Hydrogen Deuterium Exchange as a Stochastic Process

Michele Stofella

December 2019

Principles of hydrogen deuterium exchange When a protein is put in solvent containing D_2O, the amide hydrogens of the protein spontaneosly exchange with deuterium contained in solution (5). Since deuterium is heavier than hydrogen, the overall mass of the protein increases and this change in mass can be monitored by mass spectrometry (9).
Hydrogen deuterium exchange has been modeled (5; 6) as a two steps process

$$A \underset{k_c}{\overset{k_o}{\rightleftharpoons}} B \xrightarrow{k_i} C \tag{1}$$

where states A, B and C correspond respectively to *closed*, *opened* and *deuterated* states of the residue. The transition between these states occurs with opening rate k_o, closing rate k_c and intrinsic exchange rate k_i. The intrinsic rates depend on the sequence of the protein analysed, on the pH and temperature of the solution and are assumed to be known (2; 3).
The model in equation 1 is nothing but a Michaelis Menten model and can be solved analitically. In fact, it corresponds to the system of equations

$$\begin{aligned} \dot{A} &= -k_o A + k_c B \\ \dot{B} &= k_o A - (k_c + k_i) B \\ \dot{C} &= k_i B \end{aligned} \tag{2}$$

that can be written in matrix form as

$$\dot{\mathbf{X}} = K\mathbf{X} \tag{3}$$

where $\mathbf{X} = (A, B, C)$ and

$$K = \begin{pmatrix} -k_o & k_c & 0 \\ k_o & -(k_c + k_i) & 0 \\ 0 & k_i & 0 \end{pmatrix} \tag{4}$$

The solution of the equation is

$$\mathbf{X} = \sum_{i=1}^{3} \alpha_i e^{\lambda_i t} \mathbf{u_i} \tag{5}$$

where λ_i are the eigenvalues

$$\begin{aligned} \lambda_1 &= 0 \\ \lambda_2 &= -\tfrac{1}{2}\left((k_c + k_i + k_o) + \sqrt{(k_c + k_i + k_o)^2 - 4k_i k_o}\right) \\ \lambda_3 &= -\tfrac{1}{2}\left((k_c + k_i + k_o) - \sqrt{(k_c + k_i + k_o)^2 - 4k_i k_o}\right) \end{aligned} \tag{6}$$

and eigenvectors $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$ are

$$\begin{aligned}
\mathbf{u_1} &= (0, 0, 1) \\
\mathbf{u_2} &= (-\frac{-k_c+k_i-k_o-\sqrt{(k_c+k_i+k_o)^2-4k_ik_o}}{2k_i}, -\frac{k_c+k_i+k_o+\sqrt{(k_c+k_i+k_o)^2-4k_ik_o}}{2k_i}, 1) \\
\mathbf{u_3} &= (-\frac{-k_c+k_i-k_o+\sqrt{(k_c+k_i+k_o)^2-4k_ik_o}}{2k_i}, -\frac{k_c+k_i+k_o-\sqrt{(k_c+k_i+k_o)^2-4k_ik_o}}{2k_i}, 1)
\end{aligned} \tag{7}$$

and the constants α_i are set by the conditions $A + B + C = 1$, $C(0) = 0$ and $C(\infty) = 1$. The exact solution for $C(t)$ is

$$\begin{aligned}
C(t) = 1 \quad &+ \quad \frac{k_o + k_i + k_c - \sqrt{(k_o + k_i + k_c)^2 - 4k_ik_o} - 2k_ik_o/(k_i + k_o)}{2\sqrt{(k_o + k_i + k_c)^2 - 4k_ik_o}} e^{\lambda_2 t} \\
&- \quad \frac{k_o + k_i + k_c + \sqrt{(k_o + k_i + k_c)^2 - 4k_ik_o} - 2k_ik_o/(k_i + k_o)}{2\sqrt{(k_o + k_i + k_c)^2 - 4k_ik_o}} e^{\lambda_3 t}
\end{aligned} \tag{8}$$

Wagner (12) states that the solution is a single exponential $C(t) = 1 - \exp(-k_x t)$ with $k_x = -\lambda_3$ and that the relation usually used for the rate, namely

$$k_x = \frac{k_i k_o}{k_c + k_i + k_o} \tag{9}$$

derives from the Taylor expansion for $k_c \gg k_o$ (native conditions). This last step of the derivation is here omitted[1]. In the case $k_c \gg k_i$, known as EX2 limit, even considering the double exponential solution, one obtains a single exponential with

$$k_x = \frac{k_i k_o}{k_c} = \frac{k_i}{P} \tag{10}$$

where $P = k_c/k_o$ is called protection factor. In the case $k_i \gg k_c$, known as EX1 limit, one can derive that

$$k_x = k_o \tag{11}$$

The single exponential approximation is true only under the assumption of native conditions and EX2 limit, otherwise the coefficient of the exponential associated to λ_2 would not go to 0 and at the same time the coefficient associated to λ_3 would not go to 1.
The residues in the state C correspond to the residues that have exchanged and the evolution in time of is called deuterium uptake and is assumed to have the shape of a single exponential decay as previously derived. For a peptide j formed by N residues, the deuterium uptake at time t, given rates $k_{x,i}$ for each residue, is written as

$$D_j(t) = \frac{1}{N} \sum_{i=2}^{N} 1 - e^{-k_{x,i} t} \tag{12}$$

where the sum starts from the second residue since the first one exchange so fast that cannot be detected (5). Exchange rates are normally written in the form given by equation 10 or 11. Fitting experimental $D_j(t)$ to find exchange rates is not trivial since the problem is underdetermined (11).

[1] Hint for derivation: use the first order Taylor series $\sqrt{1-x} \approx 1 - x/2$ which holds for $x \to 0$.

Experimental problem Raw data are mass spectra of peptides (formed by few residues) initially centred at the mass of the analysed peptide and moving in time towards a maximum value given by the initial value plus the number of residues involved in hydrogen deuterium exchange (since the weight of deuterium is twice the weight of hydrogen and each residue has only one amide hydrogen). An example of raw data is shown in figure 1.

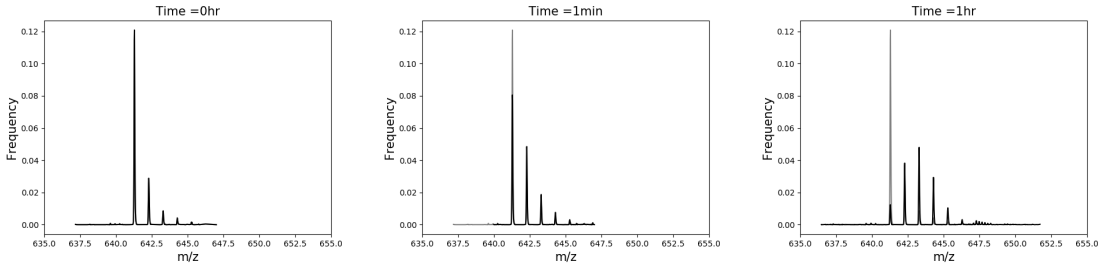

Figure 1: Example of mass spectra measured by mass spectrometry while hydrogen deuterium exchange occurs. The mass spectrum (black spectrum) moves in time towards higher mass value with respect to the initial spectrum (gray line). Spectra for mouse prion protein are shown (10).

The data in figure 1 can be clearly explained. In fact, the mass spectrum can be thought as a probability distribution (11) corresponding to the probability that k residues out of N have exchanged at time t. Such a probability distribution is a binomial multivariate distribution:

$$\Pi(k,t) = \sum_{A\in\{1,\dots n\}}^{|A|=k} \prod_{i\in A} D_i(t) \prod_{j\notin A} 1 - D_j(t) \tag{13}$$

The problem arises when the analysed peptide is under EX1 conditions, which can be experimentally reached via increasing the temperature of the solution, adding *subdenaturant* concentrations of denaturant (e.g.: urea) in solution or increasing the pH towards alkaline values (7; 8). Under such conditions, experimental data show bimodality in mass spectra than cannot be predicted by equation 13. Examples of such data can be found in (1; 4; 13; 14).

Hydrogen deuterium exchange as a stochastic process Modeling hydrogen deuterium exchange as a stochastic process helps understanding qualitatively why bimodality occurs in mass spectra and gives the key to find where it hides in the model previously introduced (equation 1).
The stochastic differential equation associated to a general stochastic process is given by

$$dX_t = \mu(X_t,t)dt + \sigma(X_t,t)dw_t \tag{14}$$

where X_t is a random variable, the first term represents the drift term and is associated to the deterministic part of the system while the second term is a stochastic term, dw_t being a Wiener process. For sake of simplicity, suppose $\mu(X_t)$ and $\sigma(X_t)$ do not depend on time.
In the case of hydrogen deuterium exchange, the random variable is a binomial variable that is 0 if the residue has not exchanged and 1 if exchange occured; $\mu(X_t)$ is clearly linked to the velocity of exchange of the residue while σ is a coefficient that takes into account stochastic phenomena that may occur, back exchange for instance.
The probability distribution of the random variable X_t subjected to evolution in equation 14 evolves

in time according to the Fokker Planck equation

$$\frac{\partial \rho(x,t)}{\partial t} = -\mu \frac{\partial \rho(x,t)}{\partial x} + D \frac{\partial^2 \rho(x,t)}{\partial x^2} \tag{15}$$

where $D = \sigma^2/2$ is called diffusion term. The probability distribution whose evolution is described in equation 15 is analogous to the mass spectrum of a peptide with infinite length since the deuterium uptake would never stops. Notice that boundary conditions could be set in order to prevent this. The Fokker Planck equation 15 can be numerically integrated, imposing as initial condition that no residue has exchanged, i.e. a Dirac Delta centered at zero (it would be analogous to center the initial distribution at any other value of m/z). Results of the integration at different times can be found in figure 2. Notice that the finite differences method used for the numerical integration introduces dissipation that is kept under control by monitoring that the integral of the the distribution is 1.

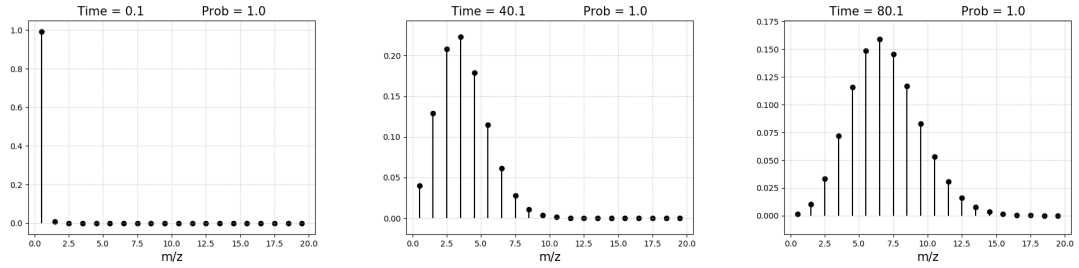

Figure 2: Numerical integration of Fokker Planck equation 15 shown at different times t=0.1,40.1 and 80.1. The integral of the distribution is shown to monitor dissipation introduced by the finite differences algorithm. Parameters: μ=0.08, D=1e-6, dt=0.1. Initial condition: Dirac Delta at zero.

The integration of equation 15 returns an evolution of a probability distribution that is qualitatively similar the experimental mass spectra in figure 1. An exact agreement with experimental data is beyond the purpose of the project whose aim is to qualitatively explain why bimodality occurs in EX1 conditions. So why bimodality occurs? In the case of stochastic modeling, the answer is straightforward: two populations with different drift velocities must exist (figure 3).

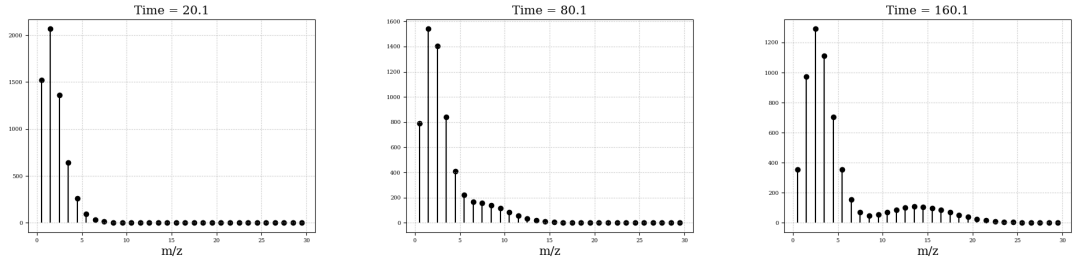

Figure 3: Numerical integration of Fokker Planck equation for two populations with different drift velocities $\mu_1 = 0.08$ and $\mu_2 = 0.02$ shown at different times: t=20.1, 80.1 and 160.1. Other parameters: D=1e-6, dt=0.1. Initial condition: Dirac Delta at zero. Population with drift velocity μ_1 is 5 times more crowded than population with velocity μ_2.

The stochastic modeling of hydrogen deuterium exchange here introduced leads to the important

conclusion that two populations must be distinguished within the residues of the peptide in order to experience bimodality in the experimental mass spectra. Such a conclusion, that seems trivial in the case of stochastic modeling, has a key role in reading the Michaelis Menten model in equation 1.

Is the second population hidden somewhere in the Michaelis Menten model? We already noted that the approximation of single exponential is only true under both native ($k_c \gg k_o$) and EX2 ($k_c \gg K_i$) conditions, while under EX1 conditions the correct solution is the double exponential in equation 8. Howevere, setting $A(0) = 1$ and $B(0) = C(0) = 0$ as initial conditions for the numerical integration of the model in equation 2, the numerical solution is always found to be a single exponential (figure 4A). Only switching on the initial condition on B the double exponential emerges (figure 4B).

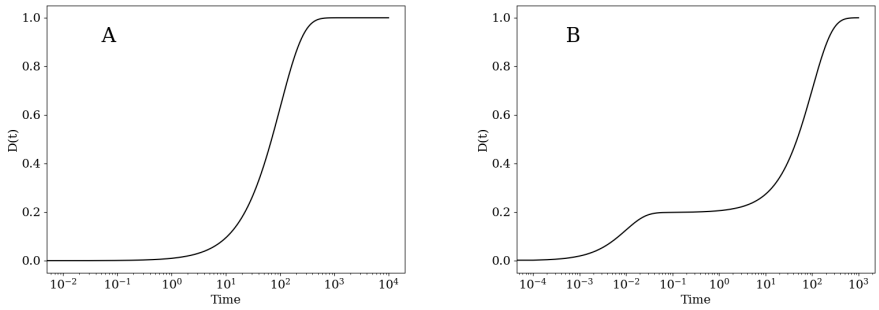

Figure 4: Numerical integration of Michaelis Menten model in equation 2 under native and EX1 conditions, namely $k_i = 100$, $k_o = 0.01$ and $k_c = 1$. A) Initial conditions: $A(0) = 1$; $B(0) = C(0) = 0$; other parameters: $dt = 0.01$ B) Initial conditions: $A(0) = 0.8$; $B(0) = 0.2$ and $C(0) = 0$; other parameters: $dt = 0.0001$.

.

It is strongly evident from figure 4 the dependence of the presence of the double exponential on the initial condition of the opened state B. In fact, the condition $B(0) = 0$ is not an assumption of the model that has as solution equation 8. We can write down the solution for $B(t)$ following from equation 5:

$$B(t) = \alpha_2 e^{\lambda_2 t} u_2^y + \alpha_3 e^{\lambda_3 t} u_3^y \tag{16}$$

where u_2^y and u_3^y are the second components of eigenvectors $\mathbf{u_2}$ and $\mathbf{u_3}$ (equation 7), λ_2 and λ_3 are the eigenvalues in equation 6, while α_2 and α_3 can be calculated from equation 8:

$$\begin{aligned} \alpha_2(k_i, k_c, k_o) &= \frac{k_o + k_i + k_c - \Delta - \frac{2k_i k_o}{k_i + k_o}}{2\Delta} \\ \alpha_3(k_i, k_c, k_o) &= -\frac{k_o + k_i + k_c + \Delta - \frac{2k_i k_o}{k_i + k_o}}{2\Delta} \end{aligned} \tag{17}$$

where $\Delta = \sqrt{(k_o + k_i + k_c)^2 - 4k_i k_o}$ was introduced. The solution for B(t) is valid for every time point and at time $t = 0$ it reads

$$B(0) = \alpha_2 u_2^y + \alpha_3 u_3^y \tag{18}$$

and since $\alpha_2, \alpha_3, u_2^y$ and u_3^y are functions of the rates k_o, k_c and k_i, the initial condition on B cannot be arbitrarily set, but is itself a function of the rates. In particular, the behaviour of the initial condition on B with respect to EX1 and EX2 conditions can be shown by plotting the value of B(0)

in equation 18 varying the ratio k_c/k_i (figure 5). Fixed the value of $k_i = 1$, the behaviour of B(0) is studied varying the ratio k_c/k_i while the ratio k_c/k_o is kept fixed in order to be under native conditions.

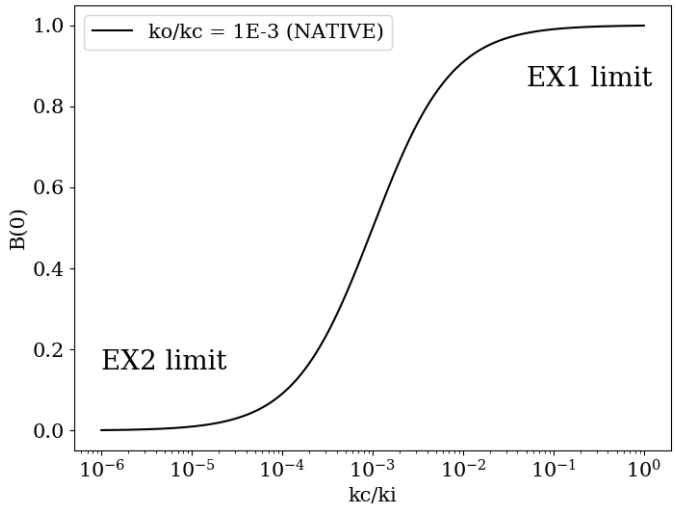

Figure 5: Behaviour of fraction of residues in the open state B at time $t = 0$ as a function of the ratio k_c/k_i and under native conditions $k_o/k_c = 1\text{E}{-3}$.

As shown in figure 5, the switch between EX2 and EX1 limit under native conditions is strongly correlated with the initial fraction of opened states B(0), suggesting that the second population that gives rise to bimodality is hidden in the population of this B state.
If $B(0) \neq 0$, then the model in equation 1 with initial conditions $A(0) = A_0$ and $B(0) = B_0 = 1 - A_0$ is equivalent to a model formed by two models each expressed by equation 1, one having as initial condition $A(0) = 1$, the second $B(0) = 1$, where the final deuterium uptake is opportunely weighted. Such an equivalence can be demonstrated numerically (see Supplementary Material). The net consequence is that the mass spectrum of the peptide would be the sum of the theoretical spectra that can be calculated out of these two processes (opportunely weighted) via equation 13.

Conclusions The goal of the project was to qualitatively explain why bimodality can be experimentally seen in the spectra of peptides under EX1 conditions. A simple stochastic model was introduced to show that in order to have bimodality in the mass spectrum, the two modes of the probability distribution must be associated to two different statistical populations.
Such a toy model, by far too simplistic to fit experimental data, gives the key to understand the important role played by the initial condition on the opened state B in equation 1. In fact, such initial condition is a function of the rates k_i, k_o and k_c that correlates strongly with the ratio k_c/k_i (figure 5), i.e. the quantity associated to the switch from EX2 to EX1 conditions.
Thus, the initial condition on B cannot be arbitrarily chosen and when different from 0, leads to a solution that can be expressed only as a sum of two exponential terms (equation 8): the single exponential approximation does not hold anymore. If the solution is a double exponential, the model in equation 1 can be equivalently written as a model formed by two Michaelis Menten processes

(opportunely weighted) with the same rates, one starting from initial condition A(0)=1, the second from B(0)=1. Such an equivalence can be easily demonstrated numerically. The mass spectrum corresponding to such process is thus the sum of the mass spectra associated to the two equivalent processes.
Concluding, the bimodality found in experimental spectra under EX1 conditions can be explained by the Michaelis Menten model in equation 1 via setting an initial condition on the opened state B different from 0. The process can be then decoupled to obtain a double weighted process from which bimodal mass spectra can be evaluated. Of course, an experimental agreement should validate the model here introduced.

Supplementary material The code for the numerical integration of Fokker Planck equation used to generate figures 2 and 3 is available at the following GitHub repository:

https://github.com/michelestofella/fokkerplanck

Codes to solve the system of differential equations in 2 and to generate figures 4 and 5 are also available. The scripts are all written in python.

References

[1] Adhikary, S., Deredge, D. J., Nagarajan, A., Forrest, L. R., Wintrode, P. L., and Singh, S. K. Conformational dynamics of a neurotransmitter:sodium symporter in a lipid bilayer. *Proceedings of the National Academy of Sciences 114*, 10 (2017), E1786–E1795.

[2] Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. Primary structure effects on peptide group hydrogen exchange. *Proteins: Structure, Function, and Bioinformatics 17*, 1 (1993), 75–86.

[3] Connelly, G. P., Bai, Y., Jeng, M.-F., and Englander, S. W. Isotope effects in peptide group hydrogen exchange. *Proteins: Structure, Function, and Bioinformatics 17*, 1 (1993), 87–92.

[4] Ferraro, D. M., Lazo, N. D., and Robertson, A. D. Ex1 hydrogen exchange and protein folding. *Biochemistry 43*, 3 (2004), 587–594. PMID: 14730962.

[5] Fersht, A. *Structure and Mechanism in Protein Science*. WORLD SCIENTIFIC, 2017.

[6] Hvidt, A., and Nielsen, S. O. Hydrogen exchange in proteins. vol. 21 of *Advances in Protein Chemistry*. Academic Press, 1966, pp. 287 – 386.

[7] Lapidus, L. Protein unfolding mechanisms and their effects on folding experiments [version 1; peer review: 2 approved]. *F1000Research 6*, 1723 (2017).

[8] Malhotra, P., Jethva, P. N., and Udgaonkar, J. B. Chemical denaturants smoothen ruggedness on the free energy landscape of protein folding. *Biochemistry 56*, 31 (2017), 4053–4063. PMID: 28714672.

[9] Masson, G., Burke, J., N.G., A., and al. Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (hdx-ms) experiments. *Nat Methods 16* (2019), 595–602.

[10] Moulick, R., Das, R., and Udgaonkar, J. Partially unfolded forms of the prion protein populated under misfolding-promoting conditions: characterization by hydrogen exchange mass spectrometry and nmr. *The Journal of biological chemistry 290* (08 2015).

[11] Skinner, S. P., Radou, G., Tuma, R., Houwing-Duistermaat, J. J., and Paci, E. Estimating constraints for protection factors from hdx-ms data. *Biophysical Journal 116*, 7 (2019), 1194 – 1203.

[12] Wagner, G., and Wüthrich, K. Structural interpretation of the amide proton exchange in the basic pancreatic trypsin inhibitor and related proteins. *Journal of Molecular Biology 134*, 1 (1979), 75 – 94.

[13] Weis, D. D., Wales, T. E., Engen, J. R., Hotchko, M., and Eyck, L. F. T. Identification and characterization of ex1 kinetics in h/d exchange mass spectrometry by peak width analysis. *Journal of the American Society for Mass Spectrometry 17*, 11 (2006), 1498 – 1509.

[14] Zhou, J., Yang, L., DeColli, A., Freel Meyers, C., Nemeria, N. S., and Jordan, F. Conformational dynamics of 1-deoxy-d-xylulose 5-phosphate synthase on ligand binding revealed by h/d exchange ms. *Proceedings of the National Academy of Sciences* (2017).