are significant outliers. The standard error of this group of mutations from the linear free energy relationship is 0.25, with a mean displacement of 0.068.

(iii) Mutations Implicated in Transition State Binding from Studies of the Kinetics of E·Tyr-AMP Formation from Tyrosine and ATP. Although this third class of mutation involves residues that are too distant to interact directly with the transition state in the model (Leatherbarrow et al., 1985), their conversion from lysine to either alanine or asparagine severely lowers the rate at which tyrosyl adenylate is formed from ATP and tyrosine (eq 1). The mechanistic interpretation (Fersht et al., 1988) for this is that in the transition state for the production of the E·Tyr-AMP complex the enzyme wraps itself around the pyrophosphate portion of the transition state. Both Lys230 and Lys233 are on a mobile loop which moves during this process, closing off the active site from bulk solvent. Mutation of the enzyme at both of these positions causes no major alteration in K_{ta}; however, k_{ta} is lowered by more than 7-fold in Lys → Asn233 and by 2.5-fold in the other three mutants. When the data are included on the linear free energy plot, the points all lie below the line. The inclusion of these points in the data fitting biases the result with β increasing to 1.22. Points corresponding to this class of enzyme show a much larger deviation from the linear free energy relationship calculated on the basis of the type i data. One explanation of these data is that there is something different between the two groups. The mean deviation here is +0.445 with a standard error of 0.082. By use of the equation $t = (d_1 - d_2)/(S_1^2/n_1 + S_2^2/n_2)$, the Student's t statistic is 2.63, significant to $p < 0.025$ at 13 degrees of freedom. An alternative explanation here is that only Lys → Asn233 is an outlier. To test this, the data from groups i and iii must be combined, to yield a mean deviation of $d = 0.074$ and a standard error of 0.195. Compared with this population, Lys → Asn233 is an outlier at a probability of $p < 0.01$.

(iv) Mutations Designed To Produce Small-Scale Structural Changes within the Active Site. The fourth class of mutation (Gly → Ala47) represents a different rationale to investigate structure–function relationships in enzymes (A. Vrielink, T. N. C. Wells, M. Panico, P. Carter, and D. M. Blow, unpublished results). Here a conserved glycine has been replaced by a larger amino acid, to investigate the effect of this possible structural perturbation on the function of the enzyme. The kinetically derived result is a 9-fold decrease in K_{ta} with a compensating change in k_{-ta}. The overall change in k_{ta} is insignificant, and the point for Gly → Ala47 lies well within the scatter of the LFER. This corresponds well with the minimal structural changes observed in the difference Fourier analysis of the X-ray crystallographic data.

Discussion

The analysis of the structure–function relationships of an enzyme using linear free energy relationships is a powerful technique in developing an understanding of enzyme catalysis. LFERs have been used recently to describe the energetics of the reaction transition state in tyrosyl-tRNA synthetase (Fersht et al., 1986, 1987) and to elucidate substrate specificity in subtilisin (Estell et al., 1986). Here, they have been used to analyze the methods by which the enzyme can use the binding energy of the intermediate, tyrosyl adenylate to protect against hydrolysis. The rationale behind both plots is an extension of the traditional ideas of LFERs from physical-organic chemistry.

From the initial experiments on tyrosyl-tRNA synthetase, we speculated that there might be a link between the binding energy of the E·Tyr-AMP complex and its rate of hydrolysis on the enzyme. To understand the mechanistic rationale behind this idea, it is necessary to consider the overall scheme

for binding and hydrolysis of tyrosyl adenylate:

$$E + \text{Tyr-AMP} \overset{K_{ta}}{\rightleftharpoons} E\cdot\text{Tyr-AMP} \underset{k_{-he}}{\overset{k_{he}}{\rightleftharpoons}} E\cdot\text{Tyr}\cdot\text{AMP} \quad (4)$$

Ideally, we should like to study the relationships of log k_{he} against log (k_{he}/k_{-he}) and ask how much of the overall change in binding energy in the hydrolysis of E·Tyr-AMP to E·Tyr·AMP is released in the transition state. This would be a similar question to that asked previously for the synthesis of E·Tyr-AMP (Fersht et al., 1987). However, the rate of formation of E·Tyr-AMP from enzyme-bound tyrosine and AMP (k_{-he}) is undetectable. The only question we can ask to probe the energetics of the hydrolysis transition state is what fraction of the binding energy released in going from free enzyme and Tyr-AMP to the hydrolysis transition state is released in the E·Tyr-AMP complex. The required LFER is thus

$$-\log K_{ta} = \beta' \log (k_{he}/K_{ta}) + \text{constant} \quad (5)$$

Figure 3 is a plot of log K_{ta} against log k_{he}, with slope $\beta = 0.25 \pm 0.07$. Since $\beta' = 1/(1 - \beta)$, $\beta' = 1.33$. The interpretation of this value is that, for every kilocalorie per mole of stabilization energy the enzyme puts into the hydrolysis transition state (relative to the free enzyme), it puts 1.33 kcal·mol^{-1} of stabilization energy into the E·Tyr-AMP complex. The finding of a β value greater than unity is not unexpected in light of this interpretation. The enzyme is able to discriminate between binding Tyr-AMP and binding the transition state for an unwanted side reaction.

There has been some debate about the validity of LFERs such as eq 5 where one parameter, in this case K_{ta}, appears on both sides of the equation (Fersht, 1987; Estell, 1987). It is stressed that this approach is valid provided that the two parameters (log K_{ta} and log k_{he}, here) correlate significantly. This is true both in this case ($r = 0.77$; $p < 0.01$) and in the disputed case of log k_3 against log k_{-3}, where $r = -0.67$ ($p = 0.01$) (Fersht et al., 1987).

The deviation of the point for Gly48 from the LFER for hydrolysis lends support to the hypothesis that His48 plays a role in protection of enzyme-bound tyrosyl adenylate against hydrolysis. The wild-type protein has a histidine at position 48, which is suggested from computer models to provide some steric hindrance to the pathway of nucleophilic attack of the hydroxide. When the histidine is replaced by a smaller amino acid such as glycine, such hindrance is removed, allowing a much greater rate of hydrolysis than would be expected from the LFER of the other mutants. Further experimental evidence for this hypothesis comes from two areas. First, the mutant His → Ala48 tyrosyl adenylate complex shows anomalously fast hydrolysis kinetics (Wells, 1987). On the other hand, the mutant His → Asn48 (Table I) and TyrTS (His → Gln48) (where the space previously occupied by the histidine imidazole ring is filled by the mutated side chain) do not show this enhanced hydrolysis rate. Second, the structure of the mutant His → Gly48 has recently been solved by X-ray crystallography (K. A. Brown and D. M. Blow, personal communication). In the space previously occupied by the imidazole ring, there is a new, well-defined water molecule. One of the roles of His48 is therefore to protect the enzyme-bound tyrosyl adenylate from hydrolysis by excluding this water from the active site.

It has often been suggested that substrate discrimination is in terms of k_{off}, the dissociation rate constant (Fersht, 1985), rather than the association rate constant. The dissociation LFER for the enzyme–tyrosyl adenylate complex dramatically confirms this idea. The slope $\beta = 1$ is interpreted as meaning for every incremental increase in binding energy contributed by a side chain to tyrosyl adenylate there is an equal increase in the activation energy of dissociation. The two processes, one thermodynamic and the other kinetic, are coupled, even though there is no a priori reason for expecting this. This interpretation holds for at least three of the four classes of mutation considered: first, all the residues which are seen to bind the tyrosyl adenylate in the X-ray crystallographic structure (Rubin & Blow, 1981); second, all those residues which interact directly with the pentacoordinate transition state model built from the coordinates of the E·Tyr-AMP complex (Leatherbarrow et al., 1985); third, the mutant Gly → Ala47 (A. Vrielink, T. N. C. Wells, M. Panico, P. Carter, and D. M. Blow, unpublished results) which was designed to perturb slightly the binding of adenylate by replacing a conserved glycine with a larger side chain.

The exceptions to this trend are mutations of Lys230 and Lys233. These residues are in a region of exceptionally high mobility, as judged from the crystallographic *B* factors (Fersht et al., 1988). From the kinetic analysis, it appears that these residues are on a loop which moves toward the substrate on formation of tyrosyl adenylate. In other words, they are implicated in the conformational change that occurs when the E·Tyr-AMP complex is formed from enzyme and tyrosyl adenylate. These mutants show only a small difference in the binding energy of tyrosyl adenylate binding but show a decrease in both the association and dissociation rates.

All of these results can be explained if the binding of tyrosyl adenylate is viewed as a two-step process:

$$E + \text{Tyr-AMP} \underset{k_{-bind}}{\overset{k_{bind}}{\rightleftharpoons}} E'\cdot\text{Tyr-AMP} \underset{k_{-close}}{\overset{k_{close}}{\rightleftharpoons}} E\cdot\text{Tyr-AMP} \quad (6)$$

The closing of the active site is the slowest step of the sequence in eq 6. Mutations in the loop region (i.e., Lys230 and Lys233) cause a change in k_{close}, resulting in a change in the observed overall association constant, k_{ta}. In all the other mutant proteins studied here, the effect of the mutation may be to alter k_{bind}. Since this is not the rate-determining step in the hydrolysis process, any change here will not alter k_{ta}. The direct result of this is that any changes in K_{ta} must be directly reflected in k_{-ta}, causing a β value of unity for the LFER. This result has been confirmed by modeling the isomerization in eq 6 into the pre-steady-state scheme of Wells et al. (1986). (T. N. C. Wells and A. R. Fersht, unpublished results).

Tyrosinyl 5′-Adenylate as an Analogue. Because of the reactivity of tyrosyl adenylate toward hydrolysis, many earlier ligand binding (Grosse et al., 1979) and crystallographic (Monteilhet & Blow, 1978) studies used the alkylamino adenylate tyrosinyl 5′-adenylate as a probe of the active site of tyrosyl-tRNA synthetases. Tyrosinyl 5′-adenylate, originally synthesized as a multisubstrate inhibitor, contains an alkyl linkage to the adenosine, in place of the carbonyl in tyrosyl adenylate (Figure 3). However, there are significant differences between the kinetics of binding of tyrosinyl adenylate and the "consensus" kinetics of tyrosyl adenylate binding as determined from the LFER. Measurements of the binding of tyrosinyl adenylate to *E. coli* tyrosyl-tRNA synthetase found values of $K_d = 2.5 \times 10^{-8}$ M and $k_{off} = 0.09$ s^{-1} (Grosse et al., 1978). From the standard formula, $\Delta\Delta G_b = RT \ln (K_{substrate}/K_{inhibitor})$, the inhibitor analogue binds 3.8 kcal·mol^{-1} less tightly than the true intermediate. This energy difference must partly result from the loss of the interaction made between the carbonyl of tyrosyl adenylate and the enzyme. The reduction in the rate constant for association to 3.6×10^6 s^{-1} places this point well off the binding LFER. This may imply

that either of the interactions between the substrate carbonyl and the enzyme stabilize the transition state for isomerization. Alternatively, the difference could be caused by differences in the geometry of the two molecules, one having an sp^2 carbon center and the other an sp^3.

These studies have shown that the usefulness of linear free energy relationships (in understanding enzyme-catalyzed processes) can be extended well beyond organic chemistry and can be applied to understanding enzyme-catalyzed processes. The importance of interpreting the β slope values in terms of relative stabilization and free energy is underlined since β values of greater than unity will arise in many enzymic LFERs.

Also, mutant enzymes whose characteristics deviate from the LFER have been identified or proposed in both cases. The interpretation of the data from these mutants both reinforces the prevailing ideas on the enzyme mechanism and suggests new approaches to understanding the catalytic process. The development of a statistical approach to outlier identification extends the utility of quantitative structure–function analysis in protein engineering. However, as always, it must be emphasized that the blind application of such techniques is extremely dangerous. Their conclusions will only be validated when backed up by intelligent experimentation and good structural data. As such, the heuristic cycle of producing a model, testing that model, and then refining the model continues.

ACKNOWLEDGMENTS

We thank Dr. R. J. Leatherbarrow and Jack W. Knill-Jones for helpful discussions on the data, Dr. Anthony Davidson of the Department of Mathematics, Imperial College, for reading the manuscript, and J. Wall for typing the manuscript.

APPENDIX

Statistical Analysis of the Linear Free Energy Relationship. (a) Regression Rather Than Correlation Is Required. The mathematical treatment of any quantitative structure–activity relationship should attempt to answer two fundamental questions: what is the best fit of a proposed analytical function to the data (the best straight line in the case of the linear free energy relationship); how precise is this estimate of the best fit?

At first sight, the problem of finding the value of the slope for the best-fit straight line to the data appears to be one of linear regression. Traditionally, this is the method by which the problem has been solved [pp 18–19 of Chapman and Shorter (1972); see references cited in Jencks (1985)]. However, in calculating the slope of a best-fit line by linear regression, the fundamental assumption is made that all the error lies on one axis, usually the Y axis. This implies that the values for the X axis are known precisely. While this may indeed be the case in many experimental situations, it is certainly not the case for the data used to construct linear free energy relationships. In these cases, the ratio of the error variance of the X axis to the Y axis can often be close to unity. In this case there is said to be a *functional relationship* between the two observables. The equation used to find the maximum likelihood estimator b, for the slope β, is given by [p 410 of Kendall and Stewart (1984)] $b^2 + b(\lambda S_{xx} - S_{yy}) - S_{xy} = 0$. The solution to this is

$$b = \{(S_{yy} - \lambda S_{xx}) + [(S_{yy} - \lambda S_{xx})^2 + 4S_{xy}^2]^{1/2}\}/2S_{xy} \quad \text{(A1)}$$

Here, λ is the ratio of the error variances on the Y axis and the X axis, σ^2_y/σ^2_x. The value of λ is known from the experiments to determine the binding and equilibrium constants. The terms prefixed by S in this expression are the sums of the differences from the respective means, for example, $S_{xy} = \sum_i^n (x_i - \bar{x})(y_i - \bar{y})$. When the error variance ratio tends to infinity, all the error is on the Y axis, and the solution to the equation becomes that for linear regression of y upon x; $b = S_{xy}/S_{xx}$. Similarly, when the error variance ratio tends to zero, the solution is the slope of the regression line of x upon y. Therefore, the value for b from the linear functional relationship expression will vary between the slopes of the two possible regression lines.

In order to assess the quality of the proposed functional relationship, it is necessary to calculate confidence limits for the maximum likelihood estimator. Confidence limits, β_1 and β_2, can be calculated from the somewhat complicated formula (Creasey, 1956)

$$\beta_1, \beta_2 = \lambda^{1/2} \tan [\tan^{-1} (b\lambda^{-1/2}) \pm 1/2 \sin^{-1} (2t\phi^{1/2})] \quad \text{(A2)}$$

In this expression, t is the appropriate Student's deviate at $n - 2$ degrees of freedom. The angle ϕ is defined by

$$\phi = \frac{\lambda S_{xx} S_{yy} - S_{xy}^2}{(n-2)[(S_{yy} - \lambda S_{xx})^2 + 4S_{xy}^2]} \quad \text{(A3)}$$

It should be stressed that the error structure for b will follow an angular distribution and that the value b will not necessarily be the midpoint of β_1 and β_2.

Throughout the literature on linear free energy relationships in physical-organic chemistry there has been a disturbing tendency to assess the quality of a proposed relationship by calculating the correlation coefficient r

$$r = S_{xy}/(S_{xx}S_{yy}) \quad \text{(A4)}$$

and testing its significance at the appropriate number of degrees of freedom. Clearly, this is not a useful approach, since the important parameter in these types of studies is the value β, and no information on the accuracy of the estimate is given by this approach. However, there is a another fundamental flaw, in that the hypothesis which is tested by this statistical approach is whether or not there is a linear free energy relationship. Correlation coefficients are used statistically to test whether or not there is a detectable linear relationship between variables. It is possible for a very good linear free energy relationship to exist when r is very small. For example, consider any linear free energy relationship where the value of β approaches zero. Then by definition a correlation coefficient of zero is formed. Under these circumstances there will be no significant correlation unless the variance ratio is already known (Fersht, 1987). The correct statistical test is to calculate the confidence interval for the estimate of the slope, β. If the argument of the inverse sine function in eq A2 is greater than 1, then the conclusion is that no value for the slope can be ruled out at the chosen confidence interval. In other words, any linear free energy relationship can be claimed to exist! For all other real values of the confidence limits, the quality of the linear free energy relationship can be assessed by the width of these confidence limits.

The other extreme, when β approaches infinity, is an interesting example. If linear regression methods are used on the data from a plot of log k_3 against log (k_3k_{pp}/k_{-3}) for the mutations in the transition state binding region of tyrosyl-tRNA synthetase [Figure 4 of Fersht et al. (1986)], then a clearly erroneous value of $\beta = 0.51$ is obtained for the best-fit slope if linear regression is used. The correlation coefficient is insignificant at $r = 0.08$. This is because linear regression assumes all the errors are parallel to the Y axis. However,

if the linear functional reationship is used rather than linear regression, with $\lambda = 2.0$, then the value for β is 83, with 95% confidence limits placed at −7.2 and +8.7 and passing through infinity. The angular dependence function calculated by Creasy (eq A2) means that any value of the slope more negative than −7.2 and more positive than 8.7 will satisfy 95% confidence limits.

(*b*) *Nature of the Error Structure.* The error structure in a linear free energy analysis consists of two components. There is error in the experimentally observed data, and there is also deviation from the regression line caused by small local perturbations in the enzyme structure. Behind this second type of "error" is the concept that even if the experimental observations had no technical error, the points would not necessarily all lie on the straight line. Before embarking on a detailed analysis of LFERs it is important to have a correctly defined statistical model, since in most cases incorrect statistics are worse than no statistics at all. Here the underlying assumption is that the errors in the free energies are normally distributed. This is most likely to be the case for the free energy changes caused by small local perturbations in structure upon mutation. However, the assumption that the experimental errors are normally distributed as free energies requires some explanation, since the measurements are made on kinetic and thermodynamic constants. The traditional form of the linear free energy relationship is

$$\ln Y = \alpha + \beta \ln X \qquad \text{(A5)}$$

Y and X cannot be measured directly; the values observed (η_i and ξ_i) are subject to experimental errors:

$$\xi_i = X_i + \delta_i'; \ \eta_i = Y_i + \epsilon_i' \quad i = 1, 2, ..., n \qquad \text{(A6)}$$

The notation is based on that of Kendall and Stuart (1984). Observations about any "true" value are distributed in a frequency distribution on the "error" random variable. It is our experience that this error is linearly dependent upon the size of the true value for kinetics experiments (Wells, 1987). This dependence can be eliminated by the definitions

$$\delta_i = \delta_i'/X_i; \ \epsilon_i = \epsilon_i'/Y_i \qquad \text{(A7)}$$

This allows a further simplification. It follows from eq A5 that the functional relationship demanded is between the logarithms of X_i and Y_i. Taking logarithms of eq A6 produces

$$\ln \xi_i = \ln (X_i + \delta_i'); \ \ln \eta_i = \ln (Y_i + \epsilon_i') \qquad \text{(A8)}$$

Provided that $\delta_i' \ll X_i$ and $\epsilon_i' \ll Y_i$ (in other words the observed deviation is small), a first-order Taylor series approximation can be applied to expand the brackets:

$$\ln \xi_i = \ln X_i + \delta_i'/X_i; \ \ln \eta_i = \ln Y_i + \epsilon_i'/Y_i \qquad \text{(A9)}$$

which by eq A8 becomes

$$\ln \xi_i = \ln X_i + \delta_i; \ \ln \eta_i = \ln Y_i + \epsilon_i$$

The error structure is therefore such that the logarithmic transform stabilizes the variances. Practically, this means that, provided the experimental errors are small and linearly related to the magnitude of the true variable, a linear functional analysis can be carried out on the logarithms of the observed data.

(*c*) *Detection of Outliers.* One of the most important aspects of the linear free energy relationship is the ability to detect outliers, that is, reactions which show significantly different thermodynamic properties to those forming the rest of the series. Although outlier detection is well developed for systems with errors only on one axis (Barnett & Lewis, 1984) and for the linear structural model (Barnett, 1985), no definitive statistical work has been carried out on the linear functional model. In the absence of a rigorous test for outliers, a variation on the method of Snedecor and Cochran (1980, p 168) has been applied. The regression was recomputed with the proposed outlier omitted. The distances of all the points from the regression line were calculated, and a Student's t test was carried out to see whether the proposed outlier was significantly outside of the normal error distribution. One caveat must be stressed when this method of outlier analysis is used. The most serious overestimate of statistical significance will occur when the proposed outlier is on the extreme left or right of the graph. In this situation, extreme caution should be observed since there is an alternative hypothesis to explain the data, that of a nonlinear free energy relationship.

Registry No. His, 71-00-1; Lys, 56-87-1.

References

Albery, W. J., & Knowles, J. R. (1976) *Biochemistry 15*, 5621–5631.

Barnett, V. (1985) *Austral. J. Statist. 27*, 151–162.

Barnett, V., & Lewis, T. (1984) *Outliers in Statistical Data*, 2nd ed., Wiley, New York.

Chapman, N. B., & Shorter, J. (1972) *Advances in Linear Free Energy Relationships*, Plenum, London.

Creasey, M. A. (1956) *J. R. Statist. Soc. B 18*, 65–71.

Estell, D. A. (1987) *Protein Eng. 1*, 441–446.

Estell, D. A., Graycar, T. P., Miller, J. V., Powers, D. B., Burnier, J. P., Ng, P. G., & Wells, J. A. (1986) *Science 233*, 659–663.

Fersht, A. R. (1974) *Proc. R. Soc. London B187*, 397–414.

Fersht, A. R. (1975) *Biochemistry 14*, 5–12.

Fersht, A. R. (1985) *Enzyme Structure and Mechanism*, 2nd ed., W. H. Freeman, London.

Fersht, A. R. (1987) *Protein Eng. 1*, 442–446.

Fersht, A. R., Leatherbarrow, R. J., & Wells, T. N. C. (1986) *Nature 322*, 284–286.

Fersht, A. R., Leatherbarrow, R. J., & Wells, T. N. C. (1987) *Biochemistry 26*, 6030–6038.

Fersht, A. R., Knill-Jones, J. W., Bedouelle, H., & Winter, G. (1988) *Biochemistry 27*, 1581–1587.

Grosse, F., Kraus, G., Kownatzki, R., & Maass, G. (1979) *Nucleic Acids Res. 6.*, 1631–1638.

Jencks, W. P. (1969) *Catalysis in Chemistry and Enzymology*, McGraw-Hill, New York.

Jencks, W. P. (1985) *Chem. Rev. 85*, 511–517.

Kendall, M. G., & Stuart, A. (1984) *The Advanced Theory of Statistics. Inference and Relationship*, 5th ed., Vol. 2, Griffin, London.

Kern, D., Lorber, B., Boulanger, Y., & Giege, R. (1985) *Biochemistry 24*, 1321–1332.

Leatherbarrow, R. J., & Fersht, A. R. (1987) *Biochemistry 26*, 8524–8528.

Leatherbarrow, R. J., Winter, G., & Fersht, A. R. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 7840–7844.

Lowe, G., & Tansley, G. (1984) *Tetrahedron 40*, 113–117.

Mejdoub, H., Kern, D., Giege, R., Ebel, J.-P., Boulanger, Y., & Reinholt, J. (1987) *Biochemistry 26*, 2054–2059.

Monteilhet, C., & Blow, D. M. (1978) *J. Mol. Biol. 122*, 407–417.

Mulvey, R. S., & Fersht, A. R. (1977) *Biochemistry 16*, 4005–4013.

Rapaport, E., Yogeeswaran, G., Zamecnik, P. C., & Remy, P. (1985) *J. Biol. Chem. 260*, 9509–9512.

Rubin, J., & Blow, D. M. (1981) *J. Mol. Biol. 145*, 489–500.

Snedecor, G. W., & Cochran, W. G. (1980) *Statistical Methods* 7th ed., Iowa State University Press, Ames, IA.
Wells, T. N. C. (1987) Ph.D. Thesis, University of London.
Wells, T. N. C., & Fersht, A. R. (1986) *Biochemistry 25*, 1881–1886.
Wells, T. N. C., & Fersht, A. R. (1987) *Protein Eng. 1*, 261.
Wells, T. N. C., Ho, C. K., & Fersht, A. R. (1986) *Biochemistry 25*, 6603–6608.

Role of Cationic Residues in Cytolytic Activity: Modification of Lysine Residues in the Cardiotoxin from *Naja nigricollis* Venom and Correlation between Cytolytic and Antiplatelet Activity†

R. Manjunatha Kini and Herbert J. Evans*
Department of Biochemistry and Molecular Biophysics, Medical College of Virginia, Virginia Commonwealth University, Richmond, Virginia 23298

Received May 15, 1989; Revised Manuscript Received July 13, 1989

ABSTRACT: Cardiotoxins and postsynaptic neurotoxins from snake venoms have similar primary, secondary, and tertiary structures. Cardiotoxins, however, in contrast to neurotoxins, exhibit general cytotoxicity. Comparison of the distribution of hydrophobic and charged amino acid residues in the three-dimensional structures of lytic cardiotoxins and nonlytic neurotoxins indicates the presence of a cationic site associated with a hydrophobic surface in cardiotoxins, but not in neurotoxins. A cationic site flanked by a hydrophobic site is a common structural feature shared by a wide variety of unrelated cytolysins and is predicted to determine the lytic activity of a large group of cytolysins. To determine the essential nature of the cationic site in cardiotoxin CTX-1 from *Naja nigricollis crawshawii* venom, we modified the positive charges of nine Lys residues to negative, neutral, or positive charges by succinylation, carbamylation, or guanidination, respectively. Circular dichroism studies indicated that these modifications did not affect the conformation of the cardiotoxin. Binding of the modified cardiotoxins to phospholipids was demonstrated by changes in the intrinsic fluorescence of native and modified CTX-1 after binding to phospholipid vesicles, and by resonance energy transfer with anthracene–phospholipid vesicles. Phospholipid binding was not affected by these modifications, but their binding preference was determined by the electrostatic interactions between the polypeptide and phospholipid. Both positively charged native and guanidinated CTX-1 showed direct lytic activity on human erythrocytes and platelets, whereas the succinylated or carbamylated derivatives did not show lytic activity. The loss of lytic activity cannot be related to conformational changes or phospholipid binding abilities of the modified cardiotoxins. Thus, these results suggest a significant role of the positive side chains of Lys residues in determining the lytic activity of CTX-1. We demonstrated earlier that the antiplatelet effects shown by the cardiotoxins are due to their lytic ability. This is further supported by the observed antiplatelet effects of lytic native or guanidinated CTX-1. Nonlytic succinylated or carbamylated CTX-1 failed to affect platelet aggregation, supporting the proposed lytic mechanism for antiplatelet effects.

Snake venoms are complex mixtures of toxic proteins and polypeptides. Among these toxins are neurotoxins and cardiotoxins, the family of polypeptides of molecular weights 6000–7500, which have been well characterized (Dufton & Hider, 1983, 1988). Neurotoxins, which block neurotransmission postsynaptically at the nicotinic site of acetylcholine receptors, belong to 2 classes: short-chain neurotoxins with 60–64 amino acid residues and 4 disulfide bridges and long-chain neurotoxins with 65–74 amino acid residues and 5 disulfide bridges. These two groups have high structural homology and share common postsynaptic activity (Dufton & Hider, 1983, 1988). Cardiotoxins, which are isolated only from elapid snake venoms, closely resemble short-chain neurotoxins, with 60 amino acid residues and 4 disulfides. Despite the structural homology, cardiotoxins are more basic and more hydrophobic than the neurotoxins, and differ from neurotoxins in their pharmacological properties. Cardiotoxins exhibit depolarization of cardiac, skeletal, and smooth muscles resulting in muscle contracture, depolarization of nerve cells leading to loss of excitability, general cytolysis, and a weak anticoagulant effect (Dufton & Hider, 1983, 1988; Harvey, 1985; Condrea, 1974; Teng et al., 1987; Kini et al., 1987).

We purified four nonenzymatic, anticoagulant, and antiplatelet polypeptides from *Naja nigricollis crawshawii* venom (Kini et al., 1987) and identified these polypeptides as cardiotoxins on the basis of their amino acid composition, amino-terminal sequences, and direct lytic activity on human erythrocytes (Kini et al., 1988). The cardiotoxins inhibit aggregation of platelets, and we demonstrated that the observed inhibition is merely a reflection of the lytic ability of these cardiotoxins (Kini & Evans, 1988). Lysis of blood cells and release of inhibitory substances, and lysis of platelets coated on electrodes, lower the impedance, and thus result in apparent inhibition of aggregation on a whole blood aggregometer. In support of this, we also showed that neurotoxin II from *N. naja oxiana* venom fails to exhibit any platelet

† This work was supported by Research Grant HL-24281 from the National Institutes of Health, U.S. Public Health Service.
* To whom correspondence should be addressed.

0006-2960/89/0428-9209$01.50/0 © 1989 American Chemical Society