Myths and verities in protein folding theories

Part II: From Kauzmann's conjecture to the dominance of the hydrophobic effect

Arieh Ben-Naim
Department of Physical Chemistry
The Hebrew University of Jerusalem
Edmond J. Safra Campus
Givat Ram, Jerusalem 91904
Israel
Email: arieh@fh.huji.ac.il

Abstract

We start with Kauzmann's conjecture regarding the factor which is most likely to determine the stability of the 3D structure of the proteins. We show first that Kauzmann's model-process, on which the original conjecture was based, is inadequate for the process of protein folding. The main reason for this is that protein folding involves the change in the *conditional* solvation of non-polar groups, not the solvation itself. The "condition" here is the presence of the protein backbone. Secondly, it is argued that various hydrophilic effects, both solvation and interactions are more likely to be dominant both in stabilizing the 3D-structure of proteins, as well as in determining the speed and the specificity of the folding process.

Key words: Protein folding, Kauzmann's conjecture, hydrophobic, hydrophilic effects

Council for Innovative Research

Peer Review Research Publishing System

I. Introduction

In Part I of this series[1] we showed how the Anfinsen thermodynamic hypothesis was misinterpreted leading to excessive, yet futile endeavors in searching for a global minimum in the Gibbs energy landscape. This article presents another myth which stemmed not from misunderstanding of a hypothesis, but rather from over exaggerating and over estimating a conjecture made by Kauzmann.[2] This has to do with the magnitude and relative importance of the various hydrophobic effects in protein folding.[3]

Lest the message of this article be misunderstood, it should be emphasized from the outset that I do not "deny the existence of the hydrophobic effect." On the contrary, I believe that the various hydrophobic ($H\phi O$) effects are very important in processes such as micelles formation, stability of membranes and most importantly in understanding the unusual properties of water.

The main message of this article is that the role of various $H\phi O$ effects in processes such as protein folding, protein-protein association and molecular recognition were over exaggerated. This has led to the mystification of the protein folding problem.[3] Besides the problem of prediction of the 3D structure of proteins which was dealt with in part I[1], there are two main problems in which $H\phi O$ effects are commonly believed to play a dominant role. What makes the folding process so fast, and what makes the 3D structure of the protein stable?

In this article we start with several quotations from Kauzmann's article. We then show that the model Kauzmann used to suggest that $H\phi O$ "bonding" might be important to the stability of the proteins, is inadequate to describe the contribution of the various $H\phi O$ effects on the stability of the native structure of proteins. We shall then show that other solvent-induced effects involving hydrophilic ($H\phi I$) groups are much more important than the corresponding $H\phi O$ effects, both on the stability of the native structure of the protein, as well as on the speed of the folding process. Thus, in effect the various $H\phi I$ affects answers the two main questions regarding the speed of folding and the stability of the 3D structure of proteins.
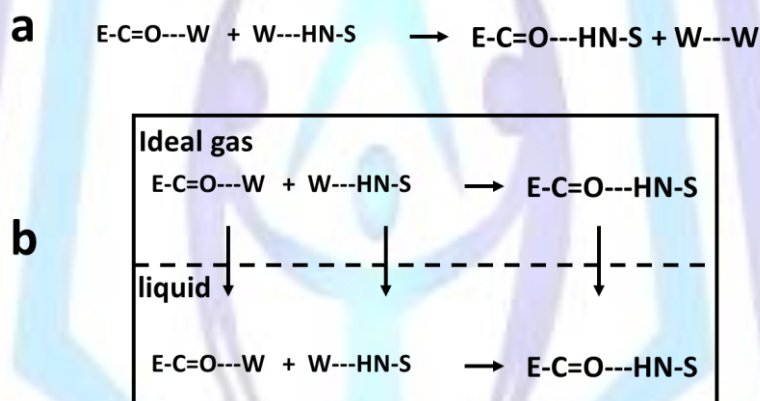

**Figure 1.a.** **The stoichiometric reaction as drawn originally by Schellman.[4]**

b. **A formation of hydrogen-bond between two hydrophilic groups in an ideal gas and in water. Dashed lines indicate hydrogen-bonds.**

II. Kauzmann's model and Kauzmann's conjecture

Let us begin with a few quotations from Kauzmann's article.[2]

(i)   *"Hydrogen bonds, taken by themselves, give a marginal stability to ordered structures"*

(ii)   *"This tendency of non-polar groups of proteins to adhere to one another in aqueous environments has been referred to as hydrophobic bonding. The hydrophobic bond is probably one of the more important factors involved in stabilizing the folded configuration in many native proteins."*

In the first quotation (i) Kauzmann summarized some conclusions reached earlier by Schellman.[4] The basic idea is that when hydrogen bonds between two solutes are formed in aqueous solutions, there are some cancellation effects, two HBs are broken and two HBs are formed. In the original article by Schellman (see Figure 1a), the two water molecules that are released into the pool of water are shown as forming a HB. The correct stoichiometric reaction is shown in Figure 1b. Hence, hydrogen bonding between hydrophilic groups in proteins was deemed to contribute "marginally" to the stability of the protein. This conclusion was contrary to the prevailing view at that time that intra-molecular HBing in proteins are important.[3] If HBs contribute marginally to the stability of the proteins, then the question that arises is: What other factors contribute significantly to the stability of the proteins? The second quotation (ii) contains Kauzmann's answer to this question.
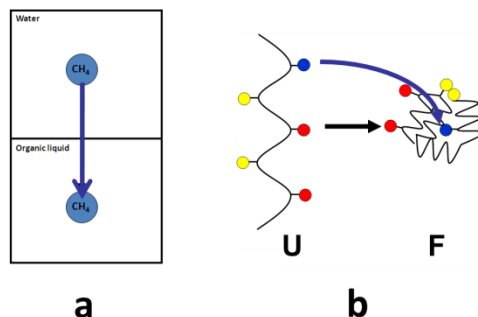
Figure 2. Kauzmann's model for the hydrophobic effect (a) and its application to the process of protein folding (b).

Kauzmann was cautious in his statement that the $H\phi0$ bond is "*probably*" one of the *more important factors* in stabilizing the folded form of proteins. His main argument was based on the Gibbs energy of transferring non-polar solutes from water into some organic liquids, Figure 2. These Gibbs energies were known to be large and negative. Kauzmann's idea was that the transfer of a non-polar group attached to the protein into the interior of the protein might be as favorable as the transfer of a non-polar solute from water into an organic liquid.

If each non-polar *group* that is initially exposed to the water, when transferred to the interior of the protein, contribute between −3 and −5 kcal/mol, as much as the Gibbs energy change for the transfer of a non-polar *solute* from water to an organic liquid, then one can expect that the combined effect of many such non-polar groups would give rise to a huge "driving force" for the process of protein folding.

Using the available data on the thermodynamics of transferring a non-polar molecule from water to a non-polar solvent, Kauzmann drew the following *tentative* conclusions:

(i) The $H\phi O$ bonds are stabilized largely by entropy effects.

(ii) The free energy change in the transfer from water to the non-polar environment is exergonic to the extent of about 3000 cal per mole of groups.

(iii) Hydrophobic bonds involving aliphatic side-chains are more stable at room temperature than they are at 0 C.

Kauzmann emphasized that these conclusions are only *tentative* "because they are based on the behavior of small hydrocarbon molecules in which no polar groups are present." Kauzmann's conclusions were based on the process of transfer of non-polar *solutes* from water to organic liquid. It was not clear at all that these conclusions also apply to the transfer of non-polar *groups* hung on a protein from being exposed to water to being exposed to the environment in the interior of the protein (Figure 2).

In spite of the uncertainty as to the relevance of Kauzmann's model to the process of protein folding, the idea that the $H\phi0$ effects play a dominant role in protein folding has become a commonly accepted fact. This idea thrives not so much on evidence but simply due to a lack of other competing factors. It is not uncommon to find statements in textbooks in chemistry and biochemistry referring to the $H\phi0$ bond, or the $H\phi0$ interaction, as "a major driving force in folding of macromolecules, the binding of substrate to enzymes and most other molecular interactions in biology." Even in recent articles one can find statements like "It is commonly accepted that hydrophobic effects plays a major role in protein folding." Indeed, it is commonly accepted but is it true?

In 1980, I wrote in the preface of my book "Hydrophobic Interactions":[5]

"In spite of my research in this field over almost ten years, I cannot confirm that there is at present either theoretical or experimental evidence that unequivocally demonstrates the relative importance of the $H\phi0$ interactions over other types of interactions in aqueous solutions."

This was in 1980. Since then I became convinced that Kauzmann's model, based on transferring a *solute* from water into an organic liquid, is inadequate for estimating the contributions of $H\phi0$ *groups* to the overall driving forces in protein folding and protein-protein association. The "evidence" most frequently cited, aside from the facts about the Gibbs energy of solvation, is that the $H\phi0$ groups are found *inside* the protein. "Could anything be more convincing?" Tanford and Reynolds (2001) asked rhetorically.[6] The fact that $H\phi0$ groups are found in the interior of the protein is *not* evidence that the hydrophobic effect is the driving force for protein folding. This apparently "convincing" argument is similar to the argument given in connection to the entropy of mixing of two different ideal gases; we *see* mixing, therefore, what could be more convincing than the conclusion that the mixing is also the *driving force* for the mixing process. The fallacy of this argument is discussed in detail elsewhere.[7]

Regarding Kauzmann's model, it is easy to see that in the process of folding it is not the Gibbs energy of solvation of non-polar *molecules* in water but the *conditional* solvation Gibbs energy of non-polar groups are involved. Furthermore,

after re-examining the entire question of solvent-induced effects on biochemical process, it became clear that hydrophilic ($H\phi I$) rather than hydrophobic ($H\phi O$) effects are probably much more important in biochemical processes.[8]

Since 1959 Kauzmann's ideas were quoted thousands of times. From a "probable" important factor, it grew into a "major driving force" and a "dominating force," adding layers upon layers of superlatives to the importance of the various $H\phi O$ effects.

In 1980 I have summarized all the evidence that existed regarding the strength of the hydrophobic interactions, and the magnitude of the hydrophobic solvation. There was no doubt that water is unique in exhibiting outstanding "solvophobic" effect – a term that I coined to contrast the behavior of water as compared to other solvents. Like many others I was fascinated by the various $H\phi O$ effects and their relationship to other outstanding properties of liquid water.

Doubts about the importance of the $H\phi O$ effects to protein folding were based on lack of evidence. Lack of evidence does not mean no evidence, nor does it mean evidence, to the contrary. I have realized that protein folding is an extremely complicated process involving many different groups, some hydrophobic, some hydrophilic, some charged. How can one reach a conclusion that one effect is the "most important," or "dominating force," without having a complete inventory of all solvent-induced effects?

Kauzmann's ideas were bold and brilliant. Given that the role of HBs fell out of grace, and having no other known solvent-induced effect, it was almost natural to assume that the spectacular $H\phi O$ solvation of non-polar molecule known at that time might also contribute significantly to the process of protein folding.

It should be noted that Kauzmann's model was used to explain the stability of the 3D structure of the protein. It offered no explanation whatsoever to the question so eloquently posed by Levinthal:[9] What are the factors that "guide" and "speed"" the folding process?

Yet, this fact did not stop people from using the $H\phi O$ effect in explaining the speed of the *process* of folding. The literature abounds with concepts such as "dominant forces" which are anything but forces.[3]

These ideas have pervaded all textbooks of biochemistry, biophysics and molecular biology. Very few questioned their validity, and no one offered any alternative factor that is important, let alone dominant, in either the stabilization of the 3D structure of the protein, or to explain the "speed" of the folding.

A turning point occurred in the late 1980s and early 1990s, when I undertook the project of creating an inventory of all possible solvent-induced interactions. This was not an easy task, nor did it offer any immediate alternatives to the dominance of the $H\phi O$ effect. However, two important conclusions emerged from this analysis. First, it was proved that the so-called HB inventory argument was seriously faulty.[10] This fact removed the basis on which Kauzmann proposed his conjecture; i.e. that intra-molecular HBs do not contribute significantly to the stability of the protein (quotation (i) above). Second, and more disturbing, it was found that the model used by Kauzmann to propose his conjecture was inadequate to represent protein folding. More precisely, the Kauzmann model involved the process of transferring of a non-polar *molecule*, say methane from water into an organic liquid. The analysis of all possible solvent-induced effect showed that the closest factor in protein folding was the transfer of a non-polar *group* from water into the interior of the protein. This process involved the *conditional* solvation Gibbs energy rather than the solvation Gibbs energy of the group. In other words, the very presence of the protein backbone could in principle change the distribution of water molecules around the non-polar group. This *condition* could have a profound effect on the solvation Gibbs energy of the non-polar group.[10] Figure 3.
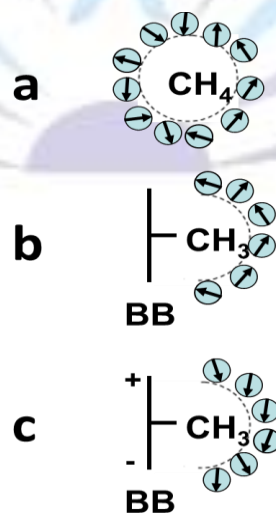
Figure 3. A schematic demonstration of the effect of the backbone on the distribution of water molecules around a non-polar group.

Not only was it found that the $H\phi O$ effects were far weaker than what they were believed, but other $H\phi I$ effects, including direct hydrogen bonding were found to be quite strong. These finding led me to conclude:[11]

"This finding leads us to conclude that intra-molecular hydrophilic interactions are quite strong, highly dependent on orientation, and very specific to the properties of the solvent. For these reasons, we believe that these interactions are probably more important than hydrophobic interactions, especially in highly specific processes such as protein folding and molecular recognition."

The discovery of the various $H\phi I$ interactions removes one mystery from the problem of protein folding. It provides a convincing evidence for the major factor that stabilizes the 3D of the native structure of protein. It does not "deny" the existence, or the contribution of all of the $H\phi O$ effects. It only claims that these are weaker and fewer compared with the corresponding $H\phi I$ effects. It left out the question of the "speed" of the folding process unanswered. We shall see in the next section that the $H\phi I$ forces also provide a convincing explanation for the "speed" of the folding, and largely answered the Levinthal question.[9]

III. 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. 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, 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. 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.[9] Here is a quotation from Levinthal's articles:[9]

"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 *force* 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.

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 $G(T, P, N; \boldsymbol{R}_1, \cdots, \boldsymbol{R}_n)$. The latter is more convenient for our purpose. $\boldsymbol{R}_i$ is the locational vector of the center of the i th group or the atom of the protein.

Assuming that the system is classical we can always write the GEL as

$$G(T, P, N; \boldsymbol{R}_1, \cdots, \boldsymbol{R}_n) = E(\boldsymbol{R}_1, \cdots, \boldsymbol{R}_n) + \Delta G^*(\boldsymbol{R}_1, \cdots, \boldsymbol{R}_n) - \Delta G^*(\boldsymbol{R} = \infty) \quad (3.1)$$

Where $E(\boldsymbol{R}_1, \cdots, \boldsymbol{R}_n)$ is the energy landscape (EL), $\Delta G^*(\boldsymbol{R}_1, \cdots, \boldsymbol{R}_n)$ is the solvation Gibbs energy of the protein being at a specific configuration $\boldsymbol{R}_1, \cdots, \boldsymbol{R}_n$. $\Delta G^*(\boldsymbol{R} = \infty)$ is the solvation Gibbs energy of the protein at some reference state. We may

choose this as the configuration such that all particles are at infinite separations from each other. This is indicated by $R = \infty$.

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

$$-\nabla_1 G = -\nabla_1 E - \nabla_1 \Delta G^* \qquad (3.2)$$

Note that the second solvation quantity $\Delta G^*(R = \infty)$ in (3.1) is the sum of the solvation Gibbs energies of all the groups at infinite separation from each other. Each of these is not a function of the corresponding location R_1.

From equation (3.2) it follows that the solvent-induced force is the gradient of the solvation Gibbs energy of the protein at the specified conformation $R_1, \cdots, R_n$. In order to analyze the conditions under which we may get a strong force, we convert the second term on the rhs of equation (3.2) 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[10]

$$F_1^{SI} = -\nabla_1 G^*(R_1, \cdots, R_n) = \int [-\nabla_1 U(R_1, X_w)] \rho(X_w | R_1, \cdots, R_n) dX_w \qquad (3.3)$$

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 $\rho(X_w | R_1, \cdots, R_n) dX_w$ 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, Figure 4.
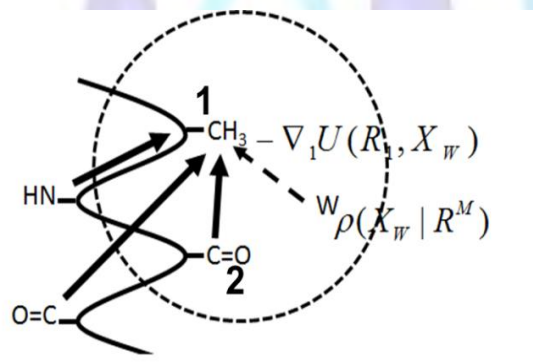
Figure 4. **The solvent induced force on the methyl group is affected only by groups in the neighborhood of the methyl group. Here, one carboxyl group C = O and one methyl group.**

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.
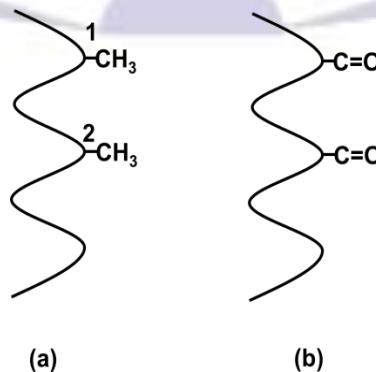
Figure 5. Two configurations : $H\phi O - H\phi O, \text{ and } H\phi I - H\phi I$.

Consider the two extreme cases shown in Figure 5:

(a) <u>The force on a $H\phi O$ group at R_1 in a $H\phi O$ environment</u>

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) <u>The force on a $H\phi I$ group at R_1 in a $H\phi I$ environment</u>

In this case the force $-\nabla_1 U(R_1, X_W)$ will be as large as in case (a). However, because of the presence of *two $H\phi I$* 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 $4.5 \overset{\circ}{A}$, and the correct orientation of the two $H\phi I$ 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 (a).

Thus, the general conclusion is that the strongest solvent induced force is expected to be exerted on a *$H\phi I$* group, when this group is also surrounded by other *$H\phi I$* groups in its immediate neighborhood. We have discussed in details only the case of *two $H\phi I$* groups, but clearly, the argument may be extended to include the effect of the presence of more *$H\phi I$* groups. The force produced by such a one-water bridge is operative in the range of distances between the two *$H\phi I$* groups of about $4.5 \overset{\circ}{A}$. Long range *$H\phi I$* forces are also possible. These are discussed in reference 10.

Stronger forces are expected when there are two (or more) groups in the neighborhood of R_1, see Figure 6. In this case, we may have either $H\phi O$ or $H\phi I$ groups at R_1, and we might have either $H\phi O$ or $H\phi I$ at either R_2 or R_3. It is clear that the strongest forces are exerted on a $H\phi I$ group at R_1 when its surrounding also contains hydrophilic groups. Other even stronger forces are discussed in reference 10.
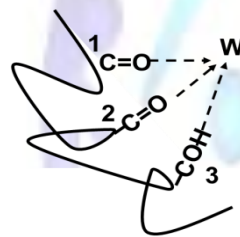
Figure 6   Two $H\phi I$ groups (2 and 3) in the vicinity of a $H\phi I$ group (1).

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 $H\phi I$ group.[10] The strongest direct force between two water molecules is shown in Figure 7. We also compare this with the maximum force exerted by a neon atom on another neon atom, Figure 8. 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.
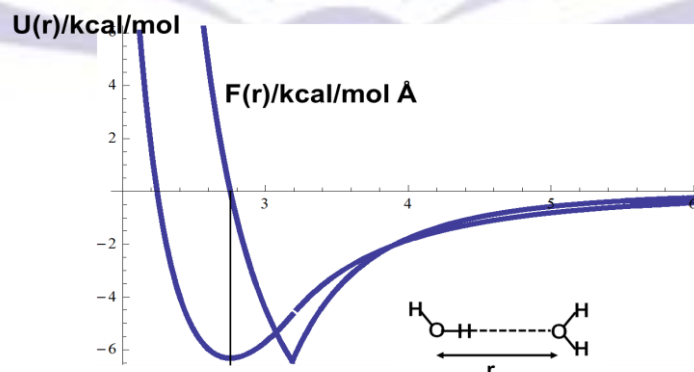
Figure 7   The strongest force exerted by a water molecule on a second water molecule. The orientation of the two water molecules is shown below the curves.
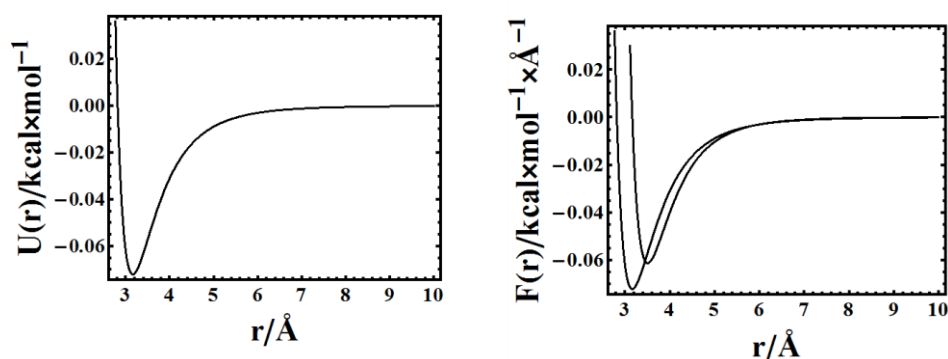
Figure 8.   The strongest force exerted on a neon atom by a second neon atom.

IV.   Conclusion and relevance to protein folding

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

When we start with a completely unfolded polypeptide we expect that there will be, on average about one $H\phi I$ group in the neighborhood of any specific $H\phi I$ group. Thus, initially there will be strong forces exerted on all the $H\phi I$ groups, far stronger than on $H\phi O$ groups. As the protein starts to fold, it becomes more compact, and therefore the neighborhood of each $H\phi I$ group becomes richer in other $H\phi I$ groups. The larger the number of the $H\phi I$ groups in the neighborhood, the larger will be the forces. Therefore, we expect that due to $H\phi I$ 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 that can be reached is that $H\phi I$ forces might be far more important than the corresponding $H\phi O$ 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 $H\phi O$ groups which are found in a $H\phi O$ environment.

References

1. Ben-Naim, A., Part I (2013)

2. Kauzmann, W., *Advances in Protein Chemistry*, 14, 1-63 (1959)

3. Ben-Naim, A., *The Protein Folding Problem and Its Solutions*, World Scientific, Singapore (2013)

4. Schellman, J. A., *Compt. Rend. Lab. Carlsberg Ser Chim*. 29, 230-59 (1955)

5. Ben-Naim, A., *Hydrophobic Interactions*, Plenum Press, New York (1980)

6. Tanford, C. and Reynolds, J., *Nature's Robots, A History of Proteins*, Oxford University Press, Oxford (2001)

7. Ben-Naim, A., *A Farewell to Entropy*, World Scientific, Singapore (2008)

8. Ben-Naim, A., *Biopolymers*, 29, 567-596 (1990)

9. Levinthal, C., *J. de Chimie Physique*, 65, 44 (1968)

10. Ben-Naim, A., *Molecular Theory of Water and Aqueous Solutions, Part II: The Role of Water in Protein Folding, Self-assembly and Molecular Recognition*, World Scientific, Singapore (2011)

11. Ben-Naim, A., *J. Chem. Phys.* 90, 7412-25 (1989)

12. Durell, S. and   Ben-Naim, A. (Submitted 2013)