Article can be accessed online at http://www.publishingindia.com

ProCOOP, A Program that Predicts the Cooperatives of Hydrogen Exchange in Proteins

K. Sarvanan*

Abstract

Each protein adopts a particular, well-defined, unique three-dimensional (3D) structure, which is directed to do certain functions. 3D structures of proteins can be well-determined at atomic level resolutions using x-ray crystallography and nuclear magnetic resonance (NMR) techniques. These high resolution structures of proteins are essential for understanding their structure-function relationships. However, there is a divergent correlation between these experimental outcomes and requirements of current research in structural biology. In this background, ProCOOP algorithm was developed to predict population of secondary structures in proteins based on molecular mass of their deuterated-forms. By taking into consideration of many different structural and environmental factors, the ProCOOP validates its outputs and gives suggestion to improve experimental conditions for better predictions. The applications of ProCOOP for the data analysis in proteomics and genomics have also been brought into fore in detail.

Keywords: Proteins, Secondary Structure, Hydrogen Exchange, Rate Constant, Algorithm

1. Introduction

Proteins, one of the major governing bodies of bio-systems, play significant roles on structural and functional aspects of all living systems. Each protein has unique secondary structures, which regulate its functions. Understanding the features of these secondary structures is essential for *denovo* protein/peptide designing. In general, motions in proteins range from microsecond to seconds. The conformational heterogeneity and macromolecular flexibility are essential determinants of protein functions. As macroscopic methods detect only the conformations that exist in majority, the results from the experiments may not provide a clear picture on the conformational motions of proteins. In order to help and speed-up the processes of structural analysis on proteins; we have developed a novel algorithm ProCOOP. The structure and applications of the ProCOOP has been discussed in this article.

2. Prerequisites for ProCOOP Analysis

The molecular mass (MM) of deuterated conformations of proteins under interest should be estimated by mass spectrometry (MS) and the MM should be given as one of the inputs to ProCOOP. Usually, proteins can be deuterated by Hydrogen-Deuterium (*H/D*) exchange process (Murphy, 2009). H/D exchange is a process in which labile protons in proteins are substituted by solvent deuterons. By dissolving a test protein in deuterium oxide (D_2O) for a limited time at defined pH, temperature and ionic strength, a deuterated protein can be prepared (Bai *et. al.*, 1993). The H/D exchange process is represented as shown below.

$$Closed\ (NH) \underset{k_{cl}}{\overset{k_{op}}{\leftrightarrow}} Open\ (NH) \xrightarrow{k_{ch}} Exchanged\ (ND)$$

* Dean, Faculty of Computer Science, PRIST University, Thanjavur, Tamil Nadu, India. E-mail: ks_tnj@yahoo.co.in

In this model, exchange competent (Open NH) and exchange incompetent (Closed NH) conformations interconvert with rate constants k_{op} and k_{cl}, respectively. Exchange happens only from the open conformation with a rate constant of k_{ch}. The observed rate constant of exchange is described as,

$$k_{obs} = (k_{op} * k_{ch}) / (k_{ch} + k_{cl})$$

Under conditions, where k_{ch} is greater than k_{cl}, the observed rate constant of exchange k_{obs} becomes k_{op}. This is known as EX1 exchange (Arrington *et. al.*, 1999). However, if k_{cl} is greater than k_{ch}, then,

$$k_{obs} = (k_{op} * k_{ch}) / k_{cl} = K_{op} * k_{ch}$$

Where, K_{op} is the equilibrium constant. This is known as EX2 exchange (Gohima.and Selvaraj, 2008). The deuterated proteins should be prepared by EX2 exchange mechanism for determining its secondary structural content by ProCOOP. This follows from the fact that EX2 exchange probes the equilibrium of a protein between its folded and unfolded states (Ferraro *et. al.*, 2004; Murphy, 2009; Gohima.and Selvaraj, 2008).

The MM of a deuterated protein can be calculated using MS (Arrington *et. al.*, 1999; Fisherman and Honag, 2003). The advantage of this method is many folds: i) the MS requires protein sample only in microgram level ii) MM can be calculated at higher resolution of 0.001 Dalton (Da) iii) the time required to collect the data is in the time span of minutes iv) experiments are simple and inexpensive. The amino acid sequence of the protein, MM of deuterated protein, the number of amide protons, opening reaction rate constants for each amide proton, time constant, the number of groupings and cut-off values for each group are the inputs to ProCOOP. The ProCOOP analyses will reveal the population of secondary structural content of given test protein in short time. The implications of the outcomes have been discussed in the later section of this article. The language 'Visual Basic' is used to write the ProCOOP. The language facilitates the use of GUI based controls as the ProCOOP requires them at enormous level.

3. The ProCOOP Algorithm

The ProCOOP algorithm has been developed to predict population of secondary structures in proteins based on MM of their deuterated conformations. In outline, the program first counts number of amino acids in the input protein sequence and then their molecular masses are summed. From this value, the MM of the input protein sequence is determined from the following equation.

$$MMIS = \sum_{i=1}^{n} MMA - [(n-1)\ (MMH_2O)]$$

Where, MMIS – Molecular Mass of input sequence

n – Number of amino acid

MMA – Molecular Mass of amino acid

MMH_2O – Molecular Mass of H_2O

The respective atomic and molecular mass corresponding to various atoms and molecules used in ProCOOP were referred using the OPUS^{TM} and IsoPro software (Arrington *et. al.*, 1999; Berman, 2007; Gohima.and Selvaraj, 2008). Next, all labile protons in the given sequence are counted and they are further classified into side chain- and backbone-labile protons. The MM of deuterated form of the protein is calculated assuming all labile protons (H) of the protein are exchanged to solvent deuterons (D). Lastly, the population of secondary structures of the input protein is estimated using the following relationship.

$$PSS = e^{(-1 * kop * t)}$$

Where, PSS – Population of Secondary Structures

kop – Rate constant of opening reaction of amide proton

t – Time Constant

The algorithm will be illustrated in more detail by a schematic diagram for a hypothetical protein molecule having eight amino acids (Fig.1). The primary sequence is given in Fig.1(a) (P - Proline, R - Arginine, I - Isoleucine, S - Serine, M-Methionine and C-Cysteine). Molecular formula and MM of the sequence are shown in Fig.1(b) based on the number and types of amino acids that are building the sequence. Fig.1(c) depicts total number of labile protons in each amino acid of the sequence. It is important to mention that all α-L-amino acids have only one backbone labile proton except proline, which has no backbone labile proton (Jackson, 1988, Maity *et. al.*, 2005). The MM of completely deuterated form of given sequence has also been shown in Fig.1(c).

4. Special Features of ProCOOP

The ProCOOP requires seven inputs: i) the primary sequence of a protein for which secondary structures need to be predicted ii) MM of the deuterated protein as determined by MS iii) the number of amide protons of the test protein sequence iv) the opening reaction rate constant for each amide proton v) the value of time constant vi) the number of groupings for exchange and vii) the cut-off value for each group. The ProCOOP delivers three outputs for each complete run: i) Conformation of each exchange ii) Molecular mass of each conformation and iii) Population of secondary structures (PSS). The outputs from the ProCOOP for a hypothetical protein (MISPR) are depicted in Table 1.

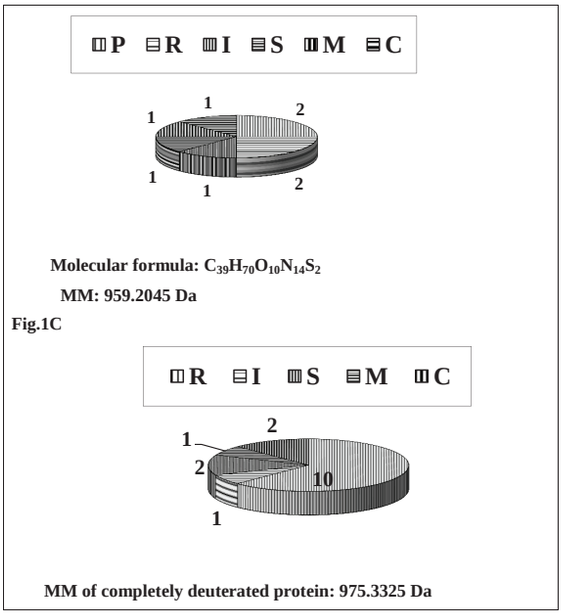

Figure 1: The schematic diagram outlines the major steps involved in ProCOOP The given polypeptide sequence is in Fig.1(a). The molecular mass (MM) of 607.558594 was taken as if the MM determined by MS for the input sequence. The Fig.1(b) and 1(c) illustrate how the ProCOOP counts the number of amino acids and labile protons of the given sequence, respectively.

The unique feature of ProCOOP is that it gives outputs for oxidized as well as reduced conformations of proteins input. In reduced state, proteins are bereft of disulfide bonds whereas in oxidized state, cysteine residues of proteins become cystine. The ProCOOP accounts the both forms of proteins and this particular aspect pave a way to derive structural parameters pertaining to disulfide bonds in proteins provided their secondary structures are known. Moreover, information about various types of labile protons in proteins is useful in designing kinetics of H/D exchange in proteins.

Table 1: Inputs and Outputs of Pro COOP for a hypothetical protein

Inputs:		
Enter the no. of amide protons & molecular weight : 7,968.56		
Enter the rate constant k_{op} for each proton		
Amide proton 1 : .5		
Amide proton 2 : .1		
Amide proton 3 : .8		
Amide proton 4 : .3		
Amide proton 5 : .9		
Amide proton 6 : .6		
Amide proton 7 : .4		
Enter the time constant : .9		
Enter the no. of groupings : 4		
Enter the cut-off value 1 : .32		
Enter the cut-off value 2 : .54		
Enter the cut-off value 3 : .78		
Outputs:		
Conformations	**MM**	**Population**
H H H H	968.558594	0.151071808950453
H H D H	969.566406	0.108168459944019
D H H H		
H D H H		
H H H D		
D H D H	970.574219	0.278805166335477
H D D H		
H H D D	971.582031	0.199626427170377
D D H H		
D H H D		
H D H D	972.589844	0.135778578922826
D D D H		
D H D D		
H D D D	973.597656	0.097218401484068
D D H D	974.605469	0.017092680063247
D D D D	975.613281	0.012238477129533

The initial conformation of output shows that there is no exchange taken place. The PSS value estimated accordingly with no exchange. The exchange is occurred for each amide proton, opening reaction rate constant value at each conformation of test protein. In some cases,

PSS as predicted by ProCOOP for given input might be misdirected. For instance, if the time required for H/D exchange process is not optimized, the protein molecules will be partly deuterated wherein some side chain labile protons may not be completely exchanged. In this circumstance, the ProCOOP overestimates the PSS and gives warning messages. Hence, the warning messages are obviously helpful to revisit/refine the experimental conditions. To our best knowledge, though there are many algorithms [Jing and Song, 2009; Lim, 1994; Krishna.*et. al.*, 2006; Schulz, 1988) for prediction of population of secondary structures in proteins, ProCOOP is the first software tool for predicting population of secondary structural contents of proteins based on their deuterated forms.

5. Conclusion

It has been clearly demonstrated that ProCOOP predicts the population of secondary structures of proteins and rationalize its results by taking into consideration of factors from many facets. In the near future, the ProCOOP will be improved for defining various types of secondary structures in proteins along with PSS. This particular goal can be achieved by coupling a program predicting population of secondary structures in proteins with ProCOOP. Foreseeing the potential applications of ProCOOP in structural biology, it is anticipated that a great scope to improve the software tool at many different angles.

References

1. Arrington, C. B., Tecsch, L. M. & Robertson, A. D. (1999). Defining protein ensembles with native state NH exchange: Kinetics of inter-conversion and cooperative units from combined NMR and MS analysis. *Journal of Molecular Biology*, 285(3), 1265-1275.
2. Bai, Y., Milne, J. S., Mayne, L. & Englander, S.W. (1993). Primary structure effects on peptide group hydrogen exchange. *Proteins: Structure, Function and Genetics*, 17(1), 75-86.
3. Berman, H., Herrick, K., Nakamura, H. & Markley, N. (2007). The worldwide Protein Data Bank, ensuring a single, uniform archive of PDB data. *Nucleic Acids Research*, 35, 301-303.
4. Ferraro, D. M., Lazo, N. D. & Robertson, A. D. (2004). EX1 hydrogen exchange and protein folding. *Biochemistry*, 43(3),587-594.
5. Fisherman, S. G. & Honag, B. (2003). Structural genomics: Computational methods for structural analysis. *Protein Science*, September, 12(9), 1813-1821.
6. Ghanty, P. & Pal, N. R. (2009). *Prediction of protein folds*. IEEE transactions on Nanoscience, 8(1), 432-442.
7. Gromiha, M. M. & Selvaraj, S. (2008). Bioinformatics approaches for understanding and predicting protein folding rates. *Current Bioinformatics*, 3, 1-8.
8. Jackson, S.E. (1998). How do small single domain proteins fold? *Folding and Design*, 3(4), R81-R91.
9. Jing, Z. & Song, L. (2009). *A simulating algorithm for protein folding process*. International Conference on Communication and Mobile Computing, 343-347.
10. Krishna, M. M. G., Maity, H., Rumbley, J. N., Lin, Y. & Englander, S. W. (2006). Order of steps in the cyto-chrome c folding pathway: Evidence for a sequential stabilization mechanism. *Journal of Molecular Biology*, 359, 1411-1420.
11. Lim, V. I. (1994). Algorithms for prediction of α-helical and β-structural regions in globular proteins. *Journal of Molecular Biology*, 88, 873-941.
12. Maity, H., Maity, M., Krishna, M. M. G., Mayne, L, & Englander, S. W. (2005). *Protein folding: The stepwise assembly of foldon units*. Proceedings of the National Academy of Sciences, 102(13), 4741-4746.
13. Murphy, K. P. (2001). *Protein Structure, Stability and Folding Editor*. Totowa, New Jersey: Humana Press.