Supplementary Information
Estimating The Mean First Passage Time of Protein Misfolding†

Vishal Singh,[a] and Parbati Biswas[a]*

[a] *Address, University of Delhi, Delhi, India. Tel: +911127666646; E-mail: pbiswas@chemistry.du.ac.in*

1 Supporting Information

1.1 Mathematical derivation of MFPT

From Eqn.(11)

$$-1 = W(q_M(t)+1, q_M(t))[\tau(q_M+1)-\tau(q_M)] + W(q_M(t)-1, q_M(t))[\tau(q_M-1)-\tau(q_M)] \tag{S1}$$

The transition rates of the misfolded contacts are defined as

$$W(q_M(t)+1, q_M(t)) = q_N(t)k_{nm}W(q_M(t)-1, q_M(t)) = q_M(t)k_{mn} \tag{S2}$$

Let,

$$U(q_M) = [\tau(q_M) - \tau(q_M+1)] \tag{S3}$$

and

$$S(q_M) = \frac{U(q_M)}{\phi(q_M)} \tag{S4}$$

with

$$\phi(q_M) = \prod_{i=1}^{q_M(t)} \frac{i}{[q_N(t)-(q_M(t)-i)]}\frac{k_{mn}}{k_{nm}} = \binom{q_N(t)}{q_M(t)} K_{eq2}^{-q_M} \tag{S5}$$

where $K_{eq2} = k_{nm}/k_{mn}$ Using these definitions, for large $q_N(t)$, Eqn.(S1) can be written as

$$q_N(t)k_{nm}[-U(q_M)] + q_M(t)k_{mn}[U(q_M-1)] = -1 \tag{S6}$$

or

$$q_M(t)k_{mn}\phi(q_M-1)[S(q_M-1)-S(q_M)] = -1 \tag{S7}$$

From Eqn.(S7), we find that

$$S(q_M) = S(0) + \sum_{i=q_M(t)}^{1} \frac{1}{k_{mn}i\phi(i-1)} \tag{S8}$$

We take the reflecting boundary condition as $q_M(t) = 1$. Thus, $\tau(0) = \tau(1)$, so from Eqn.(S3) and Eqn.(S4), $S(0) = 0$. Therefore,

$$\tau(q_M) = \phi(q_M) \sum_{i=q_M(t)}^{1} \frac{1}{k_{mn}i\phi(i-1)} + \tau(q_M+1) \tag{S9}$$

We apply the absorbing boundary condition at M, i.e., the critical number of misfolded contacts, q_{MC}, in misfolded state. Thus, $\tau(q_{MC}) = \tau(q_{MC}+1)$. Hence, Eqn.(S9) may be written as

$$\tau(q_M) = \sum_{j=q_M(t)}^{q_{MC}} \phi(j) \sum_{i=j-1}^{0} \frac{1}{k_{mn}(i+1)\phi(i)} \tag{S10}$$

Using Eqn.(S5) in Eqn.(S10), MFPT can be estimated as

$$\tau(q_M) = \frac{K_{eq2}^{-1}}{k_{mn}} \sum_{j=q_M(t)}^{q_{MC}} \frac{1}{[q_N(t)-(j-1)]} \sum_{n=0}^{j} \frac{(-1)^n(j)_n}{([q_N(t)-(j-2)])_n} K_{eq2}^{-n} \tag{S11}$$

where, $(j)_n$ is known as Pochhammer symbol[1] which is defined as

$$(j)_n = \begin{cases} 1, & n = 0 \\ \prod_{k=0}^{n-1}(j+k), & n > 0 \end{cases} \tag{S12}$$

Eqn.(S11) can be written in terms of Hyper-geometric function $[{}_2F_1(\alpha,\beta;\gamma;z)]$

$$\tau(q_M) = \frac{1}{k_{nm}} \sum_{j=q_M(t)}^{q_{MC}} \frac{1}{[q_N(t)-(j-1)]} [{}_2F_1(-j,1;q_N(t)-(j-2);-r)] \tag{S13}$$

where $r = 1/K_{eq_2} = k_{mn}/k_{nm} < 1$. One of the well known properties of the Hyper-geometric function,[2] ${}_2F_1(\alpha,\beta;\gamma;z)$ is given by

$${}_2F_1(\alpha,\beta;\gamma;z) = \frac{\Gamma(\gamma)}{\Gamma(\beta)\Gamma(\gamma-\beta)} \int_0^1 x^{\beta-1}(1-x)^{\gamma-\beta-1}(1-zx)^{-\alpha}dx; |z| < 1 \tag{S14}$$

Substituting Eqn.(S14) in Eqn.(S13) the integral form of MFPT is given as

$$\tau(q_M) = \frac{1}{k_{nm}} \int_0^1 (1-x)^{q_N(t)-q_{MC}-q_M(t)} \quad \frac{[(1-x)^{q_M(t)}(1+rx)^{1+q_{MC}} - (1-x)^{1+q_{MC}}(1+rx)^{q_M(t)}]}{(1+r)x} dx \tag{S15}$$

1.2 Selection of Proteins

Single-chain globular proteins without any ligands and missing residues are chosen from the RCSB databank (https://www.rcsb.org) with a resolution of ≤ 2.0 Å and R-factor range of $0.18-0.21$, with different lengths and sequence identity of $\leq 20\%$. The non-bonded contacts of the selected proteins are calculated upto 7.5 Å.[3,4] The unfolded, folded and misfolded states are distinguished on the basis of these non-bonded contacts. The experimental data on the protein folding and unfolding rates are extracted from the KineticDB database(https://www.kineticdb.protres.ru) and corresponding refs.[5]

1.3 Variation of the misfolding rate constant

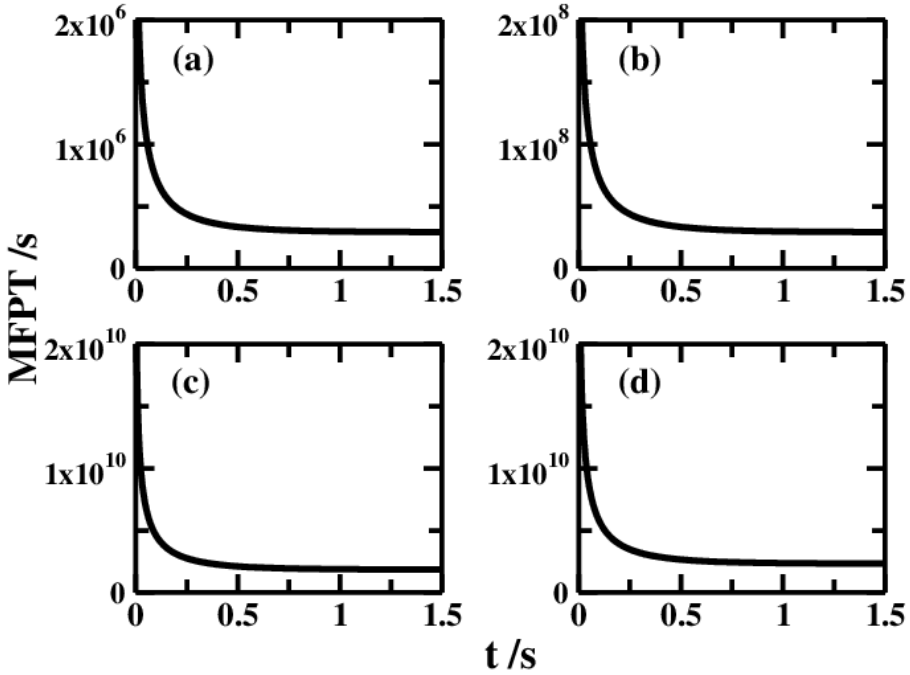

Fig. S1 MFPT of the unfolded state of α-lactalbumin with 20% native contacts and $k_{un} = 3.3s^{-1}$, $k_{nu} = 0.88s^{-1}$. The MFPT for an arbitrary value of the rate-constant $k_{mn} = 10^{-12}s^{-1}$ and different values of the rate-constant k_{nm} = (a) $10^{-6}s^{-1}$, (b) $10^{-8}s^{-1}$, (c) $10^{-9.8}s^{-1}$ and (d) $10^{-9.9}s^{-1}$ are shown. For $k_{nm} = 10^{-9.8}s^{-1}$ and $k_{nm} = 10^{-9.9}s^{-1}$ the value of MFPT is 59 and 74 years respectively, which corresponds to the range of the age of appearance of misfolding diseases in humans.

References

1 H. Srivastava, A. Çetinkaya and İ. O. Kıymaz, *Appl. Math. Compu.*, 2014, **226**, 484–491.

2 K. T. Kohl and V. H. Moll, *Scientia*, 2011, **21**, 43–54.

3 L. A. Mirny, A. V. Finkelstein and E. I. Shakhnovich, *Proc. Natl. Acad. Sci. U.S.A.*, 2000, **97**, 9978–9983.

4 A. Baruah, A. Bhattacherjee and P. Biswas, *Soft Matter*, 2012, **8**, 4432–4440.

5 N. S. Bogatyreva, A. A. Osypov and D. N. Ivankov, *Nucleic Acids Res.*, 2008, **37**, D342–D346.