Arasu Prabhavadhni et al /J. Pharm. Sci. & Res. Vol. 7(3), 2015, 159-162

ISSN:0975-1459
Journal of Pharmaceutical Sciences and Research
www.jpsr.pharmainfo.in

# A Review on Computational Methods for Predicting Residue-Specific Stabilities of Proteins

**Arasu Prabhavadhni, Tambi Richa and Thirunavukkarasu Sivaraman***

*Structural Biology Lab, Anusandhan Kendra, School of Chemical and Biotechnology, SASTRA University, Thanjavur – 613401, Tamil Nadu, India.*

**Abstract**

Residue-specific stabilities of proteins can be reliably determined by using NMR-monitored hydrogen/deuterium exchange (NMR-H/D) methods under native conditions. Notwithstanding the potential advantages of the NMR-H/D methods, they are time consuming, very expensive and technically challenging. In this review article, various computational tools and strategies reported to date in the literature on predicting the residue-specific free energy ($\Delta G_{HX}$) of proteins have been comprehensively exemplified. Moreover, scopes to develop novel computational tools for estimating $\Delta G_{HX}$ values of proteins have also been discussed.

**Keywords:**

Free energy, H/D exchange, *In silico* methods, NMR and Stability.

## 1. INTRODUCTION

Structures, stabilities and dynamics of proteins can be comprehensively studied at the level of individual residues by using NMR-assisted hydrogen/deuterium exchange methods [1-3]. In these methods, a protein is dissolved in deuterium oxide ($D_2O$) following which amide protons (NHs) of the protein exchange with deuterium, when they are exposed through an ephemeral conformational opening event. All standard amino acids but proline of proteins provide NHs for the H/D exchange processes [1,4]. In general, the H/D exchange reactions of proteins are explained by a two-state model shown in equation 1 [4, 5]. In the model, 'closed (NH)' is the folded conformation and 'open (NH)' is the unfolded conformation of proteins.

$$\text{Closed (NH)} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} \text{Open (NH)} \xrightarrow{k_{ex}} \text{Exchanged} \qquad (1)$$

The rate constant for the unfolding is $k_{op}$ and $k_{cl}$ is the folding rate constant. Exchange takes place from unfolded states of proteins as per the model. The intrinsic exchange rate constants ($k_{rc}$) of NHs can be predicted at defined experimental conditions using the method reported by Bai *et al.* [6-9]. The extrinsic exchange rate constants ($k_{ex}$) of NHs are being estimated from data obtained from NMR-HD methods and relationships among the four rate constants ($k_{ex}$, $k_{rc}$, $k_{op}$ and $k_{cl}$) discussed above are shown in equation 2.

$$k_{ex} = ( k_{op} * k_{rc})/( k_{cl} + k_{rc}) \qquad (2)$$

H/D exchange processes of proteins have two regimes: EX1 and EX2 [10, 11]. Under EX1 exchange conditions, the $k_{rc} >> k_{cl}$ and the $k_{ex}$ is estimated using equation 3. Whereas, under EX2 exchange conditions, the $k_{cl} >> k_{rc}$ and the $k_{ex}$ is estimated using equation 4, wherein, $K_{HX}$ is the residue-specific equilibrium constant for NHs in proteins. The residue-specific free energy is then determined using the relationship shown in equation 5, wherein'R' is the gas constant and 'T' is the absolute temperature. In general, the free energy of exchange ($\Delta G_{HX}$) of a protein is averaged out to three largest residue-specific $\Delta G_{HX}$ of the protein [12-14].

$$k_{ex} = k_{op} \qquad (3)$$

$$k_{ex} = ( k_{op} * k_{rc})/( k_{cl}) = K_{HX} * k_{rc} \qquad (4)$$

$$\Delta G_{HX} = - RT \ln (K_{HX}) \qquad (5)$$

The $\Delta G_{HX}$ values estimated using NMR-H/D methods are very useful on evaluating structures, stabilities, folding pathways and dynamics of proteins at high resolution. However, owing to several inherent limitations, applications of these methods are not so straightforward [15-17]. Till date (as on Jan 2015), residue specific dynamics for only 83 proteins have been studied using NMR-H/D exchange methods [18]. In these contexts, computational tools will be an excellent alternative to the H/D exchange methods provided the tools are robust and reliable on probing stabilities and folding pathways of proteins on the basis of their structural architectures [18-21]. This article reviews all the pertinent computational tools/strategies available to date for predicting the residue-specific dynamics of proteins.

## 2. COMPUTATIONAL METHODS FOR PREDICTING DYNAMICS OF NHS IN PROTEINS

'COREX/BEST' [20] and 'CamP' [22] are the computational tools available to date to predict residue-specific $\Delta G_{HX}/P$ (P – protection factor) of proteins. Other than these two tools, protection status of NHs in proteins from H/D exchange can be assessed by means of 'H-protection' computational tool [23]. In the following sections, basic principles and functioning of all the computational tools have been described.

### 2.1 COREX/BEST

COREX/BEST (Biology using Ensemble-based Structural Thermodynamics) aims at capturing conformational ensembles existing in folding/unfolding pathways of proteins [20, 24]. In outline, the program first calculates

probability for each amino acid in proteins to be in folded or unfolded states on the basis of 'global unfolding model'. Second, the program enumerates ensemble conformations of proteins and also defines free energy to each ensemble state. Finally, the program tabulates 'stability constant' for each residue present in three-dimensional (3D) structures of proteins and also outputs most probable low-energy unfolded states of proteins. However, Skinner et al. have recently demonstrated that the prediction accuracy (only around 27%) of the tool is not quite impressive [25, 26].

The COREX/BEST webserver is available at http://best.bio.jhu.edu/BEST/. In order to perform free energy predictions for proteins using the tool, users should create an account through the server under which a new workspace will be initiated, wherein the protein PDB structures to be analysed can be uploaded. At first, users should run the ensemble generator following which residue-specific stability constants can be calculated. The program also facilitates energetic profiling of proteins [27, 28], cold denaturation studies [29] and studying effect of pH on protein conformations [30].

### 2.2 CamP

Vendruscolo et al. developed the webserver CamP for predicting the protection factors of residues in proteins directly from the respective primary sequences of the proteins [22]. The strategies are based on the 'local unfolding model' of proteins. Protection factor (P) is a measure of the amount of resistance offered by a proton to exchange in the solvent and the 'P' can be calculated as shown in the equation 6. On the basis of 'P' values of residues in proteins, overall structural flexibility/rigidity of proteins can be mapped.

$$P = k_{rc}/k_{ex} \quad (6)$$

In other words, lower the protection factor, higher the flexibility of the region in the proteins and *vice versa*. CamP has been developed using neural network and following phenomenal equation 7 forms the basis of the algorithm. Wherein, $N_i^c$ and $N_i^h$ represent hydrophobic and hydrogen bond influences for the amide hydrogen of residue 'i' under calculations.

$$\ln P_i = \beta_c N_i^c + \beta_h N_i^h \quad (7)$$

The factors '$\beta_c$' and '$\beta_h$' denote free energy contributions to creating van der Waals contacts and hydrogen bonds, respectively and they were determined by fitting of experimental data derived from seven proteins for which H/D exchange rates were available [31]. However, the prediction accuracy of the model was shown to be only around 50% [25, 26] and the webserver is not publically available at present.

### 2.3 H-Protection

H-protection server predicts 'exchange protection' status of backbone amide proton of each residue in a protein using primary structure of the protein [23]. In other words, amino acid sequence (in FASTA format) is the only input to the program. The webserver is publically available at http://bioinfo.protres.ru/ogp/. The algorithm employs three predictors for successfully completing its functions: expected number of contacts of NHs, the probability of hydrogen bonding of NHs and an artificial predictor. The artificial predictor is developed on the basis of fraction of protected residues for each type of amino acid residues and the calculation could be carried out using an in-house database constructed by authors of the tool. However, the H-Protection webserver predicts only probability of a residue to be protected in the folded states of proteins and it does not provide any clues on the residue-specific stabilities or residue-specific protection factors of NHs.

### 2.4 LRWCO and $\Delta G_{HX}$

Recently, Das et al. [32] have designed a computational method to predict the residue-specific stabilities of proteins using 'long range weighted contact order (LRWCO)'. LRWCO can be calculated from the 3D structural coordinates of proteins. LRWCO denotes average sequence separation for a non-covalent contact of a given residue and is calculated using equation 8. Wherein, 'i' represents the residue subjecting to LRWCO calculation and

$$LRWCO_i = \sum_{|i-j|>12} |i-j| \Big/ \sum L_{ij}$$

$$\text{wherein, } L_{ij} = \begin{cases} 1, \text{when} |i-j| > 12 \\ 0, \text{otherwise} \end{cases} \quad (8)$$

'j' represents amino acids that are sequentially separated by more than 12 residues but structurally close in contact (within 7 Å between backbone nitrogen atoms of residues 'i' and 'j') to the residue 'i' under considerations. $L_{ij}$ represents total number of contacts between the $i^{th}$ and the all $j^{th}$ residues within the distance cut-off of the 7 Å. Using this method, the authors predicted residue-specific $\Delta G_{HX}$ values for 17 NHs of cardiotoxin III, an all β-sheet protein from venom of *Naja naja atra* and the correlation between the predicted and experimental residue-specific $\Delta G_{HX}$ values were found to be 0.82. However, reliability of the computational strategy has not yet been thoroughly validated on estimating residue-specific free energies of NHs of proteins irrespective of their class/secondary structural elements.

## 3. NMR-HD VS. COMPUTATIONAL METHODS

Cytochrome C is an all-alpha protein consisting of 104 residues. Structural stabilities and folding pathways of the protein have been extensively characterized by using NMR-H/D exchange methods at pH 7 and 303 K temperature [33]. The $\Delta G_{HX}$ determined by NMR-H/D exchange methods for various residues of the protein were in range of 3.74-13.99 kcal/mol. Figure 1A depicts the correlation between $\Delta G_{HX}$ determined by experimental methods and $\Delta G_{HX}$ predicted by COREX/BEST program (for 54 NHs that showed protections in the experimental conditions). Similarly, Figure 1B depicts the correlation between $\Delta G_{HX}$ determined by experimental methods and $\Delta G_{HX}$ predicted by LRWCO method (for 32 residues that showed non-zero LRWCO and protections in the

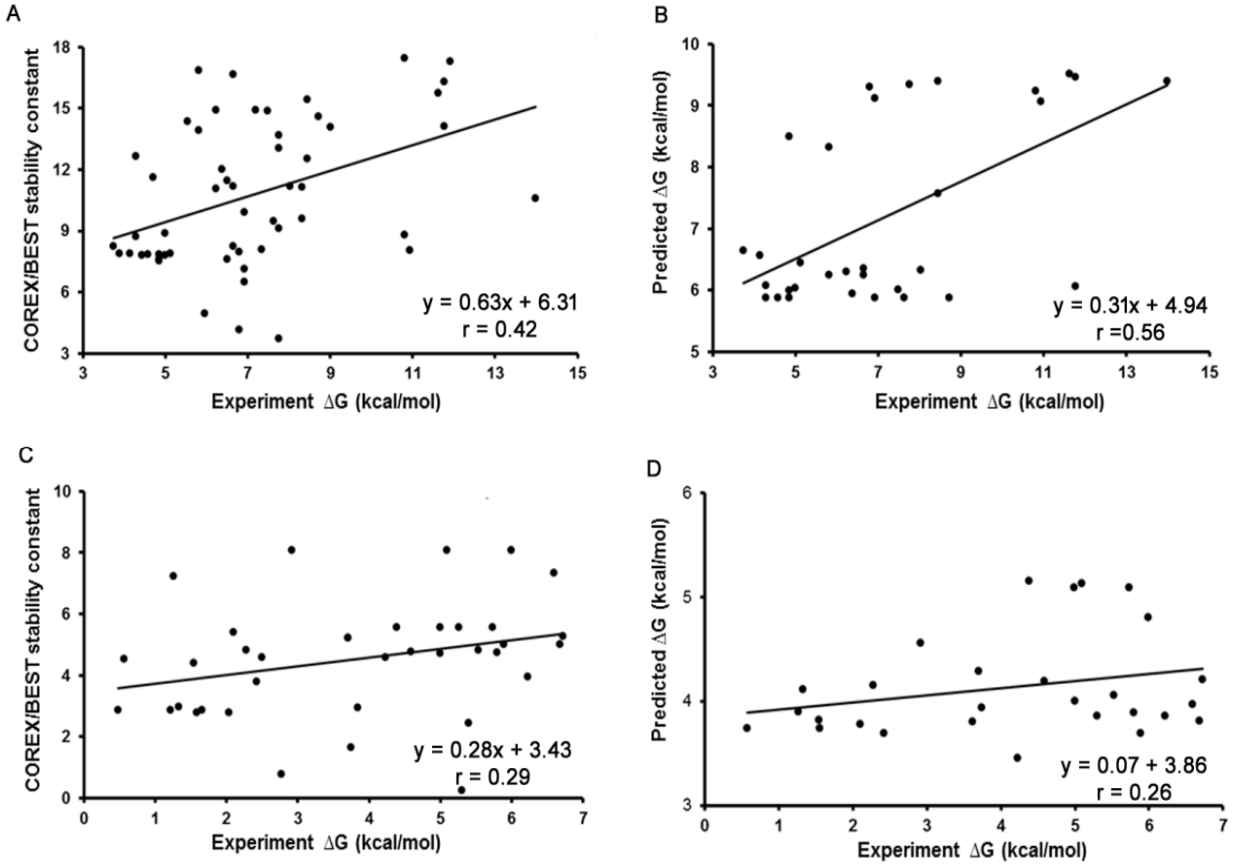


**Figure. 1:** Plot depicting relationship between residue-specific stabilities determined by NMR-HD and computational methods. Figure A & B depict correlations between $\Delta G_{HX}$ for various residues of cytochrome c (pH 7, 303K; PDB ID: 1HRC) determined by NMR-HD with that of COREX/BEST and LRWCO methods, respectively. Figure C & D depict correlations between $\Delta G_{HX}$ for various residues of CTX III (pH 3.2, 298K; PDB ID: 2CRT) determined by NMR-HD method with that of COREX/BEST and LRWCO methods, respectively.

experimental conditions). The former and later methods showed overall correlations of 0.42 and 0.56, respectively. Prediction accuracy and reliability of the COREX/BEST and LRWCO have also been tested using another protein, cardiotoxin III (CTX III). The CTX III is an all-beta sheet and single polypeptide chain consisting of 60 residues [34]. From a quick inspection to the Figure 1C & 1D, the correlation coefficients upon comparing $\Delta G_{HX}$ predicted by COREX/BEST and LRWCO with $\Delta G_{HX}$ estimated by NMR-HD methods for various residues of CTX III were found to be 0.29 and 0.26, respectively. These data analyses clearly suggest that prediction accuracy of the computational methods on predicting residue-specific $\Delta G_{HX}$ of proteins was not quite impressive.

## 4. Future Perspectives

NMR-H/D exchange methods provide unprecedented and wealth information related to protein stabilities, folding intermediates and dynamics at residue level resolution. Since these methods are usually time-taking, cost effective and tedious, developing computational tools will be an excellent alternative at many facets. Interestingly, several computational methods have also been reported in the literature to calculate rate of folding/unfolding of proteins [9, 20, 22, 23, 35, 36] and as well to predict cryptic intermediates and metastable states in the unfolding kinetics of proteins under native conditions [14, 37, 38]. In these contexts, developing a computational tool for predicting residue-specific free energies at defined conditions (such as pH, temperature, denaturants) on the basis of 3D structures of proteins would also be a cardinal task. The success on the task, in turn, will lead to computationally explore the energetic levels of residues that are unfolding and refolding by various mechanisms (global, sub-global and local structural fluctuations) under native conditions of proteins. It is also worthy to mention that NMR-H/D exchange experiments have been carried out on 83 proteins belonging to all types of classes in the time span of nearly 40 years (1985-2015) as reported in the literature and the data may be considered as an excellent platform to figure-out various types of structural and dynamic information that were mostly eluded in the macroscopic experiments of the proteins. In these backgrounds, we trust that the review will be useful for structural biologists to trigger exciting research on developing novel computational tools/strategies for addressing various structural excursions of protein molecules under native conditions.

**ACKNOWLEDGEMENTS**

This research work was supported by the research grant (09/1095/(0004)/2013/EMR-I) from the Council of Scientific & Industrial Research, New Delhi, India.

**REFERENCES**

1. Englander SW, Downer NW, Teitelbaum H. Annu Rev Biochem 1972, 41: 903-924.
2. Woodward C, Simon I, Tüchsen E. Mol Cell Biochem 1982, 48: 135-160.
3. Englander SW, Mayne L. Annu Rev Biophys Biomol Struct 1992, 21: 243-265.
4. Hvidt A, Nielsen SO. Adv Protein Chem 1966, 21: 287-366.
5. Baldwin RL. Proteins 2011, 7: 2021-2026.
6. Bai Y, Milne JS, Mayne L, Englander SW. Proteins 1993, 17: 75-86.
7. Connelly GP, Bai Y, Jeng M, Englander SW. Proteins 1993, 17: 87-92.
8. Molday RS, Englander SW, Kallen RG. Biochemistry 1972, 11: 150-158.
9. Richa T, Sivaraman T. J Pharm Sci and Res 2012a, 4: 1852-1858.
10. Englander SW, SosnickTR , Englander JJ, Mayne L. Curr Opin Struct Biol 1996, 6: 18-23.
11. Bai Y, Milne JS, Mayne L, Englander SW. Proteins 1994, 20: 4-14.
12. Huyghues-Despointes BMP, Pace CN, Englander SW, Scholtz JM (2001) In: Methods Mol Biol. Murphy KM, editor. Protein Structure, stability and folding. New Jersy: Humana Press Inc. pp. 69-92.
13. Huyghues-Despointes BMP, Scholtz JM, Pace CN. Nat Struct Biol 1999, 6: 910-912.
14. Richa T, Sivaraman T. PLoS ONE 2012b, 7: e32465.
15. Clarke J, Fersht AR. Fold Desn 1996, 1: 243-254.
16. Takei J, Chu RA, Bai Y. Proc Natl Acad Sci USA 2000, 97: 10796-10801.
17. Vu ND, Feng H, Bai Y. Biochemistry 2004, 43: 3346-3356.
18. Richa T, Sivaraman T. Int J Res Pharm Sci 2013, 4: 550-562.
19. Englander SW, Mayne L, Krishna MMG. Q Rev Biophys 2007, 40(4): 287-326.
20. Hilser VJ, E Freire. J Mol Biol 1996, 262: 756-772.
21. Compiani M, Capriotti E. Biochemistry 2013, 52: 8601-8624.
22. Tartaglia GG, Cavalli A, Vendruscolo M. Structure 2007, 15: 139-143.
23. Lobanov MY, Suvorina MY, Dovidchenko NV, Sokolovskiy IV, Surin AK, Galzitskaya OV. Bioinformatics 2013, 29(11):1375-81.
24. Hilser VJ, García-Moreno EB, Oas TG, Kapp G, Whitten ST. Chem Rev 2006, 106:1545-1558.
25. Skinner JJ, Lim WK, Bedard S, Black BE and Englander SW. Protein Sci 2012a, 21: 987-995.
26. Skinner JJ, Lim WK, Bedard S, Black BE and Englander SW. Protein Sci 2012b, 21: 996–1005.
27. Vertrees J, Wrabl JO, Hilser VJ. Methods Enzymol 2009, 455: 299–327.
28. Wrabl JO, Hilser VJ. PLoS Comput Biol 2010, 6(3): e1000722.
29. Whitten ST, Kurtz AJ, Pometun MS, Wand AJ, Hilser VJ. Biochemistry 2006, 45: 10163–10174.
30. Whitten ST, García-Moreno EB, Hilser VJ. Proc Natl Acad Sci USA 2005, 102(12): 4282-4287.
31. Best RB, Vendruscolo M. Structure 2006, 14: 97-106.
32. Das P, Richa T, Sivaraman T. Res J Pharm Biol Chem Sci 2014, 5: 1824-1831.
33. Milne JS, Mayne L, Roder H, Wand AJ, Englander SW. Protein Sci 1998, 7: 739-745
34. Sivaraman T. PhD Thesis: Investigation on the unfolding and folding kinetics of snake toxins from Taiwan cobra (*Naja naja atra*). National Tsing Hua University: Taiwan, 1999.
35. Plaxco KW, Simons KT, Baker D. J Mol Biol 1998, 277: 985-994.
36. Gromiha MM and Selvaraj S. J Mol Biol 2001, 310:27-32.
37. Richa T and Sivaraman T. RSC Adv. 2014, 4(68): 36325-36335.
38. Richa T and Sivaraman T. J Pharm Sci and Res 2011, 3(9), 1486-1490.