Heliyon

ELSEVIER

Received:
15 April 2017
Revised:
15 May 2017
Accepted:
19 May 2017

Cite as: Robert P. Bywater. A tensegrity model for hydrogen bond networks in proteins. Heliyon 4 (2018) e00307.
doi: 10.1016/j.heliyon.2017.e00307

A tensegrity model for hydrogen bond networks in proteins

Robert P. Bywater *

Computational Biology Laboratory, Francis Crick Institute, London NW1 1AT, England, UK

* Corresponding author.

E-mail address: robert.bywater@crick.ac.uk (R.P. Bywater).

Abstract

Hydrogen-bonding networks in proteins considered as structural tensile elements are in balance separately from any other stabilising interactions that may be in operation. The hydrogen bond arrangement in the network is reminiscent of tensegrity structures in architecture and sculpture. Tensegrity has been discussed before in cells and tissues and in proteins. In contrast to previous work only hydrogen bonds are studied here. The other interactions within proteins are either much stronger – covalent bonds connecting the atoms in the molecular skeleton or weaker forces like the so-called hydrophobic interactions. It has been demonstrated that the latter operate independently from hydrogen bonds. Each category of interaction must, if the protein is to have a stable structure, balance out. The hypothesis here is that the entire hydrogen bond network is in balance without any compensating contributions from other types of interaction. For sidechain-sidechain, sidechain-backbone and backbone-backbone hydrogen bonds in proteins, tensegrity balance ("closure") is required over the entire length of the polypeptide chain that defines individually folding units in globular proteins ("domains") as well as within the repeating elements in fibrous proteins that consist of extended chain structures. There is no closure to be found in extended structures that do not have repeating elements. This suggests an explanation as to why globular domains, as well as the repeat units in fibrous proteins, have to have a defined number of residues. Apart from networks of sidechain-sidechain hydrogen bonds there are certain key points at which this closure is achieved in the

http://dx.doi.org/10.1016/j.heliyon.2017.e00307

sidechain-backbone hydrogen bonds and these are associated with demarcation points at the start or end of stretches of secondary structure. Together, these three categories of hydrogen bond achieve the closure that is necessary for the stability of globular protein domains as well as repeating elements in fibrous proteins.

Keywords: Computational biology, Structural biology, Biophysics, Molecular physics

1. Introduction

Over a least a half a century, there has been an ongoing debate about the nature of the stabilizing forces that maintain the integrity of the 3D structure of proteins (Dill, 1990; Cooper, 2006; Bywater, 2013a, b; Ben-Naim, 1990; 2011). The two that are said to play the most prominent role are so-called "hydrophobic interactions" (the lipophilic effect) and hydrogen bonding. The latter are particularly important as, in addition to providing a necessary cohesive force, they confer directionality (defined below). It is important to include hydrogen bonds linking atoms within the polypeptide backbone (BB), linking backbone atoms to atoms in the sidechains (BS) and those between sidechains and other sidechains (SS). When all these are taken into account the network of hydrogen bonds can be said to resemble a "tensegrity" structure as found in the architecture of Buckminster Fuller (1961) and the sculptures of Snelson (Heartney, 1971). In the present work an analogy is made between these mechanical structural elements and those that are found in proteins. There are many parallels. The pattern of tensegrity units in proteins are asymmetric as is the case in the Snelson structures while the structures of the Fuller type generally have overall symmetry, although this is by no means a necessary requirement. The issue of greatest importance is balance, or, as referred to herein: closure. There is one further difference that needs to be clarified at the outset: the "tendons" as defined by hydrogen bonding in proteins differ from those in architectural or sculptural constructs is the sense that the latter consist of tensile elements which are internally symmetrical, while hydrogen bonding is by its very nature polarized meaning that the tension in the structural element has a sense in one direction or another. In terms of molecular interactions this distinction needs to be made, but the consequences, mechanically speaking, are independent of whether the pull is due to a polarized mechanism or not (magnets are widely used in engineering to stabilize structures, for example). Lastly, as to the issue of directionality: the underlying mechanism of hydrogen bonding is both electrostatic (a strong dipolar interaction) and quantum mechanical (overlap between n and σ^* or π^* orbitals). These forces restrict bending and torsion around the "bond" which dictate the planar and rotational angles formed by the atoms participating in the bond.

http://dx.doi.org/10.1016/j.heliyon.2017.e00307

When the structures of globular protein were first determined by X-ray crystallography they were considered to be very irregular, not exactly what the experimenters had anticipated. “Could the search for ultimate truth really have revealed so hideous and visceral looking an object?” was a classic reaction (Perutz, 1964). In reality, proteins are highly organized structures built from substructures which are used over and over again in related and even quite different proteins, mirroring the way in which segments of genes are swapped and reused in different coding sequences. An understanding of the order within proteins began to emerge with the steady accumulation of protein crystal structures. This was a result not only of the sterling efforts of the crystallographers themselves, but also of the combined work of many curators and analysts of the corpus of protein crystal structure data which led to the establishment of systematic protein structure databases, in particular CATH (Orengo et al., 1997) and SCOP (Murzin et al., 1995). Both of these databases make clear that proteins are built up of domains, defined as individual folding units which are linked by stretches of relatively unstructured polypeptide chain. A given protein consists of one or a number of domains, these may be identical (arising perhaps by a gene duplication event) or similar (a gene duplication event followed by mutation) or they can be entirely different (gene fusion). In this work the domain is the structural unit being considered. In the context of fibrous proteins which are constructed from repeating sequence fragments one may perhaps refer to these as “repeating elements”, or just “elements” (we shall encounter one such example in the case of a connective-tissue protein). The discussion that follows will focus on globular protein domains but it will be shown that these fibrous protein elements actually have the same properties of closure as defined above.

Globular protein domains are often described in terms of local structures along the polypeptide chain made up of so-called secondary structure elements (SSEs), within which stretches of polypeptide form a regular structure with repeating values of the ϕ, ψ backbone torsion angles. The two principal SSE types are the well-known α-helix and β pleated-sheet, which have been qualitatively described (Brandén and Tooze, 1999) and quantitatively defined (Kabsch and Sander, 1983) in the literature. Recent work (Bywater and Veryazov, 2015) has highlighted the preferences that different residue types have for these and other SSE classes.

There are many possible types of intramolecular interaction in protein domains (Dill, 1990; Cooper, 2006). Apart from covalent crosslinks such as disulfide bonds, these include salt bridges, hydrogen bonding and Van der Waals interactions. These can all contribute toward stabilizing the structure (Dill, 1990; Cooper, 2006). All make some enthalpic contribution (salt bridges are probably strongest in this regard but they are relatively uncommon) and all make entropic contributions (Bywater, 2013a, b), because desolvation is involved in all of these interactions, regardless of the intricacies of the physical mechanism behind the interaction.

http://dx.doi.org/10.1016/j.heliyon.2017.e00307

What ultimately stabilizes the protein is a fine balance between these forces, with the overall free energy for the folded structure being negative of the order of only 5∼10 kcal/mol (Pace et al., 1996).

Of all the possible interactions, hydrogen bonding is deemed to be highly significant (Dill, 1990; Cooper, 2006; Bywater, 2013a, b; Ben-Naim, 1990; 2011) in terms of energy and contribution to stabilization. Unlike salt bridges and "hydrophobic interactions", hydrogen bonding is, as stated above, a result of dipolar interactions and quantum mechanical orbital overlap both of which restrict bending modes. Torsional modes are further restricted due to steric encounters between nearby atoms. These restraints confer strong directionality, a feature which puts hydrogen bonding in a class of its own compared with the other aforementioned interactions. The protein architecture is here considered to consist only of the covalent backbone and sidechain structures and the hydrogen bonds that are formed between suitable donor and acceptor groups. These operate within the backbone (the principle stabilizing interaction for helical structures), between different stretches of backbone (which is the stabilizing factor for β-sheets), those that link sidechain acceptor/donor groups and the backbone (so-called "capping" structures) and those between sidechains which are very important anchoring points for preserving tertiary (3D) structure. While this is a simplified model, it accounts for the dominant proportion of the enthalpy that stabilizes the protein (Dill, 1990; Cooper, 2006; Pace et al., 1996; Bywater, 2013a, b; Ben-Naim, 1990; 2011; Seddon and Bywater, 2015), and almost all of the directionality that is required to maintain the structure within the narrow confines of the stability envelope for the correctly folded structure. This model resembles certain constructs already well known in architecture (Buckminster Fuller, 1961) and sculpture (Heartney, 1971). The word 'tensegrity' was coined by Buckminster Fuller (1961) to describe just such systems consisting of rigid though somewhat compressible rods connected at their ends (or other suitable points) by lengths of wire (or 'tendons') that are under tension. Such structures show considerable structural stability and can resist deformation, the reason why they have found a use in buildings where load-bearing strength combined with light weight is a primary requirement.

With this in mind I set out to study just how the hydrogen bonding 'tendons' in protein structures contribute to the stability of the protein. The method used is very simple, it consists of identifying all of the hydrogen bonds and then, treating them as vectors, performing a vector addition on them. In line with the declared aim to keep the analysis simple, the magnitudes of all hydrogen bond tendons are assigned a value unity. This is of course an entirely arbitrary choice and likewise, no units (e.g. energy related) are assigned. The rationale to this is that if, overall, a cancellation of contra-acting vectors is observed then this choice becomes immaterial. It only becomes of interest in the event that a resultant of significant

http://dx.doi.org/10.1016/j.heliyon.2017.e00307

magnitude is observed and if there is any interest in pursuing this issue. Their relative directions are important in the sense that they determine the ultimate resultant of the vector addition (but as shown later, the important part of the resultant is the final magnitude of the accumulated vector sum, not the "direction"). This is explained by the fact that their positions and orientations in 3D space are defined relative to an origin that is for these purposes arbitrary. Of course, for the X-ray crystallographer, these orientations are clearly not arbitrary but in the treatment here it is important to focus on features of the structures that are invariant. Therefore, it is only the relative positions and orientations that are important, the tendons are regarded as populating an affine 3D space.

In this work, it is shown that hydrogen bonding networks balance each other over the entire structure, while at the same time there is a set of "internal" balances consistent with the need to preserve substructures (SSEs in particular). The notion that there might be an optimal length for any given protein domain is strongly suggested by the data. This question has been asked before (Xu and Nussinov, 1998; Shen et al., 2005) but both groups opted for a "one size fits all" model, however there was no mutual agreement. The predicted optimal number of residues was stated in one case by (Xu and Nussinov, 1998) to be 100 while (Shen et al., 2005) concluded that a domain with a 1:1 ratio of hydrophilic and hydrophobic residues is composed of 156 residues. In contrast to these results, it is shown here that the optimal size varies and it is also explained why it varies.

Apart from showing that the tensegrity model, as proposed here, is valid for a complex (α/β class) globular protein is was further shown to apply to members of virtually all fold classes (within a diverse, nonredundant set comprehensively covering fold space − see columns 2, 3 and 5 in Table 1).

2. Methods

For the purposes of this paper, the vector sum of all the hydrogen bond tendons, as in any system with a balanced set of tendons operating on it is given by:

$$\Sigma_i \ (\Delta w_i)^T$$

where the superscript indicates the transpose and each $\mathbf{w_i}$ (i = 1, ...n) is a vector of order 3 derived from the Cartesian coordinates of the atoms participating in the ith hydrogen bond. Vector addition is associative, distributive and commutative, and is invariant to translation. The $\mathbf{\Delta x_i}$ (i = 1, ...n), where n is the upper limit of the number of hydrogen bonding pairs in the structure, are the displacements between the coordinates of the atoms forming the ends of the tendons (the above mentioned n hydrogen bonds). Although bond lengths for hydrogen bonds do vary somewhat, they are fairly uniform in proteins, populating a range from 2.5–3.0 Å peaking around 2.8 Å. The subtraction was carried out in the sense: $\mathbf{w}_{\text{acceptor\_atom}} - \mathbf{w}_{\text{donor\_atom}}$ (acceptor

http://dx.doi.org/10.1016/j.heliyon.2017.e00307

http://dx.doi.org/10.1016/j.heliyon.2017.e00307

Table 1. Identity of proteins studied in columns 1–4, cumulated vector sums in column 5, links to Figures in column 6. NMR = NMR structure, MOD = modified or modelled structure.

Protein no	Protein name	5-letter PDB I.d. (residue range)	Resolution Å	Secondary structure content	Vector sum of SS, BS and BB hydrogen bonds (Å)	Figure
1	Histone H3	1kx5e	1.94	49% α/0% β	-2.695	1
2	Haemagglutinin Ha34	1ybia	1.50	2% α/48% β	-0.630	2
3	GroEl chaperone	1xckn (2–136)	2.92	68% α/5% β	-2.872	3a
		1xckn (141–408)		41% α/23% β/2% 3_{10}	0.000	3b
		1xckn (412–515)		57% α/8% β	-3.281	3c
4	Insulin receptor tyrosine kinase	2auha	3.2	34% α/16% β/5% 3_{10}	-0.870	4
		2b4 sb	2.0	34% α/18% β/5% 3_{10}	-0.998	
5	T7 RNA Polymerase	1bpx (1–139)	2.4	68% α/9% β/2% 3_{10}	-2.836	5a
		1bpx (140–326)		30% α/29% β/2% 3_{10}	0.452	5b
6	Glutamate receptor subunit 2	1fw0a	1.9	34% α-helix/18% β-strand	0.000	6
7	Pokeweed lectin C	1ulkb	1.8	10% α/10% β/12% 3_{10}	-2.708	7
8	Metabotrope glutamate receptor	1ewka (1–89)	2.2	28% α/17% β/3% 3_{10}	-2.807	8a
		1ewka (90–448)		34% α/18% β/3% 3_{10}	-2.688	8b
9	THP type 1 alpha 1 collagen fragment	2llp	NMR	collagen triple helix	-2.794	9a 9b
10	Bovine beta-lactogobulin	1beb	1.8	10% α/42% β	2.303	10a 10b
	1beb with Crystal waters removed and replaced with simulation waters then minimized		MOD		-0.632	10c
	1beb after MD dehydration experiment (see text)		MOD		-6.892	10d
11	80 residue random sequence α-helix		MOD	all α	-5.464	11
12	80 residue random sequence β-strand		MOD	all β	-8.479	12
13	80 residue random sequence with random secondary structure		MOD	Arbitrary sequence of α,β,π and 3–10	-16.626	13a
	Same as the above but in compact "quasi-fold" (see text)		MOD	Compact form, prominent 3–10	-2.631	13b
14	Prion protein in misfolded amyloid state (structure 31 from NMR ensemble)	2rnm	NMR	all β	2.903	14

atom usually oxygen, donor atom usually a nitrogen bearing a proton). Following widespread practice (Krieger et al., 2002; Markovitch and Agmon, 2007; Bikadi et al., 2007) a cutoff of 3.5 Å was applied, rather more generous than the standard 2.8 Å, to allow for weaker or strained interactions but eliminating the inclusion of interactions with very long bond lengths which would corrupt the dataset by giving such very weak interactions unwarranted prominence. A fortran program was written (available from the author on request) which calculates the accumulated vector sum of all the hydrogen bond tendons. This sum is expressed as a sort of a financial audit expressing the overall displacement resulting from the cumulative summation of the vectors (tendons). Since this is a scalar quantity, it may be considered necessary to specify the direction of the resultant also. In practice this is an uninteresting quantity since, to begin with, we are dealing with an affine space and the origin is not defined, and furthermore, the accumulated summation is always close to zero (please refer to column 6 in Table 1) (although this is not true for denatured or "unnatural" proteins where the closure referred to here is not achieved, see Table 1 again). This not only eliminates any need to discuss the significance of this question, but rather, it amounts to a *quod erat demonstrandum* for this entire work.

Calculations were conducted on a diverse, nonredundant set of proteins which comprehensively cover protein domain fold space (see Table 1 for details and results). The numbering of these proteins matches Figs. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 and is used for reference purposes in the Results and Discussion below.

Fig. 1 Histone H3.

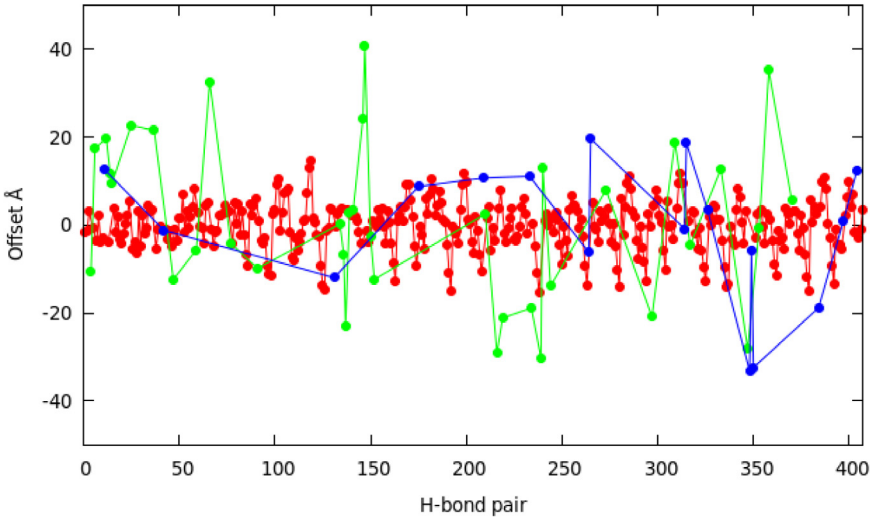

Fig. 1. Cumulative vector sums BB (red color), BS (green) and SS (blue) (BS and SS enhanced fivefold for clarity) for histone H3.

http://dx.doi.org/10.1016/j.heliyon.2017.e00307

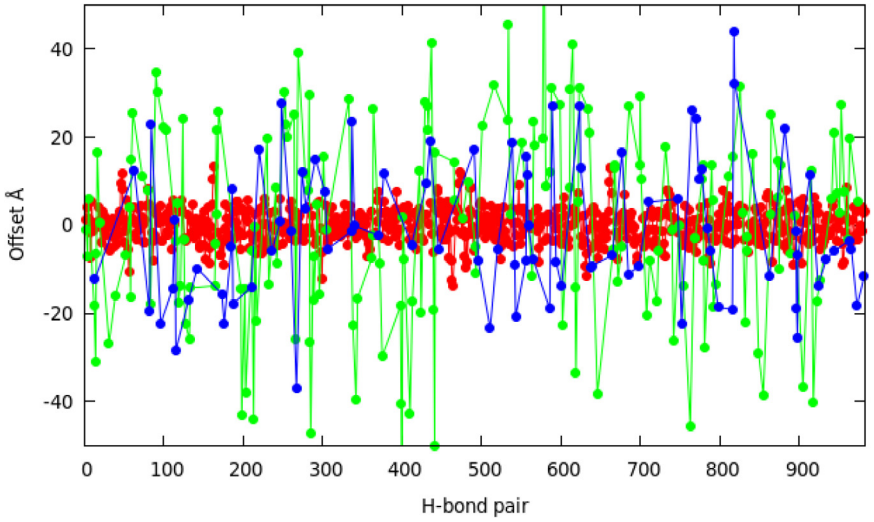

Fig. 2. Cumulative vector sums BB (red color), BS (green) and SS (blue) (BS and SS enhanced fivefold for clarity) for haemagglutinin Ha34.

Fig. 2 Haemagglutinin Ha34.

Fig. 3 GroEl chaperone (three separate domains).

Fig. 4 Insulin receptor tyrosine kinase.

Fig. 5 T7 RNA Polymerase (two separate domains).

Fig. 6 Glutamate receptor subunit 2.

Fig. 7 Pokeweed lectin C.

Fig. 8 Metabotrope glutamate receptor (two separate domains).

Fig. 9b THP type 1 alpha 1 collagen fragment.

Fig. 10b Bovine beta-lactogobulin crystal structure with crystal waters.

Fig. 10c Crystal waters replaced with simulation waters.

Fig. 10d The same after minimization followed by a MD dehydration experiment (see Seddon and Bywater (2015)).

Fig. 11 α-helix with random sequence 80 residues in length.

Fig. 12 β-strand with random sequence 80 residues in length.

In addition, as a control, a "randomised" version of constructed proteins 11 and 12 were made, numbered 13 in Table 1. This consisted of a protein with the same amino acid sequence as 11 and 12 but with arbitrary secondary structure assignments along its length. Two variants were made, an extended chain where

http://dx.doi.org/10.1016/j.heliyon.2017.e00307

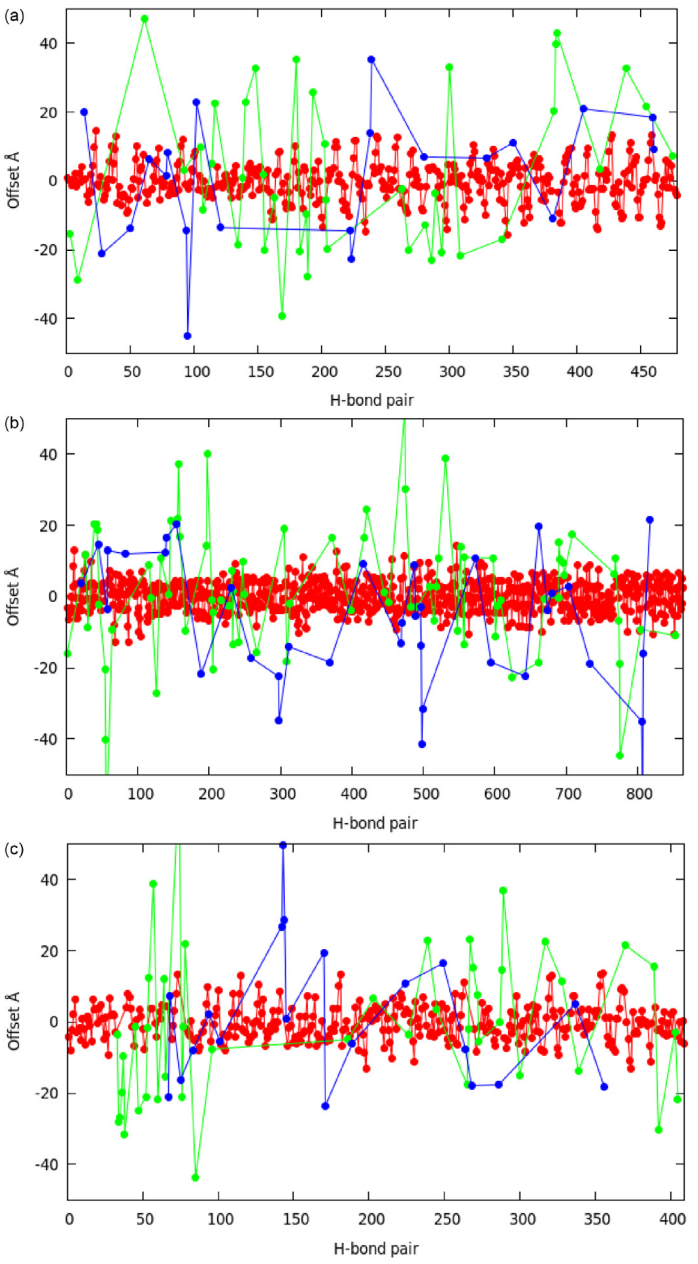

Fig. 3. (a) Cumulative vector sums BB (red color), BS (green) and SS (blue) (BS and SS enhanced fivefold for clarity) for GroEl chaperone domain I. (b) Cumulative vector sums BB (red color), BS (green) and SS (blue) (BS and SS enhanced fivefold for clarity) for GroEl chaperone domain II. (c) Cumulative vector sums BB (red color), BS (green) and SS (blue) (BS and SS enhanced fivefold for clarity) for GroEl chaperone domain III.

the structure was dominated by the (arbitrarily chosen) secondary structure and a second version where this construct was subjected to a "quasi-folding" process involving energy minimisation under a set of three (again arbitrarily chosen) constraints under the AMBER 11 force field as implemented in the YASARA

http://dx.doi.org/10.1016/j.heliyon.2017.e00307

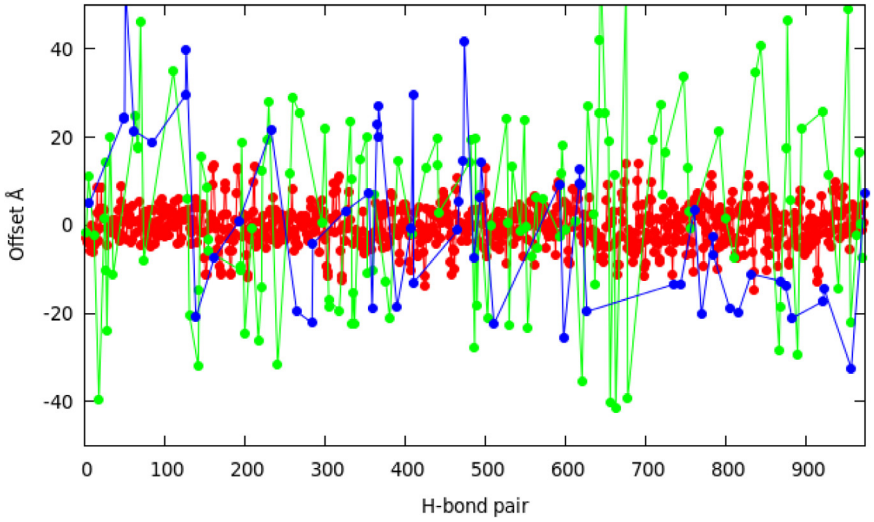

Fig. 4. Cumulative vector sums BB (red color), BS (green) and SS (blue) (BS and SS enhanced fivefold for clarity) for Insulin receptor tyrosine kinase.

program (Krieger et al., 2002) (see Fig. 13a–b). Note that the cumulative vector sums for "proteins" 13a and 13b are not shown. They are obtainable from the author. For these "proteins" it was deemd more useful to show backbone plots as in Fig. 13a–b (see legends to Figs. 9a, 10a, 13a–b). Finally, the the HET-S(218–289) prion in its amyloid form was studied (see Discussion below and entry 14 in Table 1 and Fig. 14)

The YASARA program was used to make Figs. 9a, 10a, 13a–b and 14 .

3. Results and discussion

The results of calculations of the hydrogen bond vectors (tendons) for the protein domains studied here are shown as extended plots of the cumulative hydrogen bond summation (Figs. 1–8, 9b, 10b–d and Figs. 11 and 12). In these figures the summation is plotted separately for the BB (red colour), BS (green) and SS (blue) sets and *nota bene* that both BS and SS are enhanced fivefold along the ordinate (which measures displacement) for clarity. Note further that the abscissa is not the residue number, since there will normally be more than one hydrogen bond emanating from each residue. The abscissa is the sequential number of the first participant in each hydrogen bond and is not related to atom number or residue number in any simple way. The cumulative vector sums for all three sets are listed in column 6 in Table 1.

For most of the protein domains in the series 1 to 8 (Figs. 1 –8) the accumulated vector sum is close to zero and certainly less than 2.8 Å which is equivalent of the length of a single hydrogen bond. This demonstrates that "closure" as defined above is achieved for these proteins.

http://dx.doi.org/10.1016/j.heliyon.2017.e00307

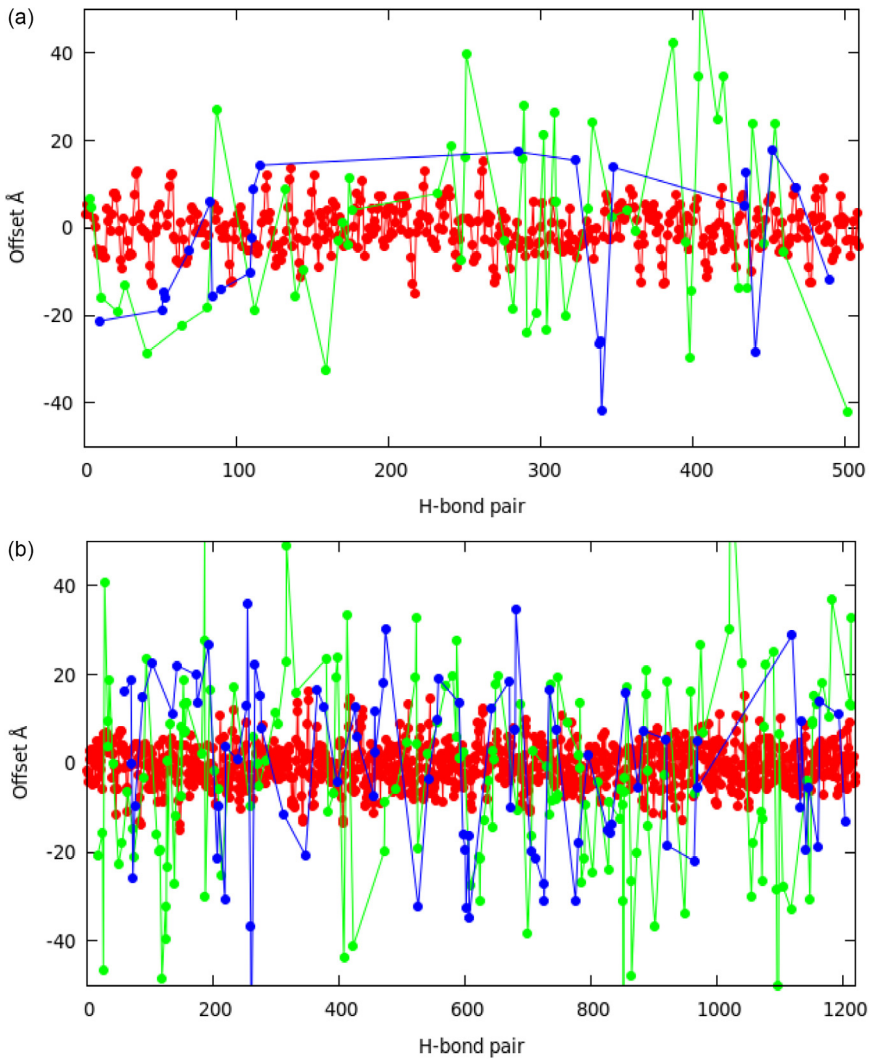

Fig. 5. (a) Cumulative vector sums BB (red color), BS (green) and SS (blue) (BS and SS enhanced fivefold for clarity) for T7 RNA Polymerase domain I. (b) Cumulative vector sums BB (red color), BS (green) and SS (blue) (BS and SS enhanced fivefold for clarity) for T7 RNA Polymerase domain II.

In protein science, as in all science, it is important not only to verify results with additional examples (here the author has taken every opportunity to cover fold space exhaustively – see columns 2, 3 and 5 in Table 1) but also to try to disprove any conjectures that are presented. In this spirit, some further studies were embarked upon, with the intention of examining the following questions:

1. The discussion so far has concentrated on globular protein domains. What about fibrous proteins? It might be supposed that fibrous proteins, being extended structures, might not exhibit the closure that is observed in globular domains, which by their very nature have an approximately (topologically speaking) spherical structure. This question is answered by the results presented in

http://dx.doi.org/10.1016/j.heliyon.2017.e00307

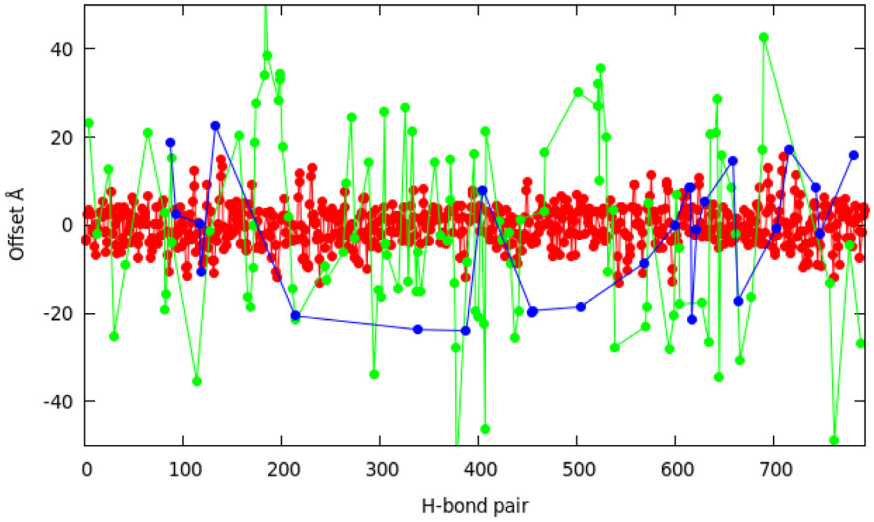

Fig. 6. Cumulative vector sums BB (red color), BS (green) and SS (blue) (BS and SS enhanced fivefold for clarity) for glutamate receptor subunit 2.

Table 1, protein number 9 (collagen triple helix). The fragment studied here (Fig. 9a) is from a crystal structure (2llb) and represents a repeating unit in the collagen triple helix structure. Its role as a repeating unit is signified by flanking regions which are proline rich (coloured orange in Fig. 9). As such, it has many of the attributes of a closed structure, and its "closure" is signified by a vector sum of −2.794, just under the "single hydrogen bond" threshold. Given that most fibrous proteins are constructed in a similar way, from repeating units, there is no reason to suppose that the "closure" model does not apply to this

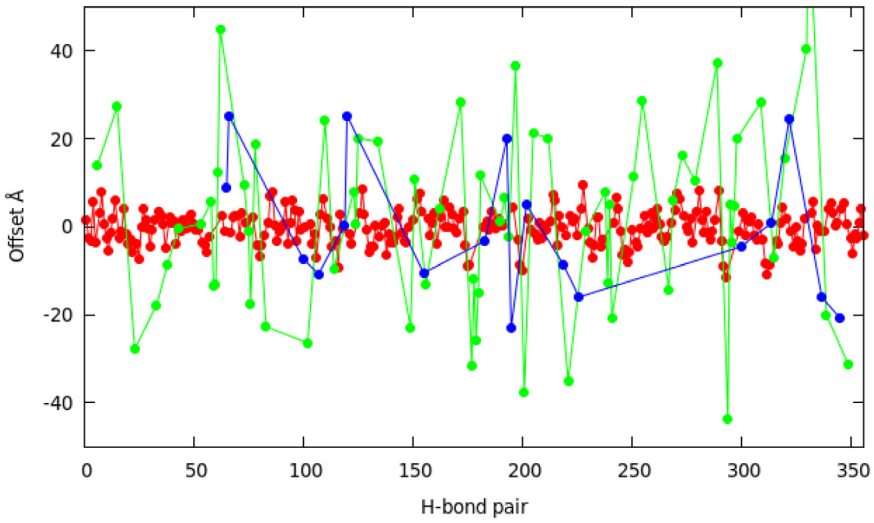

Fig. 7. Cumulative vector sums BB (red color), BS (green) and SS (blue) (BS and SS enhanced fivefold for clarity) for pokeweed lectin C.

http://dx.doi.org/10.1016/j.heliyon.2017.e00307

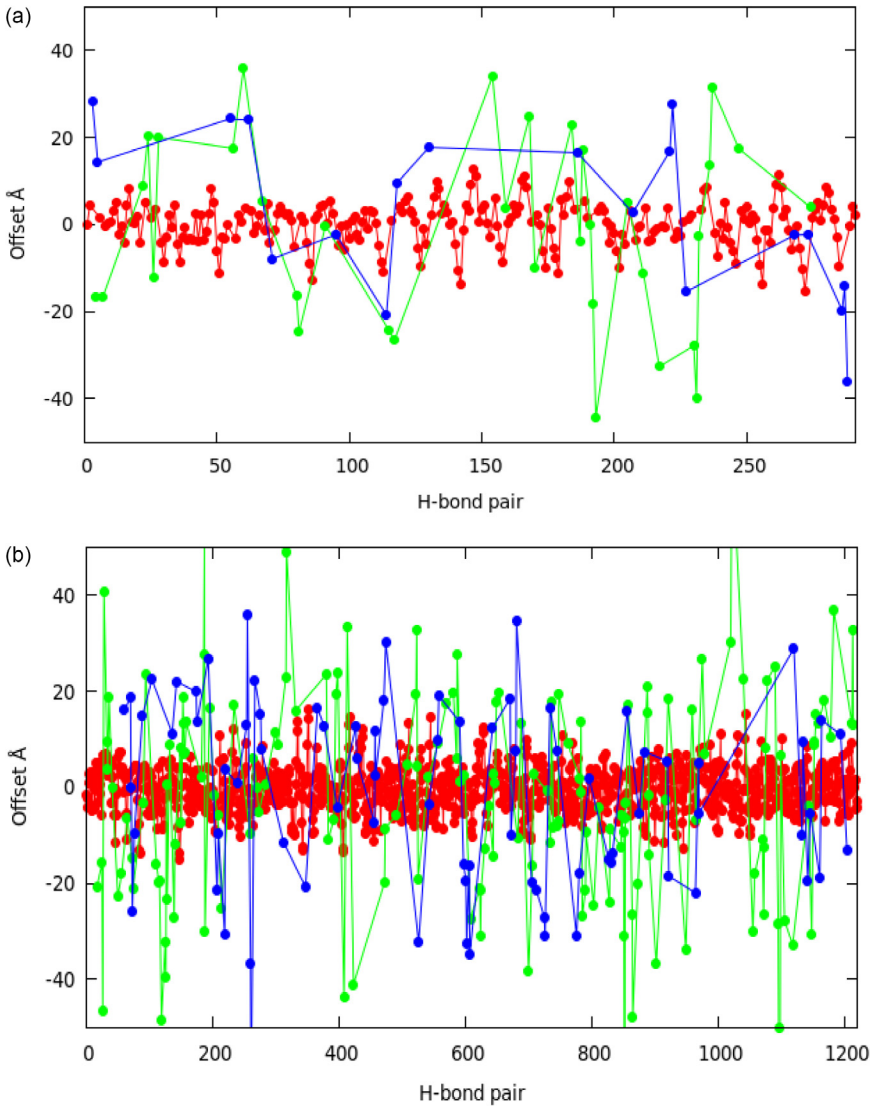

Fig. 8. (a) Cumulative vector sums BB (red color), BS (green) and SS (blue) (BS and SS enhanced fivefold for clarity) for metabotrope glutamate receptor domain I. (b) Cumulative vector sums BB (red color), BS (green) and SS (blue) (BS and SS enhanced fivefold for clarity) for metabotrope glutamate receptor domain II.

family of proteins as well in the sense that the repeating units as an analogy to the domains in globular proteins.

2. Most if not all proteins can switch from one state to another during activation or in response to interaction with ligands or environmental changes (Bywater, 2013a, b). Does this generally show up in the form of changes to this closure model? This was checked for all of the proteins 1 to 8 in Table 1, as well as some others not listed. While there were rearrangements in domain structures and in particular to the way domains pack onto each other, there were no

http://dx.doi.org/10.1016/j.heliyon.2017.e00307

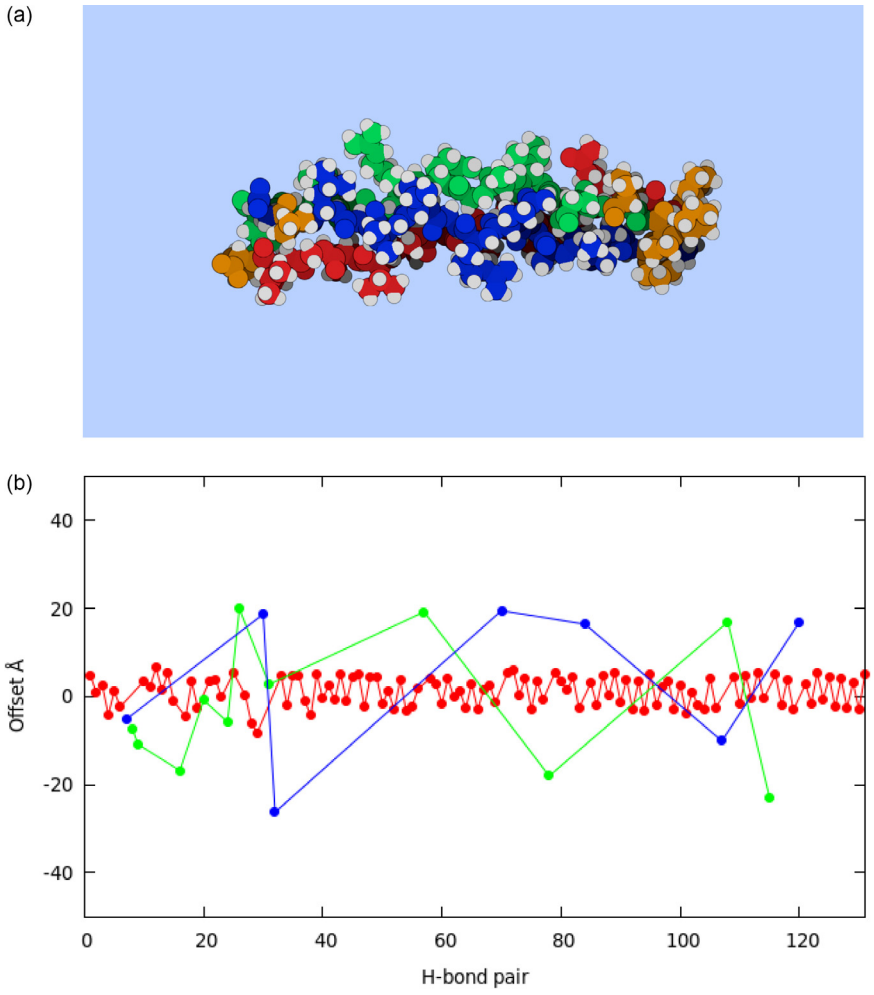

Fig. 9. (a) All atom model of THP type 1 alpha 1 collagen fragment. (b) Cumulative vector sums BB (red color), BS (green) and SS (blue) (BS and SS enhanced fivefold for clarity)) for THP type 1 alpha 1 collagen fragment.

changes in the closure scores. As an example the insulin receptor closure scores show an insignificant change from −0.870 (2auha) to −0.998 (2b4 sb) (Table 1).

3. What happens when proteins start to become denatured by increased temperature or the presence of chaotropes? This question is answered in Table 1, protein number 10 and Fig. 10b–d. This example is taken from earlier published work (Seddon and Bywater, 2015) that clearly demonstrates what happens when globular proteins become dehydrated. Dehydration will inevitably influence hydrogen bonding networks and overall structural integrity (Fig. 10a). What is important in the present context is, what, if anything, happens to the "closure" model? The answers to this question (Table 1 crystal structure: 2.303, solution structure (minimized): −0.632, dehydrated structure: −6.892) not only add credence to the closure model but illustrate an important

http://dx.doi.org/10.1016/j.heliyon.2017.e00307

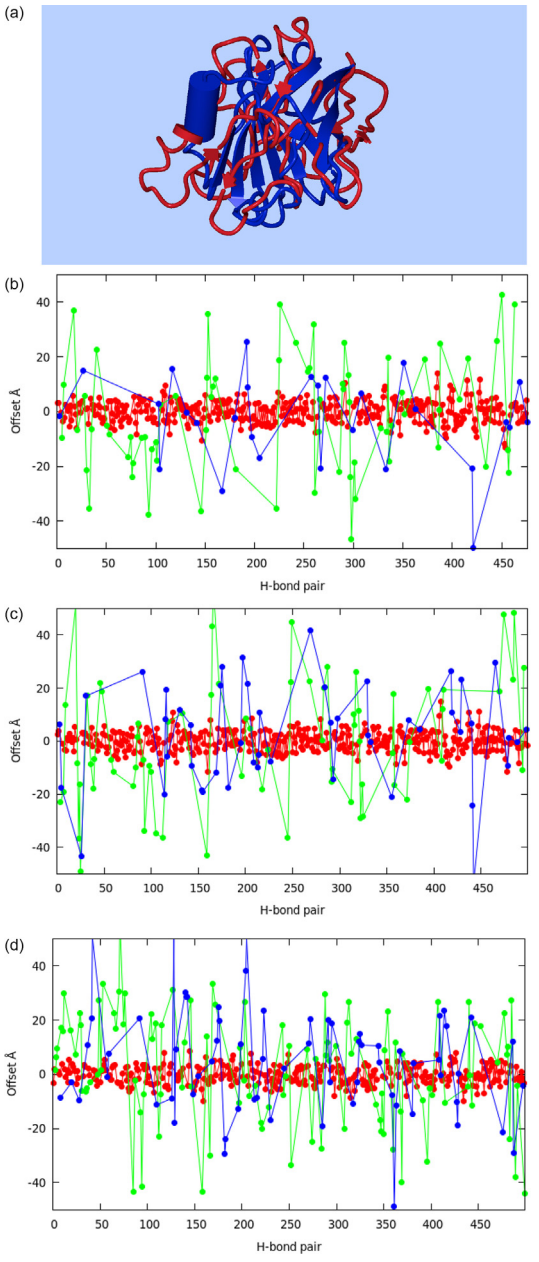

Fig. 10. (a) Backbone image of Bovine beta-lactogobulin crystal structure. (b) Cumulative vector sums BB (red color), BS (green) and SS (blue) (BS and SS enhanced fivefold for clarity) for bovine beta-lactogobulin crystal structure with crystal waters. (c) Cumulative vector sums BB (red color), BS (green) and SS (blue) (BS and SS enhanced fivefold for clarity) for the same protein as Fig. 10b with crystal waters replaced with simulation waters. (d) Cumulative vector sums BB (red color), BS (green) and SS (blue) (BS and SS enhanced fivefold for clarity) for the same protein as Fig. 10b after minimization followed by a MD dehydration experiment (see Seddon and Bywater (2015)).

http://dx.doi.org/10.1016/j.heliyon.2017.e00307

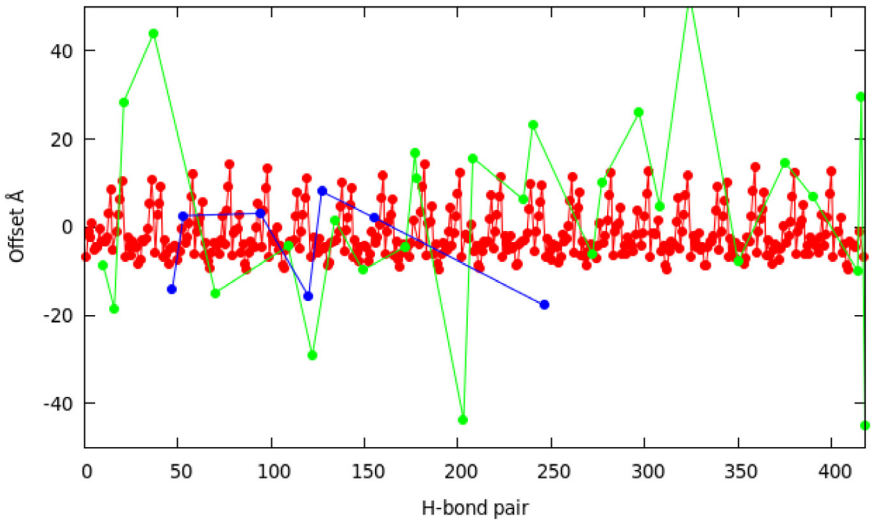

Fig. 11. Cumulative vector sums BB (red color), BS (green) and SS (blue) (BS and SS enhanced fivefold for clarity) for α-helix with random sequence 80 residues in length.

but often overlooked feature of protein structure is the fact that crystal structures contain packing constraints that cause departures from the biologically relevant structure for the protein (Peters and Bywater, 2002). This explains the "improvement" in the closure score from 2.303 to −0.632 typically observed as crystal packing deformations are relaxed and the protein assumes its conformation in an aqueous environment.

4. Fibrous proteins have been considered above, but these consist of repeating units. What happens in (effectively unlimited) extended peptides?. Fictitious

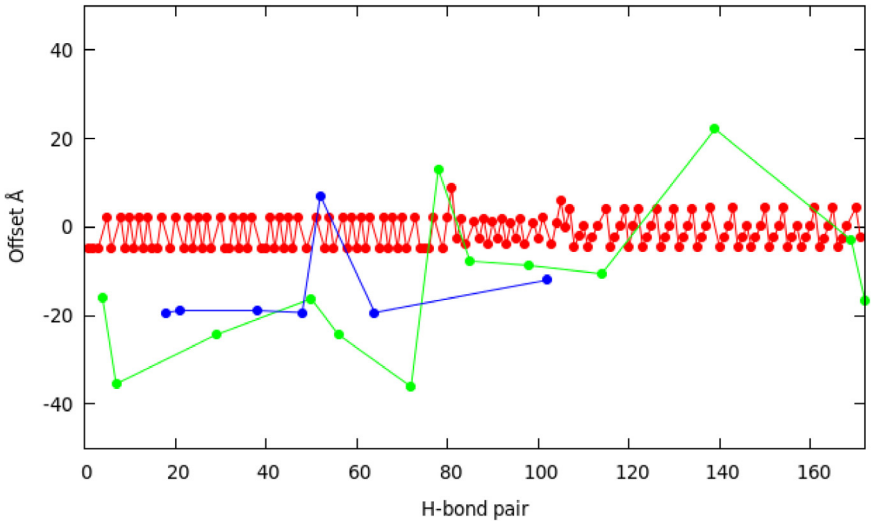

Fig. 12. Cumulative vector sums BB (red color), BS (green) and SS (blue) (BS and SS enhanced fivefold for clarity) for β-strand with random sequence 80 residues in length.

http://dx.doi.org/10.1016/j.heliyon.2017.e00307

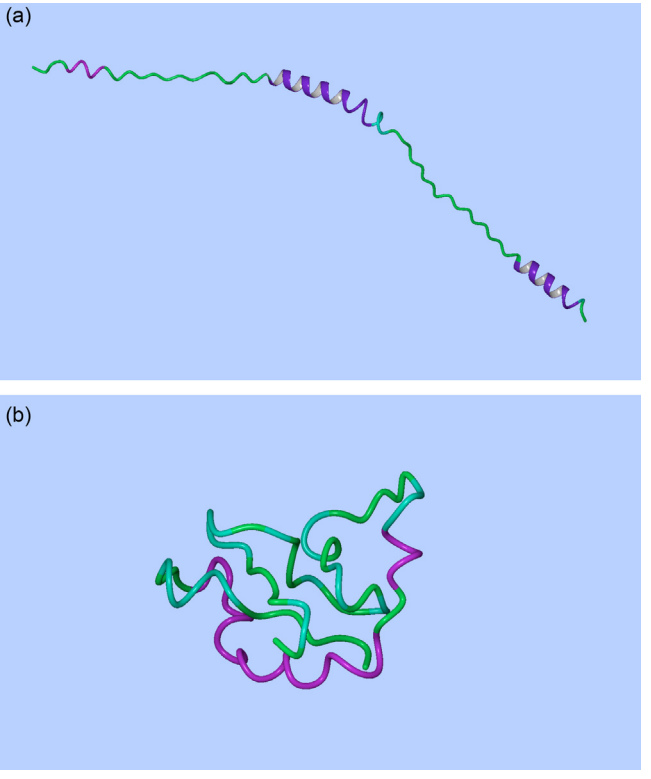

Fig. 13. (a) Backbone plot of 80 residue peptide with mixed (arbitrary arrangement) α-helix (blue), β-strand (green) and 3–10 helix (purple). (b) Backbone plot of 13.1 after "quasi folding" step to give a compact structure. Note most α-helix and β-strand has been lost but 3–10 helix is abundant.

proteins were constructed to answer these questions, an extended alpha-helix with an arbitrary sequence ADSLMFAQQWIMTACMQAICKQNWTMCQK RWSGARHPSYIFYRSQCPVGKSLFGIELVWCWMMHYLGCWCMI GWMLWCSS, protein number 11 and an extended beta strand with the same sequence (protein no. 12), results shown in Table 1 and Figs. 11 and 12 respectively. The vector sums are −5.464 and −8.479 respectively. This shows

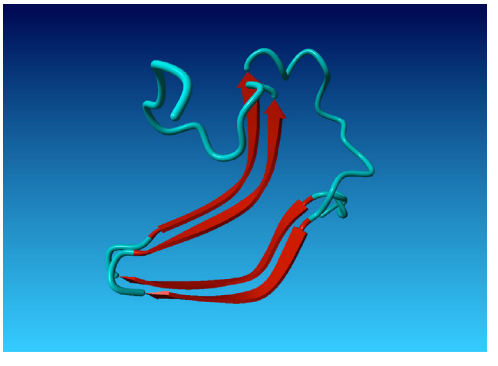

Fig. 14. Cartoon image of prion protein in misfolded amyloid state (structure 31 from NMR ensemble).

http://dx.doi.org/10.1016/j.heliyon.2017.e00307

that closure is not being observed. Different sequences were made by taking a different random number seed, the results were very similar. It is pertinent to point out that proteins 11 and 12, while being fictitious constructs can be thought of as too closely resembling certain kinds of "all α" or "all β" proteins. To this end a more robust control construct was made whereby the α/β character was more "spread out" along the chain In an arbitrary manner. This construct exhibited a major departure from the closure score: −16.626. In order to complete this control the said construct was subjected to the above-mentioned quasi-folding' procedure which produces a well-packed globular structure. The result of this procedure is a closure score of −2.631. As a control and as a *Gedankenexperiment* this surely affirms the original postulate concerning closure. (The coordinates of these constructs as well as the resulting hydrogen bonding network is available from the author on request).

5. If instead of extending the protein, the sequence is truncated what happens then? Proteins in the selected set were subjected to two kinds of "engineering": truncation of the C-terminus and elongation with a 10 residue helical tail of asparagine residues (arbitrary choice but one which permits hydrogen bonding). Results show that the closure scores are immediately compromised. Examples of this are (from Table 1) protein 6 which goes from score 0.000 (full-length) to −3.321 at 5% truncation and protein 7 which goes from −2.708 (full-length) to −5.167. Other proteins behave similarly. Addition of residues at the C-terminus is problematic because there is no *a priori* way to know what their conformation would be, but in general the consequences are always deleterious. The minimum value is found to correspond to the wild type full length (of the domain in question, not the protein in which it is situated). This immediately suggests that closure is required for the chain to be stable, and it suggests a reason why protein domains have to be of a specified length. There are of course functional reasons for this too, in terms of the various types of information that have to be "packed into" the sequence (Bywater, 2015), but the length that is defined by this closure represents an upper limit, then.

6. There are other features of protein folding that need to be considered apart from denaturaiton and unfolding. Misfolding is a not uncommon phenomenon and is almost always pathological problem *in vivo*. While denaturation results in the disruption of the balanced hydrogen bond network (entries 10.4 and 11–13 in Table 1) the amyloid state (Toyama and Weissman, 2011) is a new folded state – misfolded but not unfolded. To invesitgate this, an entry from the PDB was studied, a representative member of an ensemble of NMR structures of the HET-S(218–289) prion in its amyloid form. It would be a mistake to regard these proteins as unfolded and indeed the closure "score" (entry 14 in Table 1 with structure shown in Fig. 14) for this protein is 2.903, similar to many of the compact globular proteins in the list of those studied here.

 http://dx.doi.org/10.1016/j.heliyon.2017.e00307

4. Conclusions

Three main conclusions emerge from this work:

1. The proposed tensegrity model is upheld. Hydrogen-bond tendons operate throughout the entire structure but in a way that their net resultant is balanced out ("closure"). The seminal work in architecture and engineering by Buckminster Fuller (1961) and in sculpture by Snelson (Heartney, 1971) implies a perfect balance between the tensions and compressions in a closed, stable structure composed of rods and springs (tendons). The tensegrity model has earlier been discussed at the cellular level and has in the present work also been shown to anticipate findings within the micro-architectural world of biological macromolecules. It should be mentioned that the Buckminster Fuller structures are usually highly symmetrical (although they do not need to be as long as balance is maintained). The Snelson sculptures are asymmetrical, as indeed are proteins as well as almost all biomolecules. Symmetry is not a necessary requirement for a balanced structure.
2. Hydrogen bonding patterns explain why proteins have to be of a certain defined length. Firstly, there is the need to pack in all the information required for performing the various functions of a given protein (Bywater, 2015). This requires a certain "length". The full-length sequence represents an upper limit to this length, the only proviso is that an optimal hydrogen-bond network is put in place in order to maintain the integrity of the structure that has to cater for all these functions.
3. Other intraprotein interactions such as hydrophobic interactions (Dill, 1990) and interactions involving aromatic residues (Cockroft and Hunter, 2007) are important too, obviously, and they also "balance" mutually within their own particular functional and mechanistic group. But they act independently. The suggestion that e.g. a "hydrophobic interaction" may be used in exchange for a hydrogen bond (as in some of the correlated mutation literature) is not borne out by the data shown here.
4. Denaturation results in the disruption of the balanced hydrogen bond network (entries 10.4 and 11–13 in Table 1) but the amyloid state is a new folded state – misfolded but not unfolded (entry 14 in Table 1). It is folded structures that exhibit the balnced hydrogen bond network, unfolding is characterised by an imbalance in the hydrogen bond network.

Declarations

Author contribution statement

Robert Bywater: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

http://dx.doi.org/10.1016/j.heliyon.2017.e00307

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interest statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

Acknowledgements

The paper has been improved in response to comments received from the journal editor and two anonymous reviewers, and from Professor G Zanotti whose work is cited herein (Zanotti and Guerra, 2003).

References

Ben-Naim, A., 1990. On the role of hydrogen-bonds in protein folding and protein association. J. Phys. Chem. 95, 1437–1444.

Ben-Naim, A., 2011. The rise and fall of the hydrophobic effect in protein folding and protein-protein association and molecular recognition. Open J. Biophys. 1, 1–7.

Bikadi, Z., Demko, L., Hazai, E., 2007. Functional and structural characterization of a protein based on analysis of its hydrogen bonding network by hydrogen bonding plot. Arch. Biochem. Biophys. 461 (2), 225–234.

Brandén, C., Tooze, J., 1999. Introduction to Protein Structure, second ed. Garland Publishing, New York.

Buckminster Fuller, R., 1961. Tensegrity. Portfolio and Art News Annual, No. 4. http://www.rwgrayprojects.com/rbfnotes/fpapers/tensegrity/tenseg01.html.

Bywater, R.P., 2013a. Protein folding: a problem with multiple solutions. J. Biomol. Struct. Dyn. 31 (4), 351–362.

Bywater, R.P., 2013b. Comments on the paper "Levinthal's question revisited, and answered" by A. Ben-Naim. J. Biomol. Struct. Dyn. 31, 967–969.

Bywater, R.P., 2015. Prediction of protein structural features from sequence data based on Shannon entropy and Kolmogorov complexity. PLoS ONE.

 http://dx.doi.org/10.1016/j.heliyon.2017.e00307

Bywater, R.P., Veryazov, V., 2015. The dipeptide conformations of all twenty amino acid types in the context of biosynthesis. SpringerPlus 4, 668.

Cockroft, S.L., Hunter, C.A., 2007. Chemical double-mutant cycles: dissecting non-covalent interactions. Chem. Soc. Rev. 36, 172–188.

Cooper, A., 2006. Heat capacity of hydrogen bonded networks: an alternative view of protein folding thermodynamics. Biophys. Chem. 85, 25–39.

Dill, K.A., 1990. Dominant forces in protein folding. Biochemistry 29, 7133–7155.

Heartney, E., 1971. Kenneth Snelson: Art and Ideas. http://www.scribd.com/doc/146497147/Art-and-Ideas-Kenneth-Snelson.

Kabsch, W., Sander, C., 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22, 2577–2637.

Krieger, E., Koraimann, G., Vriend, G., 2002. Increasing the precision of comparative models with YASARA NOVA–a self-parameterizing force field. Proteins 47, 393–402.

Markovitch, O., Agmon, N., 2007. Structure and energetics of the hydronium hydration shells. J. Phys. Chem. A 111 (12), 2253–2256.

Murzin, A.G., Brenner, S.E., Hubbard, T., Chothia, C., 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. 247, 536–540.

Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B., Thornton, J. M., 1997. CATH–a hierarchic classification of protein domain structures. Structure 5, 1093–1108.

Pace, C.N., Shirley, B.A., McNutt, M., Gajiwala, K., 1996. Forces contributing to the conformational stability of proteins. FASEB J. 10, 75–83.

Perutz, M., 1964. The Hemoglobin Molecule. Scientific American 211, 65–76.

Peters, G., Bywater, R.P., 2002. Essential motions in a fungal lipase with bound substrate, covalently attached inhibitor and product. J. Mol. Recognit. 15, 393–404.

Seddon, G.M., Bywater, R.P., 2015. The fate of proteins in outer space. Int. J. Astrobiology 16, 19–27.

Shen, M.Y., Davis, F.P., Sali, A., 2005. The optimal size of a globular protein domain: a simple sphere-packing model. Chem. Phys. Lett. 405, 224–228.

Toyama, B.H., Weissman, J.S., 2011. Amyloid Structure: Conformational Diversity and Consequences. Annu. Rev. Biochem. 80, 557–585.

 http://dx.doi.org/10.1016/j.heliyon.2017.e00307

Xu, D., Nussinov, R., 1998. Favorable domain size in proteins. Fold. Des. 3, 11–17.

Zanotti, G., Guerra, C., 2003. Is tensegrity a unifying concept for protein folds? FEBS Lett. 534, 7–10.

 http://dx.doi.org/10.1016/j.heliyon.2017.e00307