ELSEVIER

Biophysical Chemistry 101–102 (2002) 167–171

Biophysical Chemistry

www.elsevier.com/locate/bpc

Methinks it is like a folding curve☆

Rajgopal Srinivasan[a,*], George D. Rose[b,1]

[a]*Jenkins Department of Biophysics, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA*
[b]*Department of Biophysics and Biophysical Chemistry, Johns Hopkins University School of Medicine, 725 N. Wolfe Street, Baltimore, MD 21205, USA*

Received 31 January 2002; received in revised form 18 February 2002; accepted 18 February 2002

Abstract

Most often, the unfolded state of peptides and proteins has been modeled as a statistical random coil. Here, we suggest an alternative model based on the presence of a significant, temperature-dependent conformational bias in the unfolded population. Conformational bias is suggested by our calculations [Proc. Natl. Acad. Sci. USA 96 (1999) 14258–14263], and it is found in recent studies of both proteins and peptides. The imposition of even a modest bias would transform our assessment of the folding problem.

Keywords: Protein folding; Conformational entropy; Random coil

1. Introduction

John Schellman is a charter member of that group of biochemists who taught us that equilibrium thermodynamics is our most powerful tool for understanding protein molecules [1]. Throughout his career, he has used thermodynamic approaches to study fundamental aspects of protein conformation and protein folding. For example, he provided early estimates of the enthalpy and entropy of helix formation [2], and today, half a century later, these estimates resemble modern values. Yet paradoxically, the problems that prompted his measurements are still being vigorously debated, such as the entropy lost upon helix formation [2], whether the peptide hydrogen bond is net stabilizing or destabilizing [3], and the influence of protein–water interactions on protein folding [4].

Why has the field failed to resolve such central questions? Perhaps it is because we still lack adequate models for the coil state of helices and the unfolded state of proteins. With more than 15 000 structures of folded proteins in the Protein Data Bank (PDB) [5], there are many examples of the native state, but the unfolded population has been a black box. The following is an illustration of the way that plausible but conflicting models can influence thought.

☆ Dedicated to John Schellman, with admiration and affection.

*Corresponding author. Tel.: +1-410-614-3970; fax: 410-614-3971.
E-mail address: raj@grserv.med.jhmi.edu (R. Srinivasan),
rose@grserv.med.jhmi.edu (G.D. Rose).
[1] Tel.: +1-410-614-3970; fax: +1-410-614-3971.

0301-4622/02/$ - see front matter
PII: S0301-4622(02)00147-3

2. The importance of models

How does order emerge from disorder in living systems? Can random trial and error followed by selection account for life's awesome complexity? This question was analyzed by Richard Dawkins [6] in the context of evolution. Starting with the proverbial question of how long it would take an ensemble of monkeys typing at random to reproduce the works of Shakespeare, Dawkins uses two contrasting models to calculate the probability that one monkey will type the 28-character phrase: *methinks it is like a weasel* (*Hamlet*, Act III, Scene ii).

To make the question precise, the typewriter is assumed to have only 26 capital letters and a space bar, and the monkey types the complete 28-character string in each successive trial. Then the likelihood that a single trial of 28 characters chosen at random will be correct is $1/27^{28} \approx 10^{-40}$, a low-probability event indeed.

Dawkins calls this procedure *single-step selection*, and he contrasts it with *cumulative selection*, in which correctly typed characters are retained during successive trials. With cumulative selection, the monkey will usually be successful in fewer than 100 trials. A typical run looks like this:

```
0.   X B T P M Z W M K I U L E R L A T T U N B W W K M C O W
1.   S F T Q E O Z D K I K B R Y P K A L G W J C W N S Y P F
2.   U G T O J   M N P I D F V R G S R G P T H H W Z O L Q X
3.   N F T C P X P L S I U W S Q V   T S R C   Z W F Q H C W
4.   C O T G G D T Q G I N E E F X I N H E Q Y C W S H O P G
5.   Z N T E C K B D P I D A W E F W V N E O H T W G U W L P
6.   C S T T S V B O F I S   S B G Q J R E O G F W U I S F X
7.   S V T Y J O N O S I W   U F V W   U E N U Q W O   S   D
8.   A S T V R X N O M I L   E G F C S R E   G T W V U S G
9.   C O T G N R Q D M I B   C P M E D J E   V I W J N S   Q
10.  X K T S B T G U N I L   H M U N V X E   O L W I Z S P V
•
•
•
50.  M E T H I N K S   I T   I S   L I K E   A   W N X S E L
51.  M E T H I N K S   I T   I S   L I K E   A   W G A S E L
52.  M E T H I N K S   I T   I S   L I K E   A   W X A S E L
53.  M E T H I N K S   I T   I S   L I K E   A   W T A S E L
54.  M E T H I N K S   I T   I S   L I K E   A   W U A S E L
55.  M E T H I N K S   I T   I S   L I K E   A   W H A S E L
56.  M E T H I N K S   I T   I S   L I K E   A   W N A S E L
57.  M E T H I N K S   I T   I S   L I K E   A   W K A S E L
58.  M E T H I N K S   I T   I S   L I K E   A   W K A S E L
59.  M E T H I N K S   I T   I S   L I K E   A   W H A S E L
60.  M E T H I N K S   I T   I S   L I K E   A   W K A S E L
61.  M E T H I N K S   I T   I S   L I K E   A   W E A S E L
```

We now extend this *monkey business* to the molecular level in order to examine the surprising degree to which entropy can facilitate organization. Entropy favors randomness in a system with a large number of equally likely states, but it can play the opposite role in a system having a small number of states with even a modest anisotropic bias. The goal of our exercise is to explore a familiar situation in which entropy favors organization.

Imagine a random monkey that operates by single-step selection, and a Dawkins monkey that operates by cumulative selection. Actually, the Dawkins monkey is too extreme for our purposes; upon discovering the correct letter by chance, the

probability of selecting it again in successive trials is unity. Accordingly, we introduce an intermediate level of bias, represented by a Metropolis monkey, in which each letter in the sequence is selected with some intermediate probability, P, i.e. $1/20 < P < 1$, for the 20 naturally occurring amino acids. We investigate the temperature-dependent behavior of the three monkey systems for the model nonapeptide: Ser–Cys–His–Glu–Leu–Leu–Met–Ala–Asn.

Starting from a randomly chosen nonapeptide, we simulate the search process for the exemplary sequence SCHELLMAN (in one-letter code), as performed by the three monkeys. A Metropolis Monte Carlo [7] search is performed using a simple scoring function, in which each correct residue is worth one point, each incorrect residue is worth zero points, and the Metropolis criterion for two sequences is given by: $e^{-\beta(E_1-E_2)} = e^{-\beta\Delta E}$, where $\beta = 1/RT$, $R = 1.987$ cal/mol K, and T is nominal temperature in K. The energy, E, of a sequence is:

$$E = \sum_{i=1}^{9} \text{points(sequence}(i))$$

The curves in Fig. 1 represent a family of simulations. Each curve plots the fraction of successes in which the monkey finds the correct sequence in 1 000 000 trials as a function of RT. The right-most curve corresponds to a random monkey, who chooses among equi-probable residues, each with likelihood $1/20$. The random monkey has no successes at 300 K ($\beta \approx 2$), but tends toward the temperature-independent Dawkins monkey as temperature decreases. The next five curves form a cluster of Metropolis monkeys, in which a bias toward the correct sequence is introduced via multiple copies of the correct residue. Here, the probability of choosing a residue correctly is not $1/20$, but $(c)/(19+c)$, where c is the number of copies of the residue in question. The curves correspond to the series $c = 1,2,4,6,8$. Finally, the remaining curve (in bold) is a Metropolis variant for $c = 8$, in which the correct sequence is rewarded with a co-operative bonus point, i.e. the scores are integers ranging from 0 to 10, but without 9.

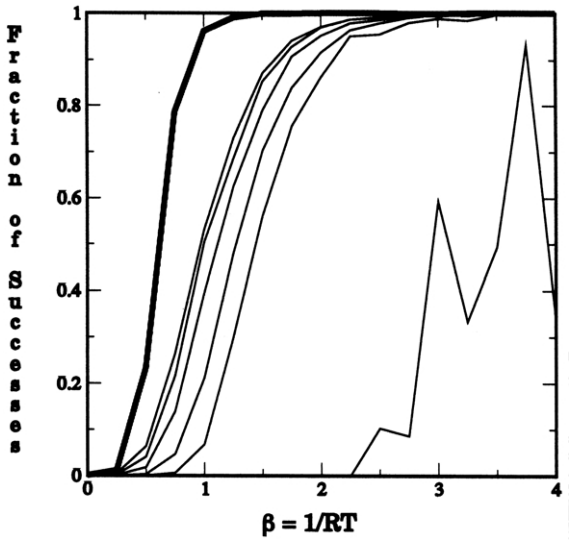

Fig. 1. The search for SCHELLMAN. Curves plot the fraction of successes against $\beta = 1/RT$. At a nominal temperature of 300 K, $RT \approx 0.6$ kcal/mol and $\beta = 1.7$. To draw each curve, β was varied in increments of 0.25, with 10^6 trials at each sample point. In the right-most curve, residues are chosen with equal probability, and the likelihood of at least one success in 10^6 trials is negligible above $\sim$225 K. As temperature decreases, results tend toward the temperature-independent Dawkins procedure, which is realized in the 0 K limit. Bias is introduced via multiple copies of the correct residue at each sequence position. The next cluster of five curves represents biased searches; progressing from right to left within the cluster, they correspond to $c = 1$, 2, 4, 6 and 8 copies, respectively. The final curve (in bold) represents a biased search with $c = 8$ and in which a 'co-operative' bonus point is awarded for the correct sequence. As bias increases, the curves increasingly resemble a classic two-state folding curve, especially so upon the inclusion of a small, co-operative energy term.

In the scene in *Hamlet* that supplies the phrase for Dawkins' exercise, Hamlet and Polonius lie on their backs staring up at the clouds, free-associating:

Hamlet: Do you see yonder cloud that's almost in the shape of a camel?

Polonius: By the mass, and 'tis like a camel, indeed.

Hamlet: Methinks it is like a weasel.

Polonius: It is backed like a weasel.

Hamlet: Or like a whale?

Polonius: Very like a whale.

In a similar vein, the co-operative variant in Fig. 1 is very like a two-state folding curve [8].

But science is more than metaphor, or should be. The Metropolis cluster in the figure illustrates how even modest bias, with no explicit energy terms, can convert a refractory search problem into a tractable procedure. Is a corresponding bias intrinsic to biophysical problems of interest, or is this merely a simulation artifice?

3. From sequence to structure

To explore further the family of curves in Fig. 1, we shift from sequence to structure. The classic two-state folding curve [8] describes the emergence of the native population, N, from the ensemble of unfolded conformers, U, in the thermodynamic reaction $U \rightleftarrows N$, but thermodynamics cannot supply the underlying mechanism. The prevailing model of the unfolded state [9] represents it as a statistical random coil, in which the backbone conformation of each residue is limited only by the steric constraints of dipeptide geometry [10] and conformational space grows exponentially with chain length, subject to long-range excluded volume considerations [11]. Accordingly, the number of conformations accessible to a polypeptide backbone—even a short one—can be huge, resulting in a seemingly intractable search problem. This view and its consequences are often invoked by alluding to the 'Levinthal paradox' [12].

For Levinthal, this conundrum was a demonstration, not a paradox. If the size of accessible conformational space is as large as implied by the random coil model, the native state cannot be found by unguided search in biological real-time. Therefore, additional constraints must exist. Levinthal's inference spawned a growth industry (e.g. [13–16]).

In our own work, these constraints take the form of conformational bias toward native-like secondary structure [17]. The bias is sterically based and sequence-specific, and it serves to pre-organize the folding process. Exhaustive analysis of short polyalanyl chains [18] shows that contracted backbone conformations are limited by additional, systematic steric clashes not observed in a dipeptide map, restricting the energy landscape to just a few, highly similar regions in multidimensional ϕ,ψ-space [19]. As solvent squeezes the polypeptide chain from its midst, only this limited set of conformational alternatives is available to be stabilized by intrachain interactions.

A recent NMR and CD study [20] of a blocked polyalanine heptamer found that the 'coil' state is predominantly left-handed polyproline II helix at 2 °C and approximately 90% polyproline II and 10% β-strand at 55 °C. These data prompt re-evaluation of the helix → 'coil' transition as an interconversion between one structural state and another, and they provide a value of the per residue entropy change on helix formation as -2.20 ± 0.37 eu. This value is close to Schellman's lower bound of 3 eu [2] and substantially smaller than most contemporary estimates [21]. With this revised number, the entropic cost of helix formation at physiological temperature (300 K) would be 0.66 kcal/mol, approximately *RT*.

The Kallenbach experiment [20] demonstrates that the conventional random coil model significantly overestimates the entropy loss on peptide helix formation because it fails to account for conformational bias in the non-helical population. This realization also changes our picture of protein folding, because polyalanine is an effective model for the protein backbone. The revised picture is consistent with recent work on unfolded staphylococcal nuclease using residual dipolar couplings, which found that the unfolded state retains structure in 8 M urea [22].

We suggest that conformational bias in the unfolded state of peptides and proteins plays a crucial organizing role by substantially reducing the entropic cost of structure formation. A little entropic monkey business can transform an otherwise intractable search problem into an accessible one.

Acknowledgments

This work was supported by grants from the NIH and the Mathers foundation.

References

[1] R.L. Baldwin, John Schellman and his scientific work, Biophys. Chem., in press.

[2] J.A. Schellman, The stability of hydrogen-bonded peptide structures in aqueous solution, C. R. Trav. Lab. Carlsburg Ser. Chim. 29 (1955) 230–259.
[3] J.A. Schellman, The factors affecting the stability of hydrogen-bonded polypeptide structures in solution, J. Phys. Chem. 62 (1958) 1485–1494.
[4] J.A. Schellman, 50 years of solvent denaturation, Biophys. Chem., in press.
[5] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The Protein Data Bank, Nucleic Acids Res. 28 (2000) 235–342.
[6] R. Dawkins, The Blind Watchmaker, W.W. Norton and Co, London, 1986.
[7] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, Equation of state calculations by fast computing machines, J. Chem. Phys. 21 (1953) 1087–1092.
[8] A. Ginsburg, W.R. Carroll, Some specific ion effects on the conformation and thermal stability of ribonuclease, Biochemistry 4 (1965) 2159–2174.
[9] A.R. Dinner, M. Karplus, Comment on the communication 'The key to solving the protein-folding problem lies in an accurate description of the denatured state' by van Gunsteren et al., Angew. Chem. Int. Ed. 40 (2001) 4615–4616.
[10] G.N. Ramachandran, V. Sasisekharan, Conformation of polypeptides and proteins, Adv. Protein Chem. 23 (1968) 283–438.
[11] K.A. Dill, Dominant forces in protein folding, Biochemistry 29 (1990) 7133–7155.
[12] C. Levinthal, How to fold graciously. Mossbauer spectroscopy in biological systems, Proc. Univ. Illinois Bull. 41 (1969) 22–24.
[13] A. Sali, E.I. Shakhnovich, M. Karplus, How does a protein fold?, Nature 477 (1994) 248–251.
[14] J.D. Bryngelson, J.N. Onuchic, N.D. Socci, P.G. Wolynes, Funnels, pathways and the energy landscape of protein folding—a synthesis, Proteins Struct. Funct. Genet. 21 (1995) 167–195.
[15] K.A. Dill, H.S. Chan, From Levinthal to pathways to funnels, Nat. Struct. Biol. 4 (1997) 10–19.
[16] W.F. van Gunsteren, R. Burgi, C. Peter, X. Daura, The key to solving the protein folding problem lies in an accurate description of the denatured state, Agnew. Chem. Int. Ed. 40 (2001) 351–355.
[17] R. Srinivasan, G.D. Rose, A physical basis for protein secondary structure, Proc. Natl. Acad. Sci. USA 96 (1999) 14258–14263.
[18] R.V. Pappu, R. Srinivasan, G.D. Rose, The Flory isolated-pair hypothesis is not valid for polypeptide chains: implications for protein folding, Proc. Natl. Acad. Sci. USA 97 (2000) 12565–12570.
[19] R.V. Pappu, G.D. Rose, 2002. A simple model for polyproline II structure in unfolded states of alanine – based peptides, Protein Science, in press.
[20] Z. Shi, C.A. Olson, G.D. Rose, R.L. Baldwin, N.R. Kallenbach, 2002. Polyproline secondary structure in a seven-alanine residue peptide, in press.
[21] J.A. D'Aquino, J. Gomez, V.J. Hilser, K.H. Lee, L.M. Amzel, E. Freire, The magnitude of the backbone conformational entropy change in protein folding, Proteins 25 (1996) 143–156.
[22] D. Shortle, M. Ackerman, Persistence of native-like topology in a denatured protein in 8 M urea, Science 293 (2001) 487–489.