Slow exchange core absolute correlation biology essay

Science, Biology

S. Manuel John1Email: sjohnmanuel@gmail. comT. Sivaraman1*Corresponding authorEmail: sivaram@scbt. sastra. edu1Department of Bioinformatics, School of Chemical and Biotechnology, SASTRA University

Abstract:

Background:

Identifying key-residues that initiate folding of proteins are indispensable to address ‘ Levinthal paradox’ of protein folding problem. The kinetic refolding and equilibrium unfolding of proteins can be monitored at residue level resolutions using hydrogen-deuterium (H/D) exchange methods in conjunction with NMR techniques. Using the methods, residue-specific refolding rate constants (kf) and exchange rate constants (kex) could be estimated and compared, for quite a few proteins. However, there had been a heated debate on analyzing and correlating the data obtained from the studies. According to the " first-in – last-out" hypothesis, the residue that folds first in the kinetic refolding of a protein will be the last to get exchanged in the unfolding kinetics of the protein under native conditions.

Results:

In the present study, four new strategies namely Absolute Correlation Percentage (ACP), Normalized Correlation Percentage (NCP), Absolute Global Correlation Percentage (AGCP) and Normalized Global Correlation Percentage (NGCP) have been formulated and successfully applied to validate the hypothesis in a bias-free manner. The values of NCP and NGCP are arbitrarily divided into four categories: poor ≤ 25%; 26% ≤ fair ≤ 50%;

51% ≤ good ≤ 75%; 76% ≤ excellent ≤ 100%. Combined analysis of the outputs from the NCP and NGCP were reliable and robust either to accept or reject the hypothesis. A dataset comprising of 11 proteins, for which values of kex and kf were available, was used to test the hypothesis using the strategies developed in the present studies. Of the 11 proteins, H/D data of apomyoglobin, hisactophilin and RNase H* were found to be ‘ favorable’ to the hypothesis; H/D data of scFV-light chain and scFV-heavy chain did not support the hypothesis. Interestingly, these results were in accordance with the experimental findings reported in the literature. However, predicted results for proteins such as CTX III, HEWL and RNase T1 were not in good agreement with conclusions drawn from manual methods and rationale for the discrepancy is discussed in detail.

Availability:

http://sblab. sastra. edu/

Keywords:

H/D exchange, Slow exchange core, Absolute correlation, Proline isomerisation

Background:

Introduction:

Proteins are macromolecules whose structure determines its function. For example, nervous system is a network of cells and the functions of each of these cells can be determined only from the properties and interactions of their proteins. Some proteins of the system might involve in transport of ions

and some others perform enzymatic functions and so on. So the diverse and specific functions of proteins depend on their shape, which in turn, depends on how efficient it had been folded []. In protein folding, the disordered polypeptide chain folds to a highly ordered native structure. []. The aminoacid sequence of the protein determines its three dimensional structure.[]. To attain the native state, if the protein searches the stable conformation in a random manner, then it would take so much time to fold. []. But most of the proteins fold within seconds []. This means that, the early stages in protein folding limits the number of conformations and thereby specifying the pathway through which it has to attain the native state.[]. Protein folding pathways were of great interest and it has been studied for various proteins. Some of the models pertaining to the protein folding pathways were molten globules, subdomain model, nucleation/growth model, framework model, hydrophobic collapse model, jigsaw puzzle model and so on. Molten globule model postulates that early stages of folding are characterized by compact intermediates with considerable formation of secondary structure elements. Subdomain model proposed that formation of stable structure involves docking of small subdomains of secondary and tertiary structure that had formed initially. [refer rnase t1 kf] According to nucleation/growth model, the rate-determining step of protein folding is the formation of smaller structural units (nucleus). In framework model, secondary structures will be formed initially during folding. [Baldwin] According to hydrophobic collapse model, hydrophobic effect is the driving force of folding.[dill 1985]. Jigsaw puzzle model denies the possibility of a unique and directed folding pathway and so it states that each protein may

follow different route to attain its native state, which is similar to solving a jigsaw puzzle in multiple ways. [harrision and dubin, 1985]One of the approaches to solve the protein folding problem is by measuring the conformational stability. The conformational stability is defined as the free energy change of unfolding reaction under ambient conditions.[book] It can be evaluated using the following relationship: ΔGu = − RT ln Ku(1)Where, ΔGu- Free energy of unfoldingR - Gas constantT - TemperatureKu - Equilibrium constant for the unfolding reactionBut there exists a problem in detecting the unfolded conformations, because protein exists predominantly in folded conformation and in addition to that protein may rarely adopt partially unfolded or completed unfolded higher energy conformation, so it is difficult to measure Ku under native conformation. One of the methods to address this problem is native state hydrogen-deuterium (H/D) exchange experiment that uses Nuclear Magnetic Resonance (NMR) technique which is used to identify the residues that exchange from globally unfolded state of protein.[book]In this experiment, the proteins that are exposed may readily exchanges with the solvent deuterium. Juxtaposed to this, an amide that is protected or buried can exchange only when the protein undergoes a local fluctuations such as the breaking of only a few hydrogen bonds or by a larger global unfolding event. So the rate of exchange for these residues is much slower than their counterpart.

kop

krcThe two-step model for hydrogen-deuterium exchange is given by,

kcl

kcl(N– H)cl (N– H)op N– D

Where, kop and kcl are the rate constants for structural opening and closing respectively. (NH)op is the open form that exchanges with solvent at the intrinsic rate constant krc, where rc stands for random coil. krc is the intrinsic rate constant for exchange when the protein is in completely unfolded state. The rate constant of exchange (kex) is given by, kex = kop krc / (kop+ kcl + krc)(2)If the protein predominantly present in the open form i. e. (kop >> kcl), then kex = krc. If the protein is especially in the closed form i. e. (kcl >> kop), then two mechanisms (EX1 and EX2) could occur depending upon the conditions, in a distinctive manner or in a combination. So, kex = kop krc / (kcl + krc)(3)(kop is ignored because protein assumes folded state predominantly).

EX1 mechanism:

According to this mechanism, Kcl << krckex = kop krc / krc (kcl is negligible) (4)So, kex = kop.(5)

EX2 mechanism:

According to this mechanism, Kcl >> krc. kex = kop krc / kcl (krc is negligible)(6)kex = K. krc (since K= kop / kcl)(7)K= kex / krc. (8)Thus under EX2 conditions, the free energy of structural opening is given by: ΔGHX = – RT ln (kex / krc)(9)Where, ΔGHX- Residue specific free energyR - Gas constantT - Absolute temperaturekex - Exchange rate constant (observed)krc - Exchange rate constant (expected)There are two different kinds of H/D exchange experiments. They are native-state H/D exchange and kinetic

refolding experiments. In the native-state H/D exchange experiment, the native protein is allowed to dissolve in the presence of deuterium and then the exchange rates will be monitored.[any kex ref]. The highly protected amide hydrogen will have higher exchange rate. In case of kinetic refolding experiment, the protein is denatured in the presence of deuterium and so all the amide hydrogens will be replaced with deuterium. And then, a fraction of time (tf) will be given for the protein to refold which is followed by the initiation of exchange in the presence of water. The residues which belong to the core, folds first and so its amide deuterium will not be replaced with hydrogen ion during refolding. According to the ‘ First-in-last-out’ hypothesis, the residue that fold first in the kinetic refolding experiment is last to exchange in the native-state unfolding experiment. But there had been a heated debate concerning whether the slow exchange core is the folding core. [debate]. In order to address this issue, we have developed a tool called FILO, incorporated with different efficient strategies, to validate the possibility of the hypothesis for a set of proteins.

Methods:

The tool FILO has been implemented using PERL scripting language. The inputs that must be supplied to the program includes the following: PDB id/ PDB file, exchange rates (kex) in min-1, folding rates (kf) in sec-1, intrinsic exchange rates (krc) in min-1 and the temperature of native state unfolding experiment in kelvin. The program calculates Absolute Correlation Percentage (ACP), Normalized Correlation Percentage (NCP), Absolute Global Correlation Percentage (AGCP), Normalized Global Correlation Percentage

(NGCP) and reports whether or not the protein follows the First-in – last-out hypothesis. Given the appropriate inputs, for calculating ACP and NCP, the algorithm will consider only the residues for which both kex and kf values were available.

Absolute Correlation Percentage (ACP):

The ACP is a straightforward procedure to correlate the two experimental data (native state unfolding and kinetic refolding experiments). The formula is as follows: ACP = [Σ Cef / n] *100(1)Where, Cef = 1, if Rex = Rf for a residue at i th position0, Otherwisen = Total number of residues for which both the kex and kf are reported. Rex = Rank of residues for which exchange rates are available. Rf = Rank of residues for which folding rates are available. The program arranges the kex and kf values in ascending and descending order respectively and then they were ranked accordingly and compared. The idea behind this approach is, if both the ranks of any residue were equal, it means, that residue exchanges and folds according to the hypothesis. And so there exists one-one correlation between the kex and kf values. So for example, if ACP value is 100, then there exists one-one correlation between the two datasets. If ACP value is 70, then there is 70% correlation. If it is 0, then there exists no correlation at all.

Normalized Correlation Percentage (NCP):

The NCP value gives the average spread of residues between the two datasets (kex and kf) and the value cannot be attributed to the % of correlation as in the case of ACP. NCP is illustrated with examples. The formula is as follows: NCP = 100 – [(NC/ Vr) *100](2)Where, Normalized

Correlation (NC) = Σ | Rex – Rf | / n(3)Rex = Rank of residues for which exchange rates are available. Rf = Rank of residues for which folding rates are available. n = Total number of residues for which both the kex and kf are reported. Vr = Variation rate (the maximum rate of variation possible for ‘ n’ number of residues). From (2), two distinctive formulae were derived successfully and they were applied by the program to calculate NCP. If the number of residues for which both kex and kf were present is odd in number, then NGCP formula would be, NCP = 100 – [(2NC * N / N2 – 1) *100](4)If the number of residues for which both kex and kf were present is even in number, then NGCP formula would be, NCP = 100 – [(2NC / N)*100](5)For getting NCP, NC should be calculated according to equation (3). Consider the following cases:

Case 1:

Table 1 Calculation of NC

#

ResiduesRexRfabs (Rex – Rf)1A1542K2423G3214S4135T532Normalized Correlation (NC) = Σ | Rex – Rf | / nNC = 2. 4

Table 2 Calculation of Vr

#

ResiduesRexRfabs (Rex – Rf)1A1542K2423G3214S4135T532Vr = Σ | Rex – Rf | / nVr = 2. 4Number of residues for which both kex and kf were present is odd. So NCP formula would beNCP = 100 – [(2NC * N / N2 – 1) *100]NCP = 0.

Case 2:

Table 3 Calculation of NC

#

ResiduesRexRfabs (Rex – Rf)1A1102K2203G3304S44056TH565600NC = 0

Table 4 Calculation of Vr

#

ResiduesRexRfabs (Rex – Rf)1A1652K2533G3414S43156TH562135Vr = 3# of residues for which both kex and kf were present is even. So NCP formula would beNCP = 100 – [(2NC / N)*100]NCP = 100

Case 3:

Table 5 Calculation of NC

#

ResiduesRexRfabs (Rex – Rf)1A1102K3303G2424S4225T550NC = 0. 8.

Table 6 Calculation of Vr

#

ResiduesRexRfabs (Rex – Rf)1A1542K2423G3214S4135T532Vr = 2. 4NCP = 66. It should be noted that if NC value is equal to 0, then there exists one-one correlation. If the same is equal to that of Vr, then there is no correlation between the two datasets. Conversely if NCP value is equal to 0, then there is no correlation and if it is 100, then there exists one-one correlation. According to case 1, NC is equal to Vr and NCP is equal to 0. The spread of residues on an average between the two datasets is 100%. So there exists

no correlation between kex and kf. According to case 2, NC is equal to 0 and NCP is equal to 100. The spread of residues on an average is 0. So there exists one-one correlation between kex and kf. According to case 3, NC is equal to 0. 8 and NCP is equal to 66. The spread of residues on an average is 34 %. If the spread is minimum, then there is a possibility for better correlation.

Calculation of Absolute Global Correlation Percentage (AGCP) and Normalized Global Correlation Percentage (NGCP):

For calculation of AGCP and NGCP, the following steps were followed: Residue specific ΔGHX was calculated. Proline isomerisation effect was taken into account and ΔGHX* was calculated. Filter was applied (0. 4 kcal/mol). Residues that cross the above filter were taken into account. Calculation of AGCP and NGCP.

Calculation of residue specific ΔGHX :

FILO calculated ΔGHX by using the equation( )ΔGHX =– RT ln (kex / krc)It has been proposed that the average of three largest ΔGHX values provides the information on the conformational stability of a protein.[] The krc values for each residue were obtained from CINTX tool []. The amide groups that has the largest stabilities are the ones which are most protected from exchange, while residues with intermediate and low stabilities exchange by either local or sub-global unfolding states.[]

Proline correction and calculation of ΔGHX*:

Proline cis-trans isomerisation is considered to be one of the rate-determining steps of protein folding.[1 proline cis-trans paper] All the aminoacids except proline were connected to one another via an amide bond whereas the preceding aminoacid connects to proline via an imide bond. In the native state, amide bond will be in trans conformation [24 ONEG] but the imide bond favours both the cis and trans conformations in an equal manner because the free energy difference between these conformers are insignificant [25 ONEG]. Based on the analysis of conformational energy calculations, trans form is dominant in all the aminoacids except proline. [3 proline cis-trans] Proline equilibrates between trans and cis conformation in denatured state. However in native state, it will be either in trans or cis conformation. The conformation of proline depends on the chemical properties of proceeding aminoacid. [24 ONEG]The Xaa - Proline is considered to have trans conformation when the distance between Cα of Xaa and Cα of Proline (Cα– Cα) is greater than Cα of Xaa and Cδ of Proline (Cα– Cδ). Similarly Xaa - Proline is considered to have cis conformation when the distance between (Cα– Cδ) is greater than (Cα– Cα). The proline isomerisation effect of all the 20 aminoacids was taken from [despointes] and the same was kept as default in FILO. FILO computes and compares the distances of (Cα– Cα) and (Cα– Cδ) of Xaa-Proline and predicts the conformation of proline based on the distance computed. And then, based on the conformation, the corresponding energy values (which was kept as default) for the aminoacids preceding each proline residues were summed

up to get the proline isomerisation effect and the same was subtracted with ΔGHX value to obtain ΔGHX*. ΔGHX* is the proline corrected free energy.

Filter:

The filter of 0. 4 kcal/mol was kept as default in FILO based on the fact that trans conformation of prolines is more predominant and it contributes about 0. 3 kcal/mol to the isomerisation effect.[8, 9 oneG]. And so the ΔGHX* was subtracted with 0. 4 kcal/mol and the same was taken as threshold. Then the algorithm scanned for residues whose ΔGHX crosses the threshold and only those residues were taken into account for the calculation of AGCP and NGCP.

Calculation of AGCP and NGCP:

From the residues that had crossed the threshold value, the algorithm selected only those residues for which both kex and kf values were available.

AGCP:

AGCP = [Σ Cef / n] *100 (17)Where, Cef = 1, if Rex = Rf for a residue at i th position0, Otherwisen = Total number of residues for which both the kex and kf are reported. Rex = Rank of residues for which exchange rates are available. Rf = Rank of residues for which folding rates are available.

NGCP:

NGCP = 100 – [(NC/ Vr) *100](18)Where, Normalized Correlation (NC) = Σ | Rex – Rf | / n (19)Rex = Rank of residues for which exchange rates are available. Rf = Rank of residues for which folding rates are available. n = Total number of residues for which both the kex and kf are reported. Vr =

Variation rate (the maximum rate of variation possible for ‘ n’ number of residues). If the number of residues for which both kex and kf were present is odd in number, then NGCP formula would be, NGCP = 100 – [(2NC * N / N2 – 1) *100](20)If the number of residues for which both kex and kf were present is even in number, then NGCP formula would be, NGCP = 100 – [(2NC / N)*100](21)The understanding of AGCP and NGCP were similar to that of ACP and NCP methods.

Validation:

The NCP and NGCP outputs were categorized as shown below:

Table 7 Correlation statuses based on NCP and NGCP values

#

NCP and NGCPCorrelation status1> = 25Poor226 to 50Fair351 to 75Good476 to 100ExcellentThe correlation status

Table 8 Comparison of NCP and NGCP correlation statuses:

#

NCP statusNGCP statusStageResults1PoorPoorINot supported2FairFairINot supported3PoorFairINot supported4FairPoorINot supported5PoorGoodIIFavorable6PoorExcellentIIFavorable7FairGoodIIFavorable8FairExcellentIIFavorable9GoodGoodIIISupported10GoodExcellentIIISupported11ExcellentExcellentIIISupported