book reviews

CrossMark

Myths and Verities in Protein Folding Theories. By Arieh Ben-Naim. World Scientific, 2015. Pp. 448. Hardback. Price GBP 45.00. ISBN 978-981-4725-98-9.

Michael Blaber*

Florida State University, College of Medicine, 1115 West Call Street, Tallahassee, FL 32306-4300, USA. *Correspondence e-mail: michael.blaber@med.fsu.edu
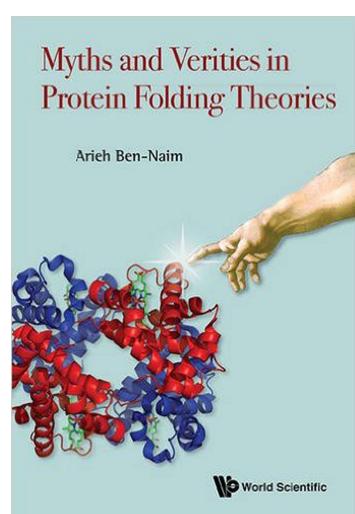
The inside of the front cover of this book states: 'This book is dedicated to the countless students and researchers, who wasted so much time, effort and funds, in a futile search of a solution to the protein folding problem in the wrong direction. And to all new students and researchers who, upon reading this book, will be saved from wasting time, effort and funds . . .'.

The book *Myths and Verities in Protein Folding Theories* by Arieh Ben-Naim might well have been titled *Protein Folding Myths – Why the Emperors have No Clothes* since the author appears to relish the opportunity to associate his criticisms of protein folding theories specifically with the investigators that published on the subject.

In Chapter 1, the *Introduction*, the belief is laid out that one cannot expect to predict the structure, let alone the dynamics and mechanism of the folding process, by reading the sequence of a protein. The author argues that the hydrophobic effect is ill defined and cannot explain the stability or kinetic problem of protein folding. However, he states that the hydrophilic effect can do both. He outlines that there are five myths associated with the protein folding problem (PFP), and these myths are the reason why the PFP is still considered to be one of the most challenging unsolved problems in molecular biology – and why searches for solutions to the PFP are going in the wrong direction.

The entropy, enthalpy and heat capacity change associated with (de)solvation of a nonpolar solute in water is discussed in detail in Chapter 2 (*From Frank and Evans' Iceberg Formation Conjecture to the Explanation of the Hydrophobic Effect*). A logical argument is given whereby nonpolar solute interactions are not likely to introduce novel (ice-like) structure to water solvent; rather, nonpolar solute interactions are only able to shift the equilibrium of pre-existing water structure. This argument is supported by diagrams and relevant equations. The figures, however, are inconsistent in the structural interpretations necessary to support the idea presented. Specifically, initial figures represent water in stick representation (one oxygen, two hydrogens) and the overall dipole. This is used to indicate how water will arrange around charged ions – correctly indicating orientation of the water dipole, and thus, the basis of entropic changes in solvent upon solvation of ions. Subsequent to this is a discussion of similar solvation of a non-polar solute (argon atoms). However, in this case, the diagram is less rigorous showing unsatisfied donors/acceptors pointing towards the solute. Subsequent figures show water as tetrahedral but with no distinction between hydrogen and oxygen lone pairs. These latter figures are used to illustrate low local density of solvent (associated with an organized lattice structure), but such figures are ambiguous as regards donor/ acceptor interactions. The discussion of Chapter 1 entirely omits any consideration of solvent structure to limit unsatisfied donor/acceptor groups orienting towards solute (an energetically costly organization when solvating nonpolar solute – and a potential basis of increased structure in water). The heat capacity of ice is significantly less than that of liquid water. Yet when nonpolar groups are solvated the heat capacity increases. Thus, an 'iceberg' definition of solvent structure forming around nonpolar solutes cannot be taken literally – it is a reference to an organized clathrate of water that minimizes destabilizing effects of unsatisfied hydrogen-bond groups. The author, as an expert in solvent structure, is clearly perturbed by the use of this metaphor; or rather, the broad lack of recognition as a metaphor. Rather, we are admonished to consider that whatever the solvent structure is around nonpolar solutes it is a pre-existing structure of solvent whose mole fraction increases. Furthermore, we are to accept that solvation of polar/ionic solutes can

increase the solvent structure (and associated entropy, enthalpy and heat capacity) just as equally as nonpolar solutes; thus, such effects when observed upon protein unfolding are not *de facto* due to exposure of hydrophobic groups. It would be easy to describe this chapter as quibbling about semantics (when is an 'iceberg' structure of water not the same thing as 'low local density'?), but the author appears to be attempting to place the narrative on a firm logical footing. However, how this issue is a 'myth' that is preventing a solution to the protein folding problem remains unclear.

The significance of the formation of hydrogen bonds to the free energy associated with the formation of protein structure (secondary and tertiary) is discussed in Chapter 3 (*From Schellman's Experiments to Fersht's Hydrogen Bond Inventory Argument*). This discussion of the importance of hydrogen bonds is directly contrasted with the importance of the 'hydrophobic bond'. John Schellman's early studies of intramolecular hydrogen bonds in urea, and the stoichiometry of this desolvation process, forms the basis of a nullifying argument for the significance of hydrogen-bond formation due to the resultant overall hydrogen-bond inventory (referenced by the author as championed by Fersht). The essence of the hydrogen-bond inventory argument is that hydrogen-bond formation in the process of protein folding is a zero-sum gain as regards both stoichiometry and energetics. As far as the protein is concerned, folding involves the breaking of hydrogen bonds with water (resulting in their release into bulk solvent) and the formation of novel intramolecular hydrogen bonds. As regards hydrating water molecules, folding involves breaking of hydrogen bonds with protein and formation of novel water⋯water hydrogen bonds with bulk solvent. As originally described, this overall process is of balanced stoichiometry as regards number of hydrogen bonds broken and formed, and also energetically (with the premise that bulk solvent hydrogen bonds, water⋯protein hydrogen bonds, and intramolecular hydrogen bonds involve equivalent Gibbs energy). The chapter could benefit from an introductory description of bulk solvent, particularly since the author distinguishes between 'true H-bond' formation between polar groups (with inferred energetic significance) and mere 'solvation' of polar groups by solvent (with an apparently less significant energetic magnitude). For the student to appreciate this argument, more explanation is needed. Such an explanation is touched upon in a discussion of number of possible hydrogen bonds by amides, carbonyls, and hydroxyls (*i.e.* 1, 2 and 3, respectively), and solvent (*i.e.* 4), and how equations of hydrogen-bond formation and desolvation must consider such effects. Such consideration identifies a favorable energetics of $\sim$1.5–2 kcal mol^{-1} for desolvation/intramolecular hydrogen-bond formation; which, when summed over the total number of such hydrogen bonds in a protein is significant. A chemical potential argument is provided to support the proposition that the entropic change in released solvent associated with hydrogen-bond formation in protein folding is negligible. An error in the stoichiometric equations used in support of the hydrogen-bond inventory argument is identified, stating that their complete description requires a solvation process from

an ideal gas. Doig & Williams (1992) described such a process for urea and identified an entropically driven (*i.e.* solvent entropy) urea dimerization (*i.e.* hydrogen-bond formation accompanied by desolvation) process. This is a highly relevant reference to address in this chapter (which is unfortunately omitted). Minor issues that can confound students is that there are switches between kJ mol^{-1} and kcal mol^{-1} units in places in this chapter. The author appears to forgive Schellman (based upon personal discussions) for contributing to the 'myth' that hydrogen bonds in proteins contribute little to stability, but holds Fersht fully accountable.

In Chapter 4 (*From Kauzmann's Conjecture to the Myth that the Hydrophobic Effect is the Dominant Factor in Protein Stability*) it is stated that the postulate that the hydrophobic effect is dominant in protein folding is a myth. To be clear, the author states 'It was only in 1989 that I realized that both the solvation of hydrophobic molecules and the pairwise potential of mean force between hydrophobic molecules are not only not dominant in protein folding, but in fact are irrelevant to protein folding. Instead, the solvation, the hydrogen bonds and the pairwise potential of mean force between hydrophilic groups are far more important'. The origin of the hydrophobic myth is traced back to a report titled *Some Factors in the Interpretation of Protein Denaturation* by Kauzmann (1959) wherein it was stated 'Hydrogen bonds, taken by themselves, give a marginal stability to ordered structures . . . The hydrophobic bond is probably one of the more important factors involved in stabilizing the folded configuration in many native proteins'. The early evidence for the free energy contribution associated with desolvation of hydrophobic groups, in support of Kauzmann's postulate, was generated from aqueous/organic solvent transfer experiments. There is criticism of such results being applied to protein folding because they do not include polypeptide backbone effects, which significantly modulates the free energy of such transfer. Indeed, the author states 'Before one can claim anything about the relative importance of one factor over another in maintaining the stability of proteins, one must have at least a complete inventory of all possible factors which contribute to the stability of proteins.' This suggests that meaningful transfer studies must include the polypeptide backbone effects, tacitly emphasizing the need for hydrophilic/hydrophobic mutagenesis studies in real proteins. However, the author is highly critical of interpretations of mutagenesis studies also, stating 'However, I do not believe that by studying mutated proteins one can reach any meaningful conclusion regarding the relative importance of the different types of interactions. Each mutation can cause changes in the many types of interactions'. Thus, we are left with a dilemma regarding how to move forward to address such fundamental questions, and must assume that the only avenue open is theoretical work. This chapter suffers from poor figures with incomplete and absent legends. Use of color in the legends (presumably distinguishing hydrophobic/hydrophilic groups) is not explained. In one figure an Ala to Gly mutation is used as an example of a hydrophobic deletion mutation; however, this is a poor illustrative example as it also substantially alters the

conformational entropy. Other figures are confusing in their use of arrows on molecules, it is not clear if such arrows are intended to show a dipole or a hydrogen-bond donor. Unfortunately, neither interpretation is consistent with the intermolecular interactions of the figures; thus, the reader is left to guess their meaning (with the conclusion that logically, the arrow has no meaning). The most unsatisfying aspect of this chapter is the complete omission of discussion of the dielectric, its modulation by hydrophobic groups, and its consequences for Gibbs energy of hydrophilic interactions. Hydrophilic and hydrophobic interactions are treated as completely separable interactions (with the conclusion that the former are key to folding and the latter are inconsequential); however, consideration of the dielectric suggests key interplay between these interactions, and they are thus not as neatly separable as suggested. The most interesting section of this chapter deals with protein–protein association. However, this is little more than a tantalizing introduction. Left out of the discussion is sequence analysis data indicating conservation of buried hydrophobic amino acids, and the basis of cold denaturation (which points to a key role for hydrophobic amino acids in protein stability). The reference frame for folding solvent appears to be essentially pure water; high salt or extremes of pH are not considered. High salt favors hydrophobic interactions and shields electrostatic interactions, and can have a major influence on the relative strengths of hydrophobic and hydrophilic interactions. Such solvent conditions may be considered 'extreme' and therefore not representative of most protein folding; however, such conditions may have been where the origin of protein folding occurred. Thus, the arguments in this chapter appear to incompletely consider varied environmental conditions key to the evolution of protein folding. Another side point is that it is likely that co-evolution occurred between the observed protein folds and the 20 common amino acids. Thus, the common protein architectures satisfy main-chain hydrogen-bond requirements and their cores can be efficiently packed using a combination of the particular hydrophobic amino acids in the set of 20 common amino acids. Thus, the hydrophobic amino acids are not as useless in protein folding as suggested.

Chapter 5 is entitled *From Levinthal's Question to Resolving Levinthal's 'Paradox'*. In the early 1970s Cyrus Levinthal concluded that protein folding cannot involve a 'random walk' conformational search (since there are too many degrees of conformational freedom), and there must be some local interactions that rapidly guide the folding process towards the native structure. This has historically been framed as Levinthal's 'Paradox', although as is pointed out, Levinthal himself never considered it a paradox. Clearly, there must be strong forces comprising the local interactions that guide and speed correct folding. According to Ben-Naim these local forces have historically been considered to be hydrophobic in origin – and this viewpoint is considered to be an egregious error on the part of many researchers. Rather, solvent interactions with hydrophilic groups of the protein are held by the author to constitute the fundamental strong force driving efficient protein folding. The 'paradox' having been solved (or rather,

never having existed) the large body of literature continually referring to or revisiting the paradox is a monumental waste of effort – effort better spent on identifying the strong force driving local interactions that guide folding. Several notable researchers are, once again, pilloried for perceived sloppiness of language or logic in work related to protein folding energy landscapes and the process of evolution. However, with notable restraint, the majority of the chapter is saved for an analysis of solvent hydrophilic interactions and their primacy in protein folding. The solvent-induced force is defined by the gradient of the solvation Gibbs energy of the protein at a given conformation. Several figures are used to explain this point; however, as is common throughout the book, such figures lack a useful legend and in some cases are incomplete and prone to ambiguity or misinterpretation. In the narrative of this chapter hydrophilic and hydrophobic groups are treated as completely separable entities, having strong (former) and weak (latter) energetic interactions. This is too simplistic a viewpoint considering that the strength of hydrophilic interactions is strongly modulated by the local dielectric – which is principally influenced by the nonpolar (*i.e.* hydrophobic) nature of neighboring groups. Thus, there is likely an inseparable cooperative interaction between these groups. Local interactions that drive correct folding typically include a combination of hydrophobic collapse and hydrophilic interaction – it is therefore likely that the strength of such hydrophilic interactions is increased by local low dielectric provided by hydrophobic collapse. Additional figures in this chapter (again, with legends too terse to be of use) appear to show conformations along a folding pathway that involve hydrogen-bond interactions passing through a solvent-excluded (*i.e.* hydrophobic) core. One must assume the dielectric of such excluded cores is critical to the strength of such hydrogen bonds, and also principally hydrophobic in nature. In this regard, dismissal of the energetic importance of the hydrophobic bond appears energetically untenable. Furthermore, earlier in the book, the idea that a 'folding code' will be found in the primary structure of a protein is disdained; however, in this chapter it is stated that 'the pattern of the hydrophilic groups determines the folding pathways of the protein'; and later on, 'Clearly the specific pattern of amino acids will determine the specific trajectory of the protein in its configurational space'. Such comments appear to be inconsistent with earlier statements.

Chapter 6, *From Anfinson's Hypothesis to the Frenetic Pursuit of the Global Minimum in the Gibbs Energy Landscape* is one of the more engaging chapters in the book, and is well written and clear. Anfinson demonstrated reversible folding of ribonuclease A. This result led to a thermodynamic description of folding whereby the folded structure is determined by the global Gibbs energy minimum of the system (protein plus solvent). This naturally led to the hypothesis that the three-dimensional structure of the folded protein was determined by the amino-acid sequence; thus, successful identification of the global minimum of the Gibbs energy landscape should identify the folded structure. The author cautions that the search for such a solution omits a key aspect

of the folding process, namely, folding kinetics restrictions due to features of the energy landscape. It is not sufficient to merely design a native structure to reside on the minimum of the Gibbs energy landscape – the sequence must also fold with useful kinetics (*i.e.* be able to efficiently traverse downhill in the landscape). At each stage in the folding pathway there must be a pattern of forces that guide and speed the folding process. This latter requirement is viewed as being more cryptic, more difficult to design, and more difficult to decipher from the primary structure, than the Gibbs energy minimum requirement. Critically, successful identification of the Gibbs energy minimum for a polypeptide chain will not necessarily identify the most populated state after folding since this depends upon the landscape. Briefly, if the denatured state comprises a random distribution of conformations, then their pathway to a lower Gibbs energy state is at the mercy of each conformation's local energy landscape (*i.e.* there may be a significant energy barrier for certain denatured conformations that trap them in a local energy well). Depending upon the energy landscape it may only be a minority of molecules that find their way to the correct native state (not necessarily residing at the global minimum) and the majority of molecules may reside within a different local energy minimum (which may be the global minimum). The author feels that a misunderstanding of Anfinson's hypothesis has led to a quixotic effort in the search for the Gibbs energy minimum. This is a bit too acerbic, as such efforts in computational protein folding have driven the development of ever more accurate energy functions and improved solvent models (these and related advances move the field forward). An analysis of the greater underlying complexity of folding leads to a conclusion that the search for a 'folding code' within the amino-acid sequence is futile. This is a defeatist idea that one must not accept – it is, perhaps, an acknowledgement of an additional layer of complexity that must be traversed to arrive at the ultimate solution. The logic laid out regarding complex energy landscapes is intriguing from the standpoint of protein evolution. Basically, the complexity of energy landscapes is directly proportional to polypeptide length. Thus, in protein evolution (progressing from simple short polypeptides to increasingly larger polypeptides *via* gene fusion) logic suggests that the earliest (*i.e.* smallest) foldable polypeptides had a greater likelihood of actually having the folded structure residing at the global Gibbs energy minimum. Thus, the need to encode 'foldability' evolved simultaneously with increasing polypeptide length.

In Chapter 7, *Some Candidates Which Can Potentially Evolve Into New Myths*, the author discusses terms, phrases and concepts recently featured in the literature of protein folding, that might be described as empty words bereft of scientific content. Among such topics is the 'landscape theory' of protein folding. Precision is critical in defining terms, and the landscape theory suffers from ambiguity in the meaning of 'landscape'. Energy landscape, Gibbs energy landscape, thermodynamic potential energy landscape, and potential energy surface, are all possible candidates for 'energy landscape'. Each term has a precise meaning; unfortunately,

authors in the field of energy landscape are not clear in their definition of landscape. Sections in this chapter detail the differences between the different energy terms and their associated 'landscape'. Of particular importance is the Gibbs energy landscape (GEL). All conformations, even those not realizable, are part of the GEL. Unlike the energy landscape (EL) the solvation Gibbs energy is not a pairwise additive function – an approximation shared by all lattice models of protein folding. In brief, the shape of the GEL is unknown for large proteins; thus, the 'general shape' of the GEL cannot be used as a basis to understand the protein folding problem (furthermore, only a small portion of the overall GEL is relevant to the process of protein folding). Solvation interaction energies are critical to understanding protein folding and are included in the GEL but not the EL. In the discussion of landscape theory the EL is often referred to when the GEL is actually intended (the EL is relevant for the protein in vacuum). In some cases a correction is applied to the EL to approximate the GEL (the correction being the potential energy surface, PES). Such correction requires knowledge of the solvation Gibbs energy; however, this is an unknown. Thus, it is argued that the PES is not an approximation of the GEL, and is therefore irrelevant to protein folding. The key point is that only the GEL is relevant to protein folding (as it describes all forces) but no one knows what the GEL looks like. Furthermore, it is pointed out that the GEL is a function, not a theory. Since the explicit form of the GEL is unknown, the 'landscape theory' is also unknown, and cannot be used to explain protein folding. This narrative leads into a discussion of the 'folding funnel', and the 'new view' and 'old view' of protein folding. The old view is associated with a single pathway and the new view is associated with multiple pathways of folding. Potential confusion in the literature is also identified regarding folding pathways and whether they describe a molecular pathway or a kinetic pathway. The main conclusion of this chapter is that it is the GEL that is critical to understand protein folding, only a small region of the total GEL is relevant to protein folding, no one knows what the energy landscape is of the GEL (due to its complexity), thus a funnel metaphor has no theoretical basis (thus, any funnel landscape is unrealistic). Later in this chapter the principle of consistency is also criticized. This principle, expressed by Go in 1983, states that in the folding process long range and short range interactions must be consistent (*i.e.* cooperative) in the folding process. Furthermore, this is interpreted in terms of a 'folding code' contained within the primary structure. At this point, the possibility of different solvent conditions, and their consequence upon protein structure is invoked. Thus, a folding code depends upon the environment, and there may be infinite possible codes for infinite possible environments. In other words, if a folding code exists, there must be separate codes to define an α-helix in high salt, low salt, acidic, basic, high temperature, low temperature, conditions and so on. The author's position is that such 'codes' simply don't exist (and it is a quixotic exercise to search for them). There seems to be a suggestion of information theory inherent to this argument, but it is not formalized further. In other words, as information

theory (*i.e.* conformational space) was used by Levinson to suggest folding pathways, information theory could be employed to exclude the possibility of a folding code. Lattice models and their utility (or lack thereof) in understanding protein folding are then criticized. The principle objection appears to be the additive nature of interaction energies intrinsic to lattice models, and this is not a feature of the critical GEL that determines protein folding. The principle of minimal frustration is discussed in a detailed section of this chapter. The term was borrowed from the theory of spin glasses. The literal description of frustration is not an applicable term to protein folding; and so, frustration requires a formal definition to be relevant for the protein folding problem. Unsurprisingly, perhaps, the author believes this is not possible. Any effort to define frustration *via* a folding code (*e.g.* proteins computing their own structures) is dismissed as previously described. Neither can energy landscapes be used to define or explain frustration – since details of the GEL are unknown, and if other energy terms are used they omit the key solvation interaction term. Thus, frustration is considered an ill-defined term, formulated upon faulty logic, and with no defined connection to the Gibbs energy minimum. The chapter is concluded with a critique of the structure–function paradigm (*i.e.* in order to function proteins must have a well defined three-dimensional structure. To critique this, one must have a definition of both structure and function. The existence of intrinsically disordered proteins (IPD) is given as an example that negates the structure–function paradigm. However, if such proteins function by adopting structure upon binding a target molecule (*e.g.* a cognate protein), then their function is dependent upon (induced fit) structure. Furthermore, protein structure is arguably defined by primary, secondary and tertiary structure. Thus, in the extreme case, even for IDP's, if scrambling their primary structure obliterates their functional role, then there is a structure–function relationship (dependent upon primary structure). If 'function' is defined as effects upon viscosity or osmolality of a solution due to specific IDP's then there may always be a functional relationship.

The book is interesting reading, and some material is enlightening, and certainly entertaining. Its utility for students is in recognizing that even experts in the field can hold strong and opposing views. This may be disconcerting, as it takes away the safety net of accepted paradigm, and forces us to think and draw conclusions for ourselves. For students, this alone is worth the read.

References

Doig, A. J. & Williams, D. H. (1992). *J. Am. Chem. Soc.* **114**, 338–343.
Kauzmann, W. (1959). *Adv. Protein. Chem.* **14**, 1–63.