MODELS FOR HYDROGEN EXCHANGE FROM FOLDED PROTEINS. II.

LYNDA M. ELLIS, *Division of Health Computer Sciences, Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, Minnesota 55455 U.S.A.*

ABSTRACT The kinetics of hydrogen exchange from folded proteins can be molded as a function of two continuous distributions of rate constants, k_{cx} and k_n, representing exchange from the folded and unfolded conformations, respectively. This model can account for the temperature dependence of soybean trypsin inhibitor at pH 3 and pH 6.5. The physical significance of this model, especially the shape and breadth of the k_n distribution, are discussed.

INTRODUCTION

A protein is composed of one or more linear polymers of α-amino acids, and in the native state is folded upon itself in a compact three-dimensional structure. In the unfolded (or denatured) state, this structure is disrupted and the protein exists as an extended chain or chains. Many workers have studied the structure and mobility of folded proteins by measuring the kinetics of hydrogen exchange of peptide amide protons. For example, we have determined the extent to which the internal regions of a protein are accessible to solvent using this technique (Ellis et al., 1975).

A conceptual model is required before the results of hydrogen exchange experiments can be interpreted. These data can often be approximated as the sum of two to four exponential terms. Thus one model frequently assumed, implicitly or explicitly, is that peptide amide protons can be divided into two to four kinetic classes, each with a single rate constant for exchange. Biophysical significance is accorded to these classes (Englander, 1975; Englander et al., 1972), though the error of this assumption is well known (Laiken and Printz, 1970).

This paper presents an alternative model for hydrogen exchange based on more valid assumptions. This model can simulate the observed hydrogen exchange kinetics of soybean trypsin inhibitor (Kunitz) (STI) as a function of temperature and pH.

BACKGROUND AND THEORY

The rate of hydrogen exchange in an unfolded protein is a function of the chemical environment of the exchanging protons. That is, it depends completely on the chemical nature of the amino acid side-chains of the residues joined by the peptide bond (Molday et al., 1972). If we assume a protein with only a single exchangeable amide proton, its rate constant for chemical exchange, k_{cx}, can be determined from model compounds.

The relative amounts of folded and unfolded protein present at any given pH and temperature depend on k_f and k_b, the rate constants for thermal refolding (renaturation) and unfolding (denaturation), respectively. The physical and chemical factors influencing exchange from the folded conformation can be lumped into another rate constant k_n, which measures the rate of exchange from the native (folded) protein. This symbol k_n has been used previously (Ellis, 1977); it corresponds to k_l of Ellis et al. (1975).

If the native (folded) protein is N, its denatured (unfolded) conformer is D and the proton to be exchanged is *; this can be diagrammed as

$$\begin{array}{ccc} \mathrm{N}^* & \xrightarrow{k_n} & \mathrm{N} + * \\ k_b \uparrow & \downarrow k_f & \\ \mathrm{D}^* & \xrightarrow{k_{cx}} & \mathrm{D} + * \end{array} \quad . \tag{1}$$

The rate constant for exchange from the folded conformation, k_n, is influenced by the proton's chemical environment, but folding imposes more important constraints, including limited solvent access.

Hydrogen exchange experiments measure the number of protons remaining unexchanged ($H_{\mathrm{rem}} = N^* + D^*$) at a given time. From diagram 1, the time-course of these measurements ($\mathrm{d}H_{\mathrm{rem}}/\mathrm{d}t$) is a function of k_n and k_{cx}: $\mathrm{d}H_{\mathrm{rem}}/\mathrm{d}t = -(k_n N^* + k_{cx} D^*)$.

Under the steady-state assumption that the rate of change of D^* equals zero,[1] $D^* = k_f N^*/(k_b + k_{cx})$. Thus $\mathrm{d}H_{\mathrm{rem}}/\mathrm{d}t = -N^* k_{\mathrm{app}}$, and $H_{\mathrm{rem}} = \exp(-k_{\mathrm{app}} t)$, where $k_{\mathrm{app}} = k_n + (k_{cx} \cdot k_f)/(k_b + k_{cx})$.

The above discussion assumes only one exchangeable proton in the protein. However, the usual hydrogen exchange experiment rarely, if ever, measures only one proton. Instead it simultaneously measures exchange of many (or all) of the amide protons. Thus k_{cx} and k_n are distributions of rate constants, one for each of n exchangeable protons, and

$$H_{\mathrm{rem}} = \sum_{i=1}^{n} \exp(-k_{\mathrm{app},i} t)$$

where the apparent rate constant of the ith proton,[2]

$$k_{\mathrm{app},i} = k_{n,i} + (k_{cx,i} \cdot k_f)/(k_b + k_{cx,i}). \tag{2}$$

---

[1]Though analytical solution of this set of differential equations is possible without this assumption, little significance can be attached to the solution because of the relative values of the rate constants. Numeric approximation methods, though lengthy, do converge, and have been used on several representative protons. Calculations based on the steady-state assumption are within 5% of those based on numeric approximation.

[2]The rate constant $k_{\mathrm{app},i}$ could alternatively be defined as $K_{eq,i} \cdot k_{dx,i}$, where $K_{eq,i}$ is an equilibrium constant between the native protein and a locally denatured intermediate which exposes the ith proton to solvent. This model would then be equivalent to a local denaturation model first proposed by Linderstrom-Lang and co-workers, with a K_{eq} distribution (see Englander et al., 1972, for a review).

Values for $k_{cx,i}$ are determined by the method of Molday et al. (1972). As mentioned earlier, they have reported that $k_{cx,i}$ is uniquely determined by the two amino acid residues joined by the peptide bond. Using their empirical rules, we can thus calculate k_{cx} for all protein amide sites, given the amino acid sequence of the protein.

The rate constants for folding and unfolding, k_f and k_b, can be obtained from the literature (e.g., Kunitz, 1948, and Pohl, 1969). Alternatively, under appropriate conditions (see Discussion), k_f can be set to zero.

Finally, if one determines $k_{n,i}$, Eq. 2 can be used to calculate $k_{app,i}$ and thus the mole (atom) fraction of hydrogen remaining at the ith site. The methods used to assign k_n values to individual amide protons involve two steps. First, the shape and breadth (maximum and minimum values) of the distribution are chosen. Second, rate constants from that distribution are assigned to individual amide protons.

Taking the second step first, each proton is assigned a classification number between 0 and 1, which is inversely proportional to the k_n value of that proton. That is, a proton whose classification number is 0 will be assigned the maximum k_n value, one whose number is 1 will be assigned the minimum value, etc. There need not be a linear relation between the number and the k_n value; that will depend on the shape of the distribution of k_n values.

The classification number may be assigned arbitrarily. For example, one may order the n amide protons by sequence and give the one closest to the amino terminus a value of 0, the next a value of $1/(n-1)$, etc. The last amide proton would then have a number of 1. This method has been used previously (Ellis, 1977; Ellis et al., 1975), but it has limited physical significance.

A more realistic method is used in this paper. As mentioned earlier, $k_{n,i}$ depends on the three-dimensional microenvironment of the ith proton. The most important attribute of this environment as it affects hydrogen exchange is the accessibility of the site to solvent. If one assumes that residues on the surface of a protein are more accessible to solvent than those buried in its center, one can calculate the probability of any amino acid being buried, based on examination of several protein X-ray crystal structures. In this manner Gates and Fisher (1971) have tabulated the probability of an amino acid being inaccessible to solvent because it is located in the interior of a protein. This probability, or accessibility coefficient, is based on side-chain, as opposed to peptide amide, accessibility. As a first approximation, we assume that the accessibility of an amide proton will be the average of that of its two adjacent side-chains. Such a classification number can be calculated as

$$\{[(B_i + B_{i-1})/2] - B_{\min}\}/(B_{\max} - B_{\min}),$$

where B_i is the accessibility coefficient of the ith residue, and $B_{\min}$ and $B_{\max}$ are the minimum and maximum B values, respectively.

The breadth of the k_n distribution is bounded at its fastest end by the largest values of k_{cx}, the rate constant for exchange from freely exposed (unfolded) amide protons. At the other end, its slowest rates approach zero or are significantly less than (k_{cx} ·

$k_f)/(k_b + k_{cx})$ and thus can be ignored in Eq. 2. The limits of this distribution were varied to best approach the observed data.

Finally, one must determine the shape of the k_n distribution. Because of the large range of values (over seven orders of magnitude), the shapes explored were limited to logarithmic distributions. A log uniform shape is one of the simplest of these, and the log normal also frequently occurs in biological distributions.

Though a k_{app} value can now be calculated for each proton in the protein, all of them will not affect the exchange observed at any given pH or temperature. This is because of the limits of the experimental method. Protons with half-times less than 5 min are completely exchanged by the time of the first H_{rem} measurement. Similarly, those with half-times over 50 h are totally unexchanged even at the last measurement. Thus, the observed k_{app} distribution varies from 0.14 min^{-1} to 2.3 $\times 10^{-4}$ min^{-1}.

Histograms of the continuous distribution of the rate constant k_{app} for soybean trypsin inhibitor at pH 3 and 20°C, using both the log uniform and the log normal distributions for k_n, are shown in Fig. 1. The shaded areas are the observable k_{app} distributions. Since $(k_{cx,i} \cdot k_f)/(k_b + k_{cx,i})$ is approximately 1.7 $\times$ 10^{-6} min^{-1} under

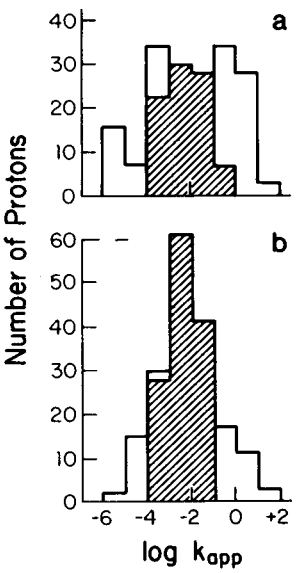

FIGURE 1

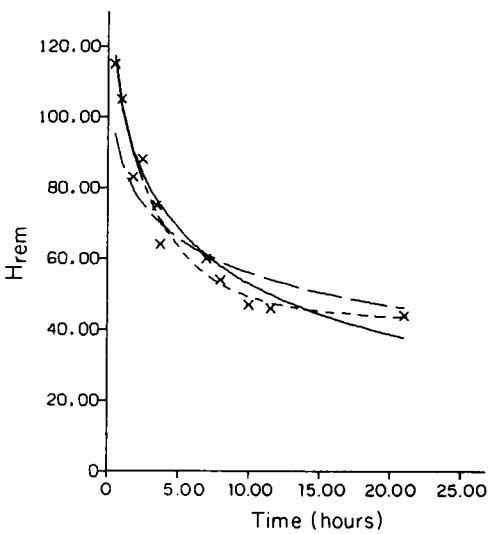

FIGURE 2

FIGURE 1 Bar graph of continuous distributions of rate constants assigned to the amide protons of soybean trypsin inhibitor (Kunitz) at pH 3 and 20°C. *a*. k_{app} values assuming a log uniform k_n distribution spanning 7.8 orders of magnitude. *b*. k_{app} values assuming a log normal k_n distribution spanning 7.7 orders of magnitude. Shaded areas indicate observable rate constants (see text). Values for k_f and k_b from Kunitz (1948).

FIGURE 2 STI hydrogen exchange kinetics at pH 3, 20°C. X, observed data (from Ellis et al., 1975); ——, line predicted by the k_{app} distribution shown in Fig. 1*a*; - - - - -, line predicted by the k_{app} distribution shown in Fig. 1*b*; . . . . . , line predicted based on the equation $H_{rem} = 51 \exp(-7.33 \times 10^{-2} t) + 72.5 \exp(-4.88 \times 10^{-3} t) + 47.5 \exp(-7.31 \times 10^{-5} t)$.

these conditions, a k_{app} distribution is identical to its parent k_n distribution down to log k values of -5 or less.

RESULTS

Models for hydrogen exchange in proteins should account for as many properties of that exchange as possible. First, they should account for exchange observed under a single pH and temperature. Both the model proposed in this paper and a model placing amide protons into three kinetic classes can do this for STI at pH 3 and 20°C, as shown in Fig. 2 and Table I. The model using a log-normal distribution for k_n (solid line) has approximately as good a fit to the data as the three exponentials model (dotted line), and both of these are significantly better than the model using a log-uniform distribution for k_n (dashed line).

To be truly useful, hydrogen exchange models should also account for the temperature and pH dependencies of this exchange. The models based on two to four kinetic classes cannot do this.

The model discussed in this paper, when using a log normal distribution for k_n, can account for the temperature dependence of STI at pH 3. To do this the maximum value for k_n was set equal to the rate constant for exchange of the side chain amide

TABLE I

MODEL PARAMETERS AND GOODNESS-OF-FIT CRITERIA FOR SOYBEAN TRYPSIN INHIBITOR

pH	Temperature	k_n breadth		n	Σ	χ^2
		Min	Max			
	°C	10^{-6} min^{-1}	10^{2} min^{-1}			
3	10	0.064	0.032	5	134	1.44
3	20	0.20	0.102	11	313	4.87
3	30	0.64	0.322	13	548	16.27
3	40	2.0	1.02	27	1,040	85.18‖
3	45	3.6	1.81	19	416	115.61‖
3	50	6.4	3.22	4	96	5.07
3	20*	0.20	0.102	11	2,490	20.67§
3	20‡			11	173	2.34
3	40	1.0	1.02	27	692	36.69
3	40	0.51	1.02	27	1,260	25.05
6.5	20	0.57	5.7	11	68	2.52
6.5	30	1.8	18.	17	104	4.07
6.5	40	5.7	57.	24	154	13.92
6.5	50	18.	180.	9	163	13.16

n is the number of values observed at the given pH and temperature; Σ is the sum of the squared residuals between the observed and calculated values; a log normal distribution is assumed for k_n except as indicated.
*Log uniform distribution assumed for k_n.
‡Calculated values obtained by the best fit to three exponential terms (see legend to Fig. 2).
§ $P < 0.05$.
‖ $P < 0.01$.

group in asparagine, since this rate, at the pH's studied, is faster than all but a very few k_{cx} values. For use in this model, the rate was calculated at 0°C (Molday et al., 1972) and then assumed to have the same temperature dependence as poly-D,L-alanine (Englander and Poulson, 1969), with an activation energy of approximately 17 kcal/mol. The breadth of the distribution, and thus its slowest rates, was varied to approximate the experimental data.

The model fit very well at 10 and 20°C and fairly well at 30 and 50°C, as shown in Fig. 3. Its agreement at 40 and 45°C was less good, but even here the correspondence shown is remarkable, since the model assumes *all* amide sites have one activation energy. Site dependent energies of activation would certainly improve the fit at varying temperatures.

The model discussed in this paper can also account for the pH dependence of hydrogen exchange. With the assumption that k_f is zero (see Discussion), this model can fit STI hydrogen exchange at pH 6.5 (Fig. 4) as well or better than at pH 3. The breadth of the k_n distribution must be increased, however, from 7.7 orders of magnitude at pH 3 to 9.0 at pH 6.5 (Table I). This is reasonable since the breadth of the k_{cx} distribution, a factor in k_n, increases away from the pH of minimum exchange (around pH 3).

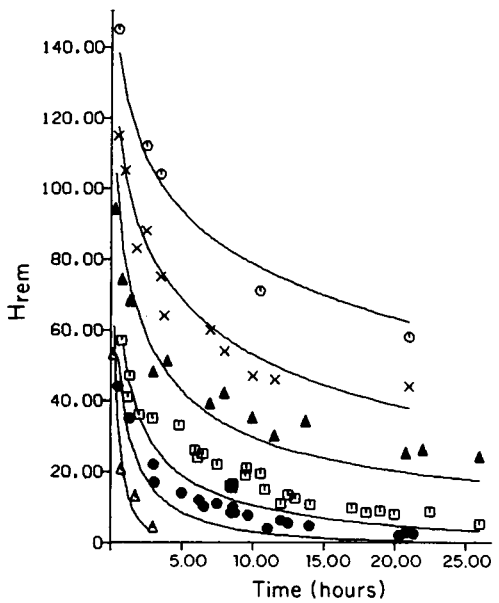

FIGURE 3

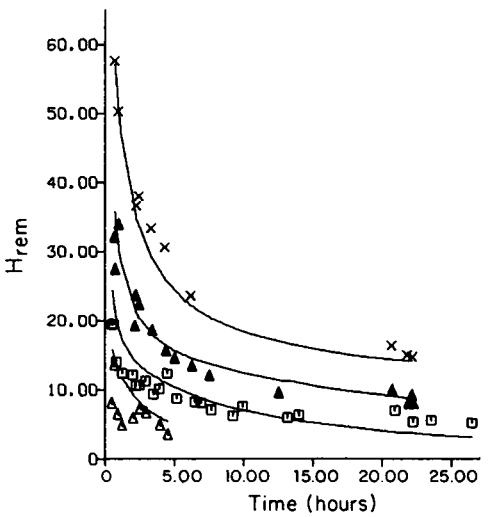

FIGURE 4

FIGURE 3 STI hydrogen exchange kinetics at pH 3. ○, 10°C; X, 20°C; ▲, 30°C; □, 40°C; ●, 45°C; △, 50°C (observed data from Ellis et al., 1975). Lines were predicted by a log-normal k_n distribution spanning 7.7 orders of magnitude. Values for k_f and k_b from Kunitz (1948).

FIGURE 4 STI hydrogen exchange kinetics at pH 6.5 X, 20°C; ▲, 30°C; □, 40°C; △, 50°C (observed data from Ellis et al., 1975). Lines predicted using a log-normal k_n distribution spanning 9 orders of magnitude. The value of k_f was assumed to be zero (see Discussion).

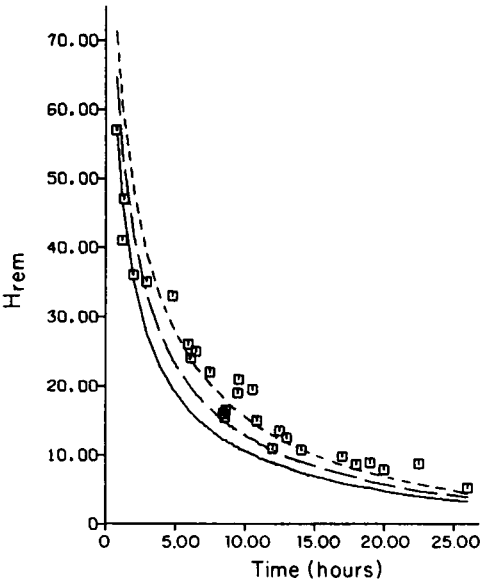

FIGURE 5 Effect of varying the breadth of the k_n distribution on hydrogen exchange kinetics. □, exchange observed for STI, pH 3, 40°C (from Ellis et al., 1975). ——, breadth 7.7 orders of magnitude; - - - - -, breadth 8.0 orders of magnitude; . . . . . , breadth 8.3 orders of magnitude. The maximum k_n value for all calculated curves was 102 min^{-1}. Values for k_f and k_b were from Kunitz (1948).

Finally, since one important objection to the sum of exponentials model is its non-uniqueness, it is important to determine this model's sensitivity to its parameters, including the shape and extent of the k_n distribution. As was seen in Fig. 2, the shape of the k_n distribution has a great effect on the kinetics calculated for the model. Fig. 5 demonstrates the effect of varying the breadth of k_n from 7.7 to 8.0 to 8.3 orders of magnitude for STI at pH 3 and 40°C. Since for convenience we have fixed the maximum rate constants, changing the extent will primarily affect the distribution of the slower rates. Thus at earlier times the curves tend to approach one another. What may be unexpected is that they also tend to come together at the longest times. This is due to the limits of detection for hydrogen exchange kinetics mentioned earlier. That is, all sites with half-times greater than approximately 50 h are not detectable and their associated rate constants are effectively zero.

DISCUSSION

Although a three-exponential model can be fitted to closely approximate the data at a single pH and temperature (Fig. 2), the model requires estimation of 6 parameters to do so. Even better agreement is possible with sums of 4 or 5 exponential terms (8 or 10 parameters).

In contrast, the model described here, based on a continuous distribution of k_n values, requires estimation of only one parameter, the breadth of this distribution.

Once the shape of the distribution is fixed, its breadth is varied (see Fig. 5) to minimize deviation from the data points as described below. Although the log-normal distribution for k_n is adequate, a distribution intermediate between log-normal and log-uniform and/or a skewed distribution might approach the data even more closely.

Fig. 5 also demonstrates the effect of different optimization strategies on the goodness-of-fit criteria. At pH 6.5 the sum of squared residuals and the χ^2 statistic could be minimized with the same model parameters. However, at pH 3, the sum of squared residuals was minimized at significantly different values of k_n distribution breadth than the χ^2 statistic. The solid line in Fig. 5 minimizes the sum of squared residuals when included with all temperatures, as in Fig. 2. The dashed line minimizes the sum of squared residuals and the dotted line minimizes the χ^2 statistic for 40°C, pH 3 data alone. Because of the experimental error inherent in hydrogen exchange measurements at low H_{rem}, it was decided to minimize the sum of squared residuals rather than the χ^2 statistic; the data shown in Fig. 2 result from this minimization.

All data discussed previously were observed or calculated at pH 3. At more neutral pH's, the model is hampered by the lack of values for k_f. However, it is reported to decrease by 2–3 orders of magnitude with every increase of one pH unit (Pohl, 1969). Since $(k_{cx} \cdot k_f)/(k_b + k_{cx}) \leq k_f$, this term can be ignored in Eq. 2 if $k_f \ll k_n$. To take the worst case, at 50°C and pH 3, STI has a k_f value of 1×10^{-2} min^{-1} (Kunitz, 1948). Thus its value should be no greater than 1×10^{-8} min^{-1} at pH 6.5. Since the k_n distribution will extend no lower than 10^{-5} min^{-1} (Table I), k_f may be assumed to be zero at pH 6.5 with no loss of accuracy in the model.

The model presented in this paper can account for hydrogen exchange in folded proteins at various temperatures and pH. It makes several assumptions in doing so, in particular assigning k_n values from a distribution on the basis of averaged accessibility coefficients. In a number of cases these assignments may result in surface amide protons being classed as inaccessible and/or interior protons classed as accessible. However, the good fit observed under most conditions demonstrates that these and similar errors will tend to cancel out when summed over the hundreds of amide protons in a protein.

The model can be improved by adding distributions other than log normal and log uniform and by allowing site-dependent energies of activation. This will increase its usefulness in predicting and/or explaining hydrogen exchange kinetics in folded proteins.

I would like to thank Dr. Clare Woodward for many helpful discussions, Dr. Myra Chern for assistance with the statistical analysis, and Dr. John Kemper for comments on the differential equations.

Received for publication 9 August 1977.

REFERENCES

ELLIS, L. M. 1977. Models for hydrogen exchange from folded proteins. Proceedings of the First International Conference on Mathematical Modeling. August 29 to September 1, 1977, St. Louis, Missouri, University of Missouri, Rolla, Mo. X. J. R. Avula, editor. Vol. 2, 1017–1026.

ELLIS, L. M., V. A. BLOOMFIELD, and C. K. WOODWARD. 1975. Hydrogen-tritium exchange kinetics of soybean trypsin inhibitor (Kunitz). *Biochemistry.* **14**:3413–3419.

ENGLANDER, S. W. 1975. Measurement of structural and free energy changes in hemoglobin by hydrogen exchange models. *Ann. N. Y. Acad. Sci.* **244**:10–27.

ENGLANDER, S. W., and A. POULSEN. 1969. Hydrogen-tritium exchange of the random chain polypeptide. *Biopolymers.* **7**:379.

ENGLANDER, S. W., N. W. DOWNER, and H. TEITELBAUM. 1972. Hydrogen exchange. *Annu. Rev. Biochem.* **41**:903–924.

GATES, R. E., and H. F. FISHER. 1971. Restrictions of sequence on the thickness of globular protein molecules. *Proc. Natl. Acad. Sci. U.S.A.* **68**:2928–2931.

KUNITZ, M. 1948. The kinetics and thermodynamics of reversible denaturation of crystalline soybean trypsin inhibitor. *J. Gen. Physiol.* **32**:241–263.

LAIKEN, S. L., and M. P. PRINTZ. 1970. Kinetic class analysis of hydrogen-exchange data. *Biochemistry.* **9**:1547–1553.

MOLDAY, R. S., S. W. ENGLANDER, and R. G. KALLEN. 1972. Primary structure effects on peptide group hydrogen exchange. *Biochemistry.* **11**:150–158.

POHL, F. M. 1969. On the kinetics of structural transition I of some pancreatic proteins. *Fed. Eur. Biochem. Soc. (FEBS) Lett.* **3**:60.