Article

On the emergence of orientational order in folded proteins with implications for allostery

Debayan Chakraborty ¹ , Mauro Lorenzo Mugnai ¹ and D. Thirumalai ¹

¹ Department of Chemistry, The University of Texas at Austin, USA

Abstract: The beautiful structures of single and multi-domain proteins are clearly ordered in some fashion but cannot be readily classified using group theory methods that are successfully used to describe periodic crystals. For this reason, protein structures are considered to be aperiodic, and may have evolved this way for functional purposes, especially in instances that require a combination of softness and rigidity within the same molecule. By analyzing the solved protein structures, we show that orientational symmetry is broken in the aperiodic arrangement of the secondary structural elements (SSEs), which we deduce by calculating the nematic order parameter, P_2 . We find that the folded structures are nematic droplets with a broad distribution of P_2 . We argue that non-zero values of P_2 , leads to an arrangement of the SSEs that can resist external stresses forces, which is a requirement for allosteric proteins. Such proteins, which resist mechanical forces in some regions while being flexible in others, transmit signals from one region of the protein to another (action at a distance) in response to binding of ligands (oxygen, ATP or other small molecules).

Keywords: secondary structure elements, nematic order parameter, allostery

1. Introduction

In his prescient and magnificent monograph, *What is Life?*, Schrödinger [1] states that, and we quote "....the most essential part of a living cell - the chromosome fiber - may suitably be called an *aperiodic crystal*". He goes on to add that the difference between the well-known periodic crystal and an aperiodic crystal is "....the same kind as that between an ordinary wallpaper in which the same pattern is repeated again and again in regular periodicity and a masterpiece of embroidery, say Raphael tapestry, which shows no dull repetition, but an elaborate, coherent, meaningful design traced by the great master", and concludes that aperiodic crystal is the "material carrier of life". Of course, similar comments are applicable to proteins, the workhorses of the cellular machinery that carry out diverse range of functional roles, by the complex organization of their three-dimensional structures. Because proteins can be crystallized, they diffract X-rays but their arrangement cannot be classified using a unit cell, which is used for describing the pattern repeated in the entire crystals. More importantly, a single folded globular protein (or a complex) does have some kind of order (certainly discernible to the eye) but is not periodic, and hence qualifies as an aperiodic crystal. The presence of well-defined scaffolds within the protein architecture, does not imply structural rigidity, and it is reasonable to suggest that nearly all proteins can access evolutionarily tuned deformation modes [2–4] in order to carry out their functions. In addition, it has been shown that coexistence of symmetrically arranged secondary structural elements (SSEs) with flexible regions within a single protein molecule could also enhance the multiplicity of folding pathways [5]. Thus, the aperiodic folds of proteins could have multiple functional consequences. Understanding the general rules underlying the packing of SSEs [6–8] is also important from the perspective of structure prediction.

In a pioneering study [6], Levitt and Chothia reached far reaching conclusions by analyzing the organization of the SSEs in 31 protein structures. They suggested that

Citation: Chakraborty, D.; Mugnai, M.; Thirumalai, D. On the emergence of orientational order in folded proteins with implications for allostery. *Symmetry* **2021**, *1*, 0. <https://doi.org/>

Received:

Accepted:

Published:

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Submitted to *Symmetry* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

β -strands that are adjacent in the sequence are ordered by forming hydrogen bonds or van der Waals contacts. It was also realized that such SSEs could form diffusing folding units, which then consolidate the native structure [6]. These ideas have been exploited in recent times in *de novo* protein design [9–11], rationalization that geometry and symmetry considerations alone restrict the formation of native folds [12], and in exploring the so-called ‘dark matter’ of the protein-universe [13,14]. In a most insightful article, Wolynes [15] recognized explicitly the importance of aperiodic crystalline nature of proteins, and the resulting approximate symmetry. Linking these ideas to the principle that the folded state should be minimally frustrated [16], he suggested a connection between folding dynamics, approximate symmetry, and the underlying energy landscape. Thus, not only structures, but functional protein dynamics is related to the aperiodic nature of the folded state of proteins.

An intriguing function of both single folded proteins or oligomers is allosteric signaling (action at a distance), which involves signal propagation through an ordered protein, usually in response to binding of a ligand [17–20]. An excellent monograph on allostery from a unique perspective has been recently published by Rob Phillips [21]. Allostery requires the ability to impact chemistry occurring at sites beyond the reach of electrostatic forces between atoms – in other words it is an emergent property [22] of the protein structure itself. The best studied example is the structural transitions that occur in Hemoglobin [23,24], when oxygen molecule binds to the heme group. Elsewhere [25], we have argued that allosteric proteins (monomers, multi-domain proteins, or molecular machines) must be aperiodic [1]. This means that these biological molecules must have an approximate symmetry, implying that at least a portion of the complex must be rigid [22] in order to resist mechanical deformations, but not overly so, thus enabling the execution of those chemical and mechanical functions, which rely upon conformational flexibility. Without an approximate symmetry, allosteric transitions would not occur. Thus, there has to be a symmetry breaking transition that must occur when proteins fold from unfolded high symmetry state. The symmetry requirements for allosteric transitions is the key premise of the much celebrated Monod-Wyman-Changeux (MWC) model [26], which explains not only the allosteric mechanism of hemoglobin [24], but also various types of signal transduction [27]. However, not all allosteric transitions satisfy the symmetry requirements imposed by the MWC theory [18]. In this context, the Koshland-Némethy-Filmer (KNF) model [28], has found numerous applications, particularly in examples of ‘induced fit’ allostery [29], and negative cooperativity [30]. An alternative viewpoint of protein allostery, inspired by the model of Cooper and Dryden [31], has recently come to the forefront. In this picture, allosteric transitions can occur due to an alteration in the thermal fluctuations (or dynamics) around the native state as a result of ligand binding, and do not require substantial conformational changes [32–34]. In all these scenarios, structures with approximate symmetry is paramount.

Here, we suggest that there is a breakage of orientational symmetry when the proteins fold. This is manifested in the emergence of nematic order, in some instances only weakly so, which is sufficient for them to behave as allosteric systems. Because allostery involves accessing alternate aperiodic states, which are excitations around the ground state, both the ground and excited states must have similar, but not necessarily identical symmetry. For biological viability, the free energy gap between the ground and the excited state cannot be too large [25], which also implies that the structured regions cannot be very rigid. We show, by analyzing the structures of a number of globular proteins and oligomeric assemblies, that they have non-zero values of the nematic order parameter, P_2 . There are large variations in P_2 . The values of P_2 for allosteric proteins are non-zero but not close to unity. As the magnitude of the order parameter measures the degree of the broken symmetry [22], intermediate values of the order parameter reflect the delicate balance between rigidity and conformational flexibility.

93 2. Methods

94 2.1. Protein Structures

95 **Database of globular proteins:** To build a representative database of globular
96 proteins, we considered only those sequences from the PDB database that meet the
97 following criteria: (1) mutual sequence identity of no more than 25%. (2) at least 40
98 amino acids in length (3) structural resolution of at least 3, Å. The list of such sequences
99 was determined using the PISCES web-server [35]. Any sequence that is known to form
100 a membrane protein has a corresponding entry in either the PDBTM [36], *mpstruc* [37], or
101 the OPM [38] database, and is excluded from our analysis. We do so out of an abundance
102 of caution because unlike globular proteins, membrane proteins are not evolved to be
103 optimally compact, and the membrane itself plays an active role in their folding and
104 assembly.

105 **Subset of allosteric domains:** A non-redundant subset of experimentally deter-
106 mined allosteric domains was obtained from the "Core Set" of the ASBench dataset
107 maintained by Zhang and coworkers [39,40]. The ASBench suite consists of 235 unique
108 allosteric sites, predominantly found in *E. Coli.* and humans. These allosteric sites are
109 distributed among diverse protein families, including transferases, hydrolases, oxidore-
110 ductases, transcription factors, and lyases [39].

111 2.2. Data Analysis

112 **Residue-specific secondary structure:** The residue-wise secondary structure of the
113 globular proteins in the curated database was determined using the DSSP algorithm [41]
114 implemented within the *mdtraj* program [42]. For simplicity, we assigned the secondary
115 structure based on a three-class system, which identifies each residue as being in helix
116 (H), β -sheet (E), or in coil/turn configuration (C). Based on the secondary structure
117 content, we segregate all the proteins in the database into three types: if the overall
118 α -helicity of a protein structure exceeds 20%, and the β -content is less than 10%, the
119 protein is classified as α -rich; if the net β -content is greater than 20%, and the α -helix
120 content is less than 10%, the protein is considered to be β -rich; If both the α -helix and the
121 β -content exceeds 12%, then we classify the protein as $\alpha + \beta$. We note that the cutoffs
122 employed for determining the different protein structural classes are quite robust, and
123 similar statistics were obtained for variations within a reasonable range of these cutoff
124 values. To further distinguish among the different topologies adopted by the three
125 protein classes, we use the CATH classification of proteins [43] as a guide.

126 **Ordering of secondary structure elements (SSE):** To probe the nematic ordering
127 of secondary structure elements (SSE) within the protein fold, we compute the nematic
128 order parameter using the WORDOM package [44,45]. The traceless second-rank tensor
129 Q , is calculated from molecular vectors, $\vec{w}_i$ as [46,47]:

$$Q_{mn} = \frac{1}{N_s} \sum_{i=1}^{N_s} \frac{3}{2} w_{im} w_{in} - \frac{1}{2} \delta_{mn}. \quad (1)$$

130 In Eq. (1), indices m and n denote the x , y and z directions, and $\vec{w}_i$ s are the molecular
131 vectors connecting the C_α atoms located at the termini of secondary structure elements,
132 such as the α -helix, or the β -sheet; N_s denotes the total number of α -helices and β -sheets
133 within the protein fold; these secondary structure elements (SSE) can be part of a single
134 or multiple different chains. Diagonalization of Q yields the nematic director, $\vec{d}$, which is
135 the eigenvector associated with the maximum nonzero eigenvalue [46,48].

From Q (Eq. 1), and the director $\vec{d}$, the nematic order parameter, P_2 is computed
as [45]:

$$P_2 = \vec{d} \cdot Q \cdot \vec{d} \quad (2)$$

136 For a near perfect ordering of SSE, $P_2 \approx 1$, while $P_2 \approx 0$ for isotropic arrangements.

Figure 1. An illustration of the vectors, $\vec{w}_i$ that are used to compute the nematic order parameter, P_2 (see Eq. 2), for a $\alpha + \beta$ protein. The protein is shown in a cartoon representation, with the α -helices depicted as red cylinders, and the extended β -sheets colored green. The irregular coils/turns connecting the different secondary structure elements (SSE) are in cyan. The vectors (for example, $\vec{h}$) connecting the C_α atoms located at the termini of α -helices are depicted as red arrows. Green arrows denote the vectors (for example, $\vec{b}$) connecting the C_α atoms located at the termini of extended β -sheets.

A diagrammatic illustration of the vectors, $\vec{w}_i$, for a $\alpha + \beta$ protein is shown in Figure 1. Here, $\vec{h}$ denotes the $\vec{w}_i$ vectors directed along the axis of the α -helical segment, and $\vec{b}$ is the $\vec{w}_i$ vectors directed along the an extended β -sheet in the protein.

To test the statistical independence of different P_2 distributions, we carried out the Pearson χ^2 test for homogeneity, as implemented within the *scipy* module.

3. Results

Single- and multi-chain proteins have different organization of secondary structural elements (SSE): The distribution of the nematic order parameter, P_2 , for the subset of single and multi-chain globular proteins that meet the structural and sequence-identity criteria (see Methods), are shown in Figure 2(a). The P_2 distribution for the multi-chain proteins peaks at relatively lower values compared to the single-chain proteins, due to the additional orientational degree of freedom dictating the relative arrangement in space of different monomeric units. In the aperiodic states, $P_2 \neq 0$, implying the ordered states are indeed nematic droplets in which orientational symmetry is broken. The relatively higher probability density in the region, $0.4 \leq P_2 \leq 0.7$, for single-chain proteins, further indicates that they adopt more folds/topologies commensurate with intermediate ordering of the SSEs. The multi-chain P_2 distribution, on the other hand, features slightly enhanced probabilities in the tail region, beyond $P_2 \approx 0.8$, suggesting that highly ordered assemblies (with only few packing defects) are more likely to result from oligomerization, rather than self-association of multiple SSE within single-chains.

Additional insights into the different folds/topologies that are adopted by single- and multi-chain proteins can be obtained from the secondary-structure dependent P_2 distributions, shown in Figure 2(b) and Figure 2(c). We classify all proteins in the database into three broad classes: α -rich, β -rich and $\alpha + \beta$ (see Methods section). Among α -rich proteins, the ordering of SSE is maximized within helical bundles and coiled-coils, featuring the helix-loop-helix type topology. A representative example of this fold includes the seven-helix bundle periplasmic sensor domain of the TorS receptors found in *E. Coli*. (Figure 3). For single-chain β -rich proteins, the probability density at $P_2 \geq 0.8$ is lower than the α -rich counterparts, suggesting that the alignment of SSEs relative to one another is somewhat less efficient. The ordering of SSEs within single-chain β -rich proteins is maximized primarily in sandwich domains having the immunoglobulin fold (involved in passive muscle elasticity), although some highly ordered structures also

Figure 2. (a) Distribution of the nematic order parameter, P_2 , for single- and multi-chain globular proteins present in the database constructed using the PISCES web-server. The secondary-structure dependent P_2 distributions in single- and multi-chain proteins are shown in (b) and (c), respectively. As outlined in the Methodology section, the residue-dependent secondary-structures were computed using the DSSP algorithm.

exhibit other topologies, such as jelly rolls, and β -prisms. A few representative examples of such folds are shown in Figure 3.

In contrast to the other two protein classes, single-chain β -proteins seem to form a large number of folds characterized by intermediate ordering of the SSEs (P_2 in between 0.4 and 0.5). We find that these folds largely have the sandwich type architecture, with either the immunoglobulin or the jelly roll type topology. Some proteins also adopt the more exotic β -propeller topology, with the number of blades varying between five and eight (Figure 3). The difference in the P_2 distributions between single-chain and multi-chain proteins can largely be explained by this ability of single-chain β -rich proteins to form such diverse folds, exhibiting intermediate ordering of SSEs. As is evident from Figure 2(b), $\alpha + \beta$ proteins are also less likely to form folds that maximize the ordering of the SSEs ($P_2 \geq 0.8$) as compared to α -rich proteins. Nonetheless, we find few topologies, such as $\alpha\beta$ -plaits, leucine-rich repeats, where SSEs could be maximally aligned (Figure 3). The peak of the P_2 distribution for $\alpha + \beta$ proteins occurs at ≈ 0.2 , consistent with a large number of architectures found in the database, such as the two-layer sandwich, three-layer aba sandwich, $\alpha\beta$ -complex, and the $\alpha\beta$ -barrel. In particular, we find that the Rossman fold (three-layer aba sandwich), $\alpha\beta$ -plaits, and TIM barrels ($\alpha\beta$ -barrel) are particularly favored (Figure 3).

Similar to their single-chain counterparts, multi-chain α -rich proteins also form a larger number of orientationally ordered structures, compared to the other two protein classes, by exploiting the helix-bundle architecture, as in the case of the central domain of bacteriophage phiX174 H protein. Multi-chain β -proteins maximize the ordering of SSEs through a diverse range of folds including aligned prisms, jelly rolls, solenoids, and immunoglobulin-like sandwich architectures. However, unlike their single-chain counterparts, the subset of multi-chain β -proteins also includes highly ordered structures ($P_2 \geq 0.9$), such as amyloid fibrils, and antifreeze proteins exhibiting the 2-solenoid topology (Figure 3). Interestingly, multi-chain β proteins do not adopt as many folds of intermediate order, suggesting that additional symmetry requirements could impose constraints on the formation of such topologies (Figure 2(c)). The P_2 distribution for multi-chain $\alpha + \beta$ proteins is quite similar to that for single-chains. A large number of the structures exhibit intermediate ordering of the SSEs (peak at ≈ 0.2), and belong to

Figure 3. Some representative protein folds found among proteins in the PDB database. (a) seven-helix bundle periplasmic sensor domain of the TorS receptors (PDB ID: 3I9Y). (b) Immunoglobulin-like fold (PDB ID: 2XSK) present in the *E. Coli*. curli protein (c) 6-blade β -propeller (PDB ID: 3A72) found within a high-resolution crystal structure of *Penicillium chrysogenum* alpha-L-arabinase in complex with arabinobiose. (d) $\alpha\beta$ plait (PDB ID: 6CBU) topology found within an immunogen. (e) Rossmann fold (PDB ID: 4LJS) present in the periplasmic binding protein. (f) Central domain of bacteriophage phiX174 H protein formed by the assembly of ten α -helices (PDB ID: 4JPN). (g) 2-solenoid topology found in antifreeze proteins (PDB ID: 4DT5) (h) jelly-roll (PDB ID: 4HFS) found in a protein from *Bacillus subtilis* subsp. *subtilis* str. 168. (i) $\alpha\beta$ -plait (PDB ID: 1S12) within the protein TM1457 from *Thermotoga maritima*. (j) TIM barrel (PDB ID: 3CBW) found in the YdhT protein from *Bacillus subtilis*.

the Rossmann fold, $\alpha\beta$ -plait, or the TIM barrel superfamily. In addition, like single-chains, multi-chain $\alpha + \beta$ proteins also primarily exploit the $\alpha\beta$ -plait topology to maximally align SSEs (Figure 3).

Overall, we find that for most single- and multi-chain proteins found in the PDB database, the arrangement of SSEs exhibit orientational order. It may be argued that nematic order is weak because P_2 values are not too close to unity. Only a few topologies are exploited to form highly symmetrical or ordered structures.

Figure 4. Distribution of the nematic order parameter, P_2 for single- and multi-chain proteins culled using the PISCES web-server, and allosteric domains present in the ASBench database. For allosteric domains, the nematic order parameter, P_2 , has an upper bound of ≈ 0.5 , and the corresponding distributions lack long tails at high P_2 values (resulting from highly ordered assemblies). In case of single-chain proteins, the Pearson χ^2 test for homogeneity yields a p -value of 0.02, indicating that the distributions computed from the ASBENCH and PISCES databases are statistically independent. The same is true for multi-chain proteins, for which a p -value of 0.0001 is obtained.

Allosteric domains display weak nematic order: In most proteins, allostery is modulated through the concerted motions of structural elements, such as strands and helices, orchestrated by flexible loops or hinges. One of the key requirements in allostery is that certain regions of the protein must be stiff enough to resist mechanical stress, and yet the overall architecture should be flexible enough to accommodate subtle changes in conformation [25]. Therefore, understanding the precise ordering of SSEs, as well as fold preferences within allosteric domains becomes crucial.

We find that in contrast to protein structures culled using the PISCES web-server, the nematic order parameter, P_2 for allosteric domains does not exceed ≈ 0.5 (Figure 4). Therefore, the arrangement of SSEs within allosteric domains cannot be overtly stiff, like in ordered assemblies, such as fibers and helix bundles, and must be flexible enough to access excited state in order to accommodate structural perturbations. In a recent study, Xie and Lai [49] argued that allosteric domains have a preference for certain topologies like the Rossmann fold, while certain folds like the immunoglobulin-like β -sandwiches are disfavored. Our results, support such a viewpoint, because we find that the distributions of the nematic order parameter for allosteric domains lack the long tails at high P_2 values, which in the case of single-chain β -rich proteins primarily result from immunoglobulin-like folds, which can withstand large mechanical forces. Furthermore, the enhanced probabilities at $P_2 \approx 0.2$ (Figure 4), also reflect the preference of allosteric domains for Rossmann folds [49].

Figure 5. Some representative examples of allosteric domains. (a) Human kinesin spindle protein, Eg5 (PDB ID: 3ZCW) (b) PKA regulatory subunit of yeast (PDB ID: 3OF1) (c) Regulatory domain of LiCMS (PDB ID: 3F6G) (d) Bullet shaped GroEL-GroES complex (PDB ID: 1PCQ).

Among single-chain allosteric domains included within the ASBench database, the human kinesin spindle protein (also known as Eg5) exhibits the maximal ordering of secondary structural elements, with a P_2 value of ≈ 0.50 (Figure 5a). Within the kinesin superfamily, Eg5 is one of the most advanced drug targets, and most designed inhibitors, such as ispinesib, specifically bind to an allosteric pocket located around 10 Å away from the ATP binding site [50,51]. Recent evidence [51,52] suggests that the binding of ispinesib and other drugs to the allosteric pocket of Eg5 only induces minimal structural changes in the arrangement of the SSEs, and much of the perturbation is accommodated by the flexible loop region. Another prototypical single-chain allosteric domain is found in the protein kinase A (PKA) regulatory subunit of yeast [53]. The enzyme consists of two cyclic adenosine monophosphate (cAMP) binding domains (Figure 5b), and the unique interface formed between the two domains modules the allosteric mechanism underlying PKA activation [53]. Substantial structural changes, which occur at the interface due to cAMP binding, are primarily driven by the movements of the helical subunits within the CNB domains. Thus, in contrast to Eg5, allosteric transitions in this protein requires a very labile network of SSEs, associated with low nematic order. We find that this is precisely the case, and the P_2 value for the yeast regulatory subunit is only around 0.15.

Within the subset of multi-chain allosteric domains, the regulatory domain of *Leptospira interrogans* citramalate synthase (LiCMS) is associated with the highest value of P_2 (≈ 0.5). The functional form of LiCMS is a homodimer (Figure 5c), and it is involved in the catalysis of the first step of the isoleucine biosynthesis pathway in *L. interrogans*. This reaction is modulated by a feedback inhibition mechanism, mediated by the binding of isoleucine to the allosteric site at the dimer interface [54]. Structural and biochemical evidence strongly suggest that the inhibition of LiCMS is likely to occur via interdomain communication between the regulatory and catalytic domains, rather than large-scale rearrangements in the quaternary structure of the protein [54]. This scenario for allosteric regulation, seems consistent with the somewhat rigid packing of SSEs found in this protein. As a final example, we consider the well-studied GroEL-GroES complex [55], an ATP-fueled nanomachine, which aids in protein folding, by undergoing

a series of complex allosteric transitions [56,57]. The different rings of the GroEL-GroES complex move in a concerted fashion, with the inter-ring motions being anticorrelated, resulting in a form of nested cooperativity [58]. It is evident that intra- and inter-subunit rearrangements underlying such complex allostery can only be attained through an arrangement of structural elements, which can be disrupted by moderate strain. Indeed, in the GroEL-GroES complex, the extent of nematic order is small ($P_2 \approx 0.06$).

The examples above suggest that the extent of nematic order found in allosteric proteins has been tuned to carry out a specific function. While allosteric inhibition could be accommodated by a rigid arrangement of SSEs, complex allosteric transitions require a somewhat flexible network of structural elements that are arranged in an aperiodic manner, but responds to stresses readily. It is worth remembering that allostery requires that the two states (usually referred to as T and R) be readily accessible upon ligand binding (oxygen to Hemoglobin (Hb) for example). In the absence of the ligand the R state would be an excited state separated from the T state by a free energy gap. During the $T \rightarrow R$ transition, although the overall symmetry might be preserved the extent of order often changes, reflecting the effect of the ligand. Indeed, in a family of Hemoglobins we find that P_2 is greater in the R state compared to the T state. Thus, the strict preservation of the approximate symmetry in the ground state is neither required nor is observed.

Nematic order in different allostery models: The MWC and KNF phenomenological models were the earliest ones to describe protein allostery, and correspond to two extreme scenarios of ligand binding (namely, 'conformational selection' and 'induced fit'), in the energy landscape picture. A third possibility was raised by Dryden and Cooper [31] wherein allostery could occur without significant conformational change. In reality, proteins may accomplish their allosteric functions through a combination of these mechanisms, and a clear-cut distinction between them (for example, the MWC and KNF scenarios) may not always be feasible, although a recipe based on Hill coefficients has recently been suggested by Horovitz and coworkers [59]. We consider a representative example for each allostery model, and show that the nematic order, P_2 could be different for the ligand-free and ligand-bound states, as in the case of Hemoglobin.

For L-lactate Dehydrogenase (LDH), an archetypal example of MWC-type allostery (Figure 6) [60], the R state is associated with a marginally higher value of P_2 as compared to the T state, implying that the protein becomes somewhat rigid upon ligand binding [61]. The change in P_2 is more dramatic in case of guanlylate kinase (GK), which accomplishes its allosteric function through an 'induced fit' (KNF-type) mechanism [62], and undergoes a large conformational rearrangement upon binding of guanosine monophosphate (GMP). Finally, we consider the PDZ3 domain, which is known to exhibit dynamic allostery [63,64]. As expected the R and T states exhibit only slightly different values of P_2 , which could be an example of a CD-type mechanism (Figure 6).

Although it is tempting to deduce the plausible allosteric mechanisms only on the basis of the P_2 values for the end-states, a note of caution is in order. We stress that the knowledge of intermediate states is essential. In this regard, a systematic comparison of how P_2 evolves along the pathways connecting the ligand-free and ligand-bound states could provide a means to distinguish between the different allosteric models. We envisage that the nematic order would change more abruptly along a KNF-type pathway, as compared to a MWC-type mechanism, while for a CD-type mechanism, which proceeds via little or no conformational change, the deviation would be minimal.

The role of packing defects in allostery: The presence of packing defects within folded protein architectures could also affect their ability to undergo conformational switch, and carry out allosteric functions. In fact, work dating back to the late 1980s shows that engineering a packing defect in an enzyme, by means of a point mutation, could induce allosteric behavior [65]. A recent study [66] also underscores the importance

Figure 6. Representative examples of proteins exhibiting distinct allosteric mechanisms. The ligand-free state is shown in magenta, and the ligand-bound state is shown in blue. (a) L-lactate Dehydrogenase (PDB ID: 1LTH), is an example of an enzyme exhibiting MWC-type allostery. The P_2 value for the T state is 0.25, and for the R state is 0.28. (b) Guanylate Kinase depicted in the ligand-free (PDB ID: 1EX6) and ligand-bound (PDB ID: 1EX7) forms. The ligand-free form has a P_2 value of 0.38, and the ligand-bound form has a P_2 value of ≈ 0.56 . In this example, the ligand binding occurs via an induced-fit (KNF-type) mechanism, involving substantial conformational change. (c) The PDZ3 domain, which exhibits dynamic allostery. The ligand-free and ligand-bound states are structurally very similar, and are associated with P_2 values of 0.13 and 0.15, respectively.

310 of packing defects in the allosteric binding of a feedback inhibitor to the Human UDP-
311 Glucose Dehydrogenase via an induced fit mechanism.

312 4. Conclusion

313 The precise arrangement of secondary structure elements within protein structures
314 holds important cues for understanding their potential as an aggregation seed, allosteric
315 scaffold, or a therapeutic target, not mention the designability of proteins and their
316 folding. In this study, we assessed the organization of secondary structure elements for
317 a subset of proteins found in the PDB database, in terms of the nematic order parameter,
318 P_2 , which has been traditionally exploited in the studies of anisotropic molecules that
319 break orientational symmetry when close packed. We find that the distributions of P_2 are
320 sensitive to the number of polypeptide chains present in the protein molecule, as well as
321 the overall secondary structure. The secondary structure elements are maximally aligned
322 in highly ordered structures, such as helix-bundles, amyloid fibers, and antifreeze
323 proteins. However, as indicated by the small to intermediate values of P_2 , in most
324 allosteric proteins the arrangement of SSEs exhibits nematic order but coexists with other
325 more flexible regions (loops for example). Such an arrangement, observed in diverse
326 fold topologies, ensures that the aperiodically arranged structures can transmit allosteric
327 signals over the length of the complex, while the flexible regions can rearrange to readily
328 access the excited states for carrying out specific functions.

329 In allosteric domains, the SSEs cannot be maximally aligned (i.e. $P_2 \approx 1$) because
330 that would make the structure too rigid. The extent of nematic ordering is functionally
331 tuned, and depends on the precise structural requirements underlying the allosteric
332 transitions. Based on the large overlap between the P_2 distributions near the peak regions
333 for proteins in the PISCES database, and the ASBench subset, we speculate that many
334 more non-fibrous proteins, than currently known, could potentially be allosteric under
335 suitable conditions.

336 **Author Contributions:** D.C., M.L.M. and D.T. designed research; D.C. performed research and
337 analyzed data; and D.C., M.L.M. and D.T. wrote the manuscript.

338 **Acknowledgments:** We are grateful to Bill Eaton, Gilad Haran, Hagen Hoffmann, and Amnon
339 Horovitz for useful comments. This work was supported by the National Science Foundation
340 (CHE19-00093) and by the Collie-Welch Chair administered through the Welch foundation (F-0019).

341 **Conflicts of Interest:** The authors declare no conflict of interest.

References

1. Schrodinger, E. *What is life?*; University Press: Cambridge, 1943.
2. Falke, J.J. Enzymology: A moving story. *Science* **2002**, *295*, 1480–1481.
3. Fraunfelder, H.; McMohan, B.H.; Austin, R.H.; Chu, K.; Groves, J.T. The role of structure, energy landscape, dynamics, and allostery in the enzymatic function of myoglobin. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 2370–2374.
4. Fraunfelder, H.; Sligar, S.G.; Wolynes, P.G. The energy landscapes and motions of proteins. *Science* **1991**, *254*, 1598–1603.
5. Klimov, D.K.; Thirumalai, D. Symmetric Connectivity of Secondary Structure Elements Enhances the Diversity of Folding Pathways. *J. Mol. Biol.* **2005**, *353*, 1171–1186.
6. Levitt, M.; Chothia, C. Structural patterns in globular proteins. *Nature* **1976**, *26*, 552–558.
7. Chothia, C.; Levitt, M.; Richardson, D. Structure of proteins: packing of alpha-helices and pleated sheets. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 4130–4134.
8. Chothia, C. Principles that determine the structure of proteins. *Annu. Rev. Biochem.* **1984**, *53*, 537–572.
9. Minami, S.; Sawada, K.; Chikenji, G. How a spatial arrangement of secondary structure elements is dispersed in the universe of protein folds. *PLoS One* **2014**, *9*, e107959.
10. Woolfson, D.N.; Bartlett, G.J.; Burton, A.J.; Heal, J.W.; Nitsu, A.; Thomson, A.R.; Wood, C.W. *De novo* protein design: how do we expand the universe of possible protein structures. *Curr. Opin. Struct. Biol.* **2015**, *33*, 16–26.
11. Huang, P.S.; Oberdorfer, G.; Xu, C.; Pei, X.; ; Nannenga, B.; Rogers, J.M.; DiMaio, F.; Gonen, T.; Luisi, B.; Baker, D. High thermodynamic stability of parametrically designed helical bundles. *Science* **2014**, *346*, 481–485.
12. Hoang, T.X.; Trovato, A.; Seno, F.; Banavar, J.R.; Maritan, A. Geometry and symmetry prescript the free-energy landscape of proteins. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 7960–7964.
13. Levitt, M. Nature of the Protein Universe. *Proc. Natl. Acad. Sci. USA* **2009**, *7*, 11079–11084.
14. Perdigo, N.; Henrich, J.; Stolte, C.; Sabir, K.S.; Buckley, M.J.; Tarbor, B.; Signal, B.; Gloss, B.S.; Hammang, C.J.; Rost, B.; Schafferhans, A.; O'Donoghue, S.I. Unexpected features of the dark proteome. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 15898–15903.
15. Wolynes, P.G. Symmetry and energy landscapes of biomolecules. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 14249–14255.
16. Onuchic, J.N.; Wolynes, P.G. Theory of protein folding. *Curr. Opin. Struct. Biol.* **2004**, *14*, 70–75.
17. Wodak, S.J.; Paci, E.; Dokholyan, N.; Berezovsky, I.N.; Horovitz, A.; Li, J.; Hilser, V.J.; Bahar, I.; Karanicolas, J.; Stock, G.; Hamm, P.; Stote, R.H.; Eberhardt, J.; Chebaro, Y.; Dejagere, A.; Cecchini, M.; Changeux, J.P.; Bolhuis, P.G.; Vreede, J.; Faccioli, P.; Orioli, S.; Ravasio, R.; Yan, L.; Brito, C.; Wyart, M.; Gkeka, P.; Rivalta, I.; Palermo, G.; McCammon, J.A.; Panecka-Hofman, J.; Wade, R.C.; Di Pizio, A.; Niv, M.Y.; Nussinov, R.; Tsai, C.; Jang, H.; Padhorny, D.; Kozakov, D.; McLeish, T.C.B. Allostery in Its Many Disguises: From Theory to Applications. *Structure* **2019**, *27*, 566–578.
18. Cui, Q.; Karplus, M. Allostery and cooperativity revisited. *Protein Sci.* **2008**, *17*, 1295–1307.
19. Campbell, E.; Kaltenbach, M.; Correy, G.J.; Carr, P.D.; Porebski, B.T.; Livingstone, E.K.; ; Afriat-Jurnou, L.; Buckle, A.M.; Weik, M.; Hollfelder, F.; Tokuriki, N.; Jackson, C.J. The role of protein dynamics in the evolution of new enzyme function. *Nat. Chem. Biol.* **2016**, *12*, 944–950.
20. Sikosek, T.; Chan, H.S. Biophysics of protein evolution and evolutionary protein biophysics. *J. R. Soc. Interface* **2014**, *11*, 20140419.
21. Phillips, R. *The Molecular Switch: Signalling and Allostery*; Princeton University Press, 2020.
22. Anderson, P.W. *Basic Notions of Condensed Matter Physics*; Westview Press, 1997.
23. Monod, J.; Wyman, J.; Changeux, J.P. On the Nature of Allosteric Transitions : A Plausible Model. *J. Mol. Biol.* **1965**, *12*, 88–118.
24. Eaton, W.A.; Henry, E.R.; Hofrichter, J.; Mozzarelli, A. Is cooperative oxygen binding by hemoglobin really understood? *Nat. Struct. Biol.* **1999**, *6*, 351–358.
25. Thirumalai, D.; Hyeon, C.; Zhuravlev, P.I.; Lorimer, G. Symmetry, Rigidity, and Allosteric Signaling: From Monomeric Proteins to Molecular Machines. *Chem. Rev.* **2019**, *119*, 6788–6821.
26. Monod, J.; Wyman, J.; Changeux, J.P. On the nature of allosteric transitions: A plausible model. *J. Mol. Biol.* **1965**, *12*, 88–118.

- 401 27. Changeux, J.P.; Edelstein, S.J. Allosteric Mechanisms of Signal Transduction. *Science* **2005**,
402 308, 1424–1428.
- 403 28. Koshland, D.E.; Nemethy, G.; Filmer, D. Comparison of Experimental Binding Data and
404 Theoretical Models in Proteins Containing Subunits. *Biochemistry* **1966**, 5, 365–385.
- 405 29. Choy, M.S.; Li, Y.; Machado, L.E.S.F.; Kunze, M.B.A.; Connors, C.R.; Wei, X.; Lindorff-Larsen,
406 K.; Page, R.; Peti, W. Conformational Rigidity and Protein Dynamics at Distinct Timescales
407 Regulate PTP1B Activity and Allostery. *Mol. Cell* **2017**, 65, 644–658.
- 408 30. Henis, Y.I.; Levitki, A. Mechanism of negative cooperativity in glyceraldehyde-3-phosphate
409 dehydrogenase deduced from ligand competition experiments. *Proc. Natl. Acad. Sci. USA*
410 **1980**, 77, 5055–5059.
- 411 31. Cooper, A.; Dryden, D.T. Allostery without conformational change. A plausible model. *Eur.*
412 *Biophys. J.* **1984**, 11, 103–109.
- 413 32. McLeish, T.C.B.; Rodgers, T.L.; Wilson, M.R. Allostery without conformation change: mod-
414 elling protein dynamics at multiple scales. *Phys. Biol.* **2013**, 10, 056004.
- 415 33. McLeish, T.C.B.; Cann, M.J.; Rodgers, T.L. Dynamic Transmission of Protein Allostery
416 without Structural Change: Spatial Pathways or Global Modes? *Biophys. J.* **2015**, 109, 1240–
417 1250.
- 418 34. McLeish, T.; Schaefer, C.; von der Heydt, A. The ‘allosteron’ model for entropic allostery of
419 self-assembly. *Phil. Trans. R. Soc. B* **2018**, 373, 20170186.
- 420 35. Wang, G.; Dunbrack, R.L. PISCES: A protein sequence culling server. *Bioinformatics* **2003**,
421 19, 1589–1591.
- 422 36. Tusnady, G.E.; Dosztanyi, Z.; Simon, I. Transmembrane proteins in the Protein Data Bank:
423 identification and classification. *Bioinformatics* **2004**, 20, 2964–2972.
- 424 37. White, S.H. Biophysical dissection of membrane proteins. *Nature* **2009**, 459, 344–346.
- 425 38. Lomize, M.A.; Lomize, A.L.; Pogozheva, I.D.; Mosberg, H.I. OPM: orientations of proteins in
426 membranes database. *Bioinformatics* **2006**, 22, 623–625.
- 427 39. Huang, W.; Wang, G.; Shen, Q.; Liu, X.; Lu, S.; Geng, L.; Huang, Z.; Zhang, J. ASBench:
428 benchmarking sets for allosteric discovery. *Bioinformatics* **2015**, 31, 2598–2600.
- 429 40. Huang, Z.; Mou, L.; Shen, Q.; Lu, S.; Li, C.; Liu, X.; Wang, G.; Li, S.; Geng, L.; Liu, Y.; Wu,
430 J.; Chen, G.; Zhang, J. ASD v2.0: Updated content and novel features focusing on allosteric
431 regulation. *Nucleic Acids Res.* **2014**, 42, D510–D516.
- 432 41. Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of
433 hydrogen-bonded and geometrical features. *Biopolymers* **1983**, 22, 2577–2637.
- 434 42. McGibbon, R.T.; Beauchamp, K.A.; Harrigan, M.P.; Klein, C.; Swails, J.M.; Hernández, C.X.;
435 Schwantes, C.R.; Wang, L.P.; Lane, T.J.; Pande, V.S. MDTraj: A Modern Open Library for the
436 Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, 109, 1528 – 1532.
- 437 43. Dawson, N.L.; Lewis, T.E.; Das, S.; Lees, J.; Lee, D.; Ashford, P.; Orengo, C.; Sillitoe, I. CATH:
438 an expanded resource to predict protein function through structure and sequence. *Nuc. Acid.*
439 *Res.* **2019**, 45, D289–D295.
- 440 44. Seeber, M.; Felling, A.; Raimondi, F.; Muff, S.; Friedman, R.; Rao, F.; Calfisch, A.; Fanelli, F.
441 Wordom: a user-friendly program for the analysis of molecular structures, trajectories, and
442 free energy surfaces. *J. Comp. Chem.* **2011**, 32, 1183–1194.
- 443 45. Cecchini, M.; Rao, F.; Seeber, M.; Calfisch, A. Replica exchange molecular dynamics simula-
444 tions of amyloid peptide aggregation. *J. Chem. Phys.* **2004**, 121, 10748.
- 445 46. Wilson, M.R. Determination of Order Parameters in Realistic Atom-Based Models of Liquid
446 Crystal Systems. *J. Mol. Liq.* **1996**, 68, 23–31.
- 447 47. Andrienko, D. Introduction to liquid crystals. *J. Mol. Liq.* **2018**, 267, 520–541.
- 448 48. Eppenga, R.; Frenkel, D. Monte Carlo study of the isotropic and nematic phases of infinitely
449 thin hard platelets. *Mol. Phys.* **1984**, 52, 1303–1334.
- 450 49. Xie, J.; Lai, L. Protein topology and allostery. *Curr. Opin. Struct. Biol.* **2020**, 62, 158–165.
- 451 50. Ulaganathan, V.; Talapatra, S.K.; Rath, O.; Pannifer, A.; Hackney, D.D.; Kozielski, F. Structural
452 Insights into a Unique Inhibitor Binding Pocket in Kinesin Spindle Protein. *J. Am. Chem. Soc.*
453 **2013**, 135, 2163–2272.
- 454 51. Yokoyama, H.; Sawada, J.C.; Katoh, S.; Matsuno, K.; Ogo, N.; Ishikawa, Y.; Hashimoto, H.;
455 Fujii, S.; Asai, A. Structural Basis of New Allosteric Inhibition in Kinesin Spindle Protein
456 Eg5. *ACS Chem. Biol.* **2015**, 10, 1128–1136.
- 457 52. Kaan, H.Y.K.; Major, J.; Tkocz, K.; Kozielski, F.; Rosenfeld, S. Snapshots of Ispinesib-induced
458 Conformational Changes in the Mitotic Kinesin Eg5. *J. Biol. Chem.* **2013**, 288, 18588–18598.

- 459 53. Rinaldi, J.; Wu, J.; Yang, J.; Ralston, C.Y.; Sankaran, B.; Moreno, S.; Taylor, S.S. Structure of
460 Yeast Regulatory Subunit: A Glimpse into the Evolution of PKA Signaling. *Structure* **2010**,
461 *18*, 1471–1482.
- 462 54. Zhang, P.; Ma, J.; Zhang, Z.; Zha, M.; Xu, H.; Zhao, G.; Ding, J. Molecular basis of the inhibitor
463 selectivity and insights into the feedback inhibition mechanism of citramalate synthase from
464 *Leptospira interrogans*. *Biochem J.* **2009**, *421*, 133–143.
- 465 55. Thirumalai, D.; Lorimer, G.H.; Hyeon, C. Iterative annealing mechanism explains the
466 functions of the GroEL and RNA chaperones. *Prot. Sci.* **2020**, *29*, 360–377.
- 467 56. Ma, J.; Karplus, M. The allosteric mechanism of the chaperonin GroEL: A dynamic analysis.
468 *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 8502–8507.
- 469 57. Hyeon, C.; Lorimer, G.; Thirumalai, D. Dynamics of allosteric transitions in GroEL. *Proc.*
470 *Natl. Acad. Sci. USA* **2006**, *103*, 18939–18944.
- 471 58. Yifrach, O.; Horovitz, A. Nested cooperativity in the ATPase activity of the oligomeric
472 chaperonin GroEL. *Biochemistry* **1995**, *34*, 5303–5308.
- 473 59. Horovitz, A.; Mondal, T. Discriminating between Concerted and Sequential Allosteric
474 Mechanisms by Comparing Equilibrium and Kinetic Hill Coefficients. *J. Phys. Chem. B* **2021**,
475 *125*, 70–73.
- 476 60. Iwata, S.; Kamata, K.; Yoshida, S.; Minowa, T.; Ohta, T. T and R states in the crystals of
477 bacterial L-lactate dehydrogenase reveal the mechanism for allosteric control. *Nat. Struct.*
478 *Mol. Biol.* **1994**, *1*, 176–185.
- 479 61. Rader, A.J.; Brown, S.M. Correlating allostery with rigidity. *Mol. Biosyst.* **2011**, *7*, 464–471.
- 480 62. Blaszczyk, J.; Li, Y.; Yan, H.; Ji, X. Crystal structure of unligated guanylate kinase from yeast
481 reveals GMP-induced conformational changes. *J. Mol. Biol.* **2001**, *307*, 247–257.
- 482 63. Petit, C.M.; Zhang, J.; Sapienza, P.J.; Fuentes, E.J.; Lee, A.L. Hidden dynamic allostery in a
483 PDZ domain. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 18249–18254.
- 484 64. Kumawat, A.; Chakrabarty, S. Hidden electrostatic basis of dynamic allostery in a PDZ
485 domain. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E5825–E5834.
- 486 65. Kuo, L.C.; Zambidis, I.; Caron, C. Triggering of allostery in an enzyme by a point mutation:
487 ornithine transcarbamoylase. *Science* **1989**, *245*, 522–524.
- 488 66. Kadirvelraj, R.; Sennett, N.C.; Polizzi, S.J.; Weitzel, S.; Wood, Z.A. Role of packing de-
489 fects in the evolution of allostery and induced fit in human UDP-glucose dehydrogenase.
490 *Biochemistry* **2011**, *50*, 5780–5789.