Current Opinion in Colloid & Interface Science 18 (2013) 502–509

Contents lists available at ScienceDirect

Current Opinion in Colloid & Interface Science

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

Solvent-induced forces in protein folding reflections on the protein folding problem

Arieh Ben-Naim

Department of Physical Chemistry, The Hebrew University of Jerusalem, Edmond J. Safra Campus, Givat Ram, Jerusalem 91904, Israel

ARTICLE INFO

Article history:
Received 28 July 2013
Accepted 4 November 2013
Available online 7 November 2013

Keywords:
Protein folding
Hydrophobic
Hydrophilic forces

ABSTRACT

It is shown that the solvent induced forces on hydrophilic groups are the strongest ones. The relevance of this finding to protein folding is discussed.

1. Introduction

In 2005, the editors of science listed 125 unsolved "What we don't we know?" questions in science. One of these is: [1]

> "Can we predict how proteins will fold? Out a near infinitude of possible ways to fold a protein picks one in just ten microseconds"

There are essentially three main problems which together comprise the so-called protein folding problem (PFP): [2]

1. Can we predict the 3D structure of a protein given its sequence of amino acids?
2. What are the main factors that determine the stability of 3D structure of a protein?
3. What are the main factors that determine the speed of the folding process?

There is another, not less a mystery, which is rarely mentioned as a problem. What are the factors that determine the solubility of globular proteins? [3] I will start by stating in advance my answers to these questions:

The first question does not have a general answer. People have concluded from Anfinsen's thermodynamic hypothesis [4], that there might be some "code" that translates from a sequence of amino acids into a 3D structure [5,6]. Unfortunately, such a code does not exist. The answer to the second question was originally suggested by Kauzmann [7]. It is now commonly believed that the so-called hydrophobic ($H\phi O$) effect [8,9] is the dominant factor which stabilizes the 3D structure of the protein. Unfortunately, the $H\phi O$ effect was over-exaggerated. It was recently found that various hydrophilic ($H\phi I$) effects are far stronger than the corresponding effects. Therefore, one can safely conclude that these effects provide a general answer to the second question [9–12].

The third question has a straightforward answer. The solvent-induced $H\phi I$ forces provide the answer to the question posed by Levinthal [13–16]. We shall briefly discuss these forces in Section 4.

Finally, the high solubility of globular protein is also determined to a large extent by various $H\phi I$ interactions between $H\phi I$ groups which are exposed to the solvent.

2. The prediction of the 3D structure

This question has been the focus of many researchers in the field. Briefly, the problem may be stated as:

> "I give you a sequence of amino acids and you give me in return a 3D structure."

Sometimes this problem is formulated in terms of a search for a "code," a "folding code," or a "thermodynamic folding code." This "code" is also mentioned in connection with the genetic code: [5,6], referring to it as the "second half" or the "second quarter". This connection strongly suggests that one can translate from the language of sequence of amino acids to coordinates of a 3D structure, similar to the translation from a sequence of bases in DNA, to a sequence of amino acids in proteins. The very idea that there is such a code comes from Anfinsen's statement: [4]

> "… that is, that the native conformation is determined by the totality of inter-atomic interactions and hence by the amino acid sequence, in a given environment."

E-mail address: arieh@fh.huji.ac.il.

1359-0294/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cocis.2013.11.001

Clearly, Anfinsen emphasized the appropriate "environment". Unlike the genetic code which translates from a short "phrase" of bases to an amino acid, which is independent of the environment (presuming that that specific phrase is translatable into an amino acid), the expected code from protein sequence to structure depends on the environment. It is known that many proteins (e.g. hemoglobin, regulatory enzyme) fold into different structures under different environments (e.g. different concentrations of oxygen). This means that there exists no unique "code" from a "phrase" in the protein to a "phrase" in a 3D structure [3].

If such a "code" existed then one could design a new protein which will perform a new job. In my view this is wishful thinking. Achieving such a goal will be as impossible as designing a new animal by reverse engineering, i.e., from a required phenotype, design the corresponding genotype.

Most people in the field extend the meaning of a "code" to include a computer algorithm, which when fed with a sequence as input, will provide a 3D structure as an output. In principle, this can be done. However, in such a case one will need a huge code-book, or dictionary containing some 20^{300} entries. In addition, one would need infinite number of such dictionaries, each pertaining to a given environment: T, P, N, i.e. for each temperature, pressure and compositions.

Clearly, such a "code" would not be found, not in the near, nor in the very far future. Nevertheless, many scientists spend huge amounts of effort (time and money as well) to find such a miraculous algorithm. In fact, many scientists are searching for an algorithm to find the global minimum of the Gibbs energy landscape (GEL) in the hope that the 3D structure resides at this global minimum [17]. Finding the global minimum in the GEL is believed to be tantamount to solving the predictability problem; feed the program with a sequence — find the location of the global minimum in the GEL, and you got the 3D structure corresponding to the input-sequence.

Unfortunately, such an algorithm even if found will be useless. There is no guarantee that the 3D structure of the protein resides in the global minimum. This is especially true for very large proteins with bulky side chains. In fact, it was recently argued that it is extremely unlikely that the 3D structure resides at the global minimum. Yet, people continue to search for such an algorithm. If you go back in time you find the origin of the motivation for searching for such an algorithm is in Anfinsen's conclusion [4].

> "The studies on the renaturation of fully denatured ribonuclease required many supporting investigations to establish, finally, the generality which we have occasionally called the 'thermodynamic hypothesis.' This hypothesis states that the three-dimensional structure of a native protein in its normal physiological milieu (solvent), pH, ionic strength, presence of other components such as metal ions, or prosthetic groups, temperature, and other is the one in which the Gibbs free energy of the whole system is lowest; that is, that the native conformation is determined by the totality of inter-atomic interactions and hence by the amino acid sequence, in a given environment. In terms of natural selection through the 'design' of the macromolecules during evolution, this idea emphasized the fact that a protein molecule only makes stable, structural sense when it exists under condition similar to those for which it was selected, the so-called physiological value."

In this quotation Anfinsen refers to the thermodynamic hypothesis, and the free energy lowest value. No doubt he meant the Second Law of Thermodynamics which states that at constant T, P, N, the system's Gibbs energy will attain a minimum. Anfinsen did not specify the variable with respect to which the Gibbs energy has a minimum. Perhaps he was not aware of the potential pitfall in this hypothesis. It is very unfortunate that most people, if not all, in the biochemical community, including well established scientists, chose the wrong variables, namely the set of internal rotational angles which determine the conformation of the protein. This is particularly unfortunate because it involves misinterpretation of the Second Law of thermodynamics.

Briefly, the Second Law does not state anything about the 3D structure of the protein residing at the global minimum of the GEL. In other words, the Second Law does not state that the function $(GEL) = G = (T,P,N;\phi_1, \cdots,\phi_n)$ has a global minimum at some configuration $\phi_1,\ldots,\phi_n$ which is identified with the native structure. What it does say is that the Gibbs energy functional has a single (hence global or absolute) minimum with respect to all possible distributions of the conformations. In other words, the functional $G = F[T,P,N;(\phi_1, \cdots,\phi_n)]$ has a single minimum with respect to the distribution of function $Pr(\phi_1, \cdots,\phi_n)$ i.e. the probability distribution of all possible conformations of the protein. A schematic example is shown in Fig. 1. More details and examples are provided in Ben-Naim [2]. It is even more unfortunate that even after publication of the first article on these pitfalls, people still failed to understand the difference between the minimum in the GEL and the minimum of the Gibbs energy functional as required by the Second Law of thermodynamics.

I realize that the pitfalls I am describing are very subtle, and perhaps even difficult to swallow in particular by those who spent a long part of their scientific career in searching for the global minimum in the GEL. I also understand the immense strong motivation for searching for such a global minimum in the GEL. Finding the 3D structure at that minimum will immediately be rewarded with a great prize for solving the protein folding problem.

The GEL is an extremely complicated function, in particular, for large proteins having many bulky side chains. This function contains all possible conformation of the protein, including all possible knots. Each of these knots will occupy a vast part of the configurational space. Such knots would divide the entire GEL into many domains which are disconnected, i.e. states in one domain will not be accessible from another domain. This means that only one of these domains will be relevant to the study of protein folding — and the vast regions of all other domains are irrelevant. This is why I have repeatedly stated that a study of the entire GEL for a large protein makes a difficult problem much more difficult. This is in sharp contrast to statements in the literature alluding to the "advances" in understanding the PFP by the GEL "theory."

Searching for a 3D structure in the global minimum of the GEL is sometimes metaphorically described as looking for a needle in a haystack. In my view the search in the entire GEL can be more appropriately described as searching in one out of many haystacks for a needle that might not even be there.

I should add that in reality most people do not study the GEL itself, but what is believed to be an "approximate" GEL, which is referred to as a potential energy surface (PES). The search of the global minimum in the PES is a more futile effort than the search in the GEL. More details about this point may be found in reference [2] The main reason for saying this is that the PES, unlike the GEL is a very complicated function — but unfortunately this function is irrelevant to proteins in solutions.

To conclude this part I would say that in my view there is no general solution to the predictability problem. The search for the global minimum in the GEL is based on a misinterpretation of Anfinsen's hypothesis, or perhaps misunderstanding of the Second Law. The search for a global minimum in the PES is totally futile and a waste of time.

We now turn to discuss briefly other parts of the PFP which have general answers.

3. The stability of the 3D structure of the protein

In the beginning everything was explained in Biochemistry in terms of hydrogen-bonds (HB) [18–25]. Then, the devastating blow, known as the HB-inventory argument came; hydrogen bonding with water molecules competed with intramolecular hydrogen-bonds. As a result, the HBs paradigm fell from grace [26–29]. The void created was immediately filled by Kauzmann's idea of hydrophobic ($H\phi O$) effect which reigned supreme in biochemical literature for over 50 years [7,9,30,31].

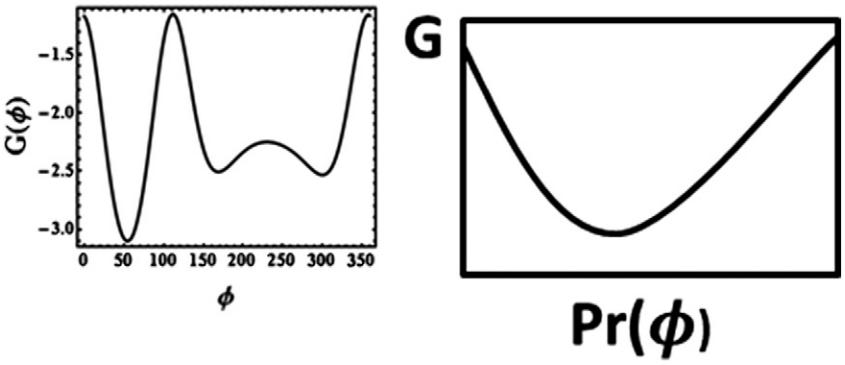


Fig. 1. Two possible minima in the Gibbs energy.

The first cracks in the HB-dogma came from the realization that though the HB energy is of the order of 24 kJ/mol, its formation in aqueous solution must have a far smaller effect on the "driving force" for the process of protein folding. The argument apparently started with the work of Schellman, summarized by Kauzmann, and eventually was encapsulated in Fersht's HB-inventory argument [26–29]. The argument seems simple, straightforward and convincing. Write the stoichiometrical reaction between a donor and an acceptor of a HB in the form

$$\mathrm{E{-}C{=}O{\cdots}}w + w\mathrm{{\cdots}HN{-}S \rightarrow E{-}C{=}O{\cdots}HN{-}S} + w\mathrm{{\cdots}}w \qquad (3.1)$$

where E and S stand for an enzyme and a substract, respectively, but can be any two molecules or parts of the same molecule, and w is a water molecule. Eq. (3.1) suggests that in the process of formation of a direct hydrogen bond between a donor (here amine group) and an acceptor (here a carbonyl group), two HBs are broken on the lhs of the equation and two HBs are formed on the rhs of the equation. Therefore, ignoring the differences in the various HB energies between the various pairs (carbonyl–water, amine–water, carbonyl–amine and water–water), we can conclude by simply counting, that the net effect of the formation of a direct HB is negligibly small. As Kauzmann summarized this argument: [7]

> "Hydrogen bonds, taken by themselves, give marginal stability to ordered structures, which may be enhanced or disrupted by interactions of side chains."

We use here the term "driving force" in the sense of $\Delta G < 0$ as commonly used in the literature.

If HBs do not provide the main "driving force" for protein folding, what factors do provide those "driving force?" This apparently conceptual vacuum was filled by the *HϕO* effect in (1959). The *HϕO* effect was known long before Kauzmann applied it to the problem of protein folding. It was applied successfully to explain surface tension of certain aqueous solutions of organic molecules, micelle formation and membranes. All these phenomena involve molecules having two moieties; a hydrophobic part, one "feared" water and tries to avoid it, the second "loved" water and mingled with water comfortably. As Tanford and Reynolds quoted from a personal communication with Kauzmann, the idea of the *HϕO* effect was hovering "in the air" for long a long time [31].

In a classical review article "Some Factors in the Interpretation of Protein Denaturation," Kauzmann applied the idea of the *HϕO* effect to protein folding [7]. For this purpose he coined the term *HϕO*-bond, and speculated that this "bond" could be the more important factor in the stabilization of the native structure of protein.

Kauzmann's idea was very simple. It was known that the Gibbs energy of transferring a small non-polar solute such as methane or ethane, from water into an organic liquid involves a large negative change in Gibbs energy. This is the same "driving force" that drives the formation of micells and membranes in aqueous solutions. Kauzmann also noticed that there are about one third of amino-acid side-chains which are *HϕO*, and most of these find themselves in the interior of the folded protein. If one can take the process of transfer, Fig. 2a, to represent the process of transfer of side chain, Fig. 2b, then we can estimate that a protein of about 150 amino acid has about 50 *HϕO* groups, and if each of these contributes between −12 and −16 kJ/mol, we get a very large "driving force" for the folding process.

Kauzmann's idea captured the imagination of many scientists including mine. Add to it the fact that the process of transferring of a non-polar solute from water into an organic liquid is "entropy driven". Add to it that entropy is a mysterious concept [32,33], not well understood, and you get a "driving force" which is enshrouded with an aura of mystery. Given all these facts, many authors could have claimed that the *HϕO* effect is the most important "driving force" in protein folding, without taking any risk of being proved wrong!

How can anyone prove anything on such a complex process as protein folding. Carried out in poorly understood solvent, involving a mysterious "entropy driven" concept?

It is not surprising therefore, that the dominance of the *HϕO* effect has prevailed for over half a century. The fact that *HϕO* groups are in the interior of the protein, and the fact that the transfer *HϕO* of molecules from water to an organic liquid is large and negative are undeniable. The former lends credibility to Kauzmann's model, while the latter provides the large negative Gibbs energy change. Although the molecular source of these large negative Gibbs energy changes were not clear, one can always say: "that is an entropy effect" and that is more than enough to silence any objections

In 1980, in the preface of my book "Hydrophobic Interactions," I wrote [9]:

> "In spite of my researches in this field over almost 10 years, I cannot confirm that there is at present either theoretical or experimental evidence that unequivocally demonstrates the relative importance of the *HϕO* interactions over other types of interactions in aqueous solutions."

My doubts were based on lack of evidence in favor of the contention that the *HϕO* effect is the most important effect in the "driving force" for protein folding. How can one claim that one factor is more important, or most important when one does not have a full inventory of all the factors involved in protein folding? Remember that Kauzmann's paper was on "some factors in the interpretation of protein denaturation" — not on all factors involved. No one knew what were all the factors especially those that are solvent-induced. The only factor that could have competed with the *HϕO* effect was the HB, but the HB-inventory argument, debilitated the effect of the HBs in aqueous media, and rendered them powerless in explaining the driving force for protein folding.

Kauzmann's model of inference from transferring of molecules from water to organic liquid, and the fact that most *HϕO* groups are found in the interior of the protein were so convincing that mere expression of

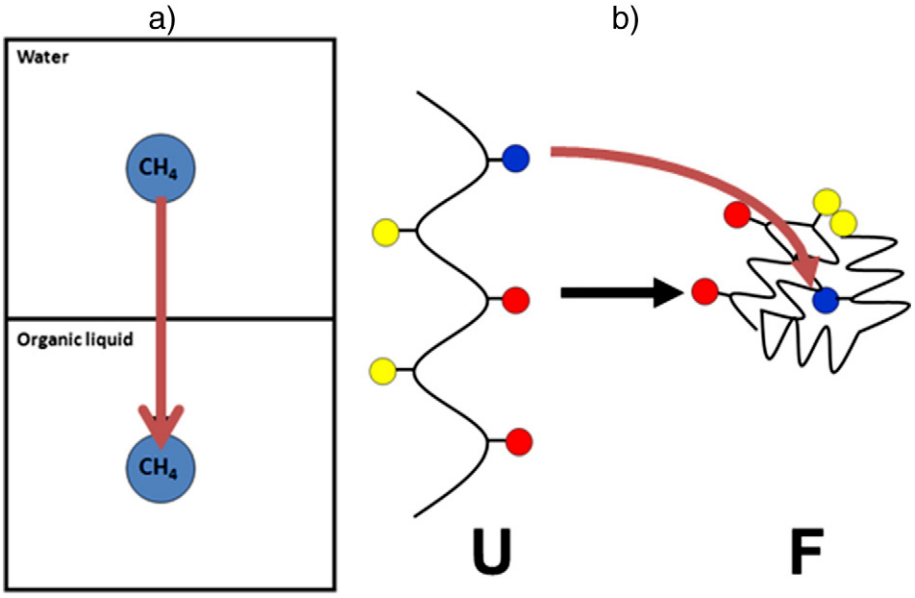


Fig. 2. (a) The process of transferring a methane molecule from water into an organic liquid. (b) The process of transferring a methyl group from water into the interior of the protein.

doubts could not have rattled the dominance of the *HϕO* dogma. One needs more than doubts. One needs facts! "Lack of evidence" for an idea cannot be used as evidence against that idea. This was the main motivation for the examination of the entire question of the solvent-induced effects on the protein folding and protein–protein association that I undertook late in the 1980s. The results of this examination were stunning; initially to me, then slowly diffusing into the literature.

The analysis of all the solvent-induced factors revealed that Kauzmann's model does not feature in the "driving force" for the process of protein folding. Instead of the Gibbs energy of solvation of a *HϕO* molecule in water, the conditional solvation Gibbs energy of a *HϕO* group features in the "driving force." These Gibbs energies are very different from the Gibbs energies of solvation in water. The main reason is that a *HϕO* group attached to the backbone (BB) of the protein is surrounded by water molecules which are perturbed by the BB.

Thus, not only the basis on which the *HϕO* model was built upon was demolished, but the Kauzmann *HϕO*-model itself was now shown to be inadequate.

The consequences of the analysis of the solvent-induced effects on protein folding, not only had undermined the foundation on which the *HϕO* dogma was erected, and not only demolished the *HϕO* dogma itself, but opened the door to a host of new solvent-induced effects that were never considered before. These effects involved *HϕI* rather than *HϕO* groups. The most important one, and so far the most studied, was the pairwise *HϕI* interactions between pairs of *HϕI* groups at a distance between 4−5 A°. For this particular *HϕI* interaction there is overwhelming evidence that it is far stronger than any of the *HϕO* effects. The evidence comes from theoretical estimates, simulations and experimental data [3]. There are also some estimates of *HϕI* effects involving three and four *HϕI* groups. These are more powerful, but probably less frequent [3].

The qualitative explanation of the pairwise *HϕI* interaction is quite simple. A *HϕI* group is characterized by a few "arms" along which HB may be formed. An amine group on the BB of the protein has one arm, a carbonyl group has two arms, a hydroxyl group three arms, and a water molecule itself has four arms. When a *HϕI* group is in water its arms are solvated by water molecule. The Gibbs energy of solvation per one arm was estimated to be of the order [3] of −9.4 kJ/mol. Note that this is quite different from a HB energy, as one might have erroneously counted in the HB-inventory argument.

When two such *HϕI* groups approach each other to a distance of about 2.8 A° they form a genuine HB. Thus, the Gibbs energy balance is: loss of the solvation Gibbs energy of two arms costs about 20 kJ/mol, and the formation of a HB provides a HB energy of about −24 kJ/mol. Therefore, the net change in Gibbs energy for this particular *HϕI* effect at a distance of about 2.8 A° is about −6 kJ/mol [3].

A more dramatic *HϕI* effect was found at a distance of about 4.5 A°, the same distance of the second nearest neighbors in ice. When two solvated arms approach each other to this distance, and with the correct orientation, they do not lose their solvation Gibbs energy as in the former *HϕI* case. They also do not gain a HB energy. Instead, the mutual solvation Gibbs energy of the pair of *HϕI* groups increase by an amount which was estimated to be between −10 and −12 kJ/mol.

The reason for such a strong *HϕI* interaction is that at this particular configuration the two arms of the two *HϕI* groups can be bridged by a water molecule. It should be stressed however that this effect is not due to a formation of long-lived HB-bridge, as some have misunderstood. Such a "permanent" bridge could provide two HB energies, i.e. about −48 kJ/mol. The real effect is a mutual solvation of the two arms of water molecules. This effect involves HB energy, but also involves probability of finding a water molecule that can form a HB-bridge between the two *HϕI* groups. The most direct evidence for the existence of such a *HϕI* effect is the second peak in the radial distribution function of pure liquid water [3]. Other experimental evidence comes from the relative solubilities of two isomers of the same molecule, having two *HϕI* groups at two different distances [3].

Because of the short range of the HB, there exists a steep gradient of the potential of mean force between two *HϕI* groups at a distance about 4.5 A°. This leads to a strong force between the two *HϕI* group, a force which plays a crucial role in the process of protein folding [2,3].

One can also think of other *HϕI* interactions; one involving one water molecule bridging three *HϕI* groups, or two water molecules forming one bridge connecting two *HϕI* groups. The former is strong, but rare, the second might be more frequent but very weak. Therefore, it is believed that the pairwise *HϕI* interaction at a distance of about 4.5 A° is the more important among the *HϕI* effects, hence probably the most important in the process of protein folding as well as in the process of protein–protein association or protein binding to DNA

4. What are the dominant forces in protein folding?

Suppose I show you a simple experiment. We start with two balls at some fixed positions in vacuum, Fig. 3. We release one of the balls, and find that it flies directly towards the second ball. Clearly, if you know

Newton's Second Law, you can deduce that a strong attractive force is operative between these two balls.

Now, we repeat the experiment with the same two solute particles not in the vacuum, but in different solvents. We find that in some solvents the released solute moves faster towards the other, in some cases it moves slower, and in others it moves completely at random. However, when we do the same experiment in water, the released solute moves faster towards the others held at a fixed position. The motion is not along a straight line, as it was in vacuum, Fig. 4b, but there is some small zigzagging along the way. The net effect of the presence of water is to decrease the time it takes the solute to move from its original position to its final position. In other words, the average speed of the solute has increased in the presence of water. What can we conclude from this experiment?

In this case the answer is a little bit more complicated. It is not only the forces that are exerted by the two solute particles which causes the motion, but there are many forces exerted by the solvent molecules on the moving solute. These forces are not always in the direction towards the second solute molecule at the fixed position.

Notwithstanding this complication, we can still say that on average there is a strong force operating on the released solute. If the average speed in the liquid is larger than the speed in vacuum, we can safely conclude that it is the solvent-induced force which is responsible for the fast motion of the solute particle. The same conclusion can be derived for the case of protein folding. If we observe a fast folding process, specifically in aqueous solutions, we should suspect that there are strong solvent-induced forces (and torques) operating on the various groups of the protein, Fig. 4. The next question is: What are these forces and how these forces are related to the $H\phi O$ or $H\phi I$ interactions? These questions were eloquently expressed by Levinthal [13]. Here is a quotation from Levinthal's articles: [13]

> "We feel that protein folding is speeded and guided by the rapid formation of local interactions which then determine the further folding of the peptide. This suggests local amino acid sequences which form stable interactions and serve as nucleation points in the folding process."

Levinthal reached the (almost) correct answer, as stated in the quotation above. Namely, that there must be preferential pathways of folding, "guided by rapid formation of local interactions." Although Levinthal did not specify what these "guiding interactions" are, his answer is almost correct. Instead of "guiding interactions," one should use the term "guiding forces." Though these forces are derived from the interactions, it is the magnitude of the forces acting on the groups of the protein that determines the speed of the folding process. The main question left unanswered by Levinthal is: What are these strong forces that guide the protein to its native structure in a relatively short time? We now know that these forces originate from the water, more specifically the solvent-induced forces exerted on the hydrophilic groups along the backbone and the side chains of the protein [3,10–12]. We shall present here a brief argument in favor of the relative importance of $H\phi I$ forces.

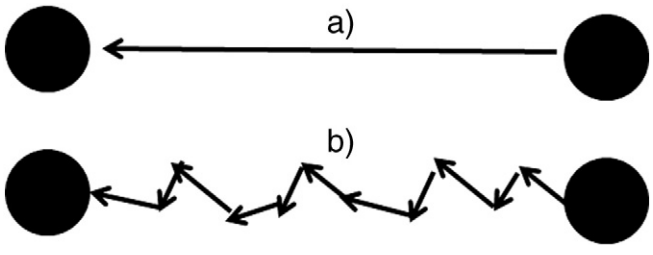


Fig. 3. Motion of one solute towards another solute at a fixed position. (a) In vacuum, (b) in water.

Suppose that we start with a protein at some specific conformation, defined by the angles $\phi_1, \cdots, \phi_M$. We observe that upon releasing the fixed conformation, the protein starts to move, and after a very short time it reaches a conformation which is stable, and besides small fluctuations the conformation remains constant. If this fast folding process occurs in aqueous solutions we should look for the solvent-induced forces acting on all the groups of the protein. The most convenient way of analyzing the solvent induced forces is to first define the Gibbs energy landscape (GEL). $G(T,P,N;\phi_1, \cdots,\phi_M)$, or equivalently $G(T,P,N;R_1, \cdots,R_n)$. The latter is more convenient for our purpose. R_i is the locational vector of the center of the ith group or the atom of the protein.

We now choose one of the groups, say R_1, and take the gradient of G with respect to R_1^3

$$\nabla_1 G = -\nabla_1 E - \nabla_1 \Delta G^* \tag{4.1}$$

From Eq. (4.1) it follows that the solvent-induced force is the gradient of the solvation Gibbs energy of the protein at the specified conformation. In order to analyze the conditions under which we may get a strong force, we convert the second term on the rhs of Eq. (4.1) to a more convenient form. We start with the (T, P, N) partition function of the N_w solvent molecules at some temperature T and pressure P, and one protein at a specific conformation $R_1, \cdots, R_n$. We take the gradient of the partition function with respect to R_1, and after some lengthy procedure we get [3]

$$F_1^{SI} = -\nabla_1 G^*(R_1, \cdots R_n) = \int [-\nabla_1 U(R_1, X_w)]\rho(X_w|R_1, \cdots, R_n)dX_w \tag{4.2}$$

This is a very convenient expression for analyzing the solvent induced force. The integrand contains two factors. The first is the direct force exerted on the group at R_1 by a water molecule at X_w (X_w is the vector specifying both the location and the orientation of a water molecule). The second factor is the *conditional* density, i.e. the density of water molecules at X_w given the specific conformation $R_1, \cdots, R_n$ of the protein. Note that the gradient ∇_1 operates only on the $U(R_1,X_w)$, and not on the conditional density. The quantity $\rho X_w(X_w|R_1, \cdots,R_n)dX_w$ may also be interpreted as the conditional probability of finding a water molecule at about X_w given the specific conformation of the protein $R_1, \cdots, R_n$.

What are the conditions under which we can expect large solvent-induced force? Clearly, a large value of the integral requires that both factors in the integrand should be large. This means that we need a strong force exerted by a water molecule, and also a relatively large density of water molecules at a position and orientation X_w from where it can exert such a strong force. We expect also that groups on the protein which are not too far from R_1, to have a significant effect on the density of water at X_w, Fig. 5. We shall analyze the solvent-induced force in several steps. First, suppose that there is only one group of the protein at R_2 which is close to R_1, and can affect the density at X_w from which a water molecule can exert a strong force on R_1. Consider the four cases in Fig. 6.

(a) The force on a $H\,\phi\,O$ group at R_1 in a $H\,\phi O$ environment
In this case the force $-\nabla_1 U(R_1,X_W)$ is expected to be weak. Furthermore, since the interaction between a water molecule and the two $H\phi O$ groups is weak even at the most favorable configuration, for the triplet at R_1,R_2 and R_W, we expect that the conditional density at X_W will be only slightly higher than the bulk density of water ρ_W.
(b) The force on a $H\phi O$ group at R_1 in a $H\phi I$ environment
In this case the force $-\nabla_1 U(R_1,X_W)$ is expected to be weak as in case (a). However, because of the presence of the $H\phi I$, group 2, there exists a distance from this $H\phi I$ group where the conditional density might be significantly enhanced. Therefore, we expect in this case, to obtain a solvent induced force larger than in case (a).
(c) The force on a $H\phi I$ group atR_1 in a HO environment
In this case the force $-\nabla_1 U(R_1,X_W)$ is expected to be larger than in cases (a) and (b). Because of the presence of one $H\phi I$ group, the

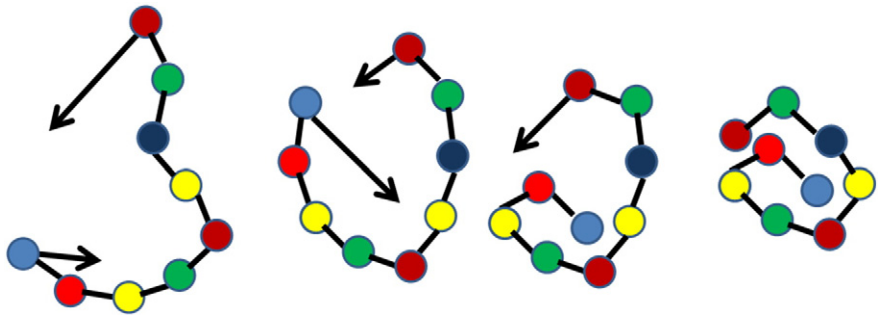


Fig. 4. A schematic fast-folding process. Arrows indicate strong forces operating on the various groups along the protein.

conditional density will also be enhanced. This has the same effect as the corresponding term in case (b), presuming the correct orientation of the *HϕI* group. Thus, in this case we expect to get a solvent induced force larger than case (b).

(d) The force on a *HϕI* group at R_1 in a *HϕI* environment

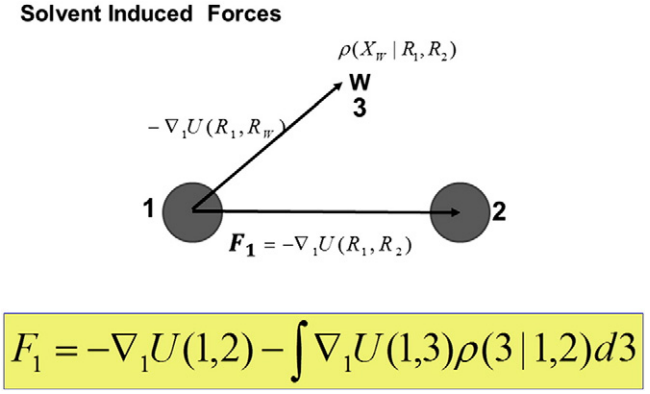
In this case the force $-\nabla_1 U(R_1,X_W)$ will be as large as in case (c). However, because of the presence of two *HϕI* groups, we might, under the right conditions, get a higher conditional density. The right conditions mean that the two hydrophilic groups are at a distance of about $4.5\,A^{\circ}$, and the correct orientation of the two *HϕI* groups so that they can be bridged by a water molecule [10]. At these orientations we expect large conditional density of water molecule. Thus, in this case both the force and the conditional density will be large and the resulting solvent-induced force is expected to be larger than in case (c).

Fig. 7 shows schematically the relative order of increasing-solvent induced force from case (a) to (d). We can conclude that the solvent induced force between the two *HϕI* groups at the correct distance and orientation is the largest of the four cases described above. Other evidence for the *HϕI* interactions and forces are available [3].

Thus, the general conclusion is that the strongest solvent induced force is expected to be exerted on a *HϕI* group, when this group is also surrounded by other *HϕI* groups in its immediate neighborhood. We have discussed in details only the case of two *HϕI* groups, but clearly, the argument may be extended to include the effect of the presence of more *HϕI* groups. The force produced by such a one-water bridge is operative in the range of distances between the two *HϕI* groups of about $4.5\,A^{\circ}$. Long range *HϕI* forces are also possible. These are discussed in reference [3].

Stronger forces are expected when there are two (or more) groups in the neighborhood of R_1, see Fig. 8. In this case, we may have either *HϕO* or *HϕI* groups at R_1, and we might have either *HϕO* or *HϕI* at either R_2 or R_3. We can repeat exactly the same argument as before with all the different cases as shown in Fig. 6. It is clear that the strongest forces are exerted on a *HϕI* group at R_1 when its surrounding also contains hydrophilic groups.

The strongest average force may be obtained when the probability of observing a water molecule at X_W is nearly one, and when X_W is the configuration of a water molecule from which it can exert a maximum direct force on the *HϕI* group [3]. The strongest direct force between two water molecules is shown in Fig. 9. We also compare this with the maximum force exerted by a neon atom on another neon atom, Fig. 10. We see that the strongest force exerted by one water molecule on a second water molecule is about two orders of magnitude as in the case of the neon atoms.



$$F_1 = -\nabla_1 U(1,2) - \int \nabla_1 U(1,3)\rho(3\mid 1,2)d3$$

Fig. 5. The solvent induced force on the methyl group is affected only by groups in the neighborhood of the methyl group. Here, one carboxyl group C═O and one methyl group.

5. Solvation and solubility of proteins

The high solubility of proteins in water is well known to any biochemist. Yet, the molecular reason for the solubility of protein is not less mysterious than the molecular reasons for protein folding and self-assembly of biological macromolecules.

The effect of the *HϕI* groups on the solubility of protein was long recognized. However, what is less known is that *HϕI* interactions are decisive in determining the high solubility of the protein.

The high solubility of protein is not only the result of the existence of *HϕI* groups on the surface of the protein. Furthermore, it is very likely that pairs and higher order correlations between *HϕI* groups on the surface of the protein contribute significantly, if not decisively in making the proteins highly soluble.

6. Conclusion and relevance to protein folding

It is now clear that that solvent-induced forces exerted on groups are much stronger than the forces exerted on *HϕI* groups. These forces will be stronger when the environment of the specific *HϕI* group is also *HϕI*. For a typical protein, say of 150–200 amino acids, we have about 50–70 *HϕO* groups, but at least 300–400 *HϕI* groups (provided by the backbone).

When we start with a completely unfolded polypeptide we expect that there will be, on average about one *HϕI* group in the neighborhood of any specific *HϕI* group. Thus, initially there will be strong forces exerted on all the *HϕI* groups, far stronger than on *HϕO* groups. As the protein starts to fold, it becomes more compact, and therefore the neighborhood of each *HϕI* group becomes richer in other *HϕI* groups. The larger the number of the *HϕI* groups in the neighborhood, the larger will be the forces. Therefore, we expect that due to *HϕI* forces, the folding process will be fast initially, but gradually becoming increasingly faster as the process of folding proceeds until it reaches the native 3D structure. Of course, the actual speed of the folding will be determined by both the direct forces exerted by groups within the protein, as well as by solvent-induced forces. Therefore, from the above argument alone one cannot estimate the actual speed of the folding process at each stage of the folding. The only conclusion is that can be reached is

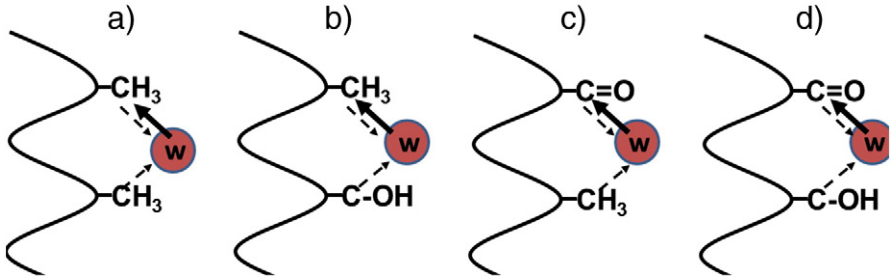


Fig. 6. A few configurations of two groups on a protein.

that *HϕI* forces might be far more important than the corresponding *HϕO* forces.

The conclusion is that the various hydrophilic forces provide a full answer to Levinthal's question regarding the speed and the specificity of the folding process. There is no reason to believe that the strong forces (let alone dominant forces) are operative on *HϕO* groups which are found in a *HϕO* environment.

Appendix A. The most complete list of "evidence" in favor of the dominance of hydrophobic interactions

For many years, the only evidence that supported Kauzmann's ideas on the role of the *HϕO* bond in protein folding was the difference in the solvation Gibbs energy of non-polar molecules in water and in organic liquid. About twenty years ago I have shown that Kauzmann's model is inadequate for two reasons. First, the solvation Gibbs energy of a molecule in water is very different from the conditional solvation Gibbs energy of a non-polar group. Second, the solvation Gibbs energy of the non polar group in an organic liquid is not a part of the solvent induced driving force in protein folding [2,3].

It is quite strange that in spite of the fact that the arguments against the role of the *HϕO* effect in protein folding have been in the literature for over twenty years, people still reiterate the same old argument, and adding new "evidence" that are anything but evidence.

In a recent review, Dill et al. write: [8]

> "There is considerable evidence that hydrophobic interactions must play a major role in protein folding. (a) Proteins have hydrophobic cores, implying nonpolar amino acids are driven to be sequestered from water. (b) Model compound studies show 1–2 kcal/mol for transferring a hydrophobic side chain from water into oil-like media, and there are many of them. (c) Proteins are readily denatured in nonpolar solvents.(d) Sequences that are jumbled and retain only their correct hydrophobic and polar patterning fold to their expected native states…, in the absence of efforts to design packing, charges or hydrogen bonding."

Unfortunately, none of these can be used as evidence in favor of the hydrophobic interaction. (a) The fact that hydrophobic groups are found in the interior of the protein does not necessarily mean that the hydrophobic interactions are responsible for bringing these groups to the interior of the protein. Such a conclusion is an illusion and cannot be supported by theory. It is similar to the conclusion that the mixing of two ideal gases is the cause of the entropy increase upon mixing, or that the "entropy of mixing" is the "driving force" for the process of mixing [33]. Similarly, one cannot say anything about the "driving force" for protein folding merely by watching the hydrophobic groups occupying the interior of the protein.(b) The Gibbs energies of transfer of small model compounds from water to an oil-like media were shown to be irrelevant to the driving force in protein folding [2,3]. (c) The fact that proteins are readily denatured in non polar solvent means that water is important. It says nothing on the relative importance of hydrophobic vis-à-vis hydrophilic effects. (d) The last evidence is even weaker than the previous one. It says nothing even on the role of water, certainly nothing about the relative importance of the hydrophobic vis-à-vis the hydrophilic effects.

It is regrettable that these non-evidence appear in the literature almost twenty years after the strong evidence in favor of the hydrophilic effects was published! Thus, we see that the first "evidence" (a), is only an illusion. The second "evidence" (b) is based on a wrong model, the third (c), and fourth (d), are "not even wrong." Altogether, there remains no single evidence in favor of the contention that the hydrophobic interactions are major factors in the process of protein folding.

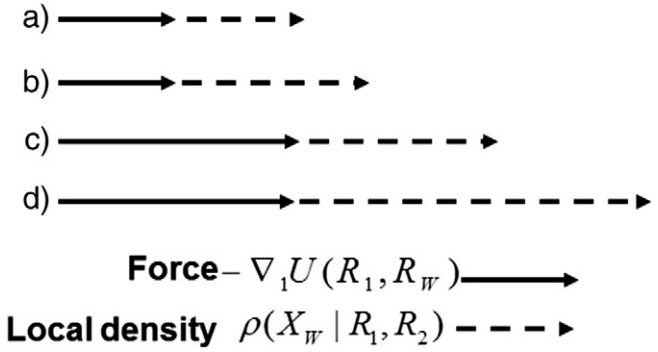


Fig. 7. The order of increasing forces from (a) to (c) in Fig. 6.

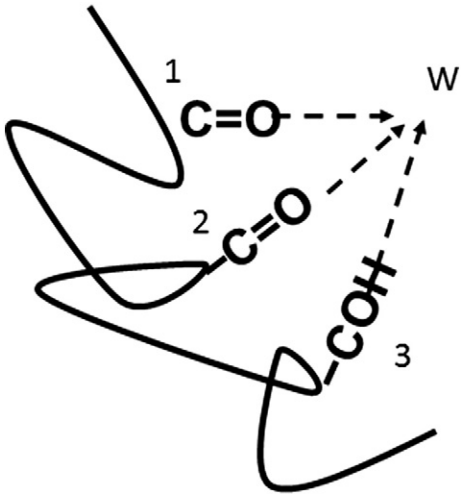


Fig. 8. Two *HϕI* groups (2 and 3) in the vicinity of a *HϕI* group (1).

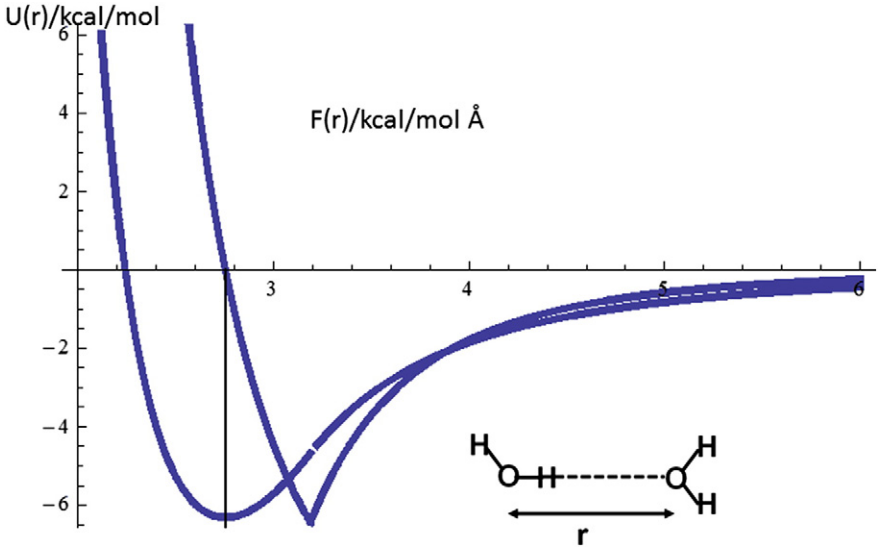


Fig. 9. The strongest force exerted by a water molecule on a second water molecule. The orientation of the two water molecules is shown below the curves.

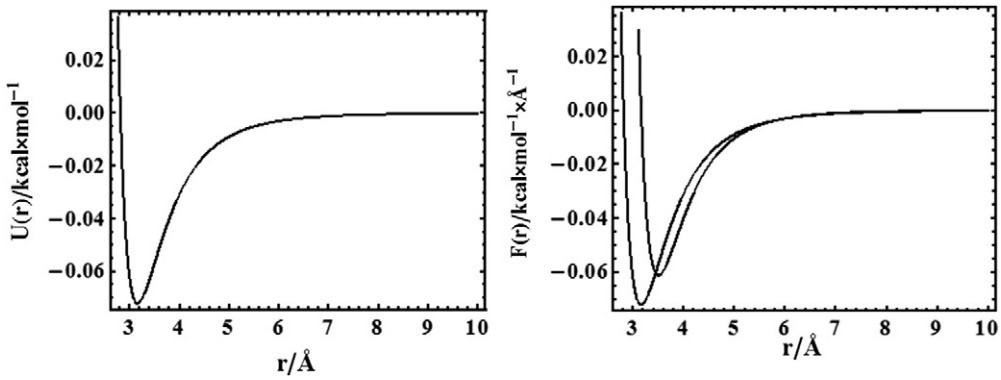


Fig. 10. The strongest force exerted on a neon atom by a second neon atom.

References

[1] Kennedy D, Norman C, Editorial. So much more to know. Science 2005;309: 5731, 75.
[2] Ben-Naim A. The protein folding problem and its solutions. Singapore: World Scientific; 2013.
[3] Ben-Naim A. Molecular theory of water and aqueous solutions, part II, the role of water in protein folding, self assembly and molecular recognition. Singapore: World Scientific; 2011.
[4] Anfinsen CB. Science 1973;181:223.
[5] Goldberg ME. TIBS 1985;10:388.
[6] Kolata G. Science 1986;233:1037.
[7] Kauzmann W. Adv Protein Chem 1959;14:1.
[8] Dill KA, Ozcan SB, Shell MS, Weikl TR. Annu Rev Biophys 2008;37:289.
[9] Ben-Naim A. Hydrophobic interactions. New York: Plenum Press; 1980.
[10] Ben-Naim A. Biopolymers 1990;29:567.
[11] Ben-Naim A. J Phys Chem 1990;95:1473.
[12] Ben-Naim A. J Chem Phys 1991;93:8196.
[13] Levinthal C. How to fold graciously. Mössbauer Spectrosc Biol Syst Proc 1969;67(41):22–6.
[14] Ben-Naim A. Levinthal's paradox revisited and dismissed. Open J Biophys 2012;2(2):22–32. http://dx.doi.org/10.4236/ojbiphy.2012.2204.
[15] Ben-Naim A. Pitfalls in Anfinsen's thermodynamic hypothesis. Chem Phys Lett 2011;511(1–3):126–8. http://dx.doi.org/10.1016/j.eplett.2011.05.049.
[16] Ben-Naim A. Levinthal's question revisited, and answered. J Biomol Struct Dyn 2012;30(1):113–24. http://dx.doi.org/10.1080/07391102.2012.674286.
[17] See for example:Goldstein M, Fredj E, Gerber RB. J Comput Chem 2011;32:1785.
[18] Pauling L. The nature of chemical bond. 2nd ed. Ithaca, New York: Cornell University Press; 1948.
[19] Pauling L. The nature of chemical bond. 3rd ed. Ithaca, New York: Cornell University Press; 1960.
[20] Anson ML, Mirsky AE. The equilibria between native and denatured hemoglobin in salicylate solutions and the theoretical consequences of the equilibrium between native and denatured protein. J Gen Physiol 1934;17:393–408.
[21] Mirsky AE, Pauling L. On the structure of native, denatured and coagulated proteins. Proc Natl Acad Sci U S A 1936;22:439–49.
[22] Pauling L, Corey RB. Atomic coordinates and structure factors for two helical configurations of polypeptide chains. Proc Natl Acad Sci U S A 1951;37:235–48.
[23] Pauling L, Corey RB. The pleated sheet, a new layer configuration of polypeptide chains. Proc Natl Acad Sci U S A 1951;37:251–6.
[24] Pauling L, Corey RB. The structure of fibrous proteins of the collagen–gelatin group. Proc Natl Acad Sci U S A 1951;37:272–81.
[25] Pauling L, Corey RB. Configurations of peptide chains with favored orientations around single bonds: two new pleated sheets. Proc Natl Acad Sci U S A 1951;37:729–40.
[26] Schellmann JA. The thermodynamics of urea solutions and the heat of Formation of the peptide hydrogen bond. C R Trav Lab Carlsberg Ser Chim 1955;29:223–30.
[27] Schellmann JA. The stability of hydrogen-bonded peptide structures in aqueous solution. C R Trav Lab Carlsberg Ser Chim 1955;29:230–59.
[28] Kauzmann W. Some factors in the interpretation of protein denaturation. Adv Protein Chem 1959;14:1–63.
[29] Fersht A. Structure and mechanism in protein science. New York: W. H. Freeman and Comp.; 1999.
[30] Dill KA. Biochemistry 1990;29:7133.
[31] Tanford C, Reynolds J. Nature's robots, a history of proteins. Oxford: Oxford University Press; 2001.
[32] Ben-Naim A. Entropy demystified. The second law reduced to plain common sense. Singapore: World Scientific; 2007.
[33] Ben-Naim A. A farewell to entropy. Statistical mechanics based on information. Singapore: World Scientific; 2008.