ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

A Computational Strategy for Predicting Residue-Specific Stabilities of Cardiotoxin III, an all β-Sheet Protein.

Pramila Das, Tambi Richa, and Thirunavukkarasu Sivaraman*

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

ABSTRACT

Residue-specific stabilities of proteins can be studied by using hydrogen/deuterium exchange methods in conjunction with nuclear magnetic resonance techniques (H/D NMR). Under EX2 exchange conditions, residue-specific equilibrium unfolding constants of residues are ratio of intrinsic exchange rate constants (k_{rc}) to extrinsic exchange rate constants (k_{ex}) of backbone amide protons (NHs) of respective amino acids in proteins. While k_{rc} of each NH in proteins can be theoretically calculated on the basis of model peptide studies, k_{ex} values are to be experimentally determined by using H/D NMR under optimized solution conditions. However, the method is technically challenging, expensive, time consuming and require sound knowledge in protein chemistry. In the present study, we demonstrate a computational strategy to predict residue-specific stabilities of proteins on the basis of residue-specific 'long range weighted contact order (LRWCO)', which can be readily calculated by using structural coordinates of proteins. Rationalization on the correspondence between residue-specific LRWCO and residue-specific free energies (determined by H/D NMR) of cardiotoxin III, an all β-sheet protein from venom of *Naja naja atra*, has been described in detail as a model system. In addition, various structural parameters that may be useful to improve the prediction accuracy of the strategy on calculating the residue-specific stabilities of proteins have also been discussed.

Keywords: Free energy of exchange, H/D NMR, long-range contact order, protein folding and residue-specific stability.

Corresponding author

ISSN: 0975-8585

INTRODUCTION

Proteins are the most important structural and functional biomolecules which play vital roles in carrying out many biological activities. The 'protein folding problem' (process by which a polypeptide chain acquires its fully folded, three-dimensional, biologically active native state) has intrigued researches for decades [1-3]. The three-dimensional structures (3D folds), stability and functions of proteins are highly correlated to each other. Thus, estimation of stability of proteins is important for understanding the folding/unfolding mechanisms of proteins [4]. The conformational stability of proteins are measured in terms of free energy of unfolding (ΔG_U), which is defined as the free energy difference between their unfolded and folded states of proteins. The ΔG_U of proteins is estimated as shown in the following equations:

$$N \rightleftharpoons U \qquad (1)$$

$$\Delta G = -RT \ln K \qquad (2)$$

wherein, 'K' is the equilibrium constant between the native (N) and the unfolded (U) state, 'R' is the gas constant and 'T' is the absolute temperature. The ΔG_U of proteins can be estimated at molecular level by using various conventional optical methods (such as fluorescence spectroscopy/circular dichroism/infrared spectroscopy [4, 5] and so on). On the other hand, NMR assisted Hydrogen-deuterium (H/D NMR) exchange methods are powerful of determining the residue-specific free energy of exchange (ΔG_{HX}) of proteins [6].

Hydrogen-deuterium (H/D) exchange method is a chemical reaction in which solvent deuterium exchange with labile protons of proteins in an irreversible manner. In this method, when a protein is dissolved in deuterium oxide (D_2O), the backbone amide proton exchange with the solvent deuterium and the rate of exchange is monitored using nuclear magnetic resonance (NMR) spectroscopy. The H/D exchange of proteins is well described using the following equation [7]:

$$\text{Closed (NH)} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} \text{Open (NH)} \xrightarrow{k_{rc}} \text{Exchanged} \qquad (3)$$

In the above equation, closed (NH) and open (NH) denote folded and unfolded states of proteins, respectively. The k_{op} and k_{cl} are the rate of exchange-competent and exchange-incompetent reactions, respectively and k_{rc} is the intrinsic exchange rate constant of labile protons in proteins. Exchange of backbone amide protons (NHs) takes place only through unfolded state of the protein. Under native conditions, the extrinsic rate of exchange (k_{ex}) can be expressed as shown in equation 4 [8-10]. According to the mathematical expression, the

$$k_{ex} = \frac{k_{op} * k_{rc}}{k_{cl} + k_{rc}} \qquad (4)$$

H/D exchange method has two limiting factors: EX1-exchange and EX2-exchange [4, 11-13]. The EX1-exchange (wherein $k_{rc} >> k_{cl}$) is pH-independent and the k_{ex} of labile protons in proteins is defined under the conditions as shown in the equation 5. Contrary, the EX2-exchange (wherein $k_{rc} << k_{cl}$) is pH-dependent and the k_{ex} of labile protons in proteins under

the conditions is estimated using equation 6. Thus, residue-specific free energy of exchange (ΔG_{HX}) can be readily calculated under the solution conditions favouring EX2-exchange as shown in equation 7.

$$k_{ex} = k_{op} \quad (5)$$

$$k_{ex} = \frac{k_{op}}{k_{cl}} * k_{rc} = K * k_{rc} \quad (6)$$

$$\Delta G_{HX} = -RT \ln (k_{ex}/k_{rc}) = -RT \ln K \quad (7)$$

As per the equation 7, residue-specific free energy of exchange can be calculated at defined experimental conditions provided the residue-specific k_{rc} and k_{ex} are known. Of the two rate constants, the k_{rc} values for amino acids in proteins at defined experimental conditions can be precisely calculated using computational tools such as SPHERE and CIntX [14, 15]. At the same time, k_{ex} values need to be determined by using H/D NMR exchange method, an only available experiment for the purpose. Notwithstanding the potential advantages of the method on estimating residue-specific k_{ex} values, the method has following inherent constraints: residue-specific k_{ex} values can be estimated only for proteins to which 3D structures/backbone assignments determined by multi-dimensional NMR techniques are readily available; moreover, the H/D NMR method is time consuming, expensive and technically challenging [16]. To our best knowledge, CamP and COREX/BEST are the only two computational methods available for predicting the residue-specific ΔG_{HX} values of proteins to date. The former and later methods predict residue-specific protection factor ($P = k_{rc}/k_{ex}$) on the basis of local unfolding and global unfolding models, respectively [17, 18]. However, the prediction accuracy (< 50%) of the methods for determining residue-specific ΔG_{HX} values from structures of proteins as a sole source is not impressive [19, 20]. In the present study, we demonstrated estimation of residue-specific 'long range weighted contact order' (LRWCO) for a given protein on the basis of its three-dimensional (3D) theoretic/experimental structure itself and rationalize the correlations between the calculated 'LRWCO' and experimental ΔG_{HX} of various residues in cardiotoxin III, an all β-sheet protein, as a model system. Merits of the strategy and as well future scopes to improve the strategy on calculating residue-specific ΔG_{HX} values have also been elaborately discussed.

METHODS

The residue-specific LRWCO for given 3D structures of proteins can be calculated by means of equation 8 as shown below, herein. In the equation, 'i' represents a residue, which is subjected to LRWCO calculation and '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) to the residue 'i' under considerations. Thus, the numerator of the equation represents sum of sequence separation between 'i' and each 'j' residues for which backbone nitrogen atoms are within 7 Å from the backbone nitrogen atom of the residue 'i'. The denominator of the equation (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. The distance between the backbone nitrogen atoms of 'i' and 'j' residues are estimated using equation 9; wherein 'D' denotes distance in angstrom; X_i, Y_i and Z_i are the atomic coordinates of backbone amide nitrogen of the i^{th} residues and X_j, Y_j and Z_j are the atomic coordinates of

backbone amide nitrogen of the j^{th} residues. Hence, the LRWCO reveals an average sequence separation for a non-covalent contact of 'i' residue, which is under the subject of the calculations. It should also be mentioned that various structure/sequence-derived descriptors (such as contact order, long range order, absolute contact order, long range contact order) have been developed by various eminent research groups around the world and the descriptors have also been successfully used to correlate with folding rates/stabilities of proteins at molecular level [21-23].

$$LRWCO_i = \sum_{|i-j|>12} |i-j| \Big/ \sum L_{ij} \qquad \text{wherein, } L_{ij} = \begin{cases} 1, \text{when} \, |i-j| > 12 \\ 0, \text{otherwise} \end{cases} \tag{8}$$

$$D = \sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2 + (Z_i - Z_j)^2} \tag{9}$$

NMR-derived 3D structures of cardiotoxin III (PDB ID: 2CRT) was retrieved from Protein Data Bank (PDB) and used for calculating residue-specific LRWCO for each residue of the protein. The residue-specific ΔG_{HX} values of various backbone NHs in the cardiotoxin III (CTX III) estimated at pH 3.2, 298K in low ionic strength solution conditions have been reported in the literature [24,25] and the reported ΔG_{HX} values in kcal/mol were used for comparative analyses in the present study throughout the text unless stated otherwise.

RESULTS AND DISCUSSION

Delineating stability of proteins at residue level is essential to identify residues that are essential for structural integrities and functional activities of proteins (24, 26). The H/D NMR exchange is the only experimental method available to date to determine residue-specific stability (refer to equation 7) of proteins under native conditions (8, 19, 20, 24-29). However, the method poses many experimental limitations and importantly the method is not suitable for proteins, which do not have NMR structures and are also prone to get degradation/aggregation at the exchange conditions. In these connections, developing a computational method to predict residue-specific stability of proteins will be an excellent alternative to the H/D NMR method. In the present study, we have explained the relationship between residue-specific LRWCO and residue-specific ΔG_{HX} of CTX III. The CTX III is a monomeric, an all β-sheet and single polypeptide chain composed of 60 amino acids. The protein offers 55 NHs as probes (as the protein has five *trans*-proline residues) for mapping dynamics of the whole structure by using H/D NMR exchange method. Interestingly, residue-specific ΔG_{HX} values for 41 residues in the protein have been reported at 298 K in pH 3.6 (24, 27-29). The residue-specific free energies of exchange for the 41 residues of the protein were found to be in the range of 0.5 - 6.7 kcal/mol.

The residue-specific 'long range weighted contact order' (LRWCO) values for 55 NHs of the protein were calculated using equation 8 (refer to methods) and essential steps involved in the calculations are enumerated in Fig. 1. The calculated LRWCO values were overspread from 0 to 56.5 for the 55 residues of the protein. Of the 55 NHs, 17 NHs depicted LRWCO of zero indicating that the 17 residues did not establish strong long range contact network interactions in the protein structure. Moreover, only 28 NHs (Lys2, Cys3, Asn4, Lys5, Leu6, Cys14, Lys18, Leu20, Cys21, Tyr22, Lys23, Met24, Phe25, Met26, Val27, Ala28, Arg36, Ile39, Leu48, Val49, Tyr51, Val52, Cys53, Cys54, Asn55, Asp57, Arg58 and

Asn60) of the protein showed both non-zero LRWCO and ΔG_{HX} (determined from H/D NMR) values. However, correlation between the LRWCO and ΔG_{HX} values for all the 28 residues was not quite impressive. Stringent analysis carried-out on the dataset suggested that 11 of the 28 residues did not show even a fair correlation between their respective LRWCO and ΔG_{HX} values. Strikingly, 17 (Lys2, Cys3, Asn4, Lys5, Leu6, Lys18, Leu20, Phe25, Ala28, Arg36, Leu48, Val49, Cys53, Asn55, Asp57, Arg58 and Asn60) of the 28 residues showed excellent correlation between the calculated LRWCO and experimentally determined ΔG_{HX} values (Fig. 2) and the positive correlation coefficient was found to be 0.82. Using the fitted parameters, the ΔG_{HX} values for the 17 residues were predicted and compared with respective ΔG_{HX} values determined by H/D NMR methods for the residues (Table 1). From a quick inspection to the table, one can easily understand that the ΔG_{HX} values for the 17 residues estimated by using theoretic and experimental methods are in good agreements. The residue-specific free energies of exchange (ΔG_{HX}) for the 17 residue determined by the H/D NMR method were found to be in the range of 0.6 - 6.0 kcal/mol. The mean ΔG_{HX} values for the 17 residues as estimated by the H/D NMR and present theoretic method were 3.2 ± 2.8 kcal/mol and 3.2 ± 1.9 kcal/mol, respectively (Table 1).

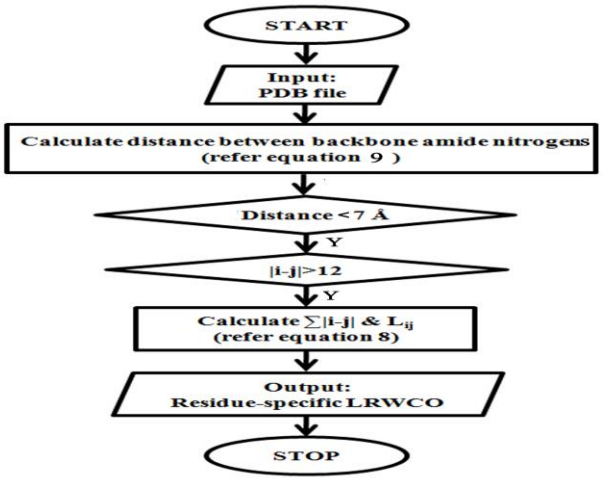
Table 1: Comparison of ΔG_{HX} values for 17 residues in CTX III estimated by the H/D NMR exchange method and theoretic method described in the present study.

Sl.No.	Residue	ΔG_{HX} (H/D NMR) kcal/mol	ΔG_{HX} (predicted) kcal/mol
1	Lys2	2.9	3.8
2	Cys3	5.1	5.3
3	Asn4	6.0	4.4
4	Lys5	4.6	2.9
5	Leu6	1.3	2.7
6	Lys18	2.1	1.8
7	Leu20	1.3	2.1
8	Phe25	3.7	2.2
9	Ala28	1.6	1.7
10	Arg36	2.4	1.6
11	Leu48	0.6	1.7
12	Val49	1.5	1.9
13	Cys53	2.3	2.8
14	Asn55	3.7	3.1
15	Asp57	5.0	5.2
16	Arg58	5.7	5.2
17	Asn60	4.4	5.3

In general, one can expect that residues having higher ΔG_{HX} should probably have higher LRWCO. Contrary to the expectation, the 11 (Cys14, Cys21, Tyr22, Lys23, Met24, Met26, Val27, Ile39, Tyr51, Val52 and Cys54) residues that exhibited poor correlation between their respective ΔG_{HX} and LRWCO depicted higher ΔG_{HX} values and extremely lower LRWCO values *vis-a-vis* other 17 residues that showed sensible linear correlation between their respective ΔG_{HX} and LRWCO (Fig. 2). We have recently shown that there were two possible foldons in the unfolding kinetics of the CTX III under native conditions. Of the 2 foldons, backbone amide protons of residues present in the most stable foldon were shown to have extraordinary protections preventing H/D exchange of the residues through structural events of global unfolding mechanism (25, personal communication). Moreover,

the foldon was shown to be composed of residues such as Cys21, Tyr22, Lys23, Met24, Phe25, Met26, Val27, Val34, Lys35, Ile39, Val52 and Cys54. Strikingly, 10 out of 11 residues mentioned above, except Cys14, were structural blocks of the foldon. These results imply that the LRWCO parameter may not be useful to predict residue-specific stabilities of residues that are not exchanging their NHs with solvent deuterium either through local or global unfolding mechanisms, the protein may adopt under native conditions. In these backgrounds, the prediction accuracy on estimating residue-specific stabilities of proteins can be presumably improved by taking into consideration of residue-specific folding/unfolding rate constants in addition to the LRWCO parameter demonstrated in the present study. Right now, the computational strategy is being applied to estimate residue-specific stabilities of 20 proteins which are globular, single domain and belonging to all classes (α, β, α+β & α/β proteins) in order to stringently validate the reliability of the strategy and as well to generalize the usefulness of the computational strategy on estimating residue-specific stabilities of proteins.

Figure 1: Flowchart depicting the key-steps used in estimation of residue-specific LRWCO of proteins.
The algorithm was used to develop an in-house computational tool for calculating residue-specific LRWCO of proteins.



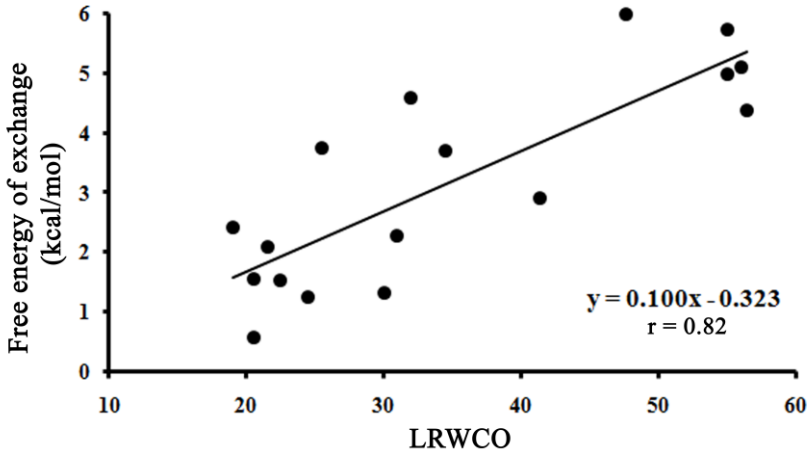


Figure 2: Plot depicting the relationship between LRWCO and ΔG_{HX} of CTX III. The solid line through the data points is the fit of the data to a linear equation.

CONCLUSION

Using CTX III as a model protein, the relationships between residue-specific LRWCO and residue-specific stabilities have been demonstrated. The computational method is prerequisite of an input only: three-dimensional structures of proteins for which residue-specific stabilities to be calculated. Interestingly, proteins having either experimental or theoretic 3D structures can be used as an input in the computational approach. Though the LRWCO parameter are reliable to estimate residue-specific stabilities of proteins, we strongly believe that the prediction efficiency can be greatly improved by combining the LRWCO and residue-specific folding/unfolding rates in the computational strategy in near future.

ACKNOWLEDGMENTS

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

REFERENCES

[1] Dobson CM, Karplus M . Curr Opin Struct Bio 1999; 9: 99-101.
[2] Dill KA, MacCallum JL. Science 2013; 338: 1042-1045.
[3] Dagget V, Fresht A. Nat Rev Mol Cell Biol 2003; 4: 497-502.
[4] Huyghues-Despointes BMP, Pace CN, Englander SW, Scholtz JM. Protein Structure, stability and folding. Humana Press Inc., New Jersy, 2001, pp. 69–92.
[5] 5 Plaxco KW, Dobson CM. Curr Opin Struct Biol 1996; 6: 630–636.
[6] Raschke TM, Marqusee S. Curr Opin Biol 1998; 9: 80-86.
[7] Hvidt A, Nielsen SO. Adv Protein Chem 1966; 21: 287–366.
[8] Englander SW, Downer NW, Teitelbaum H. Annu Rev Biochem 1972; 41: 903-924.
[9] Bai Y, Milne JS, Mayne L, Englander SW. Proteins 1994; 20: 4-14.
[10] Richa T, Sivaraman T Int J Res Pharm Sci 2013; 4: 550-562.
[11] Ferraro DM, Lazo ND, Robertson AD. Biochemistry 2004; 43: 587–594.
[12] Englander SW, Sosnick TR, Englander JJ, Mayne L. Curr Opin Struct Biol 1996; 6: 18-23.
[13] Sivaraman T, Arrington CB, Robertson AD. Nat Struct Biol. 2001; 8:331–333.
[14] Bai Y, Milne JS , Mayne L, and Englander SW. Proteins 1993; 17: 75–86.
[15] Richa T, Sivaraman T. J Pharm Sci & Res 4 2012; 4: 1852 – 1858.
[16] Englander SW, Mayne L, Krishna MMG. Q Rev Biophys 2007; 40: 287-326.
[17] Tartaglia GG, Cavalli A, Vendruscolo M. Structure 2007; 15: 139–143.
[18] Hilser VJ, Freire E. J Mol Biol 1996; 262: 756–772.
[19] Skinner JJ , Woon K. Lim, Sabrina Bedard, Ben E. Black, Englander SW. Protein sci 2012; 21: 987 – 995.
[20] Skinner JJ , Woon K. Lim, Sabrina Bedard, Ben E. Black, Englander SW. Protein sci 2012; 21: 996—105.
[21] Gromiha MM, Selvaraj S. Curr Bioinfo 2008; 3:1-9.
[22] Plaxco KW, Simons KT, Baker D. J Mol Biol 1998; 277: 985-994.
[23] Gromiha MM, Selvaraj S. Prog Biophys Mol Biol 2004; 8: 235-77.
[24] Sivaraman T. Ph.D Thesis, National Tsing Hua University: Taiwan, 1999.
[25] Richa T, Sivaraman T. PLoS ONE 2012; 7(3): e32465.

[26] Sivaraman T, Kumar TKS, Chang DK, Lin WY, Yu C. J Biol Chem 1998, 273:10181-10189.
[27] Sivaraman T, Kumar TKS, Hung KW, Yu C. Biochemistry 2000, 39(30):8705-8710.
[28] Sivaraman T, Kumar TKS, Tu YT, Peng HJ, Yu C. Arch Biochem Biophys 1999, 363:107-115.
[29] Sivaraman T, Kumar TKS, Yu C. Biochemistry 1999, 38:9899-9905.