Kinetic Class Analysis of Hydrogen-Exchange Data[*]

S. L. Laiken and Morton P. Printz

ASTRACT: Recent use of radioisotope methodology has made accurate hydrogen-exchange data readily available. From such data, quantities characterizing the structure and motility of macromolecules in solution may be calculated. In order to perform these calculations, the number of exchangeable hydrogens in each kinetic class, as well as the rate constants for exchange, must be determined. This requires an analysis of the data in terms of a linear combination of exponentials. Due to numerical and statistical difficulties, previously employed methods of analysis appear to be inadequate. We have taken a new approach and have applied the method of damped nonlinear least squares with considerable success. A computer program for this method gives class sizes, rate constants, the number of kinetic classes, and error estimates for all the parameters. Tests utilizing precise data perturbed by random errors demonstrate the reliability of this approach. However, there are limitations on any method for analyzing data in terms of a linear combination of exponentials. These limitations apply primarily to complex macromolecular systems such as proteins. We have found that the large number of simultaneous yet distinct exponential decay processes which might arise can be approximated by a much smaller number. When hydrogen-exchange curves are simulated using fictitious systems with various random distributions of rate constants, they invariably can be fit with two or three distinct kinetic classes. However, it is shown that these classes have no clear physical meaning and that the only information unambiguously obtainable from such a system is the mean rate constant and the first few moments about that mean. Therefore, unless it can be assumed safely that the exchangeable hydrogens of a macromolecule actually occur in a small number of kinetic classes, analysis may lead to serious misconceptions.

Recently, the hydrogen-exchange technique has been applied to a variety of problems involving macromolecules of biological interest. In some studies, alterations of hydrogen exchange have been used to indicate the occurrence of conformational changes, while in others, problems of a more quantitative nature involving motilities, thermodynamics of transconformational reactions, or identification of structure have been treated. In these latter cases, it is necessary to have accurate estimates of kinetic class sizes and rate constants. The importance of obtaining these parameters should not be underestimated, for without them, the amount of information obtainable from a hydrogen-exchange experiment is diminished greatly. A method which would allow extraction of this information from real experimental data is therefore essential. In the course of our investigations, we have found that most of the methods currently available appear to be inadequate and often lead to an erroneous view of the underlying physical situation. Here, we will describe a method of analysis, damped least squares, which does not share the inadequacies of previous methods and has been found quite satisfactory. In addition, the applicability of this or any other technique for analysis of hydrogen-exchange data will be related to the nature of the macromolecular system under consideration.

Previous Methods. Since the relative merits of previously available methods will be discussed elsewhere in detail, only brief comments will be made here. Most hydrogen-exchange curves may be represented by the relation (Hvidt and Neilsen, 1966)

$$H(t) = \sum_{i=1}^{n} c_i e^{-k_i t} \qquad (1)$$

where $H(t)$ is the number of hydrogens remaining unexchanged at time t, per molecule; c_i is the number of hydrogens in the ith kinetic class; k_i is the rate constant of the ith kinetic class; and n is the total number of kinetic classes. The graphical "successive subtraction" method has been the most widely used technique for determining the c_i and k_i. Such a method is, of course, ultimately subjective and the resulting pitfalls have been discussed in detail by Riggs (1963). Nongraphical methods have been proposed by Prony (Hildebrand, 1956), Cornell (1956, 1962), and Gardner *et al.* (1959). All of these methods require data points to be taken at regular intervals in time, which may be inconvenient. Failure to do so necessitates interpolation of the data, thereby increasing the uncertainty of the results. However, a more serious objection to these methods is simply that they lack a statistical basis. Consequently, standard errors on the estimates of the c_i and k_i cannot be obtained. In addition, there is no way of knowing whether a given set of class sizes and rate constants provides the best possible fit to the data for a given value of n, and there is no way to decide what value of n to use. One must resort to a visual judgment of the fit. Such a procedure can lead to serious misconceptions, as described below. The relation

$$H(t) = 1029e^{-0.0017t} + 1176e^{-0.0105t} + 1190e^{-0.0798t} \qquad (2)$$

which has rate constants typical of those found in proteins (Segal and Harrington, 1967), was employed to generate 40 values of $H(t)$, from 0 to 945 min. Then, with the aid of a random number generator, a normally distributed error of mean zero and standard deviation 5% was obtained for each point, to give a simulated $H(t)$ *vs.* t curve, as shown in Figure 1. The best possible fit with two exponentials, obtained by a

---

[*] From the Rockefeller University, New York, New York 10021. *Received October 3, 1969.* This work was supported in part by U. S. Public Health Service Grant No. AM 02493.

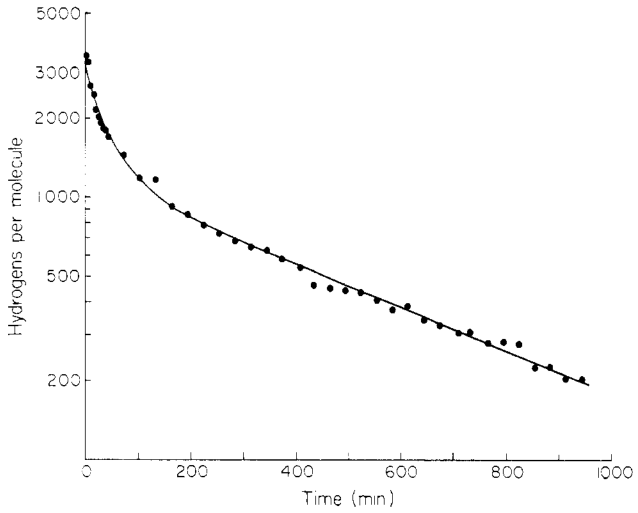

FIGURE 1: Analysis of a simulated hydrogen-exchange experiment using two kinetic classes: The filled circles represent mock data points generated from eq 2 and perturbed by normally distributed random errors of standard deviation 5%. The best fit using two classes is represented by the solid curve.

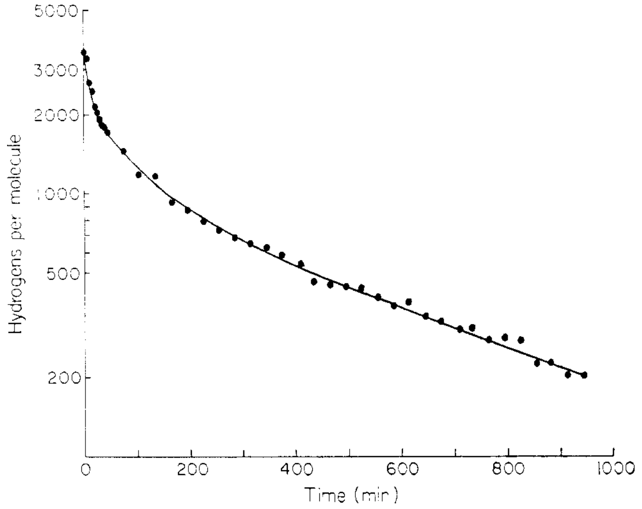

FIGURE 2: Analysis of a simulated hydrogen-exchange experiment using three kinetic classes: The mock data are identical with those of Figure 1. However, the solid curve now represents the best fit using three kinetic classes.

special nonlinear least-squares procedure to be described below, was

$$H(t) = 1180e^{-0.00188t} + 2212e^{-0.0221t}$$

Looking at Figure 1, it would appear that the data had been fit quite well, especially if a graphical method had been used. However, all of the parameters are incorrect, even though the fit using the correct number of classes, shown in Figure 2, does not look dramatically better. We are tempted to state, as is often done, that the data "could be fit with two classes." Depending on the purpose of the experiment, this conclusion could be quite misleading. For example, it might be taken to indicate the existence of two types of structure or regions of accessibility rather than three. The basic problem is that the choice between the two and three class models cannot be made by eye. However, using the damped least-squares procedure, we will show that such a choice can be made readily.

The Damped Least-Squares Method. From a statistical point of view, the best estimates of the c_i and k_i for a given number of kinetic classes can be obtained by utilizing Gauss's celebrated principle of least squares and minimizing (Keepin *et al.*, 1957; Rossing, 1966)

$$S = \sum_{j=1}^{M} w_j \left[H(t_j) - \sum_{i=1}^{n} c_i e^{-k_i t_j} \right]^2 \tag{3}$$

where w_j is the reciprocal of the variance of the jth observation, *i.e.*, its weight. The values of the k_i and c_i at the minimum will yield the best possible fit of the experimental data for a given value of n. While it would appear that any iterative technique of function minimization could be applied successfully to eq 3, we have found, as have others, that this is not the case (Pitha and Jones, 1966). Only our version of the damped least-squares method proved consistently successful. The method was formulated initially by Levenberg (1944), who noted that the commonly employed Gauss–Newton minimization procedure often tended to overshoot the minimum and diverge.[1] By inclusion of a scalar factor, p, in the iterative step, Levenberg found that progress toward the minimum could be "damped" and convergence obtained. The algorithm was later formulated independently by Marquardt (1963) and modified by Meiron (1965) to increase the rate of convergence. Successful application of the method to the fitting of overlapping spectral bands containing up to 65 variables was made by Pitha and Jones (1966) and Papouseck and Pliva (1966). The reader is referred to these authors for lucid discussions of the details and advantages of the method. Our own version differs only in the manner of assigning weights and optimizing p, the Levenberg parameter. We define a set of arrays as follows, using the notation of Pitha and Jones.[2] $X = x_1, x_2, \ldots\ldots x_{2n}$ is a column vector of variables, with odd-numbered x being the class sizes and even-numbered x the rate constants. **F** is a column vector of elements $f_1, f_2, \ldots\ldots f_M$ which are equal to the number of hydrogens remaining at the jth data joint minus the number calculated at that point. A is an $M \times 2n$ matrix of partial derivatives with

$$a_{ij} = \frac{\partial f_i}{\partial x_j}$$

B is a matrix and **G** is a column vector defined by

$$\mathbf{B} = \mathbf{A}^T\mathbf{A} \qquad \mathbf{G} = A^T F$$

The M observations *need not* have any regular spacing. The

---

[1] See Worsley and Lax (1962) for a description of the Gauss–Newton method applied to linear combinations of exponentials and its reliability.

[2] Capital letters are used here to indicate matrices and vectors, while the respective lower case letters denote their elements. The superscript T implies the transpose of a matrix.

iterative step then can be written as (Pitha and Jones, 1966)

$$\mathbf{X}_{m+1} = \mathbf{X}_m - (\mathbf{B}_m + p\mathbf{C}_m)^{-1}\mathbf{G}_m$$

where **C** is a diagonal matrix formed from **B** by including only the elements on the main diagonal of **B**. An initial guess at the **X** vector usually can be made after an inspection of a semilogarithmic plot of $H(t)$ *vs.* t. The initial value of p is taken as 1.0. If an iteration fails to give a value of S (given the current $\mathbf{B}_m$ and $\mathbf{G}_m$) lower than the previous one, p is doubled until the lowest value of S is obtained. If an iteration succeeds in giving a lower value of S, p is halved until S is improved as much as possible. The corresponding $\mathbf{X}_{m+1}$ vector then becomes $\mathbf{X}_m$ for the next iteration. In either case, the best p value is used as the next starting p value.

In order to obtain the statistically best values of the x_i, a weighting procedure must be used. We assume all samples have been assayed in a scintillation counter long enough to make counting errors negligible, and that the per cent error variance in reading the optical densities is the same for all samples. The variance of each data point is then simply proportional to the square of the true value of $H(t)$ at that point (Papouseck and Pliva, 1966; Deming, 1943). Since the true value of $H(t)$ is, in fact, unknown, we approximate it by the nonweighted least-squares estimate, *i.e.*, that obtained with the w_j set equal to one. From this value, a weighting factor is generated by taking the reciprocal of its square. The new iterative step then becomes

$$\mathbf{X}_{m+1} = \mathbf{X}_m - (\mathbf{B}' + p\mathbf{C}')^{-1}\mathbf{A}^T\mathbf{W}\mathbf{F}$$

where **W** is the diagonal matrix of normalized weights, w_j; $\mathbf{B}'$ is given by $\mathbf{A}^T\mathbf{W}\mathbf{A}$; and $\mathbf{C}'$ is found from $\mathbf{B}'$ rather than from **B**. These weights prevent the assignment of undue emphasis to early or inaccurate observations, where f_i values are large. If different data points have different counting and optical density percentage errors, individual weights may be approximated for each point by standard techniques (Deming, 1943). The nonweighted least-squares procedure is terminated, when the x_i are constant to 1% between iterations, while the weighted procedure is terminated when constancy is to 0.1%.

Finally, we note that the inverse of $\mathbf{B}'$ used in the final iteration is an approximation to the variance–convariance matrix. From the diagonal elements of $(\mathbf{B}')^{-1}$, given by b_{ii}', we may calculate the standard errors of the x_i if the experimental errors are normally distributed. The standard errors are given by

$$\sigma_{x_i} = [b_{ii}'S/(M - n)]^{1/2}$$

and approximate confidence limits are obtainable in the usual fashion (Hamilton, 1964).

We have found the technique to work well, even with poor initial guesses, on a variety of simulated and actual data. Tests of our program were made by attempting to fit data in which the points were perturbed by "random noise" as described previously, thereby simulating an actual experiment. When the standard deviation of the noise was 2 or 5%, the fits of eq 2 shown in Table I were obtained. As expected, the standard errors of the c_i and k_i in the latter case are about 2.5 times those in the former. In the course of these and other analyses, it was found that 20–40 points generally were needed for proper fitting. The reason for this is simply that the rate constants are spaced over several orders of magnitude. Therefore, a reasonable number of closely spaced points are needed at the beginning of the curve, and a reasonable number of more widely spaced points are needed to define the terminal region.

TABLE I: First-Order Rate Constants (k, min^{-1}) and Class Sizes (c, Hydrogens per Molecule) Obtained from a Simulated Hydrogen-Exchange Experiment.

Error Level (%)	Class	k	Std Error	c	Std Error
2	I	8.01×10^{-2}	0.86×10^{-2}	1242	71
5	I	8.06×10^{-2}	2.10×10^{-2}	1322	186
2	II	1.094×10^{-2}	0.072×10^{-2}	1171	56
5	II	1.155×10^{-2}	0.192×10^{-2}	1166	152
2	III	1.718×10^{-3}	0.029×10^{-3}	1039	23
5	III	1.726×10^{-3}	0.069×10^{-3}	1050	53

There are several experimental situations which arise when using damped least-squares or any other method of analysis in which a value of n must be assumed: (a) the number of kinetic classes and their sizes are known; (b) the number of kinetic classes is known; (c) the number of kinetic classes is unknown. In case a, we treat the c_i as constants and proceed with the analysis. If the data are good, accurate results for the k_i may be obtained. Case b may be treated by fitting the data using the known value of n. Such cases are often encountered with small peptides or with polymers in which the monomers can assume only a small number of conformational states. Case c presents many complications, but aided by the damped least-squares method and some simple statistical tests, we may set upper and lower bounds for the true value of n. Generally these bounds are identical and n is determined. First, consider the problem of setting the lower bound. Models such as those represented by eq 1 are classed, statistically, as intrinsically nonlinear because no transformation can remove the nonlinearity of $H(t)$ in the k_i when $n > 1$. Due to this, the usual methods for testing the significance of additional terms in linear models cannot be used. However, any terms found significant by the linear theory will be significant in the nonlinear case (Cornfield *et al.*, 1960). But the converse is not true, and it would be hazardous to call an additional exponential term insignificant. Hamilton (1964, 1965) has described a significance test, related to the F test, which appears well adapted to nonlinear models. The test may be used to determine if an additional exponential term significantly increases the fit of the model to the data. In terms of our problem, we calculate the statistic R, $R = (S_k/S_{k+1})^{1/2}$, where S_k is the S value obtained for a fit using k classes and S_{k+1} is that obtained using $k + 1$ classes. We may have $(1 - P) \times 100\%$ confidence that the fit with $k + 1$ classes is necessary if the calculated R value is greater than or equal to $R_{2,M-2k-2,P}$ (tables for R are given by Hamilton (1965)). Let us use this test to decide whether a third kinetic class is needed to fit the data from eq 2 with 2% as the standard deviation of the errors. The ratio S_k/S_{k+1} is found to be 6.58. We have, then, $R = 2.57$. Since $R_{2,34,0.05}$ is about 1.169, there is better than a 99% probability that a third class of hydrogen is present. Thus, we may readily reject the two class models, which might have been chosen by some of the previous methods. This test has been found to work well with a variety of simulated exchange curves at all reasonable error levels.

In view of the nonlinearity mentioned above, we cannot employ a significance test to determine the upper bound for n. We have found, however, the following procedure to succeed quite often. The data are fit using models with increasing values of n. Eventually, the point $n + 1$ is reached, and we find two classes with roughly the same rate constants. Generally, oscillation about this solution occurs, due to the poor conditioning of $\mathbf{B}'$. Convergence, if it occurs, may take more than the usual number of iterations (typically a maximum of 15). In addition, the standard errors of the estimates become very large. When $n + 2$ is reached, another "splitting" of one of the kinetic classes occurs. All of these phenomena make it easy to see when n is surpassed. Thus, with the data of eq 2, a splitting occurs at $n = 4$ and another at $n = 5$, and we may be assured that three classes are present within the accuracy of the experimental data.

It must be emphasized that the value of n obtained by these methods has meaning only if we have knowledge, on the basis of other information, that the molecule being studied has a small number of discrete kinetic classes. By this we mean a small number of classes whose members possess virtually identical rate constants. In the case of small peptides, for example, such information may come from nuclear magnetic resonance studies (Laiken *et al.*, 1969). When the exchangeable hydrogens of a molecule do fall in such groups, it is often assumed that the members share a particular physical or structural environment. There is reasonable justification for this assumption. If c hydrogens are really found in widely different surroundings, it is unlikely that they will possess *identical* rate constants. But it is hazardous to suppose, *a priori*, that the exchangeable hydrogens of a complex macromolecule, like a protein, are found in a small number of discrete kinetic classes. Indeed, Klotz and coworkers have indicated many factors which may influence the magnitudes of rate constants in proteins (Klotz and Frank, 1964; Klotz and Feidelseit, 1966; Leichtling and Klotz, 1966; Klotz and Mueller, 1969). The presence of regular structures such as the α helix or pleated sheet is only one of these. The protein cytochrome c, for example, possesses many slowly exchanging hydrogens and exhibits an exchange curve similar to those of other proteins (Ulmer and Kagi, 1968; Kagi and Ulmer, 1968). However, X-ray diffraction studies indicate that the molecule has little or no α helix (Dickerson *et al.*, 1967). Indeed, the rate constants of the peptide hydrogens within a given regular structure may well vary, depending on the location of the hydrogens within the macromolecule as a whole and their resultant accessibilities to the solvent. Thus, it is quite possible that the k_i found in any protein may vary continuously over the entire observable range of about 0.001 to 0.08 min^{-1}. Other situations in which a continuum of exponential decay processes occurs have been encountered (van Liew, 1967; Rossing, 1966). In these cases, attempts have been made to determine the nature of the distribution or "spectrum" of rate constants. The distribution may be regular, irregular, or "discontinuous," with the latter corresponding to the case of discrete kinetic classes. Knowledge of the distribution would yield the fraction of rate constants within any interval (k_a, k_b) which is equivalent to a kinetic class analysis of the system. In what follows, we will show the extreme difficulty involved in determining the nature of the distribution and the concomitant ambiguity introduced into the interpretation of hydrogen-exchange curves of complex systems.

Hydrogen-Exchange Curves of Complex Systems. Assume that we have a macromolecule whose exchange curve may be represented by eq 1. Let each kinetic class contain only one hydrogen and let

$$k_i = k + \epsilon_i$$

where ϵ_i is a negative or positive increment. Then

$$H(t) = \sum_{i=1}^{N} e^{-(k+\epsilon_i)t} = \sum_{i=1}^{N} e^{-kt}e^{-\epsilon_i t}$$

where N represents the total number of exchangeable hydrogens. Expanding the terms for each class and summing the results over all N classes gives

$$H(t) = e^{-kt}\left(N - t\Sigma\epsilon_i + \frac{t^2}{2!}\Sigma\epsilon_i^2 - \frac{t^3}{3!}\Sigma\epsilon_i^3 + \ldots\ldots\right)$$

Now set k equal to $\bar{k}$, the mean of all the k_i. The first-order term in the ϵ_i vanishes to yield

$$H(t) = e^{-\bar{k}t}\left(N + \frac{t^2}{2!}\Sigma\epsilon_i^2 - \frac{t^3}{3!}\Sigma\epsilon_i^3 + \ldots\right)$$

Since $(k_i - \bar{k}) = \epsilon_i$, it is easily seen that

$$\Sigma(k_i - \bar{k})^r/N = \Sigma\epsilon_i^r/N$$

Hence, the rth moment about the mean of the k_i is equal to the rth moment about the origin of the ϵ_i.

Setting

$$\mu_r = \Sigma\epsilon_i^r/N$$

we may write

$$H(t) = Ne^{-\bar{k}t}\left(1 + \sum_{r=2}^{\infty}(-1)^r\mu_r\frac{t^r}{r!}\right) \qquad (4)$$

and we have an expression for the exchange of hydrogens having *any* statistical distribution of rate constants, in terms of the moments of the distribution. It should be noted that the higher moments, μ_3, μ_4, μ_5, etc., will contribute to $H(t)$ only when the value of t is large, since the k_i generally are $\ll 1$. But $H(t)$ is small at large values of t, and the effects of these higher moments on the exchange curve will be very difficult to detect. The term μ_2 in eq 4 represents the variance of the distribution of the ϵ_i, or equivalently, of the k_i. Computations

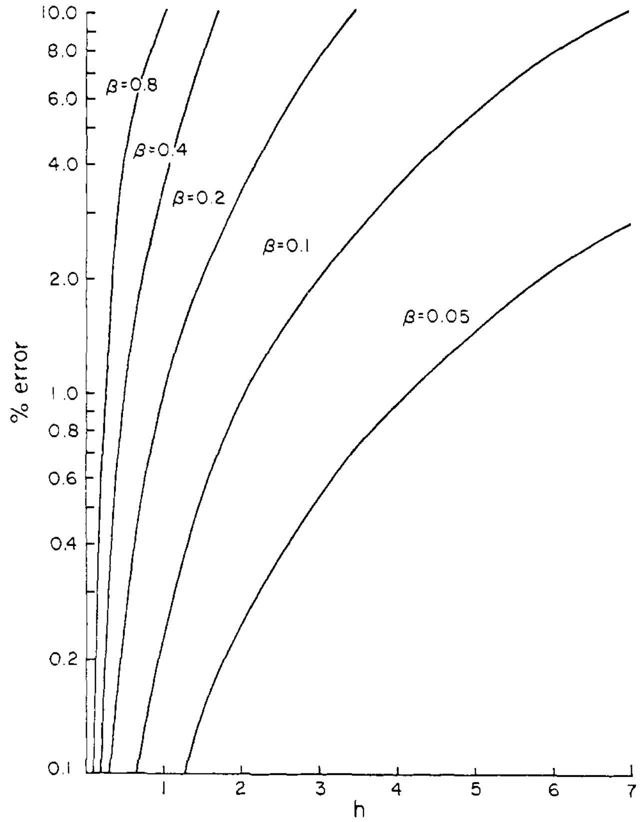

FIGURE 3: Errors due to single-class approximation. Hydrogen-exchange data resulting from a moderately symmetrical distribution of rate constants may be analyzed using a single class of rate constant, $\bar{k}$. The resulting percentage errors, in terms of the true number of hydrogens remaining, are shown. Abscissa values represent half-lives of the single class.

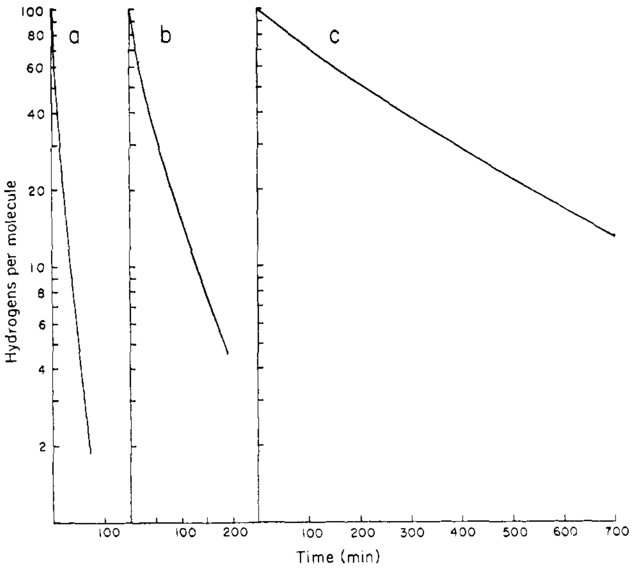

FIGURE 4: Hydrogen-exchange curves resulting from systems with uniform distributions of rate constants: In parts a, b, and c the k_i's are distributed on (0.001–0.006), (0.006–0.04), and (0.04–0.08) min^{-1}, respectively.

show that for symmetrical or moderately symmetrical distributions with $\sigma/\bar{k} < 1$, where σ is the standard deviation

$$H(t) \simeq Ne^{-\bar{k}t}\left(1 + \frac{\sigma^2 t^2}{2}\right) \qquad (5)$$

is quite accurate to at least $H(t)/N = 0.1$, because $\mu_3 = 0$ or is at least very small in these cases. One might conceive of many such distributions of rate constants possessing similar means and standard deviations. These might range from discontinuous (discrete class) distributions with different numbers of classes, to any one of a number of regular or irregular continuous distributions. However, all would yield virtually identical exchange curves by eq 5. Therefore, in the case of symmetrical or moderately symmetrical distributions of k_i, and in the presence of experimental error, not only is n indeterminate, but discrete classes cannot be distinguished from continuous distributions. If μ_3 does make significant contributions to eq 4, an analogous situation holds. There are many skewed distributions which share a common value of $\bar{k}$ and for which the first few moments are similar. van Liew (1967), in fact, has shown by the use of simulated data that all except highly skewed distributions of rate constants give very similar exponential decay curves for the same values of $\bar{k}$ and β, where β is the quantity $\sigma/\bar{k}$. This observation is verified readily by the use of eq 5. Setting $h = t/\bar{T}$, where $\bar{T}$ is the half-life of a class with rate constant $\bar{k}$, and noting that $\bar{k} = 0.693/\bar{T}$, we have $H(t) \simeq Ne^{-0.693h}(1 + 0.24h^2\beta^2)$, after substitution in eq 5. Thus, in the absence of information concerning the true nature of eq 4 for a given system, the only quantities which may be obtained unambiguously from a hydrogen-exchange experiment are $\bar{k}$ and the first few moments of the distribution of the k_i about $\bar{k}$. These may be found by fitting the curve with a reasonable number of discrete classes, as described by van Liew (1967). If the discrete class approximation fits the data well, the resulting $\bar{k}$ and moments must resemble those of the actual distribution, as shown by eq 4.

We will now illustrate the implications of these findings with some numerical examples. Equation 5 may be truncated to give

$$H(t) \simeq Ne^{-\bar{k}t} \qquad (6)$$

thereby approximating $H(t)$ with a single class of rate constant $\bar{k}$. Calculations of the error involved in using eq 6 rather than eq 5 are shown in Figure 3 for different values of β. It can be seen that the errors may not become noticeable for many values of β until $H(t)$ is quite small. This is not surprising in view of the data plotted in Figure 4. Uniform distributions of k_i on the intervals (0.001–0.006), (0.006–0.040), and (0.040–0.080) min^{-1} were assumed. These k_i cover the entire range commonly encountered in protein–hydrogen-exchange work. In a uniform distribution, the rate constants are spread evenly throughout the interval and there is no clustering. Therefore, these distributions may be taken to represent any number of hydrogens, each in a unique physical environment. Despite this, the data appear to be more or less linear over a substantial period of time. In the case of the interval (0.040–0.080), where β is 0.193, the errors resulting from the use of eq 6 to approximate the exchange curve do not become apparent

TABLE II: Representation of Uniform Distributions of Exchangeable Hydrogens by Discrete Classes.

Interval	Class	k	c
(0.001–0.006)	I	4.6×10^{-3}	58.2
	II	1.8×10^{-3}	41.8
(0.006–0.04)	I	2.7×10^{-2}	71.2
	II	9.5×10^{-3}	27.8
(0.04–0.08)	I	6.8×10^{-2}	62.5
	II	4.6×10^{-2}	37.5

(10%) until $H(t)/N$ is 0.1 or less. It should be recognized that $\bar{k}$ does not necessarily give the best single-class approximation, though. For example, the data of Figure 4c could be fit best with one class of $k = 0.003$, and the approximation is good to 500 min. Furthermore, two classes generally can be fit with great accuracy to curves resulting from distributions of k_i such as these. Table II shows two class fits to the data of Figure 4. These fits are accurate to about 0.1% up to at least $H(t)/N = 0.1$ in the case of the first two curves, and to over 1000 min in the case of the third. It may be noted that the values of β and $\bar{k}$ calculated on the basis of the two-class approximation are close to those of the underlying distribution.

In view of the above, *it appears quite possible that large numbers of hydrogens, each exchanging with a characteristic rate constant, may be cast erroneously into discrete classes with many members.* While it is impossible to state the number of classes needed to fit any exchange curve resulting from a distribution of k_i over the range 0.001–0.08 min^{-1}, our experience suggests that three classes generally will suffice. Evidence bearing on this point was obtained *via* simulation of hydrogen-exchange curves, followed by discrete class analysis, using the damped least-squares method. The simulations were based on various continuous statistical distributions of k_i or on large numbers of discrete kinetic classes. For example, random number methods were used to generate two uniform distributions of k_i on the intervals (0.001–0.020) and (0.020–0.200) min^{-1}. Each k_i then was assigned a random integral class of from one to ten members, giving a total of 489 hydrogens. The best three-class fit was $H(t) = 37e^{-0.0028t} + 235e^{-0.0081t} + 217e^{-0.1131t}$, which was virtually indistinguishable from the theoretical curve for almost 1000 min. Hvidt and Neilsen (1966) have stated, without proof, that any exchange curve can be approximated well with four or five kinetic classes with rate constants in the ratio $1:4:4^2:4^3:4^4\ldots$ While it appears that three classes are generally adequate, ratios of this type are often, but not always, encountered. Most protein hydrogen exchange curves have, in fact, been analyzed in terms of three classes. Sometimes the rate constants are spaced more widely than Hvidt and Neilsen's approximation would indicate. However, the significance of the analyses found in the literature are difficult to assess. Often, a very small number of points was used and inferior methods of treating the data were employed. Also, an "infinitely slow" class of hydrogens was subtracted. Such a class is very difficult to define, and an incorrect guess may totally change the nature of the exchange curve.

It might be argued that the constancy of kinetic class sizes, despite variation in pH or temperature, implies that such classes are real. In the absence of any conformational changes, such a variation in the environment of the protein should cause the rate of exchange of all the members of each class to change by a factor which we shall call q. It is readily shown (Segal and Harrington, 1967) that such a change merely shifts a given value of $H(t)$ to $1/q$th the time at which it occurred previously. The class sizes are not affected. A false class analysis which fits the data well will also be shifted in the same manner, without any effect on the kinetic class sizes. Thus, as long as a given model fits the data under one set of conditions, it will fit it well under another set of conditions, barring conformational changes, and this fact cannot be used as evidence of the model's physical validity.

In conclusion, when reasonably good hydrogen-exchange data are available, it is possible to do an accurate kinetic class analysis using the method of damped least squares. The analysis will have a clear interpretation only if there is reason to believe that the hydrogens of the molecule occur in a small number of groups possessing almost identical rate constants. Such groups might occur in small peptides, polyamino acids, and perhaps nucleic acids. However, in the case of complex macromolecules such as proteins, a discrete class assumption may be unwarranted and detailed quantitative interpretation of the class sizes and rate constants may lead to an erroneous view of the underlying physical situation.

References

Cornell, R. G. (1956), ORNL-2120, Oak Ridge, Tenn., Oak Ridge National Laboratory.

Cornell, R. G. (1962), *Biometrics 18*, 104.

Cornfield, J., Steinfeld, J., and Greenhouse, S. W. (1960), *Biometrics 16*, 212.

Deming, W. E. (1943), Statistical Adjustment of Data, New York, N. Y., Dover.

Dickerson, R. E., Kopka, M. L., Weinzierl, J., Varnum, J., Eisenberg, D., and Margoliash, E. (1967), *J. Biol. Chem. 242*, 3015.

Gardner, D. G., Gardner, J. C., Laush, G., and Meinke, W. W. (1959), *J. Chem. Phys. 31*, 978.

Hamilton, W. C. (1964), Statistics in Physical Science: Estimation, Hypothesis Testing, and Least Squares, New York, N. Y., Ronald.

Hamilton, W. C. (1965), *Acta Crystol. 18*, 502.

Hildebrand, F. B. (1956), Introduction to Numerical Analysis, New York, N. Y., McGraw-Hill.

Hvidt, A., and Neilsen, S. O. (1966), *Advan. Protein Chem. 21*, 287.

Kagi, J. H. R., and Ulmer, D. D. (1968), *Biochemistry 7*, 2718.

Keepin, G. R., Wimett, T. F., and Zeigler, R. G. (1957), *J. Nucl. Energy 6*, 1.

Klotz, I. M., and Feidelseit, P. L. (1966), *J. Am. Chem. Soc. 88*, 5103.

Klotz, I. M., and Frank, B. H. (1964), *J. Am. Chem. Soc. 86*, 3889.

Klotz, I. M., and Mueller, P. M. (1969), *Biochemistry 8*, 12.

Laiken, S. L., Printz, M. P., and Craig, L. C. (1969), *Biochemistry 8*, 519.

Leichtling, B. H., and Klotz, I. M. (1966), *Biochemistry 5*, 4026.

Levenberg, K. (1944), *Quart Appl. Math. 2*, 164.

Marquardt, D. W. (1963), *J. Soc. Ind. Appl. Math. 11*, 431.

Meiron, J. (1965), *J. Opt. Soc. Am. 55*, 1105.

Papouseck, D., and Pliva, J. (1966), *Collection Czech. Chem. Commun. 30*, 3007.

Pitha, J., and Jones, R. C. (1966), *Can. J. Chem. 6*, 3031.

Riggs, D. S. (1963), The Mathematical Approach to Physiological Problems, Baltimore, Md., Williams & Wilkins.

Rossing, R. C. (1966), *J. Appl. Physiol. 21*, 1907.

Segal, D. M., and Harrington, W. F. (1967), *Biochemistry 6*, 768.

Ulmer, D. D., and Kagi, J. H. R. (1968), *Biochemistry 7*, 2710.

van Liew, H. O. (1967), *J. Theoret. Biol. 16*, 43.

Worsley, B. H., and Lax, L. C. (1962), *Biochim. Biophys. Acta 59*, 1.

Kinetic Study of the Diffusion of Ribonuclease across a Liquid–Liquid Interface*

Judith England† and Girair M. Nazarian‡

ABSTRACT: The purpose of this work was to explore the possibility of observing the rate of diffusion of a protein through the undisturbed interface separating two immiscible liquid phases. Since ribonuclease is known to have appreciable solubility in both of the immiscible phases obtainable from a mixture of water, 2-ethoxyethanol (Cellosolve), and ammonium sulfate, it was anticipated that this combination of protein and solvent system would be suitable for the desired study. By choosing a composition close to the critical (plait) point, a liquid–liquid interface was obtained which was sufficiently permeable to diffusing ribonuclease molecules, without the use of shaking, to permit a kinetic investigation. The diffusion was observed with ultraviolet absorption optics in the analytical ultracentrifuge operating at minimal speed after the less dense phase of the two-phase system had been layered over the denser phase containing ribonuclease. The kinetic study was based on the densitometer traces of the photographs which indicate protein concentration *vs.* position throughout the system at successive times during the approach to partition equilibrium. Mathematical analyses of the traces for the early (free diffusion) and late (restricted diffusion) stages of the experiment indicate that the free energies of activation for penetration of the interface by ribonuclease in the two directions were at most of the order of 12 kcal/mole.

Liquid–liquid interfaces which are permeable to proteins are not common. The requirement that two liquids be immiscible, in order that an interface be formed, is almost incompatible with the requirement that protein have appreciable solubility in each. Nevertheless, such immiscible solvents have been devised and form the basis for the extraction techniques of countercurrent distribution (Craig and Craig, 1956; von Tavel and Signer, 1956) and partition chromatography (Martin and Porter, 1951) used for the fractionation of protein mixtures. The developers of these techniques have determined protein distribution coefficients between phases at equilibrium but a kinetic study of the rate of diffusion of a protein through an undisturbed liquid–liquid interface has not been previously reported. We have initiated such studies using the analytical ultracentrifuge as a diffusion apparatus to observe RNase[1] in the water–Cellosolve–ammonium sulfate system of Martin and Porter (1951).

Experimental Procedures

Water and Cellosolve are miscible in all proportions at room temperature but two liquid phases separate when sufficient ammonium sulfate is added. We chose a weight composition (59.5% water, 18.2% Cellosolve, and 22.3% ammonium sulfate) close to the plait point on the phase diagram of the system (Martin and Porter, 1951) in order that the two phases obtained would be nearly identical thereby minimizing the interfacial resistance to protein transfer. The method of preparation was to dissolve the salt in the water, add the Cellosolve, shake the system, and allow the phases to separate. The bottom phase B (density = 1.151 g/cm^3) contains more salt than Cellosolve while the opposite is true of the top phase T (density = 1.024 g/cm^3). The pH of both phases was 5.2. We used chromatoquality Cellosolve (Matheson Coleman & Bell) which we triple distilled at atmospheric pressure to remove impurities absorbing at 265 mμ, the

---

* From the Department of Chemistry, San Fernando Valley State College, Northridge, California 91324. *Received August 5, 1969.* Supported by U. S. Public Health Service Research Grant GM 10832 from the National Institute of General Medical Science and by a Special Semester Leave awarded by the California State Colleges to G. M. N. This report was taken in part from a thesis submitted to the Graduate School of San Fernando Valley State College by J. E. in partial fulfillment of requirements for the Master of Science degree.

† Present address: Division of Natural Sciences-I, University of California, Santa Cruz, Calif. 95060.

‡ To whom requests for reprints should be addressed.

[1] Abbreviation used is: RNase, bovine pancreatic ribonuclease.