Physica A 533 (2019) 122053

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Forecasting the upper bound free energy difference between protein native-like structures

Jorge A. Vila *

Instituto de Matemática Aplicada San Luis-CONICET, Universidad Nacional de San Luis, Ejército de Los Andes 950, 5700 San Luis, Argentina
Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, NY, USA

HIGHLIGHTS

- The origin of the protein marginal stability is revisit.
- The existence of an upper bound to the protein marginal stability is proved.
- The determined upper bound value is robust to molecular weight changes.
- The determined upper bound value is valid for any protein fold-class.

ARTICLE INFO

Article history:
Received 15 December 2018
Received in revised form 4 July 2019
Available online 15 July 2019

Keywords:
Free-energy
Protein
Native states
Gershgorin
Heuristic
Eigenvalues

ABSTRACT

Using a combination of statistical thermodynamics and the Gershgorin theorem we computed, in the thermodynamic limit, a plausible value for the upper bound of the free energy difference between native-like structures of monomeric globular proteins. The validity of our result is discussed herein in terms of both the observed free-energy change between the native and denatured states and the micro stability free-energy values obtained from the observed micro-unfolding tendency of nine monomeric globular proteins.

1. Introduction

An accurate determination of the free energy difference between native-like conformers of any protein is a daunting task. Indeed, the studies of ubiquitin folding through *state of the art* equilibrium atomistic simulations [1] predicted, at the melting temperature, a folding enthalpy (∼14 kcal/mol) which is several times lower than the observed value (∼84 kcal/mol) [2]. Another case in point is the longstanding evidence [3] showing that the range of microstability free energy values of native-like conformers of globular proteins is very narrow (2.5 to 7.1 kcal/mol), although, to be best of our knowledge, no theoretical proof supporting this assessment had been provided, yet. How to tackle the latter is illustrated here by analyzing the fluctuation around the native-state of monomeric globular proteins. For this purpose we made use of the following fact: the Gibbs free energy of any protein state is given, in the thermodynamic limit, by the maximum eigenvalue of the partition function. Hereafter, an upper bound to the Gibbs free energy difference between

* Correspondence to: Instituto de Matemática Aplicada San Luis-CONICET, Universidad Nacional de San Luis, Ejército de Los Andes 950, 5700 San Luis, Argentina.
E-mail address: jv84@cornell.edu.

https://doi.org/10.1016/j.physa.2019.122053
0378-4371/

native-like states can be determined by using both the Gershgorin (circle) theorem [4] and a heuristic argument. Finally, the result is judged against both the observed free energy change between the native and denatured states and the range of microstability free energy values obtained from the observed micro-unfolding tendency of nine monomeric globular proteins [5].

2. Materials and methods

It is well known that all the thermodynamics properties of any system/problem, including biomolecular ones [6], such as the helix–coil transition (induced by temperature or pH changes), loop entropy in RNA/DNA, etc., can be derived from the partition function (Q). Thus, in particular, the Gibbs free energy (G) will be given by:

$$\mathrm{G} = -\mathrm{RT}\ln Q \tag{1}$$

After a proper assignment of statistical weights the partition function (Q) can be written in terms of a matrix ($\boldsymbol{C}$) where the elements are, indeed, Boltzmann factors [6]. It is well known that [7], in the thermodynamic limit, the following equality hold:

$$\boldsymbol{Lim}_{j\to\infty}(1/j)lnQ = \boldsymbol{Lim}_{j\to\infty}(1/j)lnC_{kl}(j) = ln\lambda_{max} \tag{2}$$

where $C_{kl}(j)$ is any element of the iterate matrix $\boldsymbol{C}^j$; j is the number of residues in the chain and λ_{max} is the maximum eigenvalue of the matrix $\boldsymbol{C}$. Although derivation of Eq. (2) can be found in the literature [7] their deduction is straightforward after taking into account that: (i) the matrix $\boldsymbol{C}$ is diagonazible and, hence, its eigenvalues λ are solution of $det|\boldsymbol{C}-\lambda\mathbf{I}|=0$, where $\mathbf{I}$ is the identity matrix; (ii) the partition function can be written as $Q = \mathbf{U}\boldsymbol{C}^j\mathbf{V} = \sum_p^n\sum_r^n u_{1p}C_{pr}(j)\,v_{r1}$, where $\mathbf{U}$ and $\mathbf{V}$ appropriate end vectors and n the matrix order; and (iii) $C = \Pi\Delta\Pi^{-1}$, where $\Delta = \begin{pmatrix}\lambda_1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \lambda_n\end{pmatrix}$ with λ the eigenvalues of the matrix $\boldsymbol{C}$, and Π a matrix that satisfy $\Pi\Pi^{-1} = \mathbf{I}$; therefore, any element of the iterate matrix $\boldsymbol{C}^j$ is given by: $C_{kl}(j) = \sum_s^n t_{k,l}^s\lambda_s^j$, where t are coefficients. Then, after arranging the eigenvalues as $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ and using that $\boldsymbol{Lim}_{j\to\infty}|\lambda_i/\lambda_{max}|^j \to 0$ or $1\ \forall\, i$, Eq. (2) is obtained effortlessly.

According to the Gershgorin (circle) theorem [4], for any eigenvalue λ of the matrix $\boldsymbol{C}$ the following inequality hold $|\lambda| \leq \max\{\sum_{l=1}^n |C_{kl}|\}$ with $1 \leq k \leq n$ and n the matrix order. Taking into account that the maximum eigenvalue of $\boldsymbol{C}$ is real and positive and that any element $\boldsymbol{C}$, by definition, satisfies $|\boldsymbol{C}_{kl}| \equiv C_{kl}$, then there is a k' value such that the upper bound of the maximum eigenvalue ($\lambda_{\max}$) is given by:

$$\lambda_{\max} \leq \{\Sigma_l \boldsymbol{C}_{k'l}\} \tag{3}$$

Assuming that the native-like conformers of a given ensemble coexist in fast dynamics equilibrium [8], then an upper bound to the free energy difference, between conformers with the lowest and highest total free energy G, can be computed from Eqs. (1) to (3), as:

$$\Delta G \leq \boldsymbol{Lim}_{j\to\infty} RT\, ln[\Sigma_l \boldsymbol{C}_{k'l}/\Sigma_m \boldsymbol{C}_{t'm}]^j \tag{4}$$

where the term $[\Sigma_l \boldsymbol{C}_{k'l}/\Sigma_m \boldsymbol{C}_{t'm}]$ is strictly larger than 1 because, according with the thermodynamic hypothesis [9], the native-state is the *lowest* free-energy conformation, or conformations if this state is degenerated. Implicit in this inequality is that *all* conformers in equilibrium with the native-state will possess higher, although comparable, free energy. Indeed, if Eq. (4) were use to compute an upper bound to the free energy difference between *native* and *non-native* (denaturated) states, then it is reasonable to assume that $[\Sigma_l \boldsymbol{C}_{k'l}/\Sigma_m \boldsymbol{C}_{t'm}] >> 1$. At this point is worth noting that the protein denaturation analysis is out of the scope of this work. However, a mention to the average range of variation of the denaturation free-energy, observed at 25 °C for nine monomeric globular proteins, will be done in the next section only to illustrate the consistency of our prediction.

3. Results and discussion

There are two dominant interactions that contributed to the stability of native-like conformers in proteins regardless of the fold class, sequence or size, namely, the interactions between (i) polar groups (hydrogen bonds) and (ii) non-polar groups [10–13]. Consequently, the free energy changes between native-like conformers, given by Eq. (4), would imply variations of either one, or both, interactions. In this analysis we have assumed that the total free-energy can be computed as a sum of pairwise interactions. In other words, the many-body interactions were not included, among other reasons, because we focus on the *dominant* interactions to the protein stability [11]. While we recognized there could be many functional forms involving these two interactions it called our attention that the molecular weight (MW), of nine monomeric globular proteins [5], showed a good correlation with the total number of both the intramolecular hydrogen bonds ($R^2 = 0.98$) and the pairs of non-polar groups at distances < 4 Å ($R^2 = 0.83$). Thus, from a heuristic point of view, we conjectured that the term $[\Sigma_l \boldsymbol{C}_{k'l}/\Sigma_m \boldsymbol{C}_{t'm}]^j$ grows with j as the MW does. We can then rewrite Eq. (4) as:

$$\Delta G \leq \boldsymbol{Lim}_{MW\to\infty} RT\ ln\, MW \tag{5}$$

Considering that the largest known monomeric globular protein [14] possessed a $MW = 2.710^5$, then, at T = 298 K, Eq. (5) give us a $\Delta G \leq 7.4$ kcal/mol. This plausible value for ΔG, which is robust upon small MW changes, represents the upper bound for the free energy difference between native-like structures of monomeric globular proteins. In addition, because there is no condition on the protein other than being monomeric and globular our result for ΔG remain valid for any fold-class, sequence or whether the protein contains, or could form, (quasi) independent domains.

At this point it is worth noting that the observed free energy values (at 25 °C) of microstability (micro-unfolding) determined from nine monomeric globular proteins satisfy the inequality $\Delta G \leq 7.1$ kcal/mol[5] and that the corresponding average free energy of denaturation (macro-unfolding) is $< \Delta G > \cong 11 \pm 3$ Kcal/mol[5]. The plausible value for the upper bound free energy difference ($\Delta G \leq 7.4$ kcal/mol) is, certainly, in line with these results.

4. Conclusions

In summary, based on the use of simple statistical thermodynamics concepts, the Gershgorin theorem and a heuristic argument we have been able to compute a plausible value for the largest free energy difference between coexistent native-like structures of monomeric globular proteins. The computed value of 7.4 Kcal/mol is consistence with the experimentally observed micro-unfolding free energy changes from a set of nine globular proteins.

Considerable attention has been dedicated, during the last 40 years, to develop methods with which to compute the free energy of biological systems accurately. In this regard, the work proposed herein may spur significant progress for the development of new methods for free energy calculations aimed at solving problems of paramount importance such as an unambiguous characterization of the protein folding, misfolding and aggregation.

CRediT authorship contribution statement

Jorge A. Vila: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing.

Acknowledgments

This research was supported by the IMASL-CONICET (PIP-0087), UNSL (PROICO 3-2212), ANPCyT (PICT-0556 & PICT-0767), Argentina. I want to thank HA Scheraga, Todd Professor of Chemistry Emeritus at Cornell University, for his enduring support to the scientific research. I also want to thank to anonymous referees for suggestions that allowed me to improve the paper.

Declaration of competing interest

The author declares no competing interest.

References

[1] S. Piana, K. Lindorff-Larsen, D.E. Shaw, Atomic-level description of ubiquitin folding, Proc. Natl. Acad. Sci. USA 110 (2013) 5915–5920.
[2] P.L. Wintrode, G.I. Makhatadze, P.L. Privalov, Thermodynamics of ubiquitin unfolding, Proteins 18 (1994) 246–253.
[3] P.L. Privalov, Stability of proteins, Adv. Protein Chem. 33 (1979) 167–241.
[4] S. Gershgorin, Über die Abgrenzung der Eigenwerte einer Matrix, Izv. Akad. Nauk USSR Otd. Fiz.-Mat. Nauk 6 (1931) 749–754.
[5] P.L. Privalov, T.N. Tsalkova, Micro- and macro-stabilities of globular proteins, Nature 280 (1979) 694–696.
[6] D. Poland, H.A. Scheraga, Theory of Helix-Coil Transitions in Biopolymers, Academic Press, New York, London, 1970, (and papers therein).
[7] J. Vila, The maximum eigenvalue for a special matrix of order *n*, J. Phys. A 20 (1987) 3353–3365.
[8] J.A. McCammon, B.R. Gelin, M. Karplus, Dynamics of folded proteins, Nature 267 (1977) 585–590.
[9] C.B. Anfinsen, Principles that govern the folding of protein chains, Science 181 (1973) 223–230.
[10] P.L. Privalov, N.N. Khechinashvili, A thermodynamic approach to the problem of stabilization of globular protein structure: A calorimetric study, J. Mol. Biol. 86 (1974) 665–684.
[11] K.A. Dill, Dominant forces in protein folding, Biochemistry 29 (1990) 7133–7155.
[12] D.F. Stickle, L.G. Presta, K.A. Dill, G.D. Rose, Hydrogen bonding in globular proteins, J. Mol. Biol. 226 (1992) 1143–1159.
[13] V. Oklejas, C. Zon, G.A. Papoian, P.G. Wolynes, Protein structure prediction: Do hydrogen bonding and water-mediated interactions suffice? Methods 52 (2010) 84–90.
[14] A.H. Reisner, J. Row, H.M. Macindoe, The largest known monomeric globular proteins, Biochim. Biophys. Acta 196 (1969) 196–206.