Oriental Journal of Chemistry Vol. **25**(4), 981-986 (2009)

Computational analysis of unfolding and folding pathways of proteins from amide proton exchange

K. SHANTHI and P.S.M. KANNAN

PR Institute of Science and Technology University, (Ponnaiyah Ramajayam College)
Thanjavur - 614 904 (India).

(Received: August 25, 2009; Accepted: October 18, 2009)

ABSTRACT

The exchange rate of amide proton is dependent on primary sequence of amino acids, pH and temperature. A novel computational program has been developed to predict the exchange rates of amide protons in a protein and their relationship with unfolding and folding has been analyzed. The structure -function relationship of proteins can be well probed at atomic level resolution using Hydrogen deuterium exchange methods.

Key words: Unfolding, folding, proteins, amide proton exchange, computational analysis.

INTRODUCTION

Proteins occupy a central position in the architecture and functioning of living mater. Proteins are the polymers that are chain like molecules produced by joining a number of small units of amino acids called monomers. The amino acids are regarded as "building blocks of proteins". The amino acids found in the proteins belong to the L-alpha amino acids. Although many amino acids have been shown to be present in various plants and animals, only 20 of them are found as constituent of proteins, which are being synthesized in ribosomes.

The amino acid units are linked together through the carboxyl and amino groups to produce the primary structure of the protein chain. The bond between two adjacent amino acids is a special type of amide bond known as the peptide bond and the chain thus formed is called a polypeptide chain. Each peptide chain is of considerable length and may possess from fifty to thousands of amino acid units. To understand the functions of proteins, their structure should be considered at Primary, secondary, tertiary and quaternary levels.

Solvation energy in protein folding and binding was analyzed by David Eisenberg etal[1] to calculate the stability of protein structure in water starting from atomic coordinates. Their work is based on the earlier contribution made at different levels [2-8].Information about the structure and conformational dynamics of proteins, resolved to the level of individual amino acid residues, is contained in their hydrogen exchange (HX) behaviour. The protection against exchange imposed by protein structure is often expressed by the factor, $P = k_{rc}/k_{proc}$ derived by comparing the HX rate measured in a protein ($k_{proc)}$ with the rate expected in a random coil model (k_{rc}).[9] To extract structural information from measured protection factors, one must first know ***k,*** and understand the non-structural effects on which it is based. In aqueous solutions, peptide group HX is analyzed by hydroxide and hydronium ions[10] in pH dependent reactions and by water in a pH-independent way[11, 12].

The contribution of all other potential catalysts is negligible[13]. HX rates are, however, significantly influenced by neighboring side chains even in the absence of folded structure. Thus, k_{rc} depends not only on solution pH and temperature, but also on local amino acid sequence. Inductive effects of polar side chains on the acid and base-catalyzed HX behavior of peptide groups were demonstrated and calibrated by Molday et aL[14]These workers also demonstrated an additivity rule for predicting the summed inductive effects on a peptide NH due to its two neighboring side chains. This knowledge, although incomplete, has been widely used in protein studies to correct measured HX data for side chain inductive effects. Over the 20 years since the work of Molday and coworkers [14] various issues have arisen. For example, recent work points to a previously unsuspected effect of some a polar side[15-17]. The increasing use of HX approaches for protein studies emphasizes the need for a more complete analysis of primary structure effects. For this purpose we have examined the effects on peptide group HX due to all 20 naturally occurring amino acid side chains, determined reference rates pertinent for peptide group NH to ND exchange in proteins and oligopeptides, and considered the dependence of exchange rates on temperature.

Each protein adopts a particular, well-defined, unique three-dimensional (3D) structure, which is directed to do certain function(s). How and why does a protein adopt a particular 3D structure instead of some others is perplexing, which is popularly known as 'protein folding problem'. The protein folding problem is very much complicated from many facts such as size of proteins, primary sequences, local interactions, topology, environmental conditions and so on. However understanding the protein folding problem is important because it is believed that the knowledge gathered from protein folding would help 1) in predicting three-dimensional structures of proteins from primary sequences 2) in *de novo* peptide/ protein designs 3) in solving the mis-folding problems, which is directly linked with many human diseases.

Three-dimensional structures of proteins can be determined at atomic level resolution using x-ray crystallography and nuclear magnetic resonance (NMR) techniques. It is interesting to mention that protein folding, stability and functions of proteins are driven largely by non-covalent interactions. Consequently, native protein molecules scan many conformations under physiological conditions. These protein ensembles are manifested directly in protein folding, stability, dynamics, and legend binding and allosteric effects.

As macroscopic methods detect only the conformations that exist in majority, the results from those experiments may not provide a clear picture on the conformational motions of proteins. In this context, Hydrogen-Deuterium (H/D) exchange has become an important method to probe stability, structural fluctuations and intermolecular interactions in proteins at residue level resolution. The results from amide hydrogen exchange appear to be a robust framework for obtaining quantitative information about stabilizing forces in native and partially structured proteins.

The major objectives of the present work is to develop a novel computational program to predict exchange rates of amide protons in a protein and analyze their relationships with unfolding and folding pathways(s). The exchange rate of amide protons is dependent on primary sequence, pH, temperature and denaturant concentrations. Thus, algorithm would be designed in such a way to account all these factors affecting the exchange rate of amide protons. Moreover, combined analysis of unfolding and folding pathways of proteins monitored by macroscopic and microscopic (H/D exchange) methods will provide rich information on the protein-folding problem. However, interpretation of experimental data from H/D exchange methods is not straightforward. It requires a sound knowledge on the relationship between exchange rate and energetic conformations of proteins and a sophisticated computer algorithm to examine the complex data. In this context, a computer program is provided to full-fill the research objectives as stated above.

A novel computer algorithm to probing protein conformations is developed.. The algorithm would help to predict all types of energetic conformations of proteins. The major and minor

population of protein conformations can be effectively understood by combined analysis of experimental data with the program. Quantitative interpretation can also be precisely derived to the exchange data. Moreover, it is believed that the algorithm will contribute much on understanding the structure function relationships of the protein molecules.

Methods and Algorithms

The exchange reactions in proteins are catalyzed by acid, base and water. Equation(1) can be used to predict the Hydrogen Exchange rate of an unstructured peptide NH neighbored by particular side chains in a random chain conformation(k_{rc}).

$$k_{rc} = k_{(Acid)} + k_{(Base)} + k_{(Water)} \quad ...(1)$$

Equation(2) replaces the side chain specific acid, base and water rate constants in (k_A,k_B,k_W) with standard rate constants for the pertinent alanine reference peptide(k_A.ref, k_B.ref ,k_W.ref, Table III) multiplied by the side chain specific acid (AL,A_R) or base (B_L,B_R) factors (Table II)

$$k_{rc} = k_{A,ref}[A_L * A_R][D^-] + k_{B,ref}[B_L * B_R][OD^-] + k_{W,ref}[B_L * B_R] \quad ...(2)$$

The predicated first order rate constant for acid catalysis k($_{Acid}$) can be obtained by adding the logarithms of the appropriate terms (listed in Table 1, 2) and then taking the antilogarithm.

As the rate constants(k_A,k_B,k_W) are derived from poly- DL- alanine [PDLA]sequence, the exchange rate constant for any amide protons in the given protein sequence must be calculated by correcting the influence of neighboring amino acids to the proton under interest. The effects of primary sequence and pH can be accounted as represented below

$$\log k_{(Acid)} = \log k_A + \log A_L + \log A_R - pD \quad ...(4)$$

Analog exercised yield the rates for base and for water catalyzes

$$\log k_{(Base)} = \log k_B + \log B_L + \log B_R - pOD \quad ...(5)$$

$$\log k_{(Water)} = \log k_W + \log B_L + \log B_R \quad ...(6)$$

The expected exchange rate for any NH is the sum of its independently computed acid, base and water terms. Different solution conditions (temperature, etc) are accounted for by suitably adjusting the reference rates and catalyst concentrations. The A and B factors are insensitive to solution conditions (except for salt-sensitive charge effects). These factors are written from the perspective of the side chain. R refers to the peptide group to the right of the side chain in question and L to its left.

As an example let us calculate the expected HX rate of the oligopeptide Ile-NH-Val proton, considered just above, at pD 2.0 and 5°C in 0.5 M KC1. For these conditions, log[OD^-] = 2.00 - 15.65 = -13.65. From the oligopeptide reference rates in Table I pertinent for the high salt condition, log(k_A) = 1.56, log(k_B) = 10.20, and log(k_W) = -2.3.

The NH is R from Ile and L from Val (Table II). Therefore, the acid rate for Ile-NH Val Eq. (4) is antilog (1.56 -0.74 -0.59 -2.00) = 1.70 x 10^{-2} min^{-1}. The base-catalyzed rate is antilog (10.20 -0.70 -0.23 -13.65) = 4.16 X 10^{-5} min^{-1} The water rate is antilog (-2.30 -0.70 -0.23) = 5.9 x 10^{-4} min^{-1}. The predicted HX rate is the sum of the acid, base, and water rates.

The predicted HX rate is the sum of acid, base and water rates. For a protein at 20^0 C under normal low salt conditions the PDLA reference rates in Table III should be used.

Temperature Dependence

Reference rate constant pertinent for alanine peptides in normal low salt conditions and specific for 20^0C are listed in Table III.

To predict HX rates at other temperatures, each rate or reference rate constant in Equation(2) should be modified according to Equation(3).

$$k_{rc}(T) = k_{rc}(293).\exp((-Ea(1/T - 1/293)/R) \quad ...(7)$$

The present work weighted to account for different levels of accuracy in the various data sets.

Table 1: H to D exchange rate constants for alanine - based reference molecules at 278°K in 0.5m KCl

	$logk_A$ (M^{-1} min^{-1})	$logk_B$ (M^{-1} min^{-1})	$logk_W$ (M^{-1})
N-Ac-Ala-N'MA (L)	2.87	9.71	-2.3
N-Ac-Ala-N'MA (R)	2.81	9.01	-3.2
N-Ac-A-A-A-N'MA (A)	1.56	10.20	-.23
PDLA	1.19	9.90	-2.5

Table 2: Effects of Amino acid side chains on the HX rates of neighboring peptides

Side	$Logk_{ex}$(X) Acid catalysis		$Logk_{ex}$(Ala) Base catalysis	
chain	L	R	L	R
Ala	0.00	0.00	0.00	0.00
Arg	-0.59	-0.32	0.08	0.22
Asn	-0.58	-0.13	0.49	0.32
Asp (COO^-)	(0.9)	0.58	-0.30	-0.18
Asp (COOH)	(-0.9)	-0.12	0.69	(0.6)
Cys	-0.54	-0.46	0.62	0.55
Crys2	-0.74	0.58	0.55	0.46
Gly	-0.22	0.22	0.27	0.17
Gln	-0.47	-0.27	0.24	0.39
Glu(COO^-)	(-0.9)	0.31	-0.51	-0.15
Glu (COOH)	(-0.6)	-0.27	0.24	0.39
His			-0.10	0.14
His+	(-0.8)	-0.51	(0.8)	0.83
Ile	-0.91	-0.59	-0.73	-0.23
Leu	-0.57	-0.13	-0.58	-0.21
Lys	-0.56	-0.29	-0.04	0.12
Met	-0.64	-0.43	-0.24	0.11
Phe	-0.52	-0.43	-0.24	0.06
Pro (trans)		-0.19		-0.24
Pro (cis)		-0.85		0.60
Ser	-0.44	-0.39	0.37	0.30
Thr	-0.79	-0.47	-0.07	0.20
Trp	-0.40	-0.44	-0.41	-0.11
Tyr	-0.41	-0.37	-0.27	0.05
Val	-0.74	-0.30	-0.70	-0.14
N-term (NH^+_3)		-1.32		1.62
C-term (COO^-)	0.96		(-1.8)	
C-term (COOH)	0.05			

These are 14,17 and 19 kcal/mol for k_A, k_B and k_W respectively. In order to account the effects of temperature, Arrhenius Equation has been employed.

After a number of mathematical conversions, the Equation-4 is derived and is being used in our program , which calculate the rate constants for amide and side chain protons of a protein molecule in random coil conformations.

$$k_{rc} = 10^{(\log k_A - pD)} + 10^{(\log k_B - pOD)} + 10^{(\log k_W)} \quad ...(8)$$

ProHDEx Algorithm

In this section, we present a novel algorithm to predict the exchange rate of various amide protons present in the given protein sequence. The exchange rate constant and time constant for individual amino acids are also calculated. The exchange rate of amide protons depends on factors such as Ionic strength, pH, temperature and Primary sequence of the protein. The logarithmic values of effects of amino acid side chains on the HX rates of neighboring peptides are stored in a array. The H to D Exchange rate constant for Alanine based reference molecules at 293^{0}K are stored in another array. The temperature and pH values are added with 273 and 0.4 respectively. The new value of pH is subtracted from 15.05. The expression of Arrhenius equation is evaluated.

The Ionic Strength value is checked. If it is low, the H to D Exchange Rate constants for Alanine Based reference molecules at 293^{0}K with obtained result is added. Otherwise, the logarithmic values of H to D Exchange rate Constants for Alanine Based reference molecules at 278^{0}K with obtained result is added. The amino acid sequence is checked. The characters of the sequence are separated. Each of the character in the amino acid sequence is compared with the stored array. The corresponding left and right side chain peptide groups are added for acid, base and water. To find out the Exchange rate, the power of 10 with the three values k_A,k_B,k_W are computed.

Table 3: H to D Exchange Rate Constants for Alanine-Based Reference Molecules at 293°C

	$\log k_A$ ($M^{-1}min^{-1}$)	$\log k_B$ ($M^{-1}min^{-1}$)	$\log k_W$ (min^{-1})
N-Ac-A-A-A-N'MA	2.04	10.36	-1.5
PDLA	1.62	10.05	-1.5

RESULTS

We implemented the algorithm in C. We will test the program and compare our program with some famous existing programs. The test platform was a desktop PC with P4 2.8G CPU.

The Outputs of the program are

- Exchange rate constants for amide protons at various values of pH.
- Exchange rate constants for side chain NH protons at various values of pH.
- Optimum pH at which the overall exchange rates of amide protons is minimum.

For instance, the outputs of the program run for a mock amino acid sequence are given below:

Protein sequence : ARRSV

Amino Acids	Rate constant				Optimum
	pH2	pH3	pH4	pH5	pH
A	0.1107	0.0369	0.1428	1.2856	3
R	0.0436	0.0560	0.2831	2.5649	2
R	1.0046	1.3853	5.4406	46.0190	2
S	0.0890	0.1299	0.6636	6.0127	2
V	0.0110	0.0048	0.0206	0.1858	3

Optimum pH of the given protein sequence : 2.4000

REFERENCES

1. David Eisenberg. R and Andrew D Mathmillan S, *Nature*, **319**: 199 (1986).
2. Novotný, J., Bruccoleri, R. & Karplus, M. *J. molec. Biol.* **177**: 787"818 (1984).
3. Kauzmann, W. *Adv. Protein Chem.* **14**(1): 63 (1959).
4. Creighton, T. E. *Proteins* (Freeman, San Francisco, 1984).
5. Tanford, C. *The Hydrophobic Effect* 2nd edn (Wiley, New York, 1980).
6. Karplus, M. & McCammon, A. A. *Rev. Biochem.* **52**: 263"300 (1983).
7. Nemethy, G. & Scheraga, H. A. *Q. Rev. Biophys.* **10**: 239"352 (1977).
8. Levitt, M. & Warshel, A. *Nature* **253**, 694"698 (1975).
9. Englander, S.W., Kallenbach, N.R. Hydrogen exchange and structural dynamics of proteins and nucleic-acids. *Q.Rev. Biophys.* 16521-655 (1984).
10. Berger, A., Loewenstein, A., Meiboom, S. NMR study of the protolysis and ionization of N-methylacetamide. *J.Am.* Chem. Soc. **8. 1**: 61-67 (1959).
11. Englander, J.J., Calhoun, D.B., Englander, S.W. Measurement and calibration of peptide group hydrogen-deuterium exchange by ultraviolet spectrophotometry. *Anal. Biochem*. **92**: 517-524 (1979).
12. Gregory, R.B., Crabo, L., Percy, A.J., Rosenberg, A. Water catalysis of peptide hydrogen isotope exchange*. Biochemistry* **22**: 910-917 (1983).
13. Englander, S.W., Poulsen, A. Hydrogen-tritium exchange of the random chain polypeptide. *Biopolymers* **7**: 329-339 (1969).
14. Molday, R.S., Englander, S.W., Kallen, R.G. Primary structure effects on peptide group hydrogen exchange. *Biochemistry* **11**: 150-158 (1972).
15. Robertson, A.D., Baldwin, R.L. Hydrogen exchange in thermally denatured ibonuclease A. *Biochemistry* **30**: 9907-9914 (1991).
16. Lu, J., Dahlquist, F.W. Detection and characterization of an early folding intermediate of T4 lysozyme using pulsed hydrogen exchange and two-dimensional NMR. *Biochemistry* **31**: 4749-4756 (1992).
17. Radquist, S.E., Buck, M., Topping, K.D., Dobson, C.M.,Evans, P.A. Hydrogen exchange in denatured states of egg white lysozyme. *Proteins.* 15237-248 (1992).