Tambi Richa et al /J. Pharm. Sci. & Res. Vol.3(9), 2011,1486-1490

ISSN:0975-1459

Journal of Pharmaceutical Sciences and Research

www.jpsr.pharmainfo.in

META: A computational tool for predicting metastable states in the folding pathways of proteins

Tambi Richa and Thirunavukkarasu Sivaraman*

School of Chemical and Biotechnology, SASTRA University, Thanjavur-613401,Tamil Nadu, India.

Email address: sivaram@scbt.sastra.edu

Abstract:

The thermodynamic stabilities (ΔG, free energy change) of proteins are being measured by traditional denaturation methods (ΔG_U, free energy of unfolding) and native hydrogen-deuterium (H/D) exchange method (ΔG_{HX}, free energy of exchange), in general. Estimation of an accurate ΔG for proteins at ambient conditions is indispensable to unambiguously address the folding events of the proteins, which may pave the way of designing *de novo* therapeutic peptides with defined functions. However, the ΔG_U and the ΔG_{HX} of proteins are not in good agreement, in general. Herein, we have developed a META program, which accounts the discrepancies between the ΔG_U and the ΔG_{HX} of proteins by predicting the possible existence of higher energy metastable states in the folding kinetics of proteins under native conditions by systematically analysing residue-specific free energies of the proteins. The robustness of the program on analyzing the conformational stabilities of proteins has been validated using experimental data available in the literature and the implication of the program on medicinal chemistry is also discussed. The program is publicly available at *http://feat.sastra.edu/meta.html.*

Introduction:

Each protein adopts a specific well-defined three-dimensional (3D) structure with unique stability, which is important for its biological activities. The relationships between the conformations of such proteins and their stabilities have intrigued researchers for many decades [1]. The ΔG_U of proteins is calculated by fitting the data of the unfolded population (U) of proteins with respect to denaturant concentration or temperature, to an appropriate two-state model equation [2, 3]. The ΔG_{HX} of proteins is determined using hydrogen-deuterium (H/D) exchange method in conjunction with NMR technique [4]. Accurate estimation of ΔG_U and ΔG_{HX} for proteins at ambient conditions is indispensable to unambiguously address the thermodynamic and kinetic events of the proteins, respectively [5 - 7].

Under similar experimental conditions, it is expected that ΔG_U and ΔG_{HX} of protein should be same. But many proteins do not follow this tradition, for which the probable reason could be presence of metastable states in the folding pathways of proteins [8, 9]. Though the discrepancies in a few proteins have been experimentally addressed [10, 11], the problem left unaddressed in most cases. Moreover, unique computational programs by which such discrepancies can be systematically addressed have not been developed to date. In the present study, we describe a program (META), which addresses the discrepancies between the ΔG_{HX} and ΔG_U of proteins on predicting the possible existences of higher energy metastable states in the folding pathways of proteins under native conditions. The uniqueness of the META to address the discrepancies between the ΔG_{HX} and ΔG_U of proteins is validated using experimental data available in the literature and its applications on exploring energy landscape of proteins and on designing therapeutic peptides are also brought into fore.

Methods:

2.1 META algorithm

META algorithm has been implemented using PERL scripting language [12]. The program requires only three inputs from the user: PDB file of proteins, residue specific ΔG_{HX} and free energy of unfolding, ΔG_U of the proteins. In outline, the program has five stages for each complete cycle. In the first stage, the ΔG_{HX} of proteins is calculated by averaging out to four largest residue-specific ΔG_{HX} of the proteins [13]. In the second stage,

the program clusters the residues of proteins based on the residue-specific free energies with the tolerance limit of $\Delta G_{U,}$ when the ΔG_U is greater than ΔG_{HX} (the default tolerance limit is set to be 0.4 kcal/mol). Third, the program generates all possible residue pairs for the residues in the cluster and calculates distance in angstrom between the backbone nitrogen atoms of the two residues in each pair. Fourth, the FEAT generates a matrix in which each pair is assigned either with the value of 1 or 0: the value of 1 is given to a pair when the distance between the two residues is within 7 Å otherwise 0 is given (refer results and discussion). Fifth, the program groups the residue-pairs such that any pair in a group must have at least another pair having a residue common to each other. The program avoids redundancy in grouping the residue-pairs on the basis of 'effective contact order' [14] and generates atomic coordinate files in PDB format for residues in each group of the cluster. The program is fully automated and user-friendly in its functions and details about data formats can also be obtained from the 'help' menu of the program at *http://feat.sastra.edu/meta.html.*

Results and Discussions:

Estimation of accurate free energy change of proteins is important to understand the relationship between three-dimensional (3D) conformations and stabilities of proteins as each protein adopts a specific, well-defined 3D structure, which is important for its biological activities [15]. Moreover, free energy analysis provides clues on understanding the mechanism of unfolding of proteins (two-state/multi-state processes) and on analysing the 3D structural architectures (domains organization) of proteins [16]. Thus, it is obvious that estimation of an accurate ΔG for proteins at ambient conditions is indispensable to unambiguously address the thermodynamic and kinetic events of proteins. If there is a remarkable difference between the ΔG_{HX} and ΔG_U of a protein, the origins of the discrepancies may probably stem from many facets [17-19]. When the experimental conditions of both optical and H/D exchange methods are matching well to each other, the discrepancies arising between the ΔG_U and the ΔG_{HX} of proteins can be well rationalized. In these contexts, we have developed the META program to reconcile the discrepancy between the ΔG_{HX} and the ΔG_U by predicting the possible existence of higher energy denatured states or metastable states in the folding pathways of proteins on the basis of the distribution of residue-specific free energies in proteins. It has been shown that any relaxation of the denatured protein that occurs more slowly than refolding should give rise to higher energy metastable denatured state in the exchange experiment [20, 21]. Perhaps this relaxation process involves diffusion of a relatively compact set of conformers, crossing from the transition state, to the broader distribution of conformers that are characteristic of the denatured ground state.

In this context, residues for which $\Delta G_{HX} > \Delta G_U$ may presumably constitute the higher energy metastable denatured state of the protein [22, 23]. The META program clusters the residues of proteins based on the residue-specific free energies with the tolerance limit of ΔG_U. Existence of metastable states in the folding kinetics of proteins such as cytochrome C and OMTKY3 have been reported based on the NHs exchange studies of the proteins under native conditions [20, 24, 25]. However, the experimental methods require sound experimental skills and are tedious, time consuming and expensive. In this context, the

META is an alternative tool to those experimental methods and the outline of the program is illustrated in Fig.1. As described in the method section, after generating all possible residue pairs for the residues in the cluster, the META generates a matrix in which each pair is assigned either with the value of 1 or 0: the value of 1 (filled circle) is given to a pair when the distance between the two residues is within 7 Å otherwise 0 (cross symbol) is given and then, the program groups the residue-pairs such that any pair in a group must have at least another pair having a residue common to each other (Fig. 2).

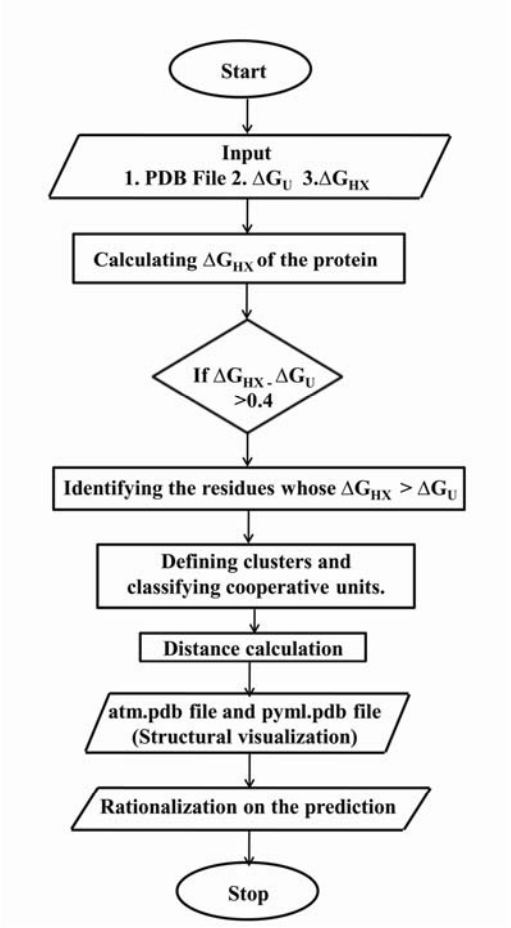

Figure 1: The flowchart depicts key-steps involved in the META algorithm on predicting higher energy metastable states in the folding kinetics of proteins.

We are herein demonstrating the possible existence of metastable states of cardiotoxin III (CTX III, a three-finger toxin isolated from Taiwan cobra *Naja naja atra; PDB ID: 2CRT*) using the META program based on the data available in the literatures on the folding kinetics of the protein [26-28].

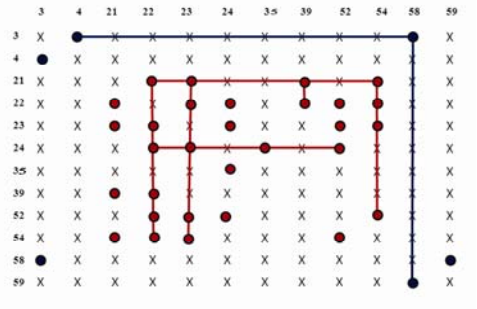

Figure 2: Matrix representing the overall strategies of META on predicting high-energy cooperative states in CTX III under native conditions. The program defines a cluster containing 12 residues (Cys3, Asn4, Cys21, Tyr22, Lys23, Met24, Lys35, Ile39, Val52, Cys54, Arg58 & Cys59) of the protein for which $\Delta G_{HX} > \Delta G_U$ and generates network contacts among them. The filled circles and cross symbols represent residue-pairs for which the distance between the backbone nitrogen atoms of the two residues is < 7 Å and > 7 Å, respectively. The program scans all the residue-pairs in each row and column of the matrix and groups them on the basis of effective contact order by avoiding redundancy in the residue-pairs. The 12 residues in CTX III are further defined into two groups by the META. The residues in the group I (Cys3, Asn4, Arg58 & Cys59) and group II (Cys21, Tyr22, Lys23, Met24, Lys35, Ile39, Val52 & Cys54) are connected by blue and red lines, respectively.

The ΔG_U and ΔG_{HX} of CTX III have been reported to be 4.5 and 6.6 kcal/mol, respectively [29]. The META program clusters 12 residues (Cys3, Asn4, Cys21, Tyr22, Lys23, Met24, Lys35, Ile39, Val52, Cys54, Arg58 and Cys59) for which values of ΔG_{HX} are higher than ΔG_U. The

residues in the cluster are analyzed on the basis of 'effective contact order' [14] and two distinct groups (non-cooperative units) are found: group I contains residues such as Cys3, Asn4, Arg58 & Cys59 and group II contains residues such as Cys21, Tyr22, Lys23, Met24, Lys35, Ile39, Val52 & Cys54 (Fig. 2). Interestingly, all the 12 residues are located in the globular head of triple stranded β-sheet domain of the protein (Fig. 3), which is highly stabilized by a network of hydrogen bonds and four cross-linked disulfide bonds.

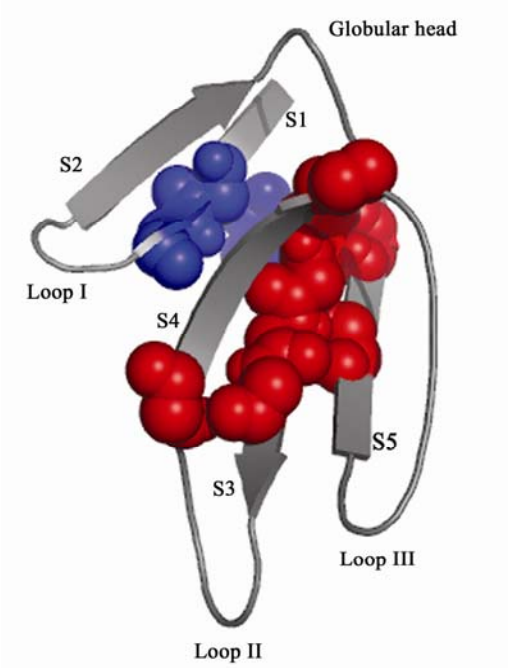

Figure 3: PyMol representation of the structure of CTX III showing overall backbone folding. The five β-strands (S1-S5), three loops and a globular head in the structure are denoted. The blue and red spheres represent residues in the group I and group II, respectively.

The residues in the group II are located in the triple stranded domain of the protein. The residues in the group I are located in the double stranded domain and C-terminal region, which is in the close proximity to the double-stranded domain of the protein. The kinetic folding pathways of CTX III have been characterized to proceed through a hydrophobic cluster due to coalescence of non-polar residues such as Ile39, Val49, Tyr51, Val52, Cys53 & Cys54 [29]. Moreover, it has also been demonstrated that the triple-stranded β-sheet segment of the protein was persistently found in the intermediate states identified along the alcohol- and acid-induced unfolding pathways of CTX III [27, 30]. To this extent, the predictions of META on the possible existence of metastable states of CTX III under native conditions are consistent with those data reported from equilibrium and kinetic studies of the protein. Moreover, the program is not only predicting the possible existence of metastable states in proteins, it also classifies them either as cooperative or non-cooperative units. In these backgrounds, we strongly believe that the program is powerful to provide hints for the possible existence of metastable denatured states in the folding kinetics of proteins under native conditions on basis of their ΔG_U and ΔG_{HX} of the proteins.

Concluding remarks:

The META program predicts possible existence of cooperative and/or non-coopperative metastable units in the folding kinetics of proteins on the basis of ΔG_U, ΔG_{HX} and effective contact orders. The program requires only three inputs to achieve the task: PDB file of the proteins, free energy unfolding, ΔG_U and free energy of exchange, ΔG_{HX}. The program is fully automated and user-friendly in its functions and it is available at *http://feat.sastra.edu/meta.html*. The robustness of the program has been well validated by predicting metastable states of CTX III for which experimental data available on its stability and folding kinetics. The program can also be effectively used to predict metastable states that may exist in the folding kinetics of any proteins

and the information may be useful to explore the energy landscapes of the proteins and to designing *de novo* therapeutic peptides in pharmaceutical sciences [31, 32]. To our best knowledge, META is a unique tool of this kind for predicting metastable states in the folding pathways of proteins.

Acknowledgements:

This work is partly supported by TRR research grants (TRR/13/Aug/2009), SASTRA University, Tamil Nadu, India.

References:

[1] Dobson, C. M., Karplus, M., *Curr. Opin. Struc. Biol.* 1999, 9, 99 – 101.

[2] Pace, C. N., *Methods Enzymol.* 1986, 131, 266 – 280.

[3] Pace, C. N., Shaw, K. L., *Proteins: Struct. Funct. Genet.* 2000, 4, 1 – 7.

[4] Woodward, C.K., *Curr. Opin. Struct. Biol.* 1994, 4, 112 – 116.

[5] Chamberlain, A. K., Marqusee, S., *Structure.* 1997, *5*, 859 – 863.

[6] Bai, Y., Milne, J. S., Mayne, L., Englander, S. W., *Proteins: Struct. Funct. Genet.*1994, 20, 4 – 14.

[7] Pace, C. N., Scholtz, C. M., in: Creighton, T. E. (Ed.), *Protein Structure – A Practical Approach*, Oxford University Press Inc., 1997, pp. 300 – 321.

[8] Baker, D., *Nat. Struct. Biol.* 1998, 5(12), 1021 – 1024.

[9] Englander, S. W., *Annu. Rev. Biophys. Biomol. Struct.* 2000, 29, 213 – 38.

[10] Huyghues-Despointes, B. M. P., Scholtz, J. M., Pace, C. N. *Nat. Struct. Biol.* 1999, 6, 910 – 912.

[11] Mayne, L., Englander, S. W., *Protein Sci.* 2000, 9, 1873 – 1877.

[12] Wall, L., Christiansen, T., Orwant, J., *Programming Perl*. O'Reilly Media Inc., Sebastopol 2000.

[13] Huyghues-Despointes, B. M. P., Pace, C. N., Englander, S. W., Scholtz, J. M.in: Murphy, K. M. (Ed.), *Protein Structure, stability and folding*, Humana Press Inc., New Jersy, 2001, pp. 69 – 92.

[14] Weikl, T. R., Dill, K. A., *J. Mol. Biol.* 2003, 329, 585 – 598.

[15] Creighton, T. E., *Biochem. J.* 1990, 270, 1 – 16.

[16] Pace, C. N., Shirley, B. A., McNutt, M., Gajiwala, K., *FASEB J.* 1996, 76, 75 – 83.

[17] Robertson, A. D., Murphy, K.P., *Chem Rev.* 1997, 97, 1251 – 67.

[18] Prakash, V., Loucheux, C., Scheufele, S., Gorbunoff, M. J., Timasheff, S. N. *Arch Biochem Biophys.* 1981, 210, 455 – 464.

[19] Reimer, U., Scherer, G., Drewello, M., Kruber, S., Schutkowski, M., Fischer, G., *J. Mol. Biol.* 1998, 279, 449 – 460.

[20] Arrington, C. B., Teesch, L. M., Robertson, A. D., *J Mol Biol.* 1999, 285, 1265 – 1275.

[21] Arrington, C. B., Robertson, A., *J. Mol. Biol.* 2000, 300, 221 – 232.

[22] Woodward, C., *Trends. Biochem. Sci.* 1993, 18(10), 359 – 360.

[23] Li, R., Woodward, C., *Protein Sci.* 1999, 8, 1571–1591.

[24] Sosnick, T. R., Mayne, L., Englander, S. W., *Proteins: Struct. Funct. Bionf.* 1996, 24(4), 413 – 26.

[25] Arrington, C. B., Robertson, A. D., *Biochemistry.* 1997, 36(39), 8686 – 8691.

[26] Sivaraman, T., Kumar, T. K. S., Hung, K. W., Yu, C., *Biochemistry.* 2000, 39, 8705 – 8710.

[27] Sivaraman, T., Kumar, T. K. S., Jayaraman, G., Han, C. C., Yu, C., *Biochem J.* 1997, 321, 457 – 464.

[28] Sivaraman, T., Kumar, T. K. S., Yu, C., *Biochemistry.* 1999, *38*, 9899 – 9905.

[29] Sivaraman, T., Kumar, T. K. S., Chang, D. K., Lin, W. Y., Yu, C., *J Biol Chem.* 1998, 273, 10181 – 10189.

[30] Kumar, T. K. S., Jayaraman, G., Lee, C. S., Sivaraman, T., Lin, W. Y., Yu, C., *Biochem. Biophys. Res. Commun.* 1995, 207, 536 – 543.

[31] Sivakumar, D., Aashis, R., Sivaraman, T., *J. Pharm. Sci. & Res.* 2011, 3(4), 1141 – 1145.

[32] Jayachandra, K., Sivaraman, T., *J. Pharm. Sci. & Res.* 2011, 3(7), 1360 – 1363.