Journal of Theoretical Biology 363 (2014) 169–187

Contents lists available at ScienceDirect

Journal of Theoretical Biology

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

Protein folding: Understanding the role of water and the low Reynolds number environment as the peptide chain emerges from the ribosome and folds

Siddhartha Sen [a], H. Paul Voorheis [b,*]

[a] CRANN, Trinity College Dublin, Dublin 2, Ireland
[b] School of Biochemistry and Immunology, Trinity College Dublin, Dublin 2, Ireland

HIGHLIGHTS

- Quantum field theory predicts that coherent domains form a water sheath around proteins.
- Each hydrophobic residue twists the sheath; eventually the sheath intersects itself.
- The intersection expels water and the reactive force drives hydrophobic collapse.
- Nanobubbles of dissolved atmospheric gasses cover exposed hydrophobic residues.
- The nanobubbles act as attracting resonators and drive folding.

ARTICLE INFO

Article history:
Received 28 February 2014
Received in revised form
20 June 2014
Accepted 25 July 2014
Available online 23 August 2014

Keywords:
Protein biosynthesis
Water sheath
Quantum field theory
Role of hydophobic amino acids
Coherent domains

ABSTRACT

The mechanism of protein folding during early stages of the process has three determinants. First, moving water molecules obey the rules of low Reynolds number physics without an inertial component. Molecular movement is instantaneous and size insensitive. Proteins emerging from the ribosome move and rotate without an external force if they change shape, forming and propagating helical structures that increases translocational efficiency. Forward motion ceases when the shape change or propelling force ceases. Second, application of quantum field theory to water structure predicts the spontaneous formation of low density coherent units of fixed size that expel dissolved atmospheric gases. Structured water layers with both coherent and non-coherent domains, form a sheath around the new protein. The surface of exposed hydrophobic amino acids is protected from water contact by small nanobubbles of dissolved atmospheric gases, 5 or 6 molecules on average, that vibrate, attracting even widely separated resonating nanobubbles. This force results from quantum effects, appearing only when the system is within and interacts with an oscillating electromagnetic field. The newly recognized quantum force sharply bends the peptide and is part of a dynamic field determining the pathway of protein folding. Third, the force initiating the tertiary folding of proteins arises from twists at the position of each hydrophobic amino acid, that minimizes surface exposure of the hydrophobic amino acids and propagates along the protein. When the total bend reaches 360°, the leading segment of water sheath intersects the trailing segment. This steric self-intersection expels water from overlapping segments of the sheath and by Newton's second law moves the polypeptide chain in an opposite direction. Consequently, with very few exceptions that we enumerate and discuss, tertiary structures are absent from proteins without hydrophobic amino acids, which control the early stages of protein folding and the overall shape of protein. Consequently, proteins only adopt a limited number of forms. The formation of quaternary structures is not necessarily prevented by the absence of hydrophobic amino acids.

1. Introduction

Proteins are typically composed of a sequence of 20 different types of amino acids linked by amide bonds in one or more polypeptide chains. Each amino acid side chain has a characteristic

* Corresponding author.

http://dx.doi.org/10.1016/j.jtbi.2014.07.025
0022-5193/

chemistry and, together with its sequence and length, defines the protein's distinct shape and function in cells and tissues. This report applies the mathematical methods of topology and quantum field theory to the protein folding process within the low Reynolds number environment of aqueous solution, part co-translationally and part post-translationally as the growing polypeptide chain emerges from the ribosomal exit tunnel. Quantum field theory is relevant for understanding protein folding because it alone allows multiple stable ground states for water. Water containing protein has two phases; the first is a bulk liquid phase of relatively unstructured water molecules of normal density and the second is a structured, stable, coherent phase of low density and large polarizability that immediately surrounds the protein as it folds and can expel non-coherent molecules if an energetically favorable path is available. The physical model that we employ suggests that the initial stages of protein folding makes use of this multiphase property of water.

A popular method of studying the process of folding has been to elevate the temperature of the system above its physiological range transiently and then lower it to its more usual value. Under these conditions, proteins unfold when the temperature rises above a critical value and in some cases they fold back to their original configuration when the temperature is lowered. However, in many cases this refolding process leads to a different conformation compared to that found originally with concomitant loss of function. There are several reasons for this behavior.

First, additional low energy states may be available when the whole protein folds compared to those available when folding occurs from the only partially complete state during biosynthesis and there may be a sufficiently high activation energy between these states to prevent interconversion under physiological conditions. In short a protein folds into the lowest energy state available *at the time of folding* and not into the global minimum energy state of the completed protein if that state does not exist when folding occurs.

Second, when the protein is folded before a post-translational modification, such as proteolytic processing of a precursor protein, the protein may fold to a higher than lowest final energy state because of a chaperone-like activity of the portion of the protein subsequently removed by processing.

Third, a number of proteins are intrinsically disordered or partially unstructured in the functional state (for a recent review see Tompa, 2012) and (for a brief overview see Babu et al., 2010). In general proteins performing a catalytic or transport role possess a well developed, relatively fixed 3-dimensional structure, whereas the conformation of signalling or regulatory proteins undergo rapid interconversion between a wide variety of conformations and are categorized as intrinsically disorganized or unstructured. Bioinformatic methods for predicting which proteins will be found to be intrinsically disordered and which regions of proteins will be unstructured have been recently reviewed (Dosztanyi et al., 2009). Analyses of disordered proteins, revealed that they possess increased amounts of proline (Theillet et al., 2013), which interrupt alpha helices, and that they are frequently highly positively (Mao et al., 2010) or negatively (Ulversky, 2013) charged, introducing sufficient electrostatic repulsion to inhibit folding (Müller-Späth et al., 2010). The relative abundance of the hydrophobic amino acids was also much reduced in these proteins (Theillet et al., 2013, see Fig. 1b), suggesting from our analysis that these proteins have a decreased tendency to nucleate the hydrophobic collapse, thus inhibiting the later stages of folding. These workers also found decreased amounts of the remaining aromatic amino acids and cysteine, which prevent stabilization of transiently folded forms of the protein by cation – π interaction and disulphide bond formation respectively.

Fourth, some proteins only fold properly in conjunction with a target structure, such as a chaperone or a permanent partner protein. When this target structure is absent, refolding does not occur in the same way or end in the same state as it does when the target structure is present.

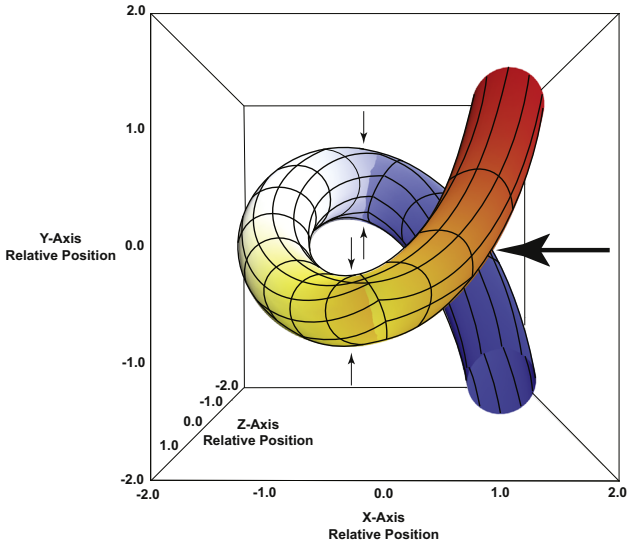

Fig. 1. A segment of lysozyme (1LYD) as it comes off of the ribosome with its attached water sheath, twisting through 360° in the region of emerging hydrophobic residues and illustrating the consequent two types of intersections of the water sheath. The view is normal to a plane in the *x*, *y* dimensions at the midpoint of the bend. Small pairs of arrows mark two of several regions where bending of the sheath causes the first type of intersection or compressions of the sheath together with coupled expansions of the sheath opposite the compressions. The type-1 intersections occur on the concave portions of the bend and equal and opposite expansions of the sheath occur on the convex portions of the bend. A single large arrow, at a position centered at the congruence of points located approximately at 30° and 330° from the beginning of the bend, marks the point at which the leading segment of the bend and the trailing segment of the bend overlap and cause the second type of intersection. Type-2 intersections are not accompanied by opposite expansions and in this case the intersection occurs along the z-dimension.

Results of this kind suggest that the folded shape of a protein is robust but also depends upon a number of additional factors beside those already listed, *e.g.* the temperature of the system, the hydrogen ion and protein concentration of the solution and the ionic strength of the solution. It is also conjectured that the folded conformation of a protein is completely determined by the sequence of its constituent amino acids under any given set of conditions (Anfinson, 1973).

In recent times researchers in bioinformatics have found that the number of basic folded protein shapes is limited (Peng and Xu, 2011). Thus, the number of possible pathways of folding that can be followed by a protein of any given sequence under any defined experimental condition, seems to be rather small and easy to define. A rather striking observation supporting this observation comes from the "Levinthal paradox" (Levinthal, 1969). Consider that a protein of just 150 residues can adopt 10^{68} possible conformations, which leads to the prediction that, effectively, folding would require about 10^{52} years if each state could be sampled in 10^{-9} s, despite the fact that the folding time of such a protein is observed to be between 10^{-1} and 10^{3} s (Glaser, 2013). Clearly, biosynthetic folding does not sample all of the possible conformations inherent in the structure of the whole protein after its complete synthesis. Levinthal further suggested that initial local sequence interactions form nucleation points from which the folding process proceeds (Levinthal, 1969).

The natural challenge inherent in the protein folding problem is, then, to determine the mechanism that imposes the observed restrictions on the number of possible protein folding pathways. However, a robust picture of such a mechanism is missing.

Numerical simulation methods can lead to a partial understanding of the factors influencing this mechanism but the simplicity of the bioinformatic result is yet to be fully understood. A convenient metaphor that is widely used to symbolize the way the number of pathways are limited emerges from the idea that the space containing the folding pathways is shaped like a funnel (Cheung et al., 2002). The experimental consequences suggested by such a view have been found to be correct. However an *ab initio* reason for the existence of these restrictions, the funnel itself, is not known.

In this note we restrict the discussion to typical cytoplasmic proteins of roughly globular shape and to the time frame between the point where the protein chain traverses the ribosomal tunnel into the bulk aqueous phase and the point where the hydrophobic collapse of the molten globule is complete. We propose a specific model for those steps within the complete folding process that occur within this time period and explain the existence of the funnel. This model does not yet contain those late steps in the folding pathway that lead to the generation of more permanent secondary structural elements nor to the way one folding domain comes to interact with a second folding domain within a single polypeptide chain. In a limited number of cases, the transient, initial, alpha helix may be stabilized by local interactions. Nevertheless, even when the initial helix is truly transient, the model developed is based on strict physical principles and the ideas that underpin it are described. In essence the model rests on two key principles. First, are the concepts inherent in low Reynolds number physics (Purcell, 1977) that are relevant for small objects moving in water. Second, are the two types of water structure (Cheung et al., 2002; Chaplin, 2006) that immediately surround a protein.

We illustrate these ideas by looking at monomeric, cytoplasmic proteins that fold independently of any chaperone requirement and possess relatively random but periodic distributions of hydrophobic and hydrophilic amino acids. For this type of simple example physical arguments suggest a specific folding pathway and suggest a specific final shape. Testable general consequences of this model are listed in the Conclusions Section.

2. Results and discussion

2.1. Experimental observations

Let us list the experimental observations and theoretical assumptions that underlie the proposed folding model: It has been well established experimentally that

1. Monomeric globular proteins are small objects, usually between a molecular weight of 10^3 to 10^6 daltons with an average surface area of

   $$A_s = 6.3(M_W)^{0.73}\ \mathring{A}^2 \cdot \text{daltons}^{-1}$$

   which is about twice the surface area of a sphere of the same volume because the surface of the protein is rough, and a volume of

   $$V = 1.27(M_W)\ \mathring{A}^3 \cdot \text{daltons}^{-1}$$

   (see p. 229 in Creighton, 1993).
2. During folding, the volume of water displaced from the surface of the polypeptide chain is about half of the volume occupied by the polypeptide chain itself and amounts to 43.3% in the case of lysozyme (see Appendix for calculations).
3. We postulate that the initial stages of the hydrophobic collapse begins when about 6 or 7 hydrophobic amino acids, not necessarily contiguous within the protein sequence, have been extruded into the aqueous environment outside the ribosomal exit tunnel.
4. The initial hydrophobic collapse is very rapid and biphasic, as found for the kinetics of refolding after mild denaturation (for a recent review see Udgaonkar, 2013). The first phase occurs in the scale of 4–120 microseconds and the second in the scale of 150–500 microseconds.
5. The formation of stable secondary structures, including α-helices and β-sheets are relatively late events, either occurring to a significant but incomplete extent during the second phase of hydrophobic collapse (*e.g.*, apomyoglobin and ribonuclease A) or well after (>1 millisecond) the hydrophobic collapse (*e.g.*, cytochrome C and monellin) *ibid*.

2.2. Theoretical assumptions

The theoretical assumptions in our model are divided into two categories. First, are the special features associated with small objects moving in water, which are quite remarkable and very different from the motion of particles described by Newton's laws. These features are

1. Small objects moving in water reside in the domain of low Reynolds numbers. In this realm there are no dynamics and no inertial effects because the frictional forces opposing the movement of molecules in water are much larger than the inertial forces propelling the molecules.
2. These objects can move only when they change their shape or when an external force acts on them and they do not move alone but carry a considerable mass of water with them.
3. When the change of shape ceases or the force is halted, the movement quickly stops (within 1 Angstrom).
4. When the force or the change in shape is initiated, the movement of the object that results is instantaneous and is of the same magnitude for small or large objects. This behavior reflects the absence of inertial effects within the low Reynolds number domain.

The second set of assumptions concern the physical role of the two water structures surrounding the emerging polypeptide chain during folding. In fact there is increasing experimental evidence that water structures play an important role during folding (Dobson, 2000). The specific assumptions that involve these water structures are

1. There is a highly structured shell of water surrounding a polypeptide chain as it folds, for which some experimental evidence exists (for reviews see Klotz, 1958; Kuntz, 1971; Franks and Eagland, 1975; Otting et al., 1991; Mattos, 2002). There are also theoretical arguments advanced by Del Guidice and Vitiello (2006); Del Guidice et al. (2010) that suggest the existence of these water structures.
2. The first type of water structure is a sheath surrounding the polypeptide chain of a certain size, usually a monolayer of water molecules but up to three molecules thick, which responds to and is deformed by the shape changes that take place in each amino acid within the emerging protein. There is also some experimental evidence for this assumption (Huang et al., 2009).
3. This water structure is destabilized in the vicinity of certain types of developing twist within the polypeptide chain and, as a result, becomes locally unstructured.
4. Consequently, the structure of the water that surrounds the emerging chain has two components: the structured component,

C_s, that immediately surrounds the polypeptide chain, and any newly formed, relatively small, unstructured region within the structured component as well as the major unstructured component, C_u, that is recognized as the bulk water phase.

5. There is a diffusion of water molecules between these two unstructured components in the direction of the bulk phase whenever there is a breakdown of a local region within the structured component.

2.3. Qualitative consequences of this picture

Several events occur as the polypeptide chain emerges from the ribosomal exit tunnel into the bulk aqueous phase of the cytoplasm.

1. The polypeptide begins to adopt a transient α-helical structure either in the ribosomal exit tunnel or in the portion of the emerging polypeptide as soon as it emerges into the bulk water phase. The formation of this structure is required for any thread-like object to move efficiently in a low Reynolds number environment (Purcell, 1977; Shapere and Wilczek, 1988). Consequently, this structural feature is a consequence of the movement of the emerging polypeptide chain.
   Furthermore, the diameter of the exit tunnel (2 nm largest radius, 1 nm smallest radius) is sufficient to accommodate a random coil structure throughout its length and an alpha helical structure in at least the terminal third of its length, while other secondary structural elements cannot be accommodated due to their larger size (Voss et al., 2006; Bhushan et al., 2010). Once outside of the ribosomal exit tunnel the forward motion of the nascent polypeptide chain is maintained in the α-helical configuration as long as the force generated by the GTPase activity of the elongation factors, EF-Tu and EF-G is acting. As soon as the peptide has been extruded from the ribosomal exit tunnel, the force moving the peptide in the forward direction ceases. Consequently, the transient alpha helical structure will relax, no longer maintained by the need to move efficiently in the low Reynolds number environment along a straight vector through its extended backbone. At this point the protein will be free to adopt a different structure, thermodynamically driven by transition to the lowest conformational energy state. However, present views of the folding process consider the formation of more permanent secondary structures, such as the beta sheet conformation, usually to be a relatively slow and, therefore late process, occurring after or during the late stages of the hydrophobic collapse of the molten globule (for review see Udgaonkar, 2013). Nevertheless, there are known examples where permanent secondary structures form very quickly after synthesis of the polypeptide chain is complete (for review see Buchner et al., 2011) and the speed record is currently held by a beta-structure (Xu et al., 2006).
2. Water becomes structured around the emerging polypeptide chain by interaction with the chain itself and is, thereby, forced to reflect the helical structure of the polypeptide inside it during the very early period immediately before and during emergence from the ribosomal exit tunnel.
3. When a hydrophobic amino acid emerges from the ribosomal exit tunnel, its side chain cannot interact with the water in its vicinity in the same way or to the same extent as do the side chains of hydrophilic amino acids.
4. The water structure surrounding the amino acid chain responds to the decrease in affinity of the emerging hydrophobic amino acid compared to preceding or following hydrophilic amino acids by a disturbance to the arrangement of water molecules in the portions of the water sheath flanking the newly emerged hydrophobic amino acid. This disturbance introduces a relatively small twist to the emerging polypeptide chain, such that water molecules in the sheath located over hydrophilic amino acids immediately before and after the hydrophobic amino acid are compressed together on the concave side of the twist and are pulled apart on the convex side of the twist by equal and opposite changes in the volume of the sheath in these regions (Fig. 1). Consequently, the affected water molecules in these regions and those connecting these regions move position within the sheath to preserve the average density of water molecules within the sheath. Furthermore, the water molecules in this region either withdraw from or fail to approach the hydrophobic surface.
5. Dissolved atmospheric gasses then migrate into the regions over the emerging hydrophobic amino acids in order to relieve the vacuum that would otherwise be created. These gas molecules form nanobubbles (Attard, 1996; Attard et al., 2002) that provide a protective layer over the hydrophobic areas of the emerging amino acid chain and prevent these hydrophobic areas from directly contacting surface water molecules. We have, elsewhere, shown that the resulting theoretically predicted number of water molecules over only the hydrophilic surface in the water layer closest to lysozyme agrees to within 3–8% with published experimental measurements of the tightly bound surface molecules of water, while these numbers are overestimated by 24–38% when the nanobubble layer is omitted (Sen and Voorheis, unpublished).
6. If the number of hydrophobic amino acids that have emerged from the ribosomal tunnel reaches approximately 6 or 7, the two ends of the water structure, flanking this group of hydrophobic amino acids, will experience a relative rotation of 360° with respect to each other as a consequence of the sum of the disturbances to the arrangement of water molecules in the sheath that arise from the emerging hydrophobic amino acids. Consequently, this twist or rotation of the water sheath as the turn of the polypeptide is formed causes the leading segment of the water sheath to intersect with the trailing segment of the water sheath (Figs. 1 and 2). This topological arrangement forms a singularity; the water sheath breaks at

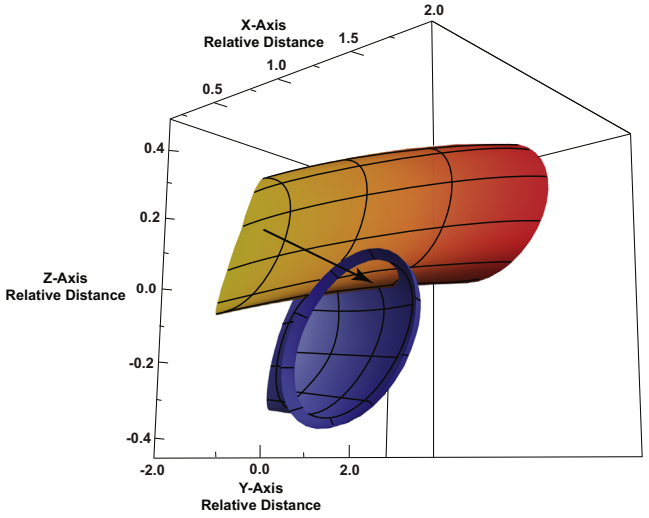

Fig. 2. The central portion of the same segment of lysozyme (1LYD) shown in Fig. 1 at twice the magnification, viewed at an oblique angle along the axis of the segment of lysozyme and its water sheath as it enters the bend and illustrating the second type of intersection of the water sheath. The arrow marks the region where the segment of the sheath entering the bend intersects the segment of the sheath exiting the bend in the *z* dimension.

this point and diffusion of water molecules away from this site takes place.

7. It can be shown that diffusion of water molecules, initiated by this singularity, occurs in waves along a helical path from the region of intersection of the portions of the water sheath, located before and after the recently emerged group of hydrophobic residues, and extend from there outwards into the bulk water phase.
8. Alternatively, the twist of the water sheath structure can be relieved by a writhe. There is a topological formula (Călugăreanu, 1961) that states: a twist + writhe must add to zero when there is no linking between components, *i.e.* in this case when a single filament, the water sheath, is not knotted. A writhe has a global effect on the protein. However, a twist + writhe of the water sheath cannot cause a sharp bend because the force caused by the first movement is immediately relieved by the second movement. In this case the structure experiencing the twist and that experiencing the writhe are not linked because they are the same unknotted structure, the continuous water sheath itself.
9. However, a breakdown of the postulated water sheath surrounding a polypeptide chain can cause the chain to bend or even double up on itself. As the sheath is narrowed at the point where breakdown in the water sheath will subsequently occur, the twist motion increases. Just before breakdown, this increased twist motion will cause a stream of water molecules to detach from the adjacent structured water component and escape to the relatively unstructured surrounding bulk water. The reaction of the amino acid chain to this helical stream of departing water molecules is to bend or, perhaps, even to double up as a consequence of the second law of thermodynamics. The bending itself is generic but the specific nature of the bending will be controlled by the type of side chain attached to each amino acid residue in the region. Thus, the hydrophobic side chains behave like steering mechanisms that determine the extent, pitch and direction of bending while the backbone of the polypeptide chain at the site of the bend in the chain determines the position from which the α-helical wave of departing water molecules first starts to form. The absence of inertial effects means that the entire polypeptide chain responds to the departing helical wave. When the bending results in a doubling up of the chain on itself, it is immediately followed by a wrapping of the chain around its doubled up half. This wrapping occurs because the helical motion of the polypeptide chain at the site of each hydrophobic side chain, responsible for the water sheath's breakdown, is still present and leads to the two chains of the doubled up structure twisting around each other. Once again these events can happen only because the response to the singular breakdown is instantaneous and because there are no inertial effects to slow the process.
10. The direct consequence of the intersection of the water sheath is the breaking of the coherence symmetry of the sheath and the breaking of the rotational symmetry of the water molecules next to amino acids at this point. Quantum field theory predicts that when a coherence symmetry is broken, a massless excitation, called a Goldstone boson, appears. This boson gives rise to a long distance coherence generating force (Preparata, 1995). The relevant symmetry breaking Goldstone boson leads to a polarization of water in the region of the intersection. A consequence of this polarization is the prediction that gas bubbles in water next to an amino acid should be charged dipoles. These dipoles are, according to our model, located at hydrophobic sites, which suggests that a long range force between hydrophobic sites of “dipole–dipole” type should appear. These forces could also contribute to the twisting and bending of the sheath at the points of intersection and to the movement of hydrophobic sites closer together.

We now illustrate the reason for an attractive resonance dipole–dipole interaction of gas molecules in different nanobubbles over the surface of different hydrophobic residues by examining the following simple model in which a collection of free dipoles interact with an oscillating electromagnetic field. This idea encapsulates the key physical principle of the model. The type of mathematical treatment we use to describe the interactions between two bosons, two molecules of atmospheric gasses as resonating oscillators, was first used by Preparata to describe the interaction between two fermions, two electrons as resonating oscillators (Preparata, 1995). The interaction Hamiltonian H is

$$H = g \int \psi^*(x,t)\psi(x,t)A(x,t)$$

where $\psi^*(x,t)$ represents the dipole creation operator and $\psi(x,t)$ the dipole destruction operator and $A(x,t)$ the oscillating electromagnetic field, where we have approximated all fields by scalar fields. The oscillatory character of $A(x,t)$ means we can write it as

$$A(x,t) = \frac{1}{\sqrt{V}}\sum[a_q e^{-i\omega_q t + \mathbf{q}.\mathbf{x}} + hc]$$

This formulation gives rise to a positive interaction of the form:

$$V_q \approx \frac{\omega_q}{(\epsilon_p - \epsilon_{p+q})^2 - (\omega_q)^2}$$

where $\epsilon = \epsilon_p - \epsilon_{p+q}$ is the energy difference between the two dipoles and ω is the photon energy. Integrating ω over the allowed momenta values of the photon with an appropriate physically determined cut off, *Omega*, gives the potential energy as a function of the difference in energies of the two dipoles. In the simple model we take the dipoles to be free air molecules in the self generated gas cavity that overlies a surface hydrophobic amino acid side chain. The quantum fluctuating momentum scale set by the quantum uncertainty principle is for a cavity size of $\Delta x \approx 1$ given by $\Delta p \approx h/\Delta x$. The corresponding energy is less than 10^{-1} eV. Carrying out the momentum integration gives the dipole-photon potential $V_{d,\gamma}(\epsilon)$ to be

$$V_{d,\gamma}(\epsilon) \approx \left(\frac{g^2}{\Omega}\right)\left(\frac{\epsilon}{2\Omega}\ln\left|\frac{1+\frac{\epsilon}{\Omega}}{1-\frac{\epsilon}{\Omega}}\right| - 1\right)$$

where $\epsilon = \epsilon_1 - \epsilon_2$ is the difference of energies between the two dipoles and Ω is the photon energy corresponding to the thermal cutoff momentum. We see this potential is attractive for $\epsilon \leq \Omega$. We note that this attractive potential, and its associated attractive force, increases as $\epsilon \Rightarrow 0$, which corresponds to the oscillators being in resonance. This new force, of quantum origin, only appears when the dipoles are oscillating and the electromagnetic field is time dependent. The oscillation of the dipoles in the model comes from the fact that they are confined in a bubble cavity of small length and are represented by oscillating waves. This new quantum force brings resonating dipoles close together. The immediate result of this movement of gas dipoles from different nanobubbles towards each other is the movement of these same nanobubbles towards each other, dragging with them their underlying hydrophobic residues. This movement can act over long distances within the electromagnetic field and is only restrained by the opposing force due to the tethering effects of the covalent structure of the polypeptide itself.

11. This dipole–dipole model calculation in the immediately preceding item also suggests that there should be a tendency for side chains of identical mass to be located close together. However, an analysis of the location of all pairs of hydrophobic

amino acids within contact distance (4.5 Å) in 10 different soluble monomeric proteins of widely varying molecular weight does not support this possibility in all cases but does in some (See Appendix B).

Thus, for a completely periodic array of amino acids, the picture of events just described suggests that a possible final configuration of the system could be a twisted one dimensional structure. It is important to note that this view of events is opposite to that proposed by Cheung et al. (2002) who consider that the movement of water molecules away from the sheath only occurs after the hydrophobic collapse. If this view were correct, the exit of water molecules could not provide any force for the collapse itself and either the new quantum force would need to be sufficient or another force that could contribute to the collapse would need to be found. In addition the structure of the water sheath after collapse of the molten globule but before exit of any water molecules would need to be determined.

In summary, the model suggests that the very earliest structure formed by the polypeptide chain, as it emerges from the ribosomal exit tunnel, is a transient α-helix. Then, subsequent but early long-distance folding occurs when the number of bends of the peptide reach approximately 6–7 due to a singularity forming that causing the water sheath to interact with itself in a type-2 intersection, expelling water molecules and initiating the molten globule collapse. Consequently, a twisted, 1-dimensional structure is produced at this point. Furthermore, the model suggests that modification of the periodic structure will alter the final structure by altering the position from which the collapse occurs and that no long distance tertiary structures are possible without the presence of hydrophobic amino acids.

We have constructed a three layer model of the water sheath to be presented in detail elsewhere. Briefly, the two layers closest to the protein contain water molecules over only the hydrophilic surface regions of the protein, while the hydrophobic surface regions are covered by two layers of atmospheric gasses, derived from those dissolved in the bulk water phase. The water molecules in the layer closest to the protein correspond to those identified experimentally as tightly bound, while those in the second layer correspond to those identified experimentally as loosely bound. The third, most distant layer is a complete layer of water molecules, covering all underlying regions, and corresponding to those identified experimentally as unbound but whose motion is disturbed by the presence of the protein. This scheme should be valid outside the ribosomal exit tunnel. Within the tunnel only a single structured water layer is likely to exist for steric reasons related to the radial dimensions of the tunnel itself. As the polypeptide chain emerges from the tunnel, the presence of gas molecules over the hydrophobic amino acids rather than water molecules, may facilitate the facile interconversion between alpha-helical and random coil conformations at these junctions.

We next turn to a simple mathematical model that gives rise to helical waves and sharp bending.

2.4. Mathematical model

The idea is to think of the breakdown of the structured water sheath due to self-intersection as initiating a process of rapid diffusion between a structured and unstructured component of water in the neighborhood of the amino acid chain. The model uses this idea and the special features of low Reynolds number physics. Let us briefly review these two topics within the framework of quantum theory before setting up the mathematical model.

2.5. Structured water

One of our key assumptions is that a water structure surrounds the amino acid chain. This idea of structured water has been suggested recently in a series of papers on the theory of liquid water based on Quantum Field Theory, (DelGuidice et al., 1988; Del Guidice and Vitiello, 2006; Del Guidice et al., 2010, 2013). We briefly summarize the basic features of this theory that we will use, closely following the account given in Montagnier et al. (2011). These ideas are also discussed in more general terms in the book by Preparata (1995).

1. The structure of water determined by quantum theory is not only due to the static forces of the system, such as the H-bonds and electric dipole–dipole interactions, but is, rather, induced by a time dependent radiative long range fluctuating electromagnetic field which is always present. This field arises from vacuum fluctuations close to the protein and is a quantum effect. Observable consequences of these vacuum fluctuations have been calculated and experimentally verified to be present with great precision for atomic systems. The insight of Del Guidice and Vitiello (2006) was to realize that these effects can, above a certain density threshold of the water molecules, make the usual perturbative ground state unstable. A new stable classical ground state where matter oscillates and rotates in phase with resonating modes of the fluctuating electromagnetic field emerges. Thus, vacuum fluctuations can spontaneously give rise to coherent oscillating classical domains for not just single atoms but for a large assembly of molecules. This formulation of the theory leads to a classical oscillating and rotating system of water molecules that form coherent domains and are spatially close together. Del Guidice and Vitiello (2006) also point out the difficulty of understanding how short range static bonds, such as H-bonds, which form and break in a time scale of 2–20 ps, can be treated as static forces responsible for the condensation of water. They avoid using hydrogen bonds in order to understand the condensation of water and then show how coherent domains can explain the process of condensation. They also show, that at any given temperature, a finite fraction of the total water exists as a coherent domain within their two phase model (DelGuidice et al., 1988). We expect this picture of Del Guidice et al. (2013) to be valid for water in direct contact with biomolecules but not for water in the bulk phase because the higher excited electronic states are known to be unstable to transitions between excited states (Shandilya et al., 2013). As a result of this instability, coherent domains that form in the bulk water phase break up in the range of 10–100 picoseconds and can, at best, form a flickering landscape (see p. 138 in Frank and Wen, 1957). However, this instability need not affect a coherent domain within a water sheath because the additional interactions between the water molecules and the protein in this case can stabilize the domain. There is also experimental evidence for such coherent structures next to biomolecules (Del Guidice et al., 2013).
2. Earlier objections to the existence of a two phase model for water, suggested in 1892 by Röntgen (1892) on theoretical grounds, were due to Bernal and Fowler (1933) and are answered by this new model. The Bernal–Fowler objection was that a basic theorem of quantum mechanics states that any system described by a Hamiltonian (which is the case for water molecules) can only have one phase. The new model evades the consequences of this theorem by using quantum field theory where multiple phases are allowed. In short the condition described by quantum mechanics emerges from a certain limit of the more general quantum field theory, just as classical dynamics emerges from a

certain limit of quantum mechanics itself. Quantum field theory is relevant in the present context because it is the theory that must be used to describe the vacuum electromagnetic fields that are always present and can influence the water molecules within the coherent domains of the water sheath surrounding the protein but not the protein itself.

3. An ensemble of molecules, interacting with a radiative electromagnetic field, acquire a new minimum energy state when they are elevated above a density threshold and below a critical temperature. In this state the molecules act in unison and form an extended domain of coherence, which in the case of bulk liquid water is of the order of 75 nm diameter (Arani et al., 1995). The size of this domain is given by the trapped electromagnetic field. Outside of this domain the electromagnetic field vanishes. Of course in the case of a water sheath, where the coherent water will also interact with the protein, the size of the domain could be significantly smaller than that found for bulk liquid water. Furthermore, the water molecules need not be strictly contiguous. They could be located in multiple foci over the hydrophilic surface areas and still behave as a coherent whole.
4. The results above apply to all liquids. The peculiarity of water is that their coherent oscillations occur between the ground state and an excited state lying at 12.06 eV, which is just below the ionization threshold of 12.60 eV. The size of the extended coherent domain of water that follows from this position is about 100 nm. The size is related to the excitation by the formula, hc/E_c, where E_c is the excitation energy. In this domain the theory predicts that there will be an ensemble of almost free electrons that are able to accept externally supplied energy and transform it into coherent excitations (vortices).
5. Coherence among molecules is reduced at any given temperature by thermal collisions. This result and cavity ionization events potentially give rise to a permanent crossover of molecules between those coherent and non-coherent regimes that are present in the bulk water, leading to spontaneous breakups of the coherent domains and, consequently, a fluctuating mixture of structured and unstructured water. However, near the surface of a protein or when water molecules are bound to membranes or to the backbone of biological macromolecules, the attractive forces present can protect the coherent structure from thermal noise and thus produce stable, extended, coherent domains. Consequently, in a water sheath that is composed of 1–3 layers of water molecules, the coherence leads to a sharing of the various binding energies and these average over the assembly of water sheath molecules to a value of 0.26 eV/water molecule, which provides a barrier to thermal disruption at physiological temperatures (Arani et al., 1995). It is this stabilized structure surrounding the polypeptide chain that we call a water sheath.
6. The presence of free electrons that rotate in unison can provide an energy source for chemical reactions and can also provide electric forces that assist the process of folding.
7. For our problem such a structure plays two essential roles. First, the structured water sheath acts as a lubricant and allows the protein to move with less hindrance in a low Reynolds number environment where viscus forces predominate while, at the same time, requiring more energy than would otherwise be the case because of the increased mass that must be moved. The lubricant action of the water sheath can be explained in part by the observation that the movement, diffusion or exchange of water molecules within the sheath has been predicted to be greatly increased theoretically compared to that observed for water molecules in the bulk phase or compared to the exchange of water molecules between the sheath and the bulk phase (Arani et al., 1995). Furthermore, if the main site of entry of water molecules from the bulk phase were to occur at the leading edge of the sheath in whatever direction the protein was moving, largely as a consequence of the increased opposing viscus force that would be present at this point compared to that at the sides of the sheath, and the main site of exit were at the trailing edge of the sheath covering the moving protein molecule for the same reason, now manifest as drag, then the protein would, in effect, move within the water sheath at a greater velocity than could be the case if it were naked within the bulk water phase. Second, its motion in response to the emergence of a hydrophobic amino acid from the ribosomal exit tunnel can introduce large bending. In this movement, which is obligatorily coupled to a change of direction, electric current flows are expected to play a role.

2.6. Low Reynolds Number physics

The basic equation of fluids is the Navier Stokes equation:

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} = \eta \nabla^2 \mathbf{v} + \mathbf{f}$$

where $\mathbf{v}$ is the fluid velocity, $\mathbf{f}$ represents the forces acting on the fluid and η is the kinematic viscosity. The presence of an object in the fluid introduces boundary conditions for the flow of the fluid. Reynolds found in 1883 that the nature of fluid flow was governed by a one dimensionless parameter, R, introduced earlier by Stokes in 1851, now called the Reynolds number, given by $R = lv/\eta$, where l was a length scale, v the velocity scale and η the kinematic viscosity. For small objects, moving with modest speeds, the Reynolds number is very small. This information translates into the statement that the coefficient of the viscosity term, $\nabla^2\mathbf{v}$, is very large in the low Reynolds number regime. Consequently, this term dominates. The other terms can be dropped and the appropriate equation to solve in this low Reynolds number regime is

$$\eta \nabla^2 \mathbf{v} = -\mathbf{f}$$

In this equation there is no time. The operator, ∇^2, depends on geometry and contains topological information as well. Thus, the fluid flow equations are very different for low Reynolds number flows. A small object in such a fluid moves when a force acts on it or if its shape changes. This behavior can be explained by considering a tangent plane at a point on the surface of a small object. Then, the motion of the point on the surface can be described by a vector, pointing in some direction in the three dimensional coordinate system at that point given by the tangent plane and its normal. A general movement of this point represents a translation plus a rotation. For an object such as a sphere, a periodic change in its shape leads to a change in the direction of its motional vector such that its motion will not return to its original direction unless a different force is applied. Shapere and Wilczek (1988) give an elegant geometrical method of working out the motion that results from changes of shape.

A small object in a low Reynolds number environment moves with difficulty because the large viscosity prevents any kind of inertial motion. Movement of an object either requires a force to act on the object or, as also explained earlier, the object must change its shape in a periodic way. Biological systems overcome this viscosity barrier in two ways: first, by changing their shape and, second, by surrounding themselves with a water sheath, which acts as a lubricant and makes movement easier as previously explained.

2.7. Diffusion equation

A differential equation for the diffusion process between the two postulated water components, one representing the structured water that forms a sheath around the polypeptide chain, and the other, representing an unstructured normal component of water which usually lies outside the sheath, can be written as

$$\frac{dC}{dt} = D\nabla^2 C + AC$$

where C is a two component vector and both D and A are two by two matrices. Thus,

$$C = \begin{pmatrix} C_1 \\ C_2 \end{pmatrix}$$

$$D = \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix}$$

$$A = [a_{ij}]$$

The D_i and a_{ij} are real constants that need to satisfy certain constraints in order to produce the types of helical wave solutions desired. The matrix, A, describes the way the two components of water mix, while the matrix, D, describes the way one component diffuses into the other component. The value of these matrix entries will depend on the nature of the amino acid that interacts with the structured water component.

We look for solutions of the form:

$$C_i = P_i^+ F^+(r) e^{-i\omega t} + P_i^- F^-(r) e^{i\omega t}$$

where F^+ and F^- are 'outgoing' and 'incoming' singular solutions of the Helmholtz equation

$$\nabla^2 F + k^2 = 0$$

in those cases where k is real.

It has been shown by Gomatam (1982), following the work of others (De Simone et al., 1973), that this system can yield singular helical waves and gives a good description of these helical waves as well as the other exotic structures that are observed. Usually, such structures emerge from a study of non-linear differential equations. However, Gomatam (1982) was able to show that these structures can emerge within the framework of a linear model because he allowed singular solutions. In our case singular solutions are very reasonable because we are interested in describing a system where singularities are expected due to self-intersections. Thus, in our model the assumptions made by Gomatam (1982) are actually physically required.

Writing ∇^2 in cylindrical coordinates, (r, ϕ, z), and setting $P_1^+ = P_1^-$, allows ∇^2 to take the form:

$$\nabla^2 = \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \phi^2} + \frac{\partial^2}{\partial z^2}$$

The singular solution for the structured component of water surrounding a polypeptide, designated C_1, that emerges from this function is

$$C_1 = A_1 \sqrt{[(J_m^2(qr) + Y_m^2(qr)]}\ \cos[\delta_m(qr) + m\phi - pz - \omega t]$$

where $q = \sqrt{(k^2 - p^2)}$ with $(|p| < |k|)$. There is a relationship between the two water components that comes from the eigenvalue problem associated with the diffusion equation. In the Debye limit, (large qr), which Gomatam (1982) showed holds even for modest values of qr, the solution becomes

$$C_1 \approx \frac{A_1}{qr}\cos[q\,r + \pi/2 - \phi - \omega t]$$

In our problem there are a sequence of r_i values corresponding to the different hydrophobic chains. One immediate observation is that the structured component will vanish when

$$qr + \pi/2 - \phi - \omega t = \pi/2$$

This result gives a time for the end of the first stage of folding. The large bending of the polypeptide chain observed during the process of folding is expected to come from a more detailed analysis of the breakdown of the structured component of water, that itself results from the self-intersection experienced by the water sheath at this point. The region of self-intersection is expected to possess a skewed and roughly elliptical form. When the breakdown of the water sheath occurs, the structured component of water in this region will diffuse rapidly into the surrounding unstructured component of water. The amino acid at this point in the polypeptide chain reacts to this event by bending in the direction opposite to that taken by the water molecules migrating away from this point. We will make an intuitive estimate of the magnitude of this event. However, a full treatment would need to take non-linear effects into account and requires numerical methods.

2.8. Understanding bending

The understanding of sharp bending we give depends upon a simple model for the water sheath as two hollow filaments. One filament is composed of the structured water sheath that surrounds the polypeptide chain on the N-terminal side of an emerging hydrophobic residue and the second filament is composed of the structured water sheath that surrounds the polypeptide chain on the C-terminal side of an emerging hydrophobic residue. Near the self-intersection region the ends of these two filaments are very close together and we can calculate the way they interact. This calculation is accomplished by first constructing a simple model for the interaction between the filaments. The nature of the interaction depends upon the special features of the low Reynolds number environment without inertial effects. For the model we construct an interaction term between the filaments that is mediated by a scale invariant force field. This interaction is modelled by considering an interaction between each filament separately, say J_i and then J_{i+1} and the force-mediating field, A_i. The interaction between the filaments can then be found using the techniques of quantum field theory. There is a unique 3-dimensional force field interaction that is scale invariant and is known as a Chern Simons interaction (Chern and Simons, 1974). Integrating over this special force mediating field, which is a standard technique of quantum field theory, the interaction energy between the two filaments is found. From this relationship it is clear that the interaction is large at short distances and causes bending of the structured water sheath in a characteristic and singular configuration. This bending of the water sheath in the absence of inertial effects causes the polypeptide chain itself to bend sharply at this point.

2.9. Justification of the use of Chern Simons field theory

The second filament is composed of the structured water sheath on the C-terminal side of the same hydrophobic residue. We have introduced the Chern Simons theory as the unique theory appropriate for use in a low Reynolds number environment. However, there are physical reasons for using this theory that we now explain. Maxwell's classical theory of electromagnetic phenomenon also exists in a quantum version: quantum electrodynamics (QED). This theory has proved very successful in describing the interaction of charged particles with electromagnetic fields. The theoretical formulation of QED uses a Lagrangian

density (Kibble and Berkshire, 2005), where the electromagnetic part has the structure, E^2-B^2, and where E and B represent the electric and magnetic vector fields respectively. Theoretically, the term, $\Sigma_{i=1}^{n} E_i B_i$, can be added to the Lagrangian. However, in most usual applications this term does not contribute to a description of the physics of electromagnetic processes. Nevertheless, in terms of the vector potential in the gauge, $A_0=0$, $E_i=-\partial_0 A_i$ and $B_i=\epsilon_{ijk}\partial_j A_k$. Here the subscript, 0, represents the time component, while the subscript, i, represents spatial components and ϵ_{ijk} is the completely antisymmetric 3 tensor. Then, in terms of these potentials:

$$\Sigma_{i=1}^{n} E_i B_i = \partial_0(\epsilon_{0ijk} A_i \partial_j A_k)$$

Integrating over space and time gives the real contribution that this term makes to the action of QED. Because of the partial time derivative, which acts on the remaining expression, the time integral can be evaluated between, say, $t=t_i$ and $t=t_f$. The answer emerges as the difference of the three dimensional spacial integral of

$$\epsilon_{0ijk} A_i \partial_j A_k$$

for $t=t$ and $t=0$. Again, ϵ_{0ijk} is the completely antisymmetric 4 tensor. However, it can be seen that these terms are just the Chern Simons expressions that we used earlier. Normally, these time dependent radiation gauge fields are set equal to zero in condensed matter physics and also ignored in QED, where the initial and final state configurations are assumed to have no radiation fields. For the water molecules in our coherent domain, this is clearly no longer true; hence, the Chern Simons term should not be ignored. Because the Chern Simons gauge fields are the gauge fields of electromagnetism, the currents introduced are electromagnetic currents that are present at the boundary of the coherent water domains. Thus, the term introduced is a basic electromagnetic interaction term which is important for coherent water domains.

2.10. Mathematical formulation of the model that describes the coherent water sheath domain

Let us now describe this model in mathematical terms. The model has two types of variable: the filaments, J_i and J_{i+1}, that are each regarded as carrying a current, and the force field itself, A_i. Correlation between the two filaments is then understood as the correlation between the two filament currents that are supported on a pair of curves, say, C_1 and C_2. The mathematical formulation describes an action, S. This action has the dimensions of energy x time. The two terms in the action correspond to the filaments interacting with the force field and the special scale invariant force field itself given by the Chern Simons term. The action is

$$S=\gamma\int d^3x\epsilon^{ijk}A_i\partial_j A_k+\sum_a\int d^3x A_i J^i_{C_a}.$$

To determine the interaction between the filaments we introduce a correlation between them, W_a, that is defined by

$$W_a=\exp\left(i\int d^3x A_i J^i(x)\right)$$

For ease of calculation a pair of exponential functions, each integrated over a different filament, will be considered and are specifically referred to later. The correlation function between the two filaments is then averaged over the force field A_i with a weight factor given by the Chern Simons term, which introduces the effects of topology into the vector force deforming the water sheath. This correlator, (W_1W_2), is then obtained by integrating out the gauge field, A_i. The methods of quantum field theory allow us to write this procedure in the following way:

$$(W_1W_2)=\exp\left(\frac{i}{2\gamma}\int dt\int d^3x\int d^3y J^i_{C_1}(\mathbf{x})G_{ij}(\mathbf{x}-\mathbf{y})J^j_{C_2}(\mathbf{y})\right)$$

where $G_{ij}(\mathbf{x}-\mathbf{y})$ denotes the Green's function (Eyges, 1972), associated to the Chern–Simons term, given by

$$L_{ij}G_{jk}(\mathbf{x}-\mathbf{y})=\delta_{ik}\delta^{(3)}(\mathbf{x}-\mathbf{y}),$$

where we defined, for compactness of notation, the operator $L_{ij}=-\epsilon^{ijk}\delta_k$. Operating with L from the left on both sides we obtain

$$(L^2)_{ij}G_{jk}=L_{ik}\delta^{(3)}(\mathbf{x}-\mathbf{y}),$$

so that the Green's function is given by

$$G_{ij}=L_{kj}(L^{-2})_{ki}\delta^{(3)}(\mathbf{x}-\mathbf{y}).$$

Now using $(L^2)_{ij}=\partial_i\partial_j-\delta_{ij}\partial^2$, we have

$$(L^2)_{ij}\frac{1}{|\mathbf{x}-\mathbf{y}|}=\partial_i\partial_j\frac{1}{|\mathbf{x}-\mathbf{y}|}-\delta_{ij}\delta(|\mathbf{x}-\mathbf{y}|),$$

which leads to

$$(L^{-2})_{ij}\delta^{(3)}(\mathbf{x}-\mathbf{y})=(L^{-2})_{ik}\partial_k\partial_j\frac{1}{|\mathbf{x}-\mathbf{y}|}-\delta_{ij}\frac{1}{|\mathbf{x}-\mathbf{y}|}.$$

Using this relationship, we obtain

$$G_{ij}=-L_{jk}\partial_k\partial_i\frac{1}{|\mathbf{x}-\mathbf{y}|}-L_{ij}\frac{1}{|\mathbf{x}-\mathbf{y}|}.$$

The first term vanishes as $L_{jk}\partial_k\partial_i=\epsilon^{jkl}\partial_l\partial_k\partial_i=0$ and, consequently, we then obtain

$$G_{ij}(\mathbf{x}-\mathbf{y})=\epsilon^{ijk}\partial_k\frac{1}{|\mathbf{x}-\mathbf{y}|}=-\epsilon^{ijk}\frac{(x-y)_k}{|\mathbf{x}-\mathbf{y}|^3}.$$

Using this expression for the Green's function, the correlator (W_1W_2) becomes

$$(W_1W_2)=\exp\left(\frac{i}{2\gamma}\int dt\int_{C_1}ds_1\int_{C_2}ds_2\epsilon_{ijk}\frac{(x-y)^k}{|\mathbf{x}-\mathbf{y}|^3}J^i(\mathbf{x}(s_1))J^j(\mathbf{y}(s_2))\right)$$

where we have introduced affine variables, s_1 and s_2, parametrizing the curves C_1 and C_2 respectively. Let us emphasize again that the interaction term which leads to the dynamics of filament interaction is very special. It is consistent with the dynamics of this process possessing an independence of size or inertia, that is expected in the low Reynolds number environment. Recalling now that $J(x)=J_{Ci}dx^i$, we see that the equation of any point on one single filament has a term

$$\frac{d\mathbf{x}(s)}{dt}=\mathbf{x}(s)\times\mathbf{v}$$

with the velocity

$$\mathbf{v}=\oint\frac{(\mathbf{x}-\mathbf{y})}{|\mathbf{x}-\mathbf{y}|^3}\times\mathbf{dr}$$

The implication of this equation is that when two separate and unconnected filament segments are very close together, sharp bending occurs. This result follows from the observation that the interaction energy term between filaments, which itself is a

consequence of the Chern Simons term, has the structure:

$$\int_{C_1} ds_1 \mathbf{B}(s_1).\mathbf{J}(\mathbf{x}(s_1))$$

where

$$\mathbf{B}(s_1) = \int_{C_2} ds_2 \frac{\mathbf{R}(s_1,s_2)\times \mathbf{J}(\mathbf{x}(s_2)}{|R(s_1,s_2|)^3}$$

and where $\mathbf{R}(s_1,s_2)=\mathbf{x}(s_1)-\mathbf{y}(s_2)$ and $R(s_1,s_2)=|\mathbf{R}(s_1,s_2)|$. From this expression we clearly see that **B** resembles a magnetic field that has been generated from a current. Thus, the equation for a current element of unit mass, due to such an interaction and representing a Lorentz type force in physics (Jackson, 1998), is

$$\frac{d\mathbf{v}}{dt} = \mathbf{v}\times\mathbf{B}$$

Thus, we have a Biot–Savart type interaction between filaments that carry currents (Jackson, 1998). The implication of this result is that when two filaments, representing neighboring strands of structured water, are close, they give rise to a rotation. This large, short distance effect will become important in the singular region and will lead to bending. The presence of the helical waves of water exiting the sheath in this region that we earlier found together with this rotation then leads to sharp bending. Thus, the qualitative model proposed can offer a physical reason for the bending. It is based firmly on the two physical principles that underpin the model: the low Reynolds number environment and the presence of a structured water sheath surrounding the polypeptide chain. The final shape for our simple protein is, therefore, a twisted one dimensional structure. The sharp bending, together with the presence of helical water waves give rise to such a shape. The special periodic nature of the protein continues to be present during such a transformation.

2.11. Possible practical applications of protein folding theory

Some possible practical outcomes of understanding protein folding better may involve the development of more effective drugs against HIV infection, in particular ones that are less subject to the development of resistance. Current efforts to understand the mechanism of resistance to HIV treatment with approved clinical inhibitors of the HIV-1 protease have been directed towards understanding the role in resistance played by known mutants of the HIV-1 protease. These mutants and polymorphisms only affect the final structure or very latest stage of protein folding and fall into two categories. First is the active site mutation, D30N, which has several diverse effects (Soares et al., 2010) and particularly affects the hydrogen bonding between aspartate-30 and the drug (Ode et al., 2005; Moseby et al., 2008; Liu et al., 2011). Second are the non-active site mutations that affect the interaction between these residues and those residues lining the active site binding pocket (Ode et al., 2005, 2007), mainly by small effects on late folding that distort and reduce the size of the binding pocket, affect the flap and in some cases actually restore the hydrogen binding between the drug and the protein.

Recently there has been interest in developing drugs for the treatment of HIV infection that affect the primary folding of the HIV protease type-1, either early when the monomer folds or late when the dimer assembles (for review see Broglia et al., 2008). This approach is in part based upon the decoupling of monomer folding and assembly of the dimer (Levy et al., 2004) and offers the hope of overcoming the development of resistance that is a problem with active site inhibitors (Broglia et al., 2005). Clearly, the effectiveness of this strategy should be increased by a better understanding of the mechanism of the early stages of protein folding. In this way synthetic efforts could be directed towards the synthesis of inhibitors that bind to or mimic those segments of the emerging polypeptide chain that direct the folding process. Of course the most effective approach would be to target any aspect of the folding process in the target protein that differs most from those found more generally in other non-target proteins. Alternatively, a peptide or small molecule could be designed to target a small, unique segment of the protein sequence that promotes hydrophobic collapse, thus restricting the inhibition of folding to the target protein alone (Broglia et al., 2008). For example, we note that the early emerging segment of the HIV-1 protease, residues 21 to 24, contains 4 residues of which 75% are hydrophobic. This segment is an excellent candidate for initiating the early stage of protein folding and, therefore, a good target for drug development. Furthermore, even though the first residue in this segment is glutamate-21, the side chain carboxyl is hydrogen bonded to two of the hydrogen atoms on the side chain methyl group of threonine-12 in the final folded protein in one of the published crystal structure (1HSI.pdb). This target emerges from the theoretical analysis of the early stages of protein folding presented here and our short analysis of the HIV-1 protease itself (Fig. C2).

2.12. Conclusions

There are a number of testable elements in the simple model:

1. The ease and speed of folding results from the physics of a low Reynolds number environment.
2. The folding process begins when the emerging polypeptide chain starts to form a transient α-helical structure as a consequence of movement within or out of the ribosomal exit tunnel and into the low Reynolds number environment of the cytoplasm.
3. A structured water sheath forms around the emerging polypeptide chain and is involved in the mechanism of folding at all further stages in the process.
4. Hydrophobic amino acids are essential for generating tertiary structure as they emerge from the tunnel. In almost all cases tertiary structures can only form when hydrophobic amino acids are present. The few exceptions either involve external components, such as divalent metal ions coordinately bound to cysteine residues or the presence of suitable electron acceptors allowing the formation of intrachain disulphide bonds, or when alanine functions as a hydrophobic amino acid (see Appendix C for an explanation of the exceptions). Quaternary structure can sometimes form in the absence of hydrophobic amino acids (see Appendix C for evidence and explanation).
5. Each hydrophobic amino acid that emerges from the tunnel acts by causing the water sheath to bend at their position in the polypeptide chain.
6. The presence of a minimum number of hydrophobic amino acids, usually 6–7, are required for long-distance folding and the collapse of the molten globule to begin.
7. The bending of the water sheath at each emerging hydrophobic amino acid sum together and, when the rotation of the sheath reaches 360°, a type-2 self-intersection point is created in the sheath that initiates the exit of a stream of water molecules from the sheath in a directional helical wave pattern into the surrounding unstructured bulk water phase.
8. The water molecules exiting the sheath cause an oppositely directed long-distance folding of the polypeptide chain as a consequence of the second law of thermodynamics.
9. Transferring the group of key hydrophobic amino acids that initiate a particular long-distance fold to a different location will also transfer the initiation site of that long-distance fold to the new location of these same hydrophobic amino acids.

10. Nanobubbles of dissolved atmospheric gasses protect hydrophobic amino acid side chains at the surface of proteins from direct contact with water molecules.
11. The quantum force between different gas molecules in widely separated nanobubbles can contribute to the forces guiding folding by pulling these nanobubbles towards each other and dragging the underlying polypeptide chain with them. We expect this force to be a many bodied force, whose two body part has been estimated in our simple model.
12. The funnel or limited number of folding patterns observed in proteins is created by the way two forces, the reaction to water exiting the sheath during self-intersection of the sheath and the dipole–dipole interaction between gas molecules in different nanobubbles, act in a time ordered sequence in the mechanisms just described during protein biosynthesis.

Acknowledgments

We thank Drs. S. Adhikari, A. Khan, K. Mok, J. O'Sullivan and G. Robinson for their suggestions and critical reading of the manuscript; Drs. Andrew Knox and Darren Fayne for assistance with the computer application, MOE and Dr. Andrew McDonald for writing the Perl script.

Appendix A. Calculation of the volume of water loss from the surface of lysozyme as a result of folding

The surface area of a single molecule of the native lysozyme polypeptide from the crystal structure in 1LYS.pdb, obtained by the rolling sphere algorithm, using the application, Molecular Operating Environment (Molecular Operating Environment (MOE), 2012.10) was found to be 163.9 nm^2, while the surface area of the completely unfolded lysozyme peptide was found to be 211.3 nm^2, based on the accessible surface area of all of the constituent amino acids as assessed in the tripeptides, Gly-X-Gly (Table 4.4, p. 142 of Creighton, 1993; Miller et al., 1987). Consequently, the folding of lysozyme into the mature native structure involves the total loss of surface exposed to the aqueous environment of some 47.4 nm^2, composed of 10.36 nm^2 over hydrophilic residues and 37.1 nm^2 over hydrophobic residues, together with the surface water bound to this surface before folding.

Computer analysis of the mature, native, folded surface of lysozyme (Molecular Operating Environment (MOE), 2012.10) revealed that 78.16% of the surface (128.1 nm^2) was over hydrophilic residues and 21.84% of the surface (35.8 nm^2) was over hydrophobic residues. Our model of the structure of the arrangement of the water molecules in the water sheath surrounding lysozyme is composed of three layers: the inner, first layer contains 216 molecules of tightly bound water located only over the exposed portions of the hydrophilic residues; the second, middle layer contains 287 molecules of loosely bound water, again over only the exposed portions of the hydrophilic residues; the third, outer layer contains 458 molecules of water that is not formally bound but whose motion is affected by the presence of the protein, forming a coherent phase over the entire surface of the protein (Sen and Voorheis, unpublished). The first and second layers over the exposed hydrophobic residues are thought to be covered by nanobubbles of dissolved atmospheric gasses. Consequently, in the native state of lysozyme there are 861 water molecules of water in the inner two water layers over the 128.2 nm^2 of surface area covering the exposed portions of the hydrophilic residues or 6.72 water molecules per nm^2 and 100 molecules of water in the outermost water layer over the 35.8 nm^2 of the surface area covering the exposed portions of the hydrophobic residues or 2.79 water molecules per nm^2. We assume that both hydrophobic and hydrophilic residues that are buried in the native structure will subsequently bind as much surface water after the protein is unfolded as their respective types of residues do at the surface of the native enzyme. However, these same residues will loose their surface water when folded to the interior of the protein. Therefore, because all of the residues are exposed in the unfolded state the total H_2O displaced by folding is equal to H_2O over hydrophilic residues displaced + H_2O over hydrophobic residues displaced. Consequently,

H_2O displaced by folding

$$= 861 \text{ molecules } H_2O\left(\frac{10.36\ nm^2}{128.2\ nm^2}\right)$$

$$+ 100 \text{ molecules } H_2O\left(\frac{37.1\ nm^2}{35.8\ nm^2}\right)$$

$$= 173 \text{ molecules } H_2O \text{ displaced by folding}$$

then the volume of the water sheath displaced by the removal of these water molecules during folding is given by

$$= 173 \text{ molecules } H_2O\left(\frac{3.28 \times 10^{-2}\ nm^3}{\text{molecule of } H_2O}\right)$$

$$= 5.68\ nm^3$$

and this volume is 43.2% of the volume of the unfolded peptide chain itself (13.16 nm^3), based on the Van der Waals volume of the constituent amino acids (Table 4.3, p. 141 of Creighton, 1993).

Appendix B. Analysis of contacts between hydrophobic amino acids

Analysis of the contacts between the side chains of hydrophobic amino acids in ten different monomeric proteins across a wide range of molecular weights and soluble in an aqueous environment, revealed the unexpected result that some hydrophobic amino acids have about a 2-fold increased tendency to occur together in the tertiary structure compared to all other possible hydrophobic side chain combinations (Figs. B1 and B2). Furthermore, all ten of the proteins had either one or two types of contact that were present at a much higher frequency than all of the remaining hydrophobic amino acid contacts. Six of the ten proteins (ribonuclease, β-lactoglobulin, carbonic anhydrase, ovalbumin, phosphorylase and β-galactosidase) had one type of contact that was higher than all other remaining contacts amongst the hydrophobic amino acids, while the remaining four proteins (insulin, lysozyme, carboxypeptidase and bovine serum albumin) had two types of contact amongst the hydrophobic amino acids that were of a higher frequency than any of those remaining.

Within all ten proteins the highest frequencies of contact were six phenylalanine – phenylalanine contacts, three isoleucine – isoleucine contacts, two valine – isoleucine contacts and one each of phenylalanine – valine, phenylalanine – leucine and valine – leucine contacts.

The index of contact frequency was defined as

$$\frac{N}{N_x(N_x-1)}$$

for an amino acid of one identity contacting another amino acid of the same identity and as

$$\frac{N_x}{N_x(N_y)}$$

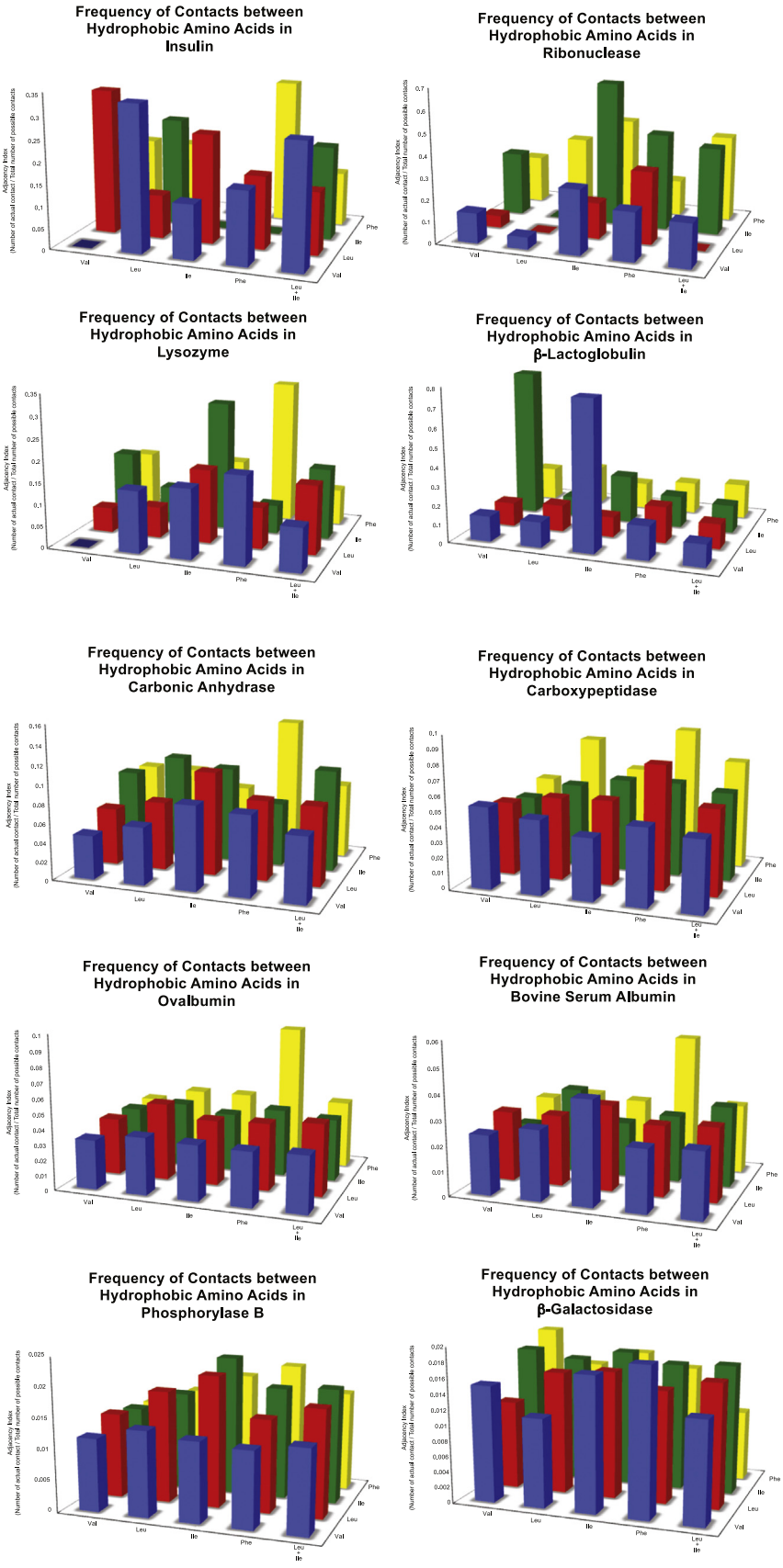

Fig. B1. Frequency with which hydrophobic amino acids contact each other in the 3-D structure of 10 different proteins.

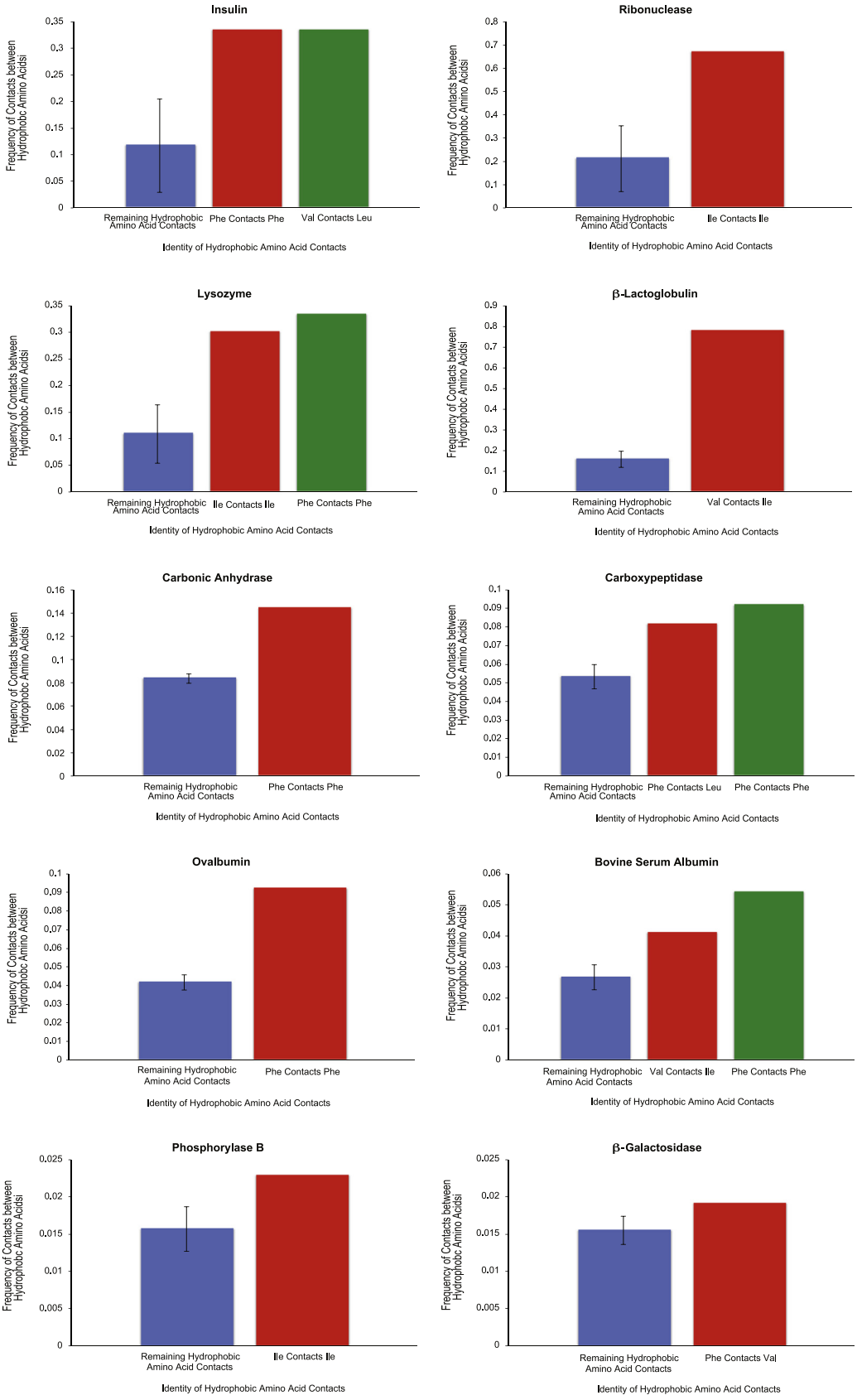

Fig. B2. Difference between the frequency of the hydrophobic amino acids with the greatest frequency of contacting each other and the frequency of the remaining hydrophobic amino acids contacting each other in the 3-D structure of 10 different proteins.

```
#!/usr/bin/perl

use strict;
# initialise variables
my $i=0;       # (non)match counter
my $j=0;       # sequence counter
my $desc="";
my $seq="";
my $regex="[FLIVX]";   # the search pattern (a regular expression)
my $msl=12;    # the minimum length of sequence to consider
# start the loop, reading one line of input at a time
while (<>) {
# $_ holds the current line
# remove newline char from $_
chomp();
# sequence description lines start (^) with a greater-than symbol (>)
if( /^>/ && /mol:protein/ ){   # it's a protein-sequence description line
# record the description in $desc
$desc = $_;
# read the next line of input
chomp($_ = <>);
# remember the sequence
$seq = $_;
# increment the protein-sequence counter
$j++;
#if (! /X/ ) { $j++ }
# test for absence or presence of the pattern held in $regex,
# excluding any sequences consisting only of 'X',
# or those of length less than $msl:
if ( (! /$regex/) && (! /^X+$/) && ( length $seq >= $msl ) ) {
# print the description and the sequence
print "$desc\n$seq\n";
# we have a (non-)match, so increment the (non-)match counter
$i++;
}
}
}
print "$i out of $j protein sequences do not match /$regex/\n"
```

Fig. C1. Perl script, written by Dr. Andrew McDonald, for selection of sequences from the RSCB Protein Data Bank that contain 12 or more amino acids and also lack hydrophobic amino acids as well as unspecified (*X*) amino acids.

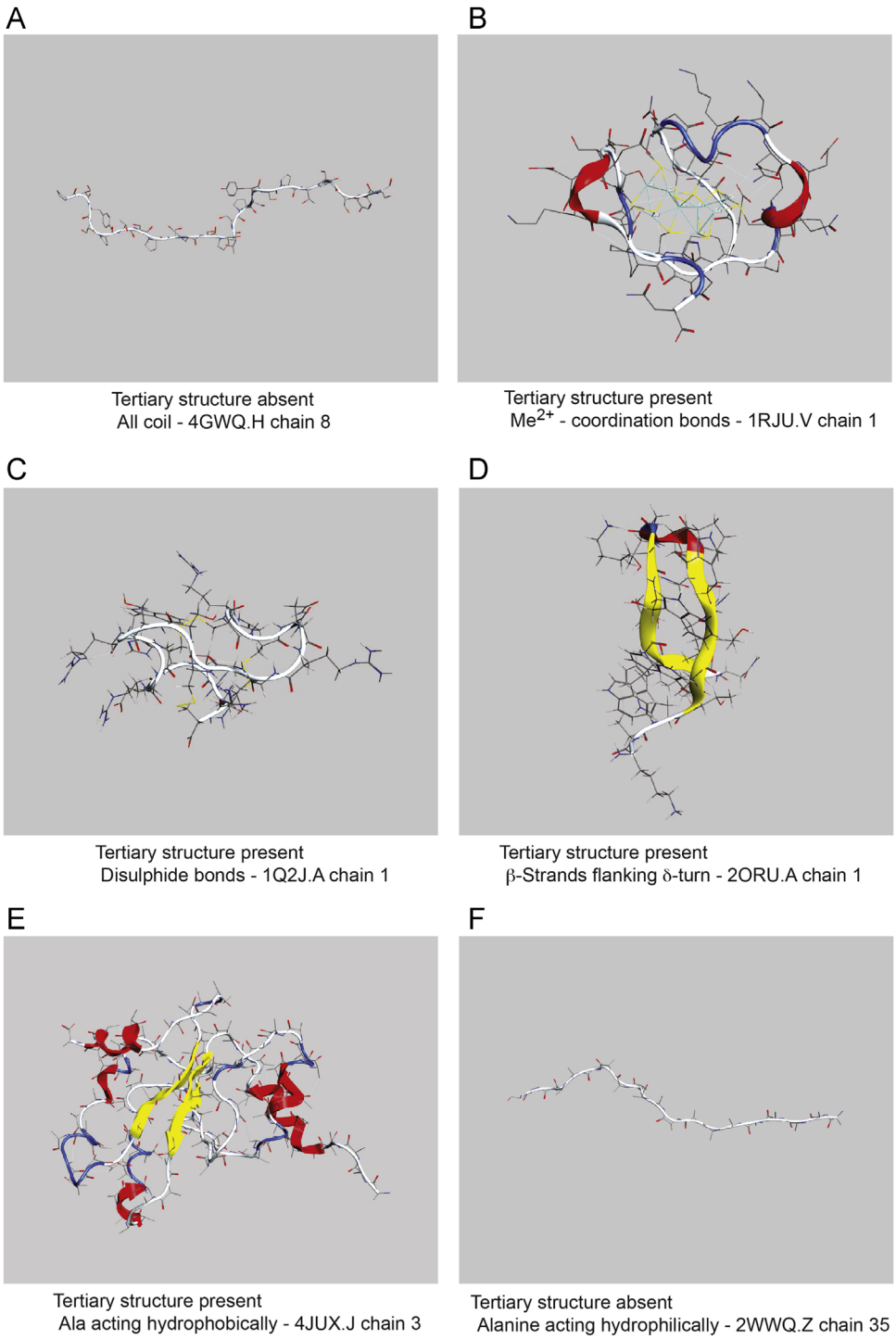

Fig. C2. Single examples of the dependence of the tertiary structure of proteins from each structural class in the list below on the lack of hydrophobic amino acids together with one of the following conditions: (Panel A) no other condition, (Panel B) the presence of Me^{2+} ions coordinately bonded to cysteine residues in a tetrahedral configuration, (Panel C) the presence of disulphide bridges, (Panel D) the presence of β-strands flanking a β-turn, (Panel E) long chain polyalanine acting as a group of hydrophobic residues and (Panel F) short chain polyalanine acting as a group of hydrophilic residues.

for an amino acid of one identity contacting an amino acid of a different identity, where N_x is the number of amino acids of identity x and N_y is the number of amino acids of identity y in a given protein. Then, $N_{x<->x}$ are the number of amino acids of identity x that contact an amino acid of the same identity in a given protein and $N_{x<->y}$ are the number of amino acids of identity x that contact an amino acid of identity y in a given protein.

Appendix C. Analysis of the dependence of protein primary structure on the presence of hydrophobic amino acids

The relationship between the tertiary structure of proteins and the presence or absence of hydrophobic amino acids was analysed in a downloaded version of the RSCB Protein Data Bank (pdb-seqres.txt). This local data bank was searched initially using a Perl script (Fig. C1), kindly written by Dr. Andrew McDonald, Post Doctoral Research Staff, School of Biochemistry and Immunology, Trinity College Dublin. The script was saved as a fasta-match.pl file and, using the Mac terminal, searched with the command "cat pdb−seqres.txt|perl fasta−match.pl|more hydrophobics.txt" to generate a text file.

There were 281,012 sequences in the database of which 263,637 were protein sequences. The script selected 426 of these protein sequences that met the following criteria:

(1) The amino acids, phenylalanine, leucine, isoleucine and valine were absent.
(2) Each sequence in an experimental structure file contained 12 or more amino acids. This criterion was based on the idea that the shortest sequence that could possibly form a tertiary structure would be two α-helical segments of 4 amino acids each (one complete helical turn) connected by a 4-amino acid β-turn to enable hydrogen bonding between the two α-helical segments.
(3) Each sequence did not contain unspecified residues (*X*).

These 426 sequences were further reduced to 73 sequences by examining each sequence in turn by hand and selecting those meeting the further criteria:

(4) At least 12 consecutive amino acids in the sequence could be identified in the structures obtained by either x-ray crystallography or by nuclear magnetic resonance.
(5) Each sequence was not a fragment of a larger sequence that was also present and that also satisfied all other criteria.
(6) Metal ions, known to be present close to the sequence, were present in the experimental structure, eliminating one structure (1T2Y.A chain 1) from the analysis.
(7) The structure did not contain a cyclic peptide, which also eliminated one structure (3OE0.I chain 2), because the resulting structure automatically formed a tertiary structure as a consequence of the cyclization reaction that depends on catalysis outside of the protein sequence itself.

Of the 73 sequences, the greatest number, 50, had no tertiary structure as predicted (List below and Panel A of Fig. C2).

There were 11 additional sequences that only formed tertiary structures because of the mediation of 3–9 divalent metal ions, which themselves formed tetrahedral coordination bonds with

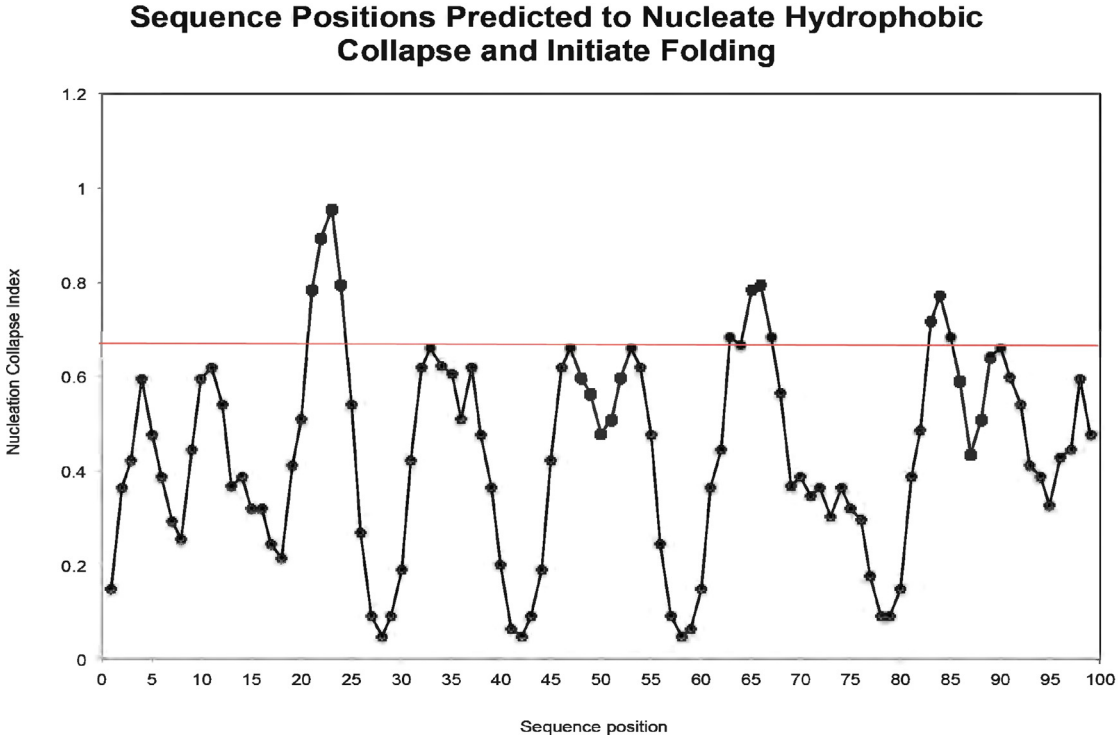

Fig. C3. Prediction of the short segments of the HIV-1 protease that will nucleate the hydrophobic collapse and initiate folding. Hydrophobic amino acids, Phe, Leu, Ile and Val, are assigned values of 1 and all other amino acids as well as imaginary amino acids flaking the sequence are assigned values of 0. Windows of 3 to 9 residues in length in turn are run along the sequence in steps of one residue and the average values from all runs at each position are averaged to create a nucleation collapse index that is plotted as a function of the mid position of each segment. The mid position is clear in the case of a window with an odd number of residues. In the case of a window with an even number of residues, the window is constructed at each position in the sequence by including the values of $(n/2)-1$ residues on each side of the central position together with 0.5 of the value of the single adjacent residues immediately flanking both sides of this segment of the sequence. The red horizontal line is placed at a value of 0.67. Sequence segments above this value are predicted to direct hydrophobic collapse and the first of these segments initiates folding. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the sulphur atoms of widely spaced cysteine residues (List below and Panel B in Fig. C2). In this case the clear prediction is that this tertiary structure will collapse if the metal ions are removed. Consequently, these tertiary structures depended on the presence of constituents, divalent metal ions, outside of the sequence itself.

A further 9 sequences formed tertiary structures as a consequence of 2 or more disulphide bonds between widely spaced cysteine residues (List below and Panel C in Fig. C2). Once again these tertiary structures are dependent on the presence of constituents, electron acceptors, outside of the protein sequence itself because the formation of disulphide bonds is an oxidative reaction.

Two sequences (2ORU.A chain 1 & 2EVQ.A chain 1) out of the 73 sequences represented genuine examples of an exception to the requirement for hydrophobic residues to form tertiary structure in a protein sequence. In both of these 2 cases β-structures were present separated by a single β-turn, which allowed the formation of a small 2-component anti-parallel β-sheet held together by several hydrogen bonds (List below and Panel D in Fig. C2). However, outside of these two strands and turn only random coil structure exists and this situation will lead to folding problems if most of the protein sequence is located in these random coil regions. The two exceptions found in the RSCB Protein Data Bank are small sequences with only a very few residues outside of the β-strands.

Finally, one sequence, polyalanine (4JUX.J chain 3), had extensive tertiary structure (List below and Panel E of Fig. C2). However, alanine has been identified previously to possess considerable hydrophobic character when incorporated into protein because of its alkyl side chain and to function in some but not all circumstances as a hydrophobic amino acid (*e.g.* see Table 2-2, p. 52 in Metzler, 2001). In this case the length of the polypeptide chain also appears to play some role because relatively long chains of polyalanine (130 residues) produce a complex tertiary structure (List below and Panel E of Fig. C2), while considerably shorter chains of polyalanine (19 or 20 residues) yield secondary structures without tertiary structure (List below and Panel F in Fig. C2: Compare the structure in 4JUX.J chain 3 with the structures in both 3RFR.D chain 2 and 2WWQ.Z chain 35.

The formation of quaternary structure will be affected if it depends on the interaction of exposed hydrophobic amino acid side chains from each monomer at their interface within a final dimer or multimer. On the other hand proteins without tertiary structure can also at times form quaternary structure. For example in all of the collagen or collagen-like proteins in the database, the simple α-helices from different protein molecules form helical supercoils containing four primary helices. *e.g.* see all of the crystal structures in the collagen and collagen-like protein entries in the "No Tertiary Structure" section of the List below.

An example of the practical outcome of the theory of protein folding could be the identification of segments of the protein sequence that are either directing folding of the monomer or are responsible for the formation of the binding interface between monomers within the quaternary dimer. Here we analyze the sequence of the HIV-1 protease for the most likely segment initiating the collapse of the molten globule at the beginning of the mature folding of the monomer with a view to predicting the most effective sequence for a synthetic peptide that could be used to disturb folding and, consequently, functionally inactivate the protein (Fig. C3).

The following list gives the identity of the 73 unique sequences from the RCSB Protein Data Bank that lack strictly hydrophobic residues (F, L, I, V) and satisfy the selection criteria above in Appendix C.

No tertiary structure

2G66.B chain 2=collagen-like peptide, 27 residues, all coil—no tertiary structure

3U29.A chain 1=triple Helical Peptide, 24 residues, all coil—no tertiary structure

3T4F.A chain 1=collagen mimetic peptide, 24 residues, all coil—no tertiary structure

3A08.A chain 1=collagen-like peptide, 23 residues, all coil—no tertiary structure

3UKZ.C chain 2=nuclear cap-binding protein subunit 1, 23 residues, coil & 1 δ-turn—no tertiary structure

4GWQ.H chain 8=50S ribosomal protein L50, 25 residues, all coil—no tertiary structure

1CJF.C chain 2=polyproline, 15 residues, all coil—no tertiary structure

2OTW.E chain 5=polyglutamine+1C-te-glycine, 11 residues, all coil—no tertiary structure

1GUW.B chain 2=histone H3 tail, 18 residues, 2 coils & 1 β-strand, no tertiary structure

2KKG.A chain 1=major prion protein fragment, 35 residues, all coil—no tertiary structure

1ETF.B chain 2=REV peptide, 23 residues, 1 α-helix—no tertiary structure

1G70.B chain 2=RSG 1.2 peptide, 16 residues, 1 α-helix & remainder coil, no tertiary structure

1T8Z.C chain 3=engineered tryptophan zipper protein, 51 residues, all α-helix, no tertiary structure

1ZR7.A chain 1=WW domain of Huntington interacting protein FBP11/HYPA, 30 residues, 3 β-strand & 1 -turn & 1 -turn & β-turn & coil, no tertiary structure

1QSU.A chain 1=collagen-like peptide, 30 residues, all coil, no tertiary structure

1K2DP.A chain 3=Myelin basic protein fragment, 12 residues, all coil, no tertiary structure

1P16.D chain 4=carboxy-terminal domain of RNA polymerase II, 17 residues, 2 coils & 1 β-turn, no tertiary structure

2M1A.A chain 1=HIV-1 Rev ARM peptide, 26 residues, all α-helix, no tertiary structure

3UL0.C chain 2=classical nuclear localization signal peptide of cap-binding protein 80, 23 residues, 2 coils & 1 δ-turn, no tertiary structure

1I9F.B chain 2=RSG 1.2 peptide, 19 residues, 1 α-helix & 1 coil, no tertiary structure

2L44.A chains 1 & 2=C-terminal zinc knuckle of the HIVNCp7, 19 residues, coil & 1 δ-turn & 1 γ-turn, tertiary structure formed by coordination bonds from Zn^{2+} ion to 3 cysteine and 1 histidine residues

3UVU.B chain 1=nuclear import signal peptide of Flap endonuclease-1, 19 residues, all coil, no tertiary structure

1OV3.C chain 3=peptide fragment of tail of the P22(PHOX) light chain of the α-subunit of NADPH oxidase, 18 residues, all coil, no tertiary structure

1EJY.N chain 1=Nucleoplasmin NLS peptide fragment, 16 residues, all coil, no tertiary structure

1EXY.B chain 2=HTLV-1 peptide, 16 residues, 2 coils linked by a single-turn α-helix, no tertiary structure

1CAG.B chain 2=collagen-like peptide, 30 residues, all coil, no tertiary structure

1EI8.A chain 1=collagen peptide with G-P-G interruption, 29 residues, all coil, no tertiary structure

2SEB.E chain 4=human collagen-II peptide, 12 residues, 1 β-strand & coil, no tertiary structure

2RSN.B chain 2=H3K9me3 peptide, 18 residues, 2 coils joined by short β-strand, no tertiary structure

2X0L.C chain 3=engineered splice variant histone-H3 peptide, 16 residues, 1 α-helix & 2 δ-turns & remainder coil, no tertiary structure

2LA5.B chain 2=RGG peptide of human fragile-X mental retardation protein, 17 residues, 1 δ-turn & remainder coil, no tertiary structure

3O0E.L chain 7=colicin peptide OBS1, 15 residues, all coil, no tertiary structure

4JML.E chain 2=colicinE9 TBE peptide, 16 residues, 2 δ-turns & remainder coil, no tertiary structure

2IVZ.F chain 6=peptide of the colicin E9 T-domain, 16 residues, 1 γ-turn & remainder coil, no tertiary structure

2ZNE.C chain 3=alix ABS peptide, 15 residues, 1 β-turn & remainder coil, no tertiary structure

3AT0.B chain 2=fibrinopeptide A, 12 residues, 1 β-strand & remainder coil, no tertiary structure

1D5H.A chain 1=S peptide fragment, 15 residues, 1 α-helix flanked by 2 coils, no tertiary structure

2LMB.A chain 1=C-terminal RAGE peptide, 15 residues, α-turn flanked by 2 coils, no tertiary structures

2ROR.B chain 2=tyrosine phosphorylated peptide from SLP76, 15 residues, all coil, no tertiary structure

1D7Q.B chain 1=translation initiation factor eIF1A, 14 residues, all coil, no tertiary structure

1HXL.C chain 3=synthetic mini-protein, 13 residues, 1 α-helix, no tertiary structure

2HUG.B chain 2=cpSRP54 peptides, 14 residues, all coil, no tertiary structure

2LTW.B chain 2=synthetic SMAD7 derived peptide, 14 residues, all coil, no tertiary structure

4F1Z.Q chain 2=synthetic peptide from keratin, 14 residues, 1 β-strand & 1 δ-turn, no tertiary structure

1O0P.B chain 2=N-terminal SF1 peptide, 13 residues, all coil, no tertiary structure

2MJV.A chain 1=di-acetylated twist peptide, 12 residues, all coil, no tertiary structure

3VE6.B chain 2=capsid protein NLS peptide, 12 residues, all coil, no tertiary structure

3E1R.C chain 3=non-canonical coiled-coil in CEP55s, 13 residues, 1 δ-turn & remainder coil, no tertiary structure

3UI2.B chain 2=CPSRP54 tail peptide, 13 residues, 1 β-strand & 1 δ-turn & remainder coil, no tertiary structure

4F27.Q chain 2=Fibrinogen -peptide, 13 residues, 1 β-strand & remainder coil, no tertiary structure

Tertiary structure formed by metal coordination bonds

1AQS.A chain 1=copper metallothionein, 40 residues, coil & 3 δ-turns, 5 Cu^{2+} ions certainly and 2 other Cu^{2+} ions probably covalently linked to 9 cysteines—extensive tertiary structure

1RJU.V chain 1=yeast copper thionein, 36 residues, coil & 2 α-helices & 4 δ-turns & 1 γ-turn, 8 Cu^{2+} ions tetrahedrally coordinately bonded to 10 cysteines—extensive tertiary structure

2KAK.A chain 1=wheat Ec1 metallothionein, 53 residues, coil & 1 α- helical segment & 2 δ-turns & 1 γ-turn, tertiary structure formed by 3 Zn^{2+} ions tetrahedrally coordinated to 9 cysteine residues and 1 Zn^{2+} ion tetrahedrally coordinately bonded to 2 cysteine and 2 histidine residues

1NCP.C chains 2 & 4=synthetic peptide fragment containing 1 of 2 Zn^{2+}- binding domains from the HIV-1 P7 nucleocapsid protein, 18 residues, all coil, tertiary structure formed by a Zn^{2+} ion tetrahedrally coordinately bonded to 3 cysteine residues & 1 histidine residue

1DFT.A chains 1 & 2=beta-domain of mouse metallothionein, 30 residues, all coil, tertiary structure formed by 3 Cd^{2+} ions tetrahedrally coordinated to 9 cysteine residues

1DMC.A chains 1 & 2=metallothionine-I from the Atlantic blue crab, 31 residues, 1 α-helix & 1 δ-turn & 1 γ-turn & remainder coil, tertiary struc ture formed by 3 Cd^{2+} ions tetrahedrally coordinately bonded to 9 cysteine residues

1J5L.A chains 1 & 2=beta-C domain from Lobster metallothionein-1, 30 residues, 2 α-helices & 2 δ-turns & remainder coil, tertiary structure formed by 3 Cd^{2+} ions tetrahedrally coordinately bonded to 9 cysteine residues

1J5M.A chains 1 & 2=(113) Cd(3) beta-domain from American lobster met allothionein, 28 residues, 1 δ-turn & remainder coil, tertiary structure formed by 3 Cd^{2+} ions tetrahedrally coordinately bonded to 9 cysteine residues

1M0J.A chains 1 & 2=metallothionein-nc from the Antarctic black cod, 28 residues, 1 α-helix & 1 γ-turn & remainder coil, tertiary structure formed by 3 Cd^{2+} ions tetrahedrally coordinately bonded to 9 cysteine residues

2MHU.A chains 1 & 2=[113CD7] human metallothionein-2, 30 residues, 1 α-helix & 1 δ-turn & remainder coil, tertiary structure formed by 3 Cd^{2+} ions tetrahedrally coordinately bonded to 9 cysteine residues

2MRB.A chains 1 & 2=rabbit liver CD-7 metallothionein-2A, 31 residues, 1 δ-turn & remainder coil, tertiary structure formed by 3 Cd^{2+} ions tetra hedrally coordinately bonded to 9 cysteine residues

Tertiary structure formed by disulphide bridges

1Q2J.A chain 1=Mu conotoxin SmIIIA, 22 residues, all coil, tertiary structure formed by 3 disulphide bridges

1GIB.A chain 1=Mu conotoxin GIIIB, 22 residues, coil &1 α-helical segment & 1 γ-turn, tertiary structure formed by 3 disulphide bridges

1Y2P.A chain 1=scorpion toxin -KTx6 family, 34 residues, coil & 1 α-helix & 1 δ-turn & 2 β-strands, tertiary structure formed by 3 disulphide bridges

1HP9.A chain 1=scorpion kappa-Hefutoxin 1 s, 22 residues, coil & 2 α-helices & 1 δ-turn & 1 γ-turn, tertiary structure formed by 2 disulphide bridges

2M6A.A chain 1=wheat antimicrobial peptide Tk-Amp-X2, 28 residues, 2 α-helices & 1 δ-turn & coil, tertiary structure formed by 2 disulphide bridges

2IFJ.A chain 1=synthetic peptide from alpha-conotoxin ImI, 12 residues, 2 coils linked by 1 δ-turn, tertiary structure formed by 2 disulphide bridges

2JUR.A chain 1=conotoxin alpha-RgiA, 13 residues, 1 α-helix & 1 δ-turn & remainder coil, tertiary structure formed by 2 disulphide bridges

2LU6.1 chain 1=synthetic midi-peptide based on *mu*-conotoxins, 14 residues, 2 coils linked by 1 δ-turn, tertiary structure formed by 2 disulphide bridges

1R8T.A chain 1=synthetic peptide displaying high affinity for streptavidin, 15 residues, 2 coil regions separated by a single turn α-helix and a γ-turn, tertiary structure formed by 2 disulphide bridges

Structure tertiary formed by hydrogen bonds between β-structures flanking a turn

2ORU.A chain 1=beta-hairpin xtz-1 peptide, 20 residues, 1 β-turn & 2 β- strands & coils, tertiary structure formed by 6 hydrogen bonds between 2.18–3.6 in length

2EVQ.A chain 1 = 12-residue beta hairpin HP7, 12 residues, 2 β-strands separated by a β-turn, possible tertiary structure stabilized by 4 hydrogen bonds

Tertiary structure formed by alanine acting as a hydrophobic residue

4JUX.J chain 3 = poly-A, 130 residues, 5 α-helices & 3 β-strands & 4 δ-turns & 2 γ-turns & 2 loops, extensive tertiary structure

References

Anfinson, C., 1973. Principles that govern the folding of protein chains. Science 181, 223.

Arani, R., Bono, J., Del Giudice, E., Preparata, G., 1995. QED Coherence and the thermodynamics of water. Int. J. Mod. Phys. B 9, 1813.

Attard, P., 1996. Bridging bubbles between hydrophobic surfaces. Langmuir 12, 1693.

Attard, P., Moody, M.P., Tyrrell, J.W.G., 2002. Nanobubbles: the big picture. Physica A 314, 696.

Babu, M.M., Kriwacki, R.W., Pappu, R.V., 2010. Versatility from protein disorder. Science 337, 1460.

Bernal, J.D., Fowler, R.J., 1933. A theory of water and ionic solution, with particular reference to hydrogen and hydroxyl ions. J. Chem. Phys. 1, 515.

Bhushan, S., Gartmann, M., Halic, M., Armache, J.-P., Jarasch, A., Mielke, T., Berninghausen, O., Wilson, D.N., Beckmann, R., 2010. α-Helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel. Nat. Struct. Mol. Biol. 17, 313.

Buchner, G.S., Murphy, R.D., Buchete, N.V., Kubelka, J., 2011. Dynamics of protein folding: probing the kinetic network of folding-unfolding transitions with experiment and theory. Biochim. Biophys. Acta 184, 1001.

Broglia, R.A., Tiana, G., Sutto, L., Provasi, D., Simona, F., 2005. Design of HIV-1-PR inhibitors that do not create resistance: blocking the folding of single monomers. Prot. Sci. 14, 2668.

Broglia, R.A., Levy, Y., Tiana, G., 2008. HIV-1 protease folding and the design of drugs which do not create resistance. Curr. Opin. Struct. Biol. 18, 60.

Călugăreanu, G., 1961. Sur les classes d'isotopie des noeuds tridimensionnels et leurs invariants [On isotopy classes of three dimensional knots and their invariants]. Czechoslov. Math. J. 11, 588.

Chaplin, M., 2006. Do we underestimate the importance of water in cell biology. Nat. Rev. Mol. Cell Biol. 7, 861.

Chern, S.-S., Simons, J., 1974. Characteristic forms and geometric invariants. Ann. Math. 99, 48.

Cheung, M.S., Garcia, A.E., Onuchic, J.N., 2002. Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. Proc. Natl. Acad. Sci. USA 99, 685.

Creighton, T., 1993. Proteins, 2nd ed. W.H. Freeman & Co., New York 507 pp.

DelGuidice, E., Preparata, G., Vitiello, G., 1988. Water as a free electric dipole laser. Phys. Rev. Lett. 61, 1085.

Del Guidice, E., Spinetti, P., Tedeschi, A., 2010. Water dynamics at the root of metamorphosis in living organisms. Water 2, 566.

Del Guidice, E., Tedeschi, A., Vitiello, G., Voeikov, V., 2013. Coherent structures in liquid water close to hydrophilic surfaces. In: Journal of Physics: Conference Series, vol. 442, 012028.

Del Guidice, E., Vitiello, G., 2006. Role of the electromagnetic field in the formation of domains in the process of symmetry-breaking phase transitions. Phys. Rev. A 74, 022105.

De Simone, J.A., Beil, D.L., Scriven, L.E., 1973. Ferroin-collodion membranes: dynamic concentration patterns in planar membranes. Science 180, 946.

Dobson, C., 2000. In: Pain, R.H. (Ed.), Mechanisms of Protein Folding, 2nd ed. Oxford University Press, Oxford, UK, pp. 1–33.

Dosztanyi, Z., Meszaros, B., Simon, I., 2009. Bioinformatical approaches to characterize intrinsically disordered/unstructured proteins. Brief. Bioinform. 2, 225.

Eyges, L., 1972. In: The Classical Electromagnetic Field. Dover Publications, New York, p. 82 (Chapter 5).

Frank, H.S., Wen, W.-Y., 1957. III. Ion–solvent interaction. Structural aspects of ion–solvent interaction in aqueous solutions: a suggested picture of water structure. Discuss. Faraday Soc. 24, 133.

Franks, F., Eagland, D., 1975. Role of solvent interactions in protein conformation. CRC Crit. Rev. Biochem. Mol. Biol. 3, 165.

Glaser, R., 2013. Biophysics, 2nd ed. Springer Verlag, Heidelberg & London p. 60.

Gomatam, J., 1982. Pattern synthesis from singular solutions in the Debye limit: helical waves and twisted toroidal scroll structures. J. Phys. A: Math. Gen. 15, 1463.

Huang, C., Wikfeldt, K.T., Tokushima, T., Nordlund, D., Harada, Y., Bergmann, U., Niebuhr, M., Weiss, T.M., Horikawa, Y., Leetmaa, M., Ljungberg, M.P., Takahashi, O., Lenz, A., Ojame, L., Lyubartsev, A.P., Shin, S., Pettersson, L.G.M., Nilsson, A., 2009. The inhomogeneous structure of water at ambient conditions. Proc. Natl. Acad. Sci. USA 106, 15214.

Jackson, J.D., 1998. Classical Electrodynamics, 3rd ed. John Wiley and Sons, New York 815 pp.

Kibble, T.W.B., Berkshire, F.H., 2005. Classical Mechanics, 5th ed. Imperial College Press, London 479 pp.

Klotz, I.M., 1958. Protein hydration and behaviour. Science 128, 815.

Kuntz, I.D., 1971. Hydration of macromolecules. III. Hydration of polypeptides. J. Am. Chem. Soc. 93, 514.

Levinthal, C., 1969. How to fold graciously. In: DeBrunner, J.T.P., Munck, E. (Eds.), Mossbauer Spectroscopy in Biological Systems: Proceedings of a meeting at Allerton House. University of Illinois Press, Monticello, Illinois, pp. 22–24, This difficult to find reference is reproduced verbatim at ⟨http://WWW-miller.ch.cam.ac.uk/levinthal/levinthal.html⟩.

Levy, Y., Caflish, A., Onuchic, J.N., Wolynes, P.G., 2004. The folding and dimerization of HIV-1 protease: evidence for a stable monomer from simulations. J. Mol. Biol. 340, 67.

Liu, X., Dai, Q., Li, L., Xiu, Z., 2011. Resistant mechanism against nelfinavir of subtype C human immunodeficiency virus type 1 proteases. J. Mol. Struct. 986, 30.

Mao, A.H., Crick, S.L., Vitalisa, A., Chicoine, C.L., Pappua, R.V., 2010. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. Proc. Natl. Acad. Sci. USA 107, 8183.

Mattos, C., 2002. Protein–water interactions in a dynamic world. Trends Biochem. Sci. 27, 203.

Miller, S., Janin, J., Lesk, A.M., Chothia, C., 1987. Interior and surface of monomeric proteins. J. Mol. Biol. 196, 641.

Molecular Operating Environment (MOE), 2012.10; Chemical Computing Group Inc., 1010.

Montagnier, l., Aissa, J., DelGuidice, E., Lavellee, C., Tedeschi, A., Vitiello, G., 2011. DNA waves and water. J. Phys.: Conf. Ser. 306, 012007.

Moseby, S., Morris, L., Dirr, H.W., Sayed, Y., 2008. Active-site mutations in the South African human immunodeficiency virus type 1 subtype C protease have a significant impact on clinical inhibitor binding: kinetic and thermodynamic study. J. Virol. 82, 11464.

Müller-Späth, S., Soranno, A., Hirschfeld, V., Hofmann, H., Rüegger, S., Reymond, L., Nettels, D., Schuler, B., 2010. Charge interactions can dominate the dimensions of intrinsically disordered proteins. Proc. Natl. Acad. Sci. U.S.A. 107, 14613.

Ode, H., Ota, M., Neya, S., Hata, M., Sugiura, W., Hoshino, T., 2005. Resistant mechanism against nelfinavir of human immunodeficiency virus type 1 proteases. J. Phys. Chem. B 109, 565.

Ode, H., Matsuyama, S., Hata, M., Neya, S., Kakizawa, J., Sugiura, W., Hoshino, T., 2007. Computational characterization of structural role of the non-active site mutation M36I of human immunodeficiency virus type 1 protease. J. Mol. Biol. 370, 598.

Otting, G., Liepinsh, E., Wuthrich, K., 1991. Protein hydration in aqueous solution. Science 254, 974.

Peng, J., Xu, J., 2011. A multiple-template approach to protein threading. Proteins 79, 1930.

Preparata, G., 1995. QED Coherence in Matter. World Scientific Publishing Co. Pt. Ltd., Singapore and London 236 pp.

Purcell, E., 1977. Life at low Reynolds number. Am. J. Phys. 45, 3.

Röntgen, W., 1892. Ueber die Constitution des flssigen Wassers. Ann. Phys. 45, 91.

Shandilya, B.K., et al., 2013. Selective bond breaking mediated by state specific vibrational excitation in model HOD molecule through optimized femtosecond IR pulse: a simulated annealing based approach. J. Chem. Phys. 139, 034310.

Shapere, A., Wilczek, F., 1988. Geometry of self-propulsion at low Reynolds number. J. Fluid Mech. 198, 557.

Soares, R.O., Batista, P.R., Costa, M.G.S., Dardenne, L.E., Pascutti, P.G., Soares, M.A., 2010. Understanding the HIV-1 protease nelfinavir resistance mutation D30N in subtypes B and C through molecular dynamics simulations. J. Mol. Graph. Mod. 29, 137.

Theillet, F.-X., Kalmar, L., Tompa, P., Han, K.-H., Selenko, P., Dunker, A.K., Daughdrill, G.H., Ulversky, V.N., 2013. The alphabet of intrinsic disorder: I. Act like a Pro: on the abundance and roles of proline in intrinsically disordered proteins. Intrin. Disord. Protein 1, e24360-1.

Tompa, P., 2012. Intrinsically disordered proteins: a 10-year recap. Trends Biochem. Sci. 37, 509.

Udgaonkar, J.A., 2013. Polypeptide chain collapse and protein folding. Arch. Biochem. Biophys. 531, 24.

Ulversky, V.N., 2013. The alphabet of intrinsic disorder: II. Various roles of glutamic acid in ordered and intrinsically disordered proteins. Intrin. Disord. Protein 1, e24684-1.

Voss, N.R., Gerstein, M., Steitz, T.A., Moore, P.B., 2006. The geometry of the ribosomal polypeptide exit tunnel. J. Mol. Biol. 360, 893.

Xu, Y., Purkayastha, P., Gai, F., 2006. Nanosecond folding dynamics of a three-stranded beta-sheet. J. Am. Chem. Soc. 128, 15836.