Competitive Model on Denaturant-Mediated Protein Unfolding

R. Murugan
Department of Chemical Sciences, Tata Institute of Fundamental Research, Colaba, Mumbai, 400005, India

ABSTRACT A denaturant-mediated protein unfolding model, which is different from already existing ones based on the assumption that denaturant competes for water molecules to interact and thus reduces water–protein interactions, which leads to unfolding phenomenon, has been developed with a detailed mathematical justification. Theoretical results suggested that the parameter (m_u) obtained from the usual linear extrapolation model must be a linear function of the number of bound water molecules (n) on protein with a zero intercept. However, application of this theory to a set of proteins for which m_u values for urea denaturation are already known showed that m_u was a linear function of n but with a nonzero intercept. Finally this nonzero intercept was attributed to binding of denaturant to protein at $n = 0$. Detailed investigation of this factor showed that average equilibrium constant for binding of urea with aromatic side chains (generally nonpolar side chains) was $k_b \approx 0.65 \pm 0.45$ mol^{-1}, which agreed well with earlier experimental estimations, and also suggested that an integrated approach was necessary to avoid discrepancy in ΔG^{H_2O} estimated from different models.

INTRODUCTION

The unfolding problem of proteins by denaturants like urea and guanidine hydrochloride has been described so far in four different ways (although the exact mechanism is still debatable), namely, the Tanford model, the binding model, the solvent-exchange model, and the linear-extrapolation model (LEM). According to Tanford's model, the unfolding free energy can be expressed as a sum of potentials of group transfer from water to denaturants (Tanford, 1968):

$$\Delta G_{unfolding} = \Delta G^{H_2O} + \sum_i n_i \alpha_i \delta g_{tr}^i. \quad (1)$$

Here, δg_{tr}^i is the transfer potential of the ith side chain, α_i is the average fractional change in the degree of exposure of the groups of type i, and n_i is total number of such groups. In this model the importance has been given to denaturant ← protein → water interaction (i.e., protein interacts independently with water and denaturant, or denaturant–water interaction is negligible), whereas the binding model assumes that denaturant has ''binding sites'' on the protein molecule (Tanford, 1970; Pace and Vanderburg, 1979) and thus, upon sequential binding, it unfolds the protein. The overall equilibrium constant for this binding can be given as:

$$K = K_0 \frac{1 + \sum_{j=0}^{n_D} L_{j,D} a_x^j}{1 + \sum_{j=0}^{n_N} L_{j,N} a_x^j} \quad (2)$$

$$L_j = l_1 l_2 \ldots l_j, \quad l_j = \frac{[NX_j]}{[N/DX_{j-1}][X]}.$$

Here $[X]$ denotes the concentration of denaturant, $[N]$ and $[D]$ denote the concentrations of native and denatured forms respectively, l_j is the equilibrium constant (mol^{-1}) for the jth step, and n_D and n_N are the total number of binding sites in denatured and native forms, respectively. When $l_i = l_j = k$, Eq. 2 reduces to the following form:

$$K = K_0 (1 + k a_x)^n. \quad (3)$$

Here, n is total number of binding sites on protein and a_x is denaturant's activity. From this model, it was also shown that (Tanford, 1968)

$$\frac{\partial \ln K}{\partial \ln a_x} = \Delta\vartheta_x - \frac{m_x}{55.5}\Delta\vartheta_w, \quad (4)$$

where $\Delta\vartheta_x$ is the maximum number of bound denaturant molecules, $\Delta\vartheta_w$ is the maximum number of bound solvent molecules (here it is water), and m_x is the molality of denaturant. In this model, the importance has been given to protein → denaturant interaction. The solvent–exchange model also yields the same type of relation as in Eq. 3. LEM (Greene and Pace, 1974) assumes a linear relation between unfolding free energy and denaturant activity. According to LEM,

$$\Delta G = \Delta G^{H_2O} - m_u[\text{denaturant}]. \quad (5)$$

Detailed experimental analysis had shown that LEM always gave an underestimate of ΔG^{H_2O}. But urea and guanidine hydrochloride gave fairly similar ΔG^{H_2O} for a particular protein (except for a few) when LEM was used and which is not so in case of two other models (binding and Tanford's model gave a higher ΔG^{H_2O} in the case of guanidine hydrochloride-mediated unfolding; Pace, 1986). All the aforementioned models were generally aimed to explain the experimentally observed sharp (two-state) transition (such abrupt transitions are usually said to be ''cooperative'') of unfolding curves. Apart from these, the water-structure-breaking model was also developed and used to explain unfolding phenomenon (Von Hippel and Wong,

Submitted May 29, 2002, and accepted for publication September 16, 2002.

Address reprint requests to R. Murugan, Homibhabha Road, Tata Institute of Fundamental Research, Colaba, Mumbai, 400005, India. Tel.: +91-22-215-2971; Fax: +91-22-215-2110 or 215-2181; E-mail: muruga@tifr.res.in.

0006-3495/03/02/770/05 $2.00

1965). In this model, it is believed that denaturants break the structure of water by interfering with its usual hydrogen-bonding network, which in turn unfolds the protein (by breaking the ordered water structure present around the protein molecule). One thing we should note is that all the abovesaid models would fail when the denaturant concentration is comparable with water concentration (i.e., 55.55 M). In that situation, the interaction type becomes denaturant → water → protein. In this article, I present one more interesting model based on an assumption that denaturant competes for water molecules to interact, thus reducing water–protein interaction, which leads to unfolding (this is different from the solvent-exchange model in the sense that, here, denaturant–protein interaction is negligible). The specialty of the model is: 1), when the denaturant concentration is in nitesimal, it converges to LEM and 2), it is purely an equilibrium kinetic model based on the law of mass action.

The competitive model and its derivation

Here the basic idea is that it is well known that denaturants signi cantly interact with water (here activities are comparable) and protein–water interactions are also signi cant (each protein molecule contains hundreds of bound water molecules as could be seen from crystal data). Therefore, unfolding by denaturants must be through denaturant → water → protein interaction. According to this model, the unfolded form of protein interacts sequentially (at speci c sites) with n number of water molecules (here n varies from a few to hundreds) to yield native form $N[H_2O]_n$, where the interactions are independent. When denaturant D is introduced in the system, each D molecule interacts with m number of water molecules to yield $D[H_2O]_m$. Therefore, the approximate scheme for folding phenomenon in the presence of denaturant becomes:

$$\begin{aligned} U + H_2O &\overset{k_1}{\leftrightarrow} U[H_2O] + H_2O \overset{k_2}{\leftrightarrow} U[H_2O]_2 \\ &\overset{k_3}{\leftrightarrow} \ldots\ldots\ldots\ldots \overset{k_n}{\leftrightarrow} N[H_2O]_n \\ D + mH_2O &\underset{k_4'}{\overset{k_3'}{\leftrightarrow}} D[H_2O]_m. \end{aligned} \tag{6}$$

Here U denotes the denatured state of protein (but not dehydrated; i.e., there may be nonspeci c binding sites for water which do not contribute to folding phenomenon), D denotes denaturant, n is the average (due to the fact that it is a uctuating quantity in solution) number of water molecules interacting with native form N, m is the average number of water molecules interacting with a denaturant molecule, k_1, $k_2 \ldots k_n$ (mol^{-1}) are the respective equilibrium constants of binding of water to protein, k_3' and k_4' are the respective rate constants for denaturant water interaction, and

$$[H_2O] \approx h_0 - my \tag{7}$$

$$k_3'(d_0 - y)(h_0 - my)^m - k_4'y = 0, \tag{8}$$

where $y = D[H_2O]_m$ and $u_0 = [U]_{t=0}$, $h_0 = [H_2O]_{t=0}$ and $d_0 = [D]_{t=0}$ are the corresponding activities. So, n and m indirectly indicate the water interaction potential of protein and denaturant respectively. At equilibrium, assuming $k_1 = k_2 = k_3 = \ldots k_i = k$ (as the average equilibrium constant for binding),

$$\begin{aligned} K_{folding} &= \frac{N[H_2O]_n}{U} = \frac{N[H_2O]_n}{N_0 - N[H_2O]_n} = K_0(1 + k(h_0 - my))^n \\ \Delta G_f &= \Delta G_f^0 - nRT\sum_{i=0}^{\infty}(-1)^{i+1}\left[kh_0\left(1 - \frac{my}{h_0}\right)\right]^{i+1} \\ &\approx (\Delta G_f^0 - nkh_0) + nRTkmy = \Delta G_f^{H_2O} + nRTkmy \\ \Delta G_u &= \Delta G_u^{H_2O} - nRTmky. \end{aligned} \tag{9}$$

Here, N_0 is the total protein concentration in the system, ΔG_u and ΔG_f are the corresponding unfolding and folding free energies, and $\Delta G_u = -\Delta G_f$. The approximate value of k can be found as follows: for the *ith* step, the equilibrium constant for binding can be given as

$$k_i = \frac{U[H_2O]_{i+1}}{h_0 \times U[H_2O]_i} = \frac{1}{h_0} \times \exp\left(-\frac{\Delta G_i}{RT}\right), \tag{10}$$

where $\Delta G_i = -RT\ln k_i\ h_0$, and ΔG_i is the folding free energy for the ith step. Inasmuch as the partitioned free energy corresponding to a single step is much less, the following approximation will always hold:

$$\lim_{\Delta G_i \to 0} k_i = k \approx \frac{1}{h_0} \tag{11}$$

Therefore, the LEM limit of the folding free energy becomes

$$\Delta G_f = \Delta G_f^{0\prime} + nRT \times \left(\frac{my}{h_0}\right). \tag{12}$$

Now the concentrations of the folded and unfolded forms can be given as

$$\begin{aligned} N[H_2O]_n &= \frac{N_0K_0(1 + k(h_0 - my))^n}{1 + K_0(1 + k(h_0 - my))^n} \\ U &= \frac{N_0}{1 + K_0(1 + k(h_0 - my))^n}. \end{aligned} \tag{13}$$

But it is necessary to solve Eq. 8 for the variable y to get the correct solutions for Eqs. 9 and 13. We can encounter two different cases that depend on the value of y.

Case I: lower concentration of denaturant

At an in nitesimal denaturant activity, $h_0 \gg my$, $y \approx d_0$ and, therefore, Eq. 9 can be approximated to

$$
\begin{aligned}
K_{folding} &= \frac{N[H_2O]_n}{U} = \frac{N[H_2O]_n}{N_0 - N[H_2O]_n} \\
&= K_0(1 + k(h_0 - m'd_0))^n \\
\Delta G_f &= \Delta G_f^0 - nRT \sum_{i=0}^{\infty}(-1)^{i+1}\left[kh_0\left(1 - \frac{m'd_0}{h_0}\right)\right]^{i+1} \\
&\approx \left(\Delta G_f^0 - nRTkh_0\right) + nRTkm'd_0 \\
&= \Delta G_f^{H_2O} + nRTkm'd_0 \\
\Delta G_u &= \Delta G_u^{H_2O} - nRTkm'd_0,
\end{aligned}
\tag{14}
$$

where

$$
m' = \frac{m}{1 + \dfrac{k_4'}{k_3' h_0^m}},
$$

inasmuch as

$$
\frac{k_4'}{k_3'} \ll h_0^m, \quad m' \approx m. \tag{15}
$$

Therefore, the LEM limit of folding free energy becomes

$$
\begin{aligned}
\Delta G_f &= \Delta G_f^{0\prime} + \frac{nRT}{h_0} m d_0 = \Delta G_f^{H_2O} + \frac{nRTm}{h_0} \times d_0 \\
\Delta G_u &= \Delta G_u^{H_2O} - \frac{nRTm}{h_0} \times d_0,
\end{aligned}
\tag{16}
$$

and the corresponding concentrations of folded and unfolded forms become

$$
\begin{aligned}
N[H_2O]_n &= \frac{N_0 K_0 (1 + k(h_0 - md_0))^n}{1 + K_0(1 + k(h_0 - md_0))^n} \\
U &= \frac{N_0}{1 + K_0(1 + k(h_0 - md_0))^n}.
\end{aligned}
\tag{17}
$$

One also should note that Eq. 17 predicts a sharp transition near $d_0 = h_0 / m'$, which resembles the usual experimental observations and suggests that the cooperativity assumption is not necessary to explain such transitions.

Case II: higher concentration of denaturant

The approximation given by Eqs. 14 and 17 will not hold when the denaturant activity is comparable with solvent. In this situation one has to solve Eq. 8 for y completely. Here the solution has been given using the perturbation method, rewriting Eq. 8 as

$$
(d_0 - y)(1 - \alpha y)^m - \beta y = 0, \tag{18}
$$

where

$$
\alpha = \frac{m}{h_0}, \quad \beta = \frac{k_4}{k_3 h_0}.
$$

And inasmuch as $\alpha < 1$, using in nite binomial expansion (Courant and John, 1989), Eq. 18 can be expanded in terms of asymptotic series as

$$
d_0 \sum_{n=0}^{\infty} {}^mC_i(\alpha y)^n - y \sum_{i=0}^{\infty} {}^mC_i(\alpha y)^i - \beta y = 0, \tag{19}
$$

where ${}^mC_i = m\ /i\ (m - i)$ are binomial coef cients. Finally, Eq. 19 can be reduced to an ordinary perturbation problem with the perturbation parameter $\varepsilon = \alpha d_0$, as,

$$
d_0 - y(m\alpha d_0 + 1 + \beta) = \varepsilon(\gamma y^2 + \delta y^3 + \ldots.), \tag{20}
$$

where

$$
\begin{aligned}
\lambda &= (m\alpha d_0 + 1 + \beta), \quad \gamma = -\left(\frac{m(m-1)\alpha}{2} + \frac{m}{d_0}\right), \\
\delta &= \left(\frac{m(m-1)(m-2)\alpha^2}{3} + \frac{m(m-1)\alpha}{d_0 2}\right).
\end{aligned}
$$

Neglecting the third-order terms on the right side of Eq. 20, we obtain

$$
d_0 - \lambda y = \varepsilon(\gamma y^2 + \delta y^3 + \ldots.) \approx \varepsilon \gamma y^2. \tag{21}
$$

Perturbation expansion of y in terms of ε yields

$$
y = a + b\varepsilon + O_2(\varepsilon^2). \tag{22}
$$

Putting Eq. 22 in Eq. 21 and solving for coef cients a and b, we obtain the rst-order corrected value of y as

$$
y \approx \frac{d_0}{\lambda}\left(1 - \frac{\gamma d_0^2 m}{h_0 \lambda^2}\right). \tag{23}
$$

Substituting Eq. 23 in Eq. 9 and performing calculations as in Case I, the folding free energy in the presence of higher activity of denaturant becomes

$$
\begin{aligned}
K_{folding} &= \frac{N[H_2O]_n}{U} = \frac{N[H_2O]_n}{N_0 - N[H_2O]_n} \\
&= K_0\left(1 + k\left(h_0 - m \times \frac{d_0}{\lambda}\left(1 - \frac{\gamma d_0^2 m}{h_0 \lambda^2}\right)\right)\right)^n \\
\Delta G_f &= \Delta G_f^0 - nRT \sum_{i=0}^{\infty}(-1)^{i+1} \\
&\quad \times \left[kh_0\left(1 - \frac{md_0}{h_0\lambda}\left(1 - \frac{\gamma d_0^2 m}{h_0 \lambda^2}\right)\right)\right]^{i+1} \\
&\approx \Delta G_f^{0\prime} + nRTkm \times \frac{d_0}{1 + \beta + \dfrac{m^2 d_0}{h_0}} \\
&= \Delta G_f^{0\prime} + nRTkm \times \left\{\frac{h_0}{m^2}\sum_{i=1}^{\infty}(-1)^{i+1}\left(\frac{m^2 d_0}{h_0(1+\beta)}\right)^i\right\}.
\end{aligned}
\tag{24}
$$

Again, the LEM limit of the folding free energy becomes

$$\Delta G_f = \Delta G_f^{0\prime} + \frac{nRT}{h_0} m d_0 = \Delta G_f^{H_2O} + \frac{nRTm}{h_0} \times d_0$$

$$\Delta G_u = \Delta G_u^{H_2O} - \frac{nRTm}{h_0} \times d_0. \quad (25)$$

Validation of the theory

Inasmuch as the theory converged to LEM at an in nitesimal activity of denaturant, to check its validity we can use the experimental m_u values already obtained for several proteins for urea-mediated denaturation. Theoretical m_u values can be calculated from Eq. 25 (taking only the linear term) as

$$m_u = 0.0212 \times n(\text{kcal} \times \text{mol}^{-1}\,\text{M}^{-1}), \quad (26)$$

where $RT = 0.592$ kcal $\times$ mol^{-1}, $h_0 = 55.55$ M, $T = 298$ K, and $m \approx 2$ (this is because urea has two NH_2 groups to interact with two molecules of water at a time) were used. Thus the theory predicts that there is a linear relation between the number of bound water molecules on protein and m_u. To check this, a set of proteins (Scholtz et al., 1995) with different n values was chosen and the corresponding m_u values were calculated using Eq. 26 (here, n was taken directly from PDB data). The linear regression analysis between m_u^{obs} (i.e., observed m_u value) and n gave a poor tting ($r = 0.2$) for Eq. 26 but a good tting ($r = 0.65$) to the following equation:

$$m_u^{obs} = (0.752 \pm 0.535) + (0.011 \pm 0.005)n. \quad (27)$$

The tted data along with error bar is shown in Fig. 1 and predicted m_u values are shown in Table 1. The intercept value in Eq. 27 clearly shows the presence of other contributions (may be the binding). Therefore, if we include the contribution from binding, by denoting the number of binding sites on protein as b and the average equilibrium constant for binding as k_b (mol^{-1}),

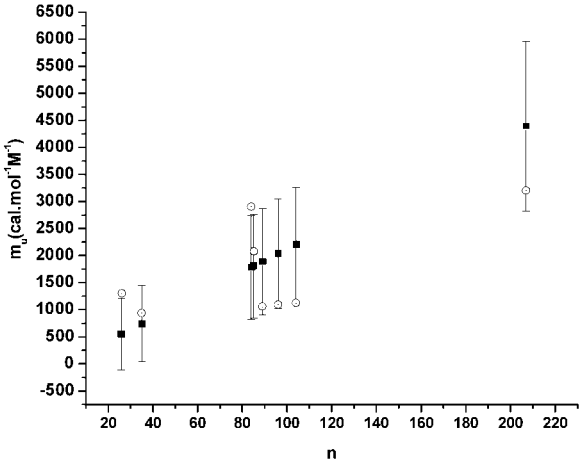

FIGURE 1 Here, n denotes the average number of bound water molecules as could be seen from crystal data, the open circles indicate the observed m_u values, and lled squares are the prediction by Eq. 27 with standard error bar.

$$m_u = RTbk_b + \frac{RTm}{h_0} \times n$$

$$= 0.592 \times bk_b + (0.011 \pm 0.005) \times n. \quad (28)$$

Thus, to a crude approximation, relation $bk_b = 1.3 \pm 0.9$ mol^{-1} (obtained by just dividing the intercept in Eq. 27 by RT) holds at an in nitesimal quantity of urea (and at $n = 0$). Inasmuch as $n = 0$ belongs to a pure aromatic side chain (and only the side chain) for which $b \approx 2$, and thus $k_b \approx 0.65 \pm 0.45$ mol^{-1}, which falls in an already observed range (Pace, 1986), the aforementioned results prove the validity of our theory and suggest that an integrated approach (i.e., we have to use binding and competitive models simultaneously as in Eq. 28) is necessary for a complete description of denaturant-mediated unfolding of proteins. Supposing that only the binding model has been used, we would then miss the contributions from denaturant water (which is dependent on m and n) interactions, and vice versa. This is the reason why contradictions arise in ΔG^{H_2O}calculated from different models and by using different denaturants. But LEM yields a fairly consistent value of ΔG^{H_2O}due to the fact that here all

TABLE 1 Predicted m_u values from Eqs. 26 and 27 for different proteins

Protein	PDB Code	No. of water molecules from crystal structure (n)	Experimental m_u value Kcal $\times$ mol^{-1} M^{-1}	Prediction by Eq. 26 Kcal $\times$ mol^{-1} M^{-1}	Prediction by Eq. 27 Kcal $\times$ mol^{-1} M^{-1}
Calbindin D9K	3ICB	35	0.940	0.742	1.137 ± 0.710
Trp aporepressor	3WRP	84	2.900	1.780	1.676 ± 0.955
Thioredoxin	1THO	26	1.300	0.551	1.038 ± 0.665
β-Lactamase	3BLM	207	3.200	4.388	3.029 ± 1.570
HPr (*Bacillus subtilis*)	2HPR	89	1.050	1.887	1.731 ± 0.980
RNase A (Bovine)	9RSA	181	1.300	3.837	2.743 ± 1.440
Lysozyme (HEW)	6LYZ	104	1.120	2.205	1.896 ± 1.055
RNase Ba	1RNB	96	1.095	2.035	1.808 ± 1.015
α-Chymotryp-Sinogen	4CHA	85	2.070	1.802	1.687 ± 0.960
RNase T1	6RNT	104	1.210	2.204	1.896 ± 1.055

contributions are put into a single m_u value. Moreover, Eq. 28 also suggests that

$$b = \frac{m_u^{obs}}{RTk_b} - \frac{mn}{h_0 k_b}. \quad (29)$$

But the drawback of this theory is that standard error in prediction is >50%. This is mainly because the number of proteins used for calculation is less (here we used only 10 proteins) and we are using n values obtained from crystal data, which is not actually true in the solution condition. If we assume that, in the solution condition, the bound water is in double-well potential $U(x)$, where x is the distance from the protein molecule, with minima at A and B (A-minima is for the protein–water form and B-minima is for water in bulk), then the population n_A at A can be approximately given as (assuming $U(A) > U(B)$ and steady state; see Gardiner, 1983):

$$n_A = n_0 p_s^A = n_0(1 - e^{-U(A)/RT}), \quad (30)$$

where $\lim_{T\to 0} n_A = n_0$ and p_s^A is the stationary probability function. Therefore, the re ned form of Eq. 28 becomes

$$m_u^{obs} = (0.752 \pm 0.535) + (0.011 \pm 0.005) n_A. \quad (31)$$

The author thanks Professors S. Mazumdar and S. Mitra for constant encouragement and help.

REFERENCES

Courant, R., and F. John. 1989. Introduction to Calculus and Analysis, Vol I. Springer-Verlag, New York. 456.

Gardiner, C. W. 1983. Handbook of Stochastic Methods. H. Haken, editor. Springer-Verlag, Berlin, Germany. 342.

Greene, R. F., and C. N. Pace. 1974. Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, alpha-chymotrypsin, and beta-lacto globulin. *J. Biol. Chem.* 249:5388–5393.

Pace, C. N., and K. E. Vanderburg. 1979. Determining globular protein stability: guanidine hydrochloride denaturation of myoglobin. *Biochemistry.* 18:288–292.

Pace, C. N. 1986. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* 131:266–280.

Scholtz, J. M., D. Barrick, E. J. York, J. M. Stewart, and R. I. Baldwin. 1995. Urea unfolding of peptide helices as a model for interpreting protein unfolding. *Proc. Natl. Acad. Sci. USA.* 92:185–189.

Tanford, C. 1968. Protein denaturation. *Adv. Protein Chem.* 23:121.

Tanford, C. 1970. Protein denaturation. *Adv. Protein Chem.* 24:1–95.

Von Hippel, P. H., and K. Y. Wong. 1965. On the conformational stability of globular proteins. The effects of various electrolytes and non-electrolytes on the thermal ribonuclease transition. *J. Biol. Chem.* 240:3909–3923.