

## Article

# Protein Fluctuations in Response to Random External Forces

Domenico Scaramozzino <sup>1,\*</sup> , Pranav M. Khade <sup>2,\*</sup>  and Robert L. Jernigan <sup>2,\*</sup> 

<sup>1</sup> Department of Structural, Geotechnical and Building Engineering, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy

<sup>2</sup> Roy J. Carver Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, IA 50011, USA

\* Correspondence: domenico.scaramozzino@polito.it (D.S.); pranavk@iastate.edu (P.M.K.); jernigan@iastate.edu (R.L.J.)

**Abstract:** Elastic network models (ENMs) have been widely used in the last decades to investigate protein motions and dynamics. There the intrinsic fluctuations based on the isolated structures are obtained from the normal modes of these elastic networks, and they generally show good agreement with the B-factors extracted from X-ray crystallographic experiments, which are commonly considered to be indicators of protein flexibility. In this paper, we propose a new approach to analyze protein fluctuations and flexibility, which has a more appropriate physical basis. It is based on the application of random forces to the protein ENM to simulate the effects of collisions of solvent on a protein structure. For this purpose, we consider both the C<sup>α</sup>-atom coarse-grained anisotropic network model (ANM) and an elastic network augmented with points included for the crystallized waters. We apply random forces to these protein networks everywhere, as well as only on the protein surface alone. Despite the randomness of the directions of the applied perturbations, the computed average displacements of the protein network show a remarkably good agreement with the experimental B-factors. In particular, for our set of 919 protein structures, we find that the highest correlation with the B-factors is obtained when applying forces to the external surface of the water-augmented ANM (an overall gain of 3% in the Pearson's coefficient for the entire dataset, with improvements up to 30% for individual proteins), rather than when evaluating the fluctuations obtained from the normal modes of a standard C<sup>α</sup>-atom coarse-grained ANM. It follows that protein fluctuations should be considered not just as the intrinsic fluctuations of the internal dynamics, but also equally well as responses to external solvent forces, or as a combination of both.

**Keywords:** elastic network model; protein flexibility; B-factors; protein fluctuations; random force application; protein surface



**Citation:** Scaramozzino, D.; Khade, P.M.; Jernigan, R.L. Protein Fluctuations in Response to Random External Forces. *Appl. Sci.* **2022**, *12*, 2344. <https://doi.org/10.3390/app12052344>

Academic Editor: Hervé Quiquampoix

Received: 1 December 2021

Accepted: 19 February 2022

Published: 23 February 2022

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



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The B-factors of a protein, the Debye-Waller factors or temperature factors, are measures of the atomic displacements about their equilibrium position, i.e., atomic fluctuations [1–3], but also the effects of multiple conformations as well as errors in the structures. They are generally accepted to be mostly the result of internal protein dynamics and any static disorder [4]. They have also been shown to be associated with protein flexibility and to correspond closely to protein mechanisms [5–9]. B-factors have been associated with protein flexibility, which is strictly related to protein action and function [10–12]. The experimental B-factors obtained from X-ray crystallography have been reproduced fairly accurately by various computational models.

One of the most widely used computational methods for investigating protein dynamics and fluctuations has been molecular dynamics (MD). MD simulations have proven their usefulness for investigations of protein folding, enzyme catalysis, and protein mechanisms in general [13–15]. Also, it has been shown that the MD-derived atomic fluctuations due to the internal protein motions show some degree of agreement with the experimental

B-factors [16,17]. However, due to the high computational burden of MD simulations, these can sometimes be expensive for investigating the large molecular complexes, especially regarding the slowest protein motions, accessible only at long simulation times. These slow motions are in fact usually the ones most closely related to the functional mechanisms of the protein and can take place on longer time scales than may be accessible in standard MD simulations. The harmonic assumption has been exploited for the extraction of the low-frequency protein dynamics [18–21]. Normal mode analysis (NMA) came into play as a simplified yet powerful tool to investigate the slower protein motions and for evaluating protein fluctuations and mechanisms [22,23], even in torsional space [24].

The seminal work of Tirion [25] showed that even a single-parameter harmonic potential, only based on the elastic properties of a network of Hookean springs connecting the protein atoms, was sufficient to reproduce the slow dynamics in good detail. All of the elastic network models are essentially entropic models since there is not usually any distinction of atom or amino acid types, i.e., all springs are taken to be similar in character. A further step towards simplification came with the coarse-graining development for these elastic network models (ENMs). Among the ENMs, the gaussian network model (GNM) was developed to obtain insights into protein dynamics and fluctuations simply by diagonalizing the Kirchhoff matrix, built by using the network contacts between close neighboring  $C^\alpha$  atoms [26–31]. Despite the remarkable correlations obtained between the GNM-based fluctuations and the experimental B-factors, the GNM lacks the information about the directions of motions, since it assumes that the motions are fundamentally isotropic in all directions [28]. The anisotropic network model (ANM) was then developed to include the three-dimensional directionality in the calculation of protein motions [32]. The ANM was then improved by various research groups to achieve higher correlations between the computed fluctuations and the experimental B-factors [33–37]. These elastic models were subsequently used to study the conformational changes of proteins arising from sets of low-frequency modes [38–48] as well as to generate feasible pathways between two known conformations [48–54].

Structural elastic models, particularly the ANM, were applied widely for the investigation of protein dynamics, fluctuations, and mechanism. However, they are also well-suited for the analysis of the protein structural responses from the application of external perturbations. Based on the work from Ikeguchi et al. [55], who showed that protein conformational changes upon ligand-binding could be analyzed based on linear response theory, the perturbation-response scanning (PRS) method was proposed by the Atilgan group [56,57]. Randomly oriented forces were applied at selected residues, and the corresponding response of the ferric binding protein [56] and another 24 proteins [57] were found to agree fairly accurately with the experimentally detected conformational change. A similar study was conducted by Gerek and Ozkan [58] to study the allosteric network in PDZ domains. A PRS-based technique, coupled with energy-based Metropolis Monte Carlo (MMC) simulations, was carried out by Liu et al. [59] to simulate the closed-to-open conformational change of a GroEL subunit due to directional forces presumed to originate from exothermic ATP hydrolysis. Interestingly, some of the apparent conformational changes being attributed to the binding of ATP or ADP may originate from the exothermic forces generated by hydrolysis. Recently, it was also shown that the application of forces in a dynamic fashion is able to drive the conformational change with a strong directionality correlation [60]. Eyal and Bahar [61] investigated the mechanical response of protein structure to external pulling forces in order to detect the anisotropic mechanical resistance to explain the outcomes of single-molecule manipulation techniques. More recently, we made use of a similar pairwise force application methodology in order to measure the overall protein flexibility by using the engineering concepts of structural compliance and stiffness [62].

Most of the works based on the coarse-grained ENMs include only one or a few representative atoms of the amino acids in the protein network, e.g., the  $C^\alpha$  atoms. Remarkably it has been seen that this geometric coarse-graining at the level of one point per amino acid yields almost exactly the same motions as from a full atomic elastic model. This

result is believed to be the result of the dense packing leading to the strong stability of protein domains [63]. However, most of these models do not explicitly account for the protein surface, which is the part most exposed to the surrounding environment. Water and small molecules can often be tightly bound to the protein surface, thereby affecting what is actually considered to be the surface of a protein structure. The role of such tightly bound crystalized waters in protein dynamics has been studied in the last few decades [64–67]. There have also been investigations of the solvent network surrounding the protein and its effect on the dynamics [68–70]. The inclusion of water molecules in the structure yields some increases in the quality of calculated enthalpies and of the residue interaction network [71,72]. This is one of the important reasons why all-atom MD simulations usually include these explicit waters.

This paper presents a novel method based on random perturbations applied to protein ENMs to assess protein fluctuations and flexibilities. Random forces are applied both throughout the complete protein elastic network and also separately to only the protein surface, which is exposed to the surrounding environment. In addition, a water-enriched ENM is considered, where the water molecules whose coordinates are given in the Protein Data Bank (PDB) files [73] are used as additional nodes in the elastic network. These latter force application simulations aim to mimic the random collisions occurring on the protein structure due to the interaction with the solvent and other solutes. From the calculation of the displacements within the protein network, i.e., the protein responses, we show that a good correlation is found with the experimental B-factors, thus leading to a good prediction of the protein flexibilities. The correlations with the usual mode-based fluctuations are also reported for comparison. It is also found that, in most cases, applying random forces on the surface of the water-enriched protein ENM leads to the highest correlation between the resulting displacements and the experimental B-factors. This demonstrates that the protein fluctuations may reflect more than the internal dynamics alone, and also include some effects from the continuous random bombardments or restraints by the surrounding solvent on the protein structure.

## 2. Methods

In this section, we briefly recount the fundamentals of the Anisotropic Network Model (ANM) [32], that is commonly used for generating the fluctuations in terms of the normal modes, and then we describe the computational framework related to the presently adopted force applications on the elastic networks.

### 2.1. Anisotropic Network Model (ANM) and the Calculation of Normal Mode-Based Fluctuations

ANM relies on the assumption that proteins can be modeled as simple elastic networks, made up of point nodes connected by linear elastic springs, allowing insights regarding fluctuations and global mechanisms [32,34,74]. For a system of  $N$  points, e.g.,  $N$  residues in the one-bead-per-residue coarse-grained representation, the  $3N \times 3N$  Hessian matrix of the system takes the following form:

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{1,1} & \dots & \mathbf{H}_{1,i} & \dots & \mathbf{H}_{1,j} & \dots & \mathbf{H}_{1,N} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \mathbf{H}_{i,1} & \dots & \mathbf{H}_{i,i} & \dots & \mathbf{H}_{i,j} & \dots & \mathbf{H}_{i,N} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \mathbf{H}_{j,1} & \dots & \mathbf{H}_{j,i} & \dots & \mathbf{H}_{j,j} & \dots & \mathbf{H}_{j,N} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \mathbf{H}_{N,1} & \dots & \mathbf{H}_{N,i} & \dots & \mathbf{H}_{N,j} & \dots & \mathbf{H}_{N,N} \end{bmatrix}, \quad (1)$$

where each  $3 \times 3$  submatrix  $\mathbf{H}_{i,j}$  contains the stiffness coefficients of the springs connecting nodes  $i$  and  $j$ . The off-diagonal submatrix  $\mathbf{H}_{i,j}$  is computed based on the harmonic potential of the elastic spring with force constant  $\gamma_{i,j}$ . The diagonal submatrices  $\mathbf{H}_{i,i}$  are calculated as the summation involving all the nodes linked to  $i$ th node with a negative sign [32]. The model, and consequently the Hessian matrix, depends on some numerical parameters: the

usual model uses a cut-off limit in distance to define the network topology. The original ANM was developed by considering equal spring constants for all connections, i.e.,  $\gamma_{ij} = \gamma$ , and a geometrical cut-off  $r_c$  was applied in order to consider springs placed only between close nodes. Typical values of  $r_c$  in the ANM are around 15 Å. Later on, distance-dependent force constants were introduced [33,34], as:

$$\gamma_{ij} \propto \frac{1}{(r_{ij}^0)^p}, \quad (2)$$

where  $p$  represents an inverse number for the decay parameter that allows connecting all points in a structure, with springs with variable strength—lower spring constants for longer inter-node distances  $r_{ij}^0$ . This distance-dependent spring network was shown to provide an improved agreement between the results and experimental data [33,34].

Once the Hessian matrix is computed based on the protein coordinates from the PDB file [73] and the spring connectivity, the  $3N$  eigenvalues  $\lambda_n$  and  $3N$  eigenvectors  $\mathbf{U}_n$  are obtained by solving the eigenvalue-eigenvectors decomposition. Due to the fact that the protein structure is usually not externally constrained, the first six eigenvalues are found to be zero, with corresponding mode shapes associated with the six rigid-body motions of the whole molecule. Therefore, these six motions are factored out and singular value decomposition is used to obtain the normal modes. Hence, the fluctuations based on the normal modes can be easily calculated as [22,75,76]:

$$B_i = \frac{8}{3} \pi^2 k_B T \sum_{n=7}^{3N} \frac{U_{i,n}^2}{\lambda_n}, \quad (3)$$

where  $B_i$  represents the computed B-factor for residue  $i$ ,  $k_B$  is the Boltzmann constant,  $T$  is the absolute temperature,  $U_{i,n}$  stands for the displacement of node  $i$  in the  $n$ th mode, and  $\lambda_n$  is the eigenvalue for the  $n$ th mode.

## 2.2. Force Application on Elastic Networks

Here, we propose a new approach for the calculation of protein flexibility and fluctuations. This approach is based on the application of random forces on protein elastic networks. These forces are intended to simulate the external perturbations that arise from the external environment, i.e., protein-solvent interactions, Brownian motions, collisions of molecules, etc. The reality, however, is that the environment is not usually known, but cryoEM has the promise of providing some information about this.

In this work, we use two different ENMs for modelling the protein structure. The first one is the parameter-free anisotropic network model (pfANM) [33], where the  $C^\alpha$  atoms are the only nodes used to build the protein network and all the  $C^\alpha$ - $C^\alpha$  connections are considered to be linked with distance-dependent springs. In the second model, the water molecules contained in the PDB file are also added to the network as additional nodes. Additional springs are correspondingly created that connect the water molecules to all the other nodes of the network. We refer to this second model as the water-pfANM (wpfANM). Both models are built by considering a decay exponent  $p$  for the spring constant equal to 3 (see Equation (2)), based on results shown to yield the best results analyzed in our previous work [62].

The response of the protein structure to external perturbations is evaluated by applying forces to the nodes of the network and consequently computing the corresponding elastic displacements. Various force application patterns are considered here. For the pfANM, the perturbations are applied both to the complete structure, i.e., on all  $C^\alpha$  atoms, and separately only to the nodes lying on the external protein surface. For the wpfANM, three different force patterns are considered: (1) forces acting on the entire network, i.e., all the  $C^\alpha$  atoms and water molecules, (2) only on the nodes lying on the protein-water network surface, and (3) only on the water molecules.

For each of the models considered (pfANM and wpfANM) and their force application patterns, the calculation is based on the generation of a random  $3 \times 1$  force vector  $F_i^s$  for each node  $i$  to be perturbed for each simulation  $s$ . The three scalar components of this force vector, i.e.,  $F_{i,x}^s$ ,  $F_{i,y}^s$  and  $F_{i,z}^s$ , are sampled from a uniform distribution  $U$  in the interval  $[-1,1]$ . The complete force vector  $F^s$  is then generated by assembling all the  $3 \times 1$  nodal vectors. Note that  $F^s$  is a  $3N \times 1$  vector, with  $N$  the number of  $C^\alpha$  atoms in the pfANM case or the total number of  $C^\alpha$  atoms plus water molecules in the wpfANM. Once the force vector  $F^s$  is defined, it is straightforward to compute the  $3N \times 1$  displacement vector  $\delta^s$  for each simulation  $s$  that contains the elastic displacements of the nodes, i.e., the protein response, as follows:

$$\delta^s = H^{-1}F^s, \quad (4)$$

where  $H^{-1}$  is the pseudo-inverse of the elastic network.  $H^{-1}$  can be computed from the eigenvalues and eigenvectors of the Hessian matrix as:

$$\tilde{H}^{-1} = \sum_{n=7}^{3N} \frac{U_n U_n^T}{\lambda_n}. \quad (5)$$

From the displacement vector  $\delta^s$ , the total displacement of each  $i$ th node can be computed as:

$$\delta_i^s = \sqrt{(\delta_{i,x}^s)^2 + (\delta_{i,y}^s)^2 + (\delta_{i,z}^s)^2}, \quad (6)$$

With  $\delta_{i,x}^s$ ,  $\delta_{i,y}^s$  and  $\delta_{i,z}^s$  being the three Cartesian components of the node displacements.

This procedure is repeated multiple times in order to generate a sample with different random force vectors  $F^s$  and evaluating the corresponding node displacements  $\delta^s$  each time. The average displacement of each node  $i$  is then evaluated as the average of all the displacements  $\delta_i^s$  over the total number of simulations  $S$ :

$$\delta_i = \frac{1}{S} \sum_{s=1}^S \delta_i^s. \quad (7)$$

In this analysis, we have generated a sample of 10,000, i.e.,  $S = 10,000$ . Then, the average displacement  $\delta_i$  of the  $i$ th residue for the sample can be compared to the experimental B-factors available in the PDB file. Pearson's correlation coefficient can finally be used to estimate the similarity between the two distributions, i.e., between the experimental B-factors and the simulated average displacements of the protein network due to the random perturbations. As a result, high Pearson's coefficients would indicate a high degree of similarity between the computed protein fluctuations and the experimental B-factors.

We mention above that different force application patterns are considered in this study. Specifically, besides applying forces to the entire protein network in the pfANM and wpfANM, and to the water molecules alone in the wpfANM, we also apply forces only to the nodes lying on the external protein surface (pfANM) and on the protein-water network surface (wpfANM). The reason for this is due to the fact that the effect of random collisions is more likely to occur on the exterior protein surface, rather than in the interior. For this purpose, the surface residues were calculated by computing the boundary geometry of the set of 3D coordinates of the network points. The external nodes were defined as those lying on the boundary surface. In this analysis, the generation of this surface was dependent on a parameter, known as the shrink factor. The shrink factor characterizes the amount of shrinkage of the boundary geometry, with values ranging from 0 to 1: zero corresponds to the convex hull, one corresponds to the maximum shrunk boundary. Note that the shrink factors used here for the generation of the external surface correspond to the normalized alpha shape, recently used by us [77] to extract hinge-domain information from protein structures. By using the approach based on the shrink factor, different external surfaces are generated by varying this numerical parameter from 0 to 1, in steps of 0.1. The surfaces obtained are finally used to select the external nodes on which the external forces will be applied.

### 2.3. Protein Dataset and Summary of the Models and Analyses

The analysis was performed on the same dataset used in our previous work [62] that includes 921 high-resolution X-ray single-chain protein structures from the PDB. The resolution of the selected crystal structures is below 1.3 Å, with a maximum sequence identity of 30%. Two structures were removed from the dataset, i.e., 1IXH and 2BK9, because of errors found in the PDB file regarding the waters. The size of the 919 final proteins range from 101 to 1174 residues.

As mentioned in Section 2.2, the pfANM and wpfANM were built for all protein structures by neglecting and considering water molecules, respectively. For both models, after evaluating the Hessian matrix, the mode-based fluctuations were evaluated according to Equation (3) and compared to the experimental B-factors. Then, for the pfANM, random forces were applied: (1) to all the nodes of the network and (2) only to the external nodes. Similarly, for the wpfANM, forces are considered to be acting on: (1) the entire network, (2) only on the external nodes and (3) only on the waters. From each of these simulations, the nodal displacements are computed from Equation (7) and Pearson correlation coefficients with the experimental B-factors are also calculated. In Table 1, the summary of the models and simulations and their designators for their Pearson correlation coefficients are given.

**Table 1.** The Designators for the Pearson Collection Coefficients for the Various Elastic Network Models.

Model		pfANM		wpfANM				
Nodes in the network		C <sup>α</sup> atoms		C <sup>α</sup> atoms + water molecules *				
Analysis		Mode-based fluctuations	Force application	Mode-based fluctuations	Force application			
Nodes perturbed		-	All nodes	External nodes	-	All nodes	External nodes **	Water nodes ***
Correlation coefficient		$\rho_{FL}$	$\rho_{FR,ALL}$	$\rho_{FR,EXT}$	$\rho_{W,FL}$	$\rho_{W,FR,ALL}$	$\rho_{W,FR,EXT}$	$\rho_{W,FR,WAT}$

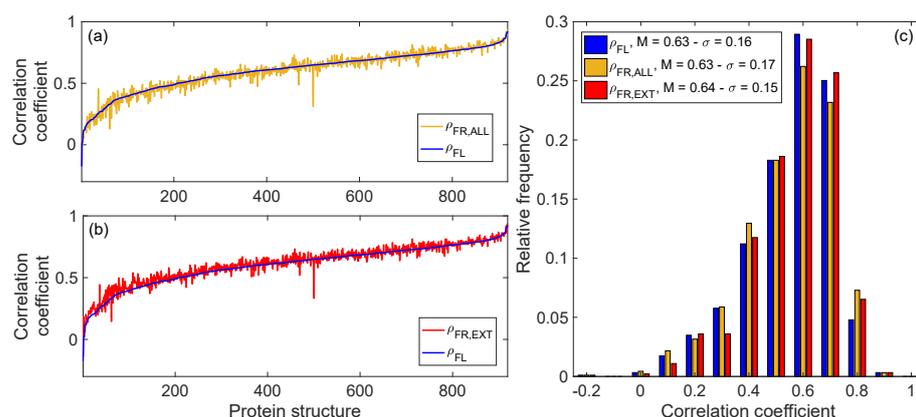
\* All waters in the pdb are included, without checking whether they are exterior. \*\* All nodes lying on the external surface (boundary geometry) of the protein-water network are perturbed by random forces; \*\*\* Only water molecules in the protein-water network are perturbed by random forces.

## 3. Results and Discussion

### 3.1. Fluctuations and Force Application on the pfANM

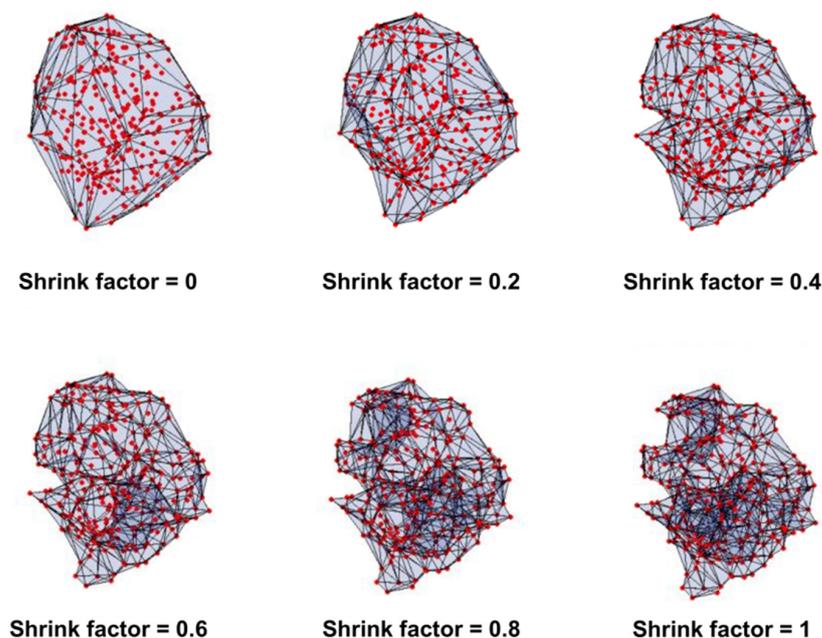
In this section, the flexibility of the protein structure will be investigated by using traditional pfANM mode-based fluctuations as well as from the outcome of the random force applications to the protein. Figure 1 shows the correlation coefficients obtained from a comparison of the experimental B-factors with the mode-based fluctuations as well as the average displacements due to the applied random forces. Figure 1a,b report the Pearson coefficients obtained for the 919 single-chain protein structures, with the values ordered by ascending values of  $\rho_{FL}$ . Figure 1c shows the statistical distributions of the correlation coefficients, whose median values and standard deviations are reported in the keys.

Figure 1a,b shows the distributions of correlation coefficients  $\rho_{FR,ALL}$  and  $\rho_{FR,EXT}$ , that are due to the random force applications, and these are observed to oscillate near the population of  $\rho_{FL}$ . This means that the average displacements of the protein elastic network due to the force application are indeed well correlated with the experimental B-factors, with a similar agreement as for the traditionally used mode-based fluctuations. The same conclusions can also be drawn by looking at Figure 1c, where the three statistical distributions exhibit the same pattern and similar median values, i.e., 0.63, 0.63 and 0.64, for  $\rho_{FL}$ ,  $\rho_{FR,ALL}$  and  $\rho_{FR,EXT}$ , respectively. Therefore, it cannot be concluded that applying random forces on the protein structures always enhances the correlation with the experimental B-factors, while it can be concluded that perturbing the protein structure by applying random forces leads to a good estimate of the experimentally determined fluctuations, at least as good as those found with the normal modes.



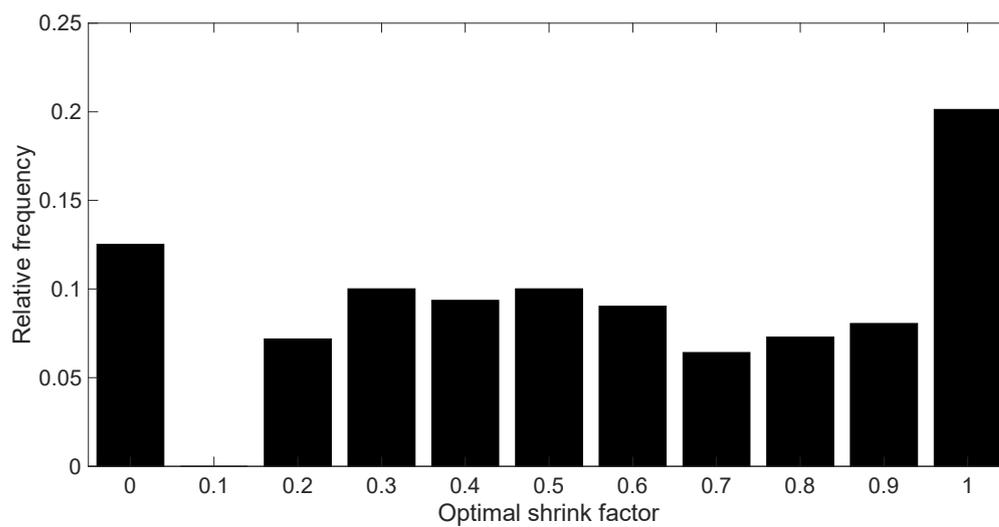
**Figure 1.** Correlations of experimental B-factors with ENM-based fluctuations and the average displacements due to random perturbations (pfANM): (a,b) correlation coefficients for the analyzed 919 protein structures (blue for  $\rho_{FL}$ , orange for  $\rho_{FR,ALL}$ , and red for  $\rho_{FR,EXT}$ ); (c) statistical distribution of the obtained correlation coefficients. The results are all very similar, showing relatively little differences among them.

As mentioned in the previous section, the application of random forces on the external protein surface requires the selection of the nodes that lie on the exterior. For this purpose, various external boundaries were generated by changing the shrink factor of the surface, varying from 0 to 1 with steps of 0.1. Figure 2 shows an example of different external surfaces generated with shrink factors of 0, 0.2, 0.4, 0.6, 0.8 and 1 for the infrared fluorescent protein (PDB: 5AJG). As can be observed, increasing the shrink factor leads to considering a higher number of nodes lying on the surface, which in turn has a more detailed structure. Since the primary determinant of a structure's dynamics is its shape, clearly the most detailed structure would be expected to be the best [78].



**Figure 2.** Dependence of the generated external boundary (external protein surface) on the value of the shrink factor. Infrared fluorescent protein (PDB: 5AJG) is reported as an example, with shrink factors equal to 0, 0.2, 0.4, 0.6, 0.8 and 1. Red points represent the nodes of the network ( $C^\alpha$  atoms), while the external surface is represented by Delaunay triangles (in light blue), which connect the nodes in the external boundary.

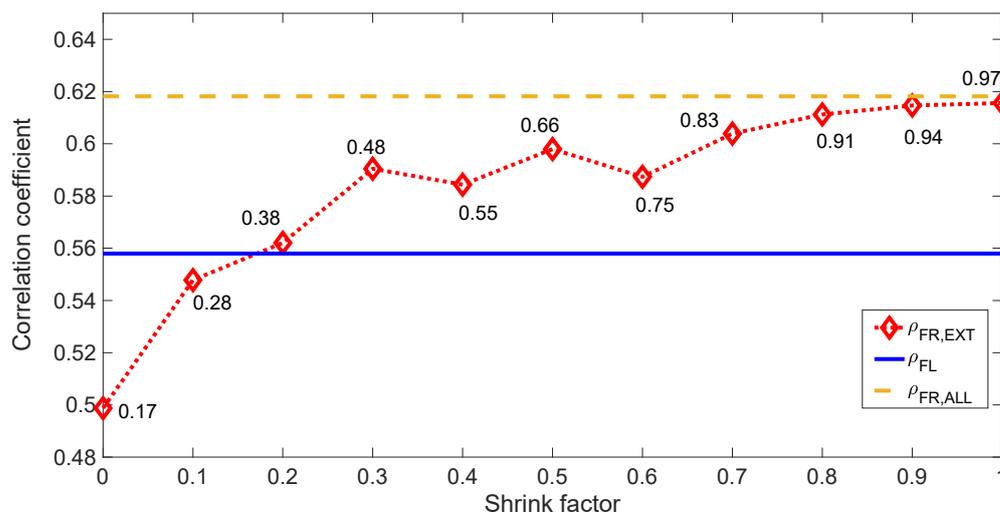
For each of these generated surfaces, the random forces were applied only to the nodes lying on the external boundary. The resulting network displacements were then evaluated from Equations (4)–(7). It follows that the correlation of the average displacements (from 10,000 samples) due to the application of the external forces, i.e.,  $\rho_{FR,EXT}$ , also depends on the adopted shrink factor. The shrink factor that leads to the maximum value of  $\rho_{FR,EXT}$  for each protein is then selected as the optimal one. Figure 3 reports the statistical distribution of the optimal shrink factors obtained for the 919 single-chain proteins. As can be observed, the optimal shrink factor assumes almost all values, meaning that it is strongly protein-specific. Nevertheless, a slight preference towards shrink factors equal to 1.0 is observed.



**Figure 3.** Statistical distribution of the optimal shrink factor, based on a comparison between experimental B-factors and average displacements due to the application of random forces on the exterior protein surface.

In Figure 4 we show results for the example of the infrared fluorescent protein (PDB: 5AJG), where the correlation coefficients  $\rho_{FL}$ ,  $\rho_{FR,ALL}$  and  $\rho_{FR,EXT}$ , are shown depending on the shrink factor for the external surface representation. From these calculations, we obtain a correlation between the B-factors and the traditional mode-based fluctuations  $\rho_{FL}$  equal to 0.56, a correlation with the displacements resulting from the application of forces to the entire structure  $\rho_{FR,ALL}$  equal to 0.62, and a correlation derived from perturbations only on the external surface  $\rho_{FR,EXT}$  which varies with the shrink factor and reaches a maximum value of 0.62 for an optimal shrink factor of 1.0. As can be seen from the results, in this case, applying random forces on the protein network enhances the correlation with the experimental B-factors of 6% (0.62 vs. 0.56) compared with the usual mode-based fluctuations. This result points out the cohesive nature of the protein structure, showing that the point of application of forces does not matter much, with the result of applying forces in all possible directions on the surface yielding nearly the same result as applying them in all directions throughout the structure when the surface representation is detailed enough.

Figures S1–S5 in the Supplementary Material report similar results, obtained by adopting different values of  $p$  for the decay exponent of the ENM spring constants ( $p = 1, 2, 4, 6$  and 12). As can be seen there, similar conclusions can be drawn for these cases. Note that, for this protein, higher values of  $p$ , e.g.,  $p = 4$  and 6, lead to a greater enhancement in the correlation with experimental B-factors when the ENM forces are applied, rather than just looking at the intrinsic dynamical fluctuations.



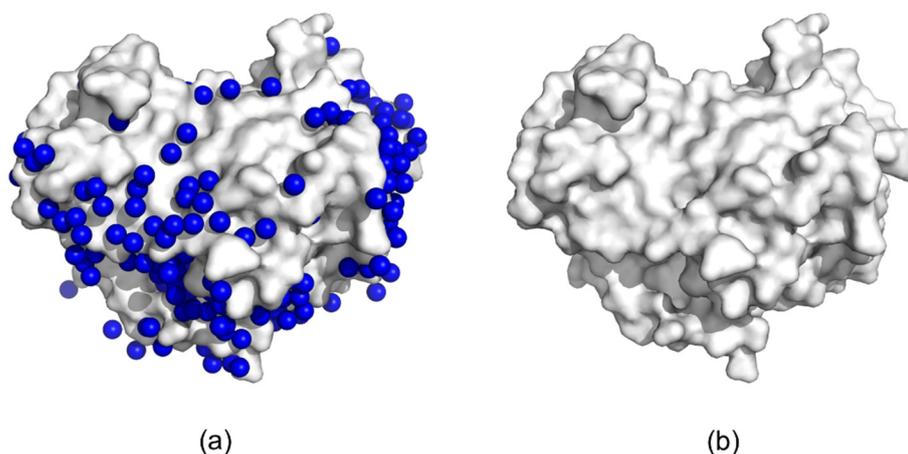
**Figure 4.** Experimental B-factors vs. mode-based fluctuations and average displacements due to random forces for the infrared fluorescent protein (PDB: 5AJG). The solid blue line refers to  $\rho_{FL}$ , the dashed orange line to  $\rho_{FR,ALL}$  and the dotted red line to  $\rho_{FR,EXT}$ . The correlation arising from the application of forces only on the external surface, i.e.,  $\rho_{FR,EXT}$ , depends on the selected shrink factor, which is in the range 0–1. For each shrink factor, the values reported close to the marker represent the fraction of external nodes, out of the total 301 nodes of the network. This shows that the best result is the structure representation with the greatest level of detail, and interestingly the most detailed structure with forces applied to the surface only leads to nearly the same result as the application of forces throughout the structure.

### 3.2. Incorporating Waters into the Computations

In this section, we show results obtained by also including the localized waters given in the crystal structure as part of the structure for defining the elastic network. There is some ongoing debate about whether or not these bound waters should be considered as an actual part of the structure. Each high-quality protein structure available in the PDB contains a substantial number of waters that were present in the crystal formed at low temperatures. The open question is whether these remain bound at higher temperatures. These molecules often typically form a network of hydrogen bonds with the side chains of polar amino acids on the protein surface and thus can appear to be quite stable. It follows that these waters may possibly cause some changes to the overall flexibility and dynamics of any given protein. Moreover, since we are interested in looking at the responses of each protein structure due to external perturbations, the inclusion of these external waters would be expected to affect the motions to some extent. Figure 5 shows a surface representation of the infrared fluorescent protein (PDB: 5AJG), with and without the addition of water molecules in Figure 5a,b, respectively. The protein structure is shown in light-gray, with the surface depiction highlighting the external surface and cavities. Water molecules available in the PDB are shown in Figure 5a as blue spheres. As can be seen from the comparison between the two figures, most of the crystallized water molecules are bound in concave parts of the protein surface, and thus act to smooth the structure [79]. This smoothing might restrict the flexibility of certain parts of the proteins that might cause problems for the mechanisms otherwise; this can be looked at as flowing liquid that represents protein motion, if a certain direction of the flow is restricted, it may change the overall flow of the water, therefore, to have an optimal flow path (specific functional protein motion), the restrictions (waters) are as important as the structure itself.

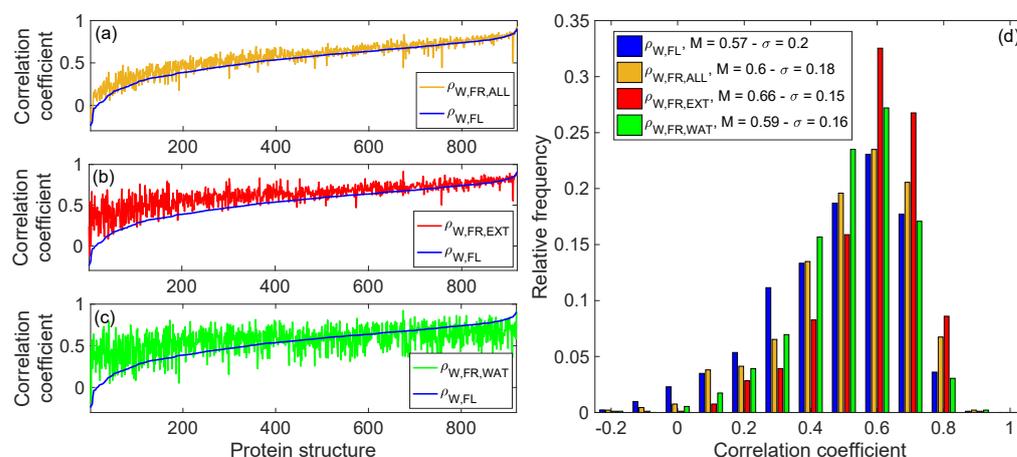
As mentioned in the previous section, the wpfANM is built as the usual pfANM with  $p$  equal to 3 (see Equation (2)), except that both the coordinates of the  $C^\alpha$  atoms and the water molecules are now considered as a part of the whole structure. Based on the resulting wpfANM, the mode-based fluctuations are evaluated from Equation (3), whereas the

average displacements resulting from the random perturbations are computed according to Equations (4)–(7). In the latter cases, as explained above, three types of force application are considered, as shown in Table 1. From the calculations, we then obtain four correlation coefficients to compare with the experimental B-factors, namely  $\rho_{W,FL}$ ,  $\rho_{W,FR,ALL}$ ,  $\rho_{W,FR,EXT}$  and  $\rho_{W,FR,WAT}$ , with these being described in Table 1.



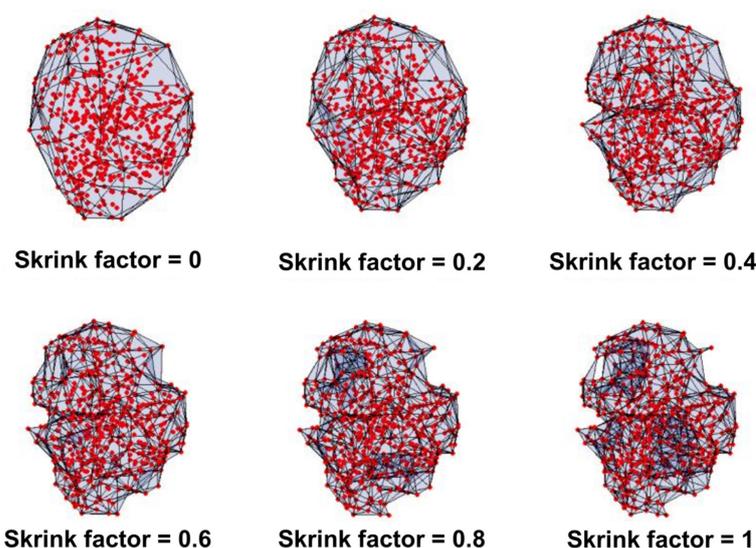
**Figure 5.** Surface representation of infrared fluorescent protein (PDB: 5AJG): (a) protein structure (light gray) + water molecules (blue spheres); (b) protein structure alone.

Figure 6 shows the correlation coefficients for the 919 proteins investigated. Figure 6a–c report the correlations ordered by increasing values of  $\rho_{W,FL}$ , whereas Figure 6d displays the statistical distribution for all four Pearson coefficients, whose median values and standard deviations are shown in the key. From the results reported in Figure 6, it follows that applying forces on the protein network slightly (<10%) enhances the prediction of the B-factors with respect to the traditional mode-based fluctuations. As a matter of fact, the median value of  $\rho_{W,FL}$  was found to be 0.57 for the selected dataset, whereas the median values of  $\rho_{W,FR,ALL}$ ,  $\rho_{W,FR,EXT}$  and  $\rho_{W,FR,WAT}$  were 0.60, 0.66, and 0.59, respectively. It is evident that applying random forces on the surface of the network (which now considers also the layer of water molecules) yields a significant 10% gain in the correlation with the experimental fluctuations, compared to the mode-based fluctuations.

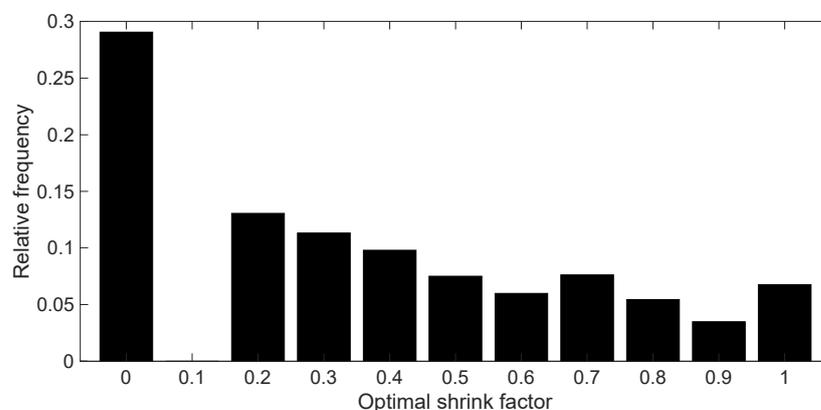


**Figure 6.** Experimental B-factors vs. mode-based fluctuations and average displacements due to random perturbations (wpfANM—water molecules included): (a–c) correlation coefficients for the 919 protein structures (blue curve for  $\rho_{W,FL}$ , orange curve for  $\rho_{W,FR,ALL}$ , red curve for  $\rho_{W,FR,EXT}$  and green curve for  $\rho_{W,FR,WAT}$ ); (d) statistical distribution of the correlation coefficients. The highest correlations are seen when perturbations are applied on the surface.

Also in the case of the wpfANM, the surface of the network is not unique but depends on the adopted shrink factor. The water molecules contained in the network play a major role in the definition of the surface since they are mostly placed on the exterior of the structure (see Figure 5). As an example, Figure 7 shows the different surfaces generated for the infrared fluorescent protein (PDB: 5AJG, with water molecules included) with shrink factors equal to 0, 0.2, 0.4, 0.6, 0.8 and 1. As can be seen by comparing Figure 7 to Figure 2, other than the selected value of the shrink factor, the shape of the external surface also depends on the presence of the water molecules within the network. Similarly to what was shown in the previous section, Figure 8 shows the optimal shrink factors (with the best correlation with B-factors) obtained for the 919 water-enriched protein structures. Again, the distribution of the optimal shrink factor is rather uniform, although in this case a slight bias towards the convex hulls surfaces appears, i.e., with a shrink factor equal to 0. This probably reflects a preference for smoother structures when water molecules are added (see Figure 5).

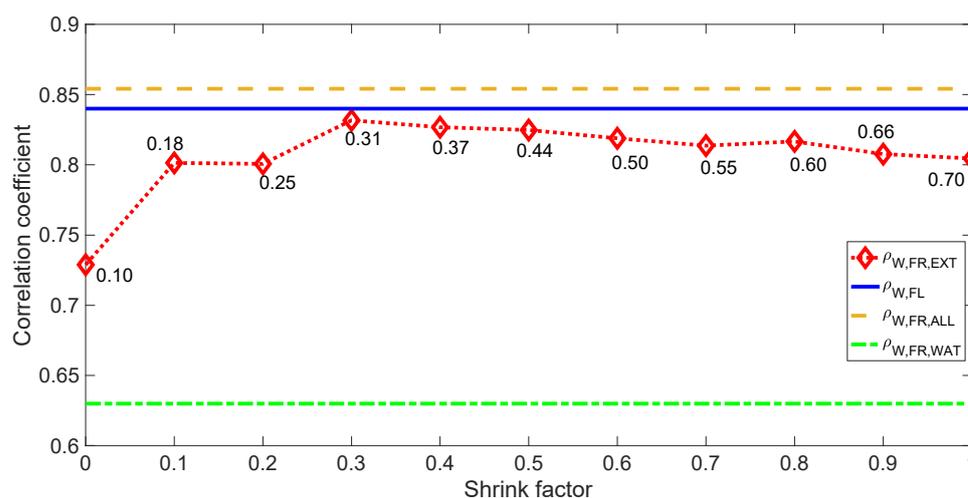


**Figure 7.** Dependence of the generated external boundary (external protein surface) on the value of the shrink factor. Infrared fluorescent protein (PDB: 5AJG), with water molecules included from the PDB structure file, is reported as an example, with shrink factors equal to 0, 0.2, 0.4, 0.6, 0.8 and 1. Red points represent the nodes of the network ( $C^\alpha$  atoms + water molecules), while the external surface is represented by Delaunay triangles (in light blue), which connect the nodes in the external boundary.



**Figure 8.** Statistical distribution of the optimal shrink factor, resulting from the comparison between experimental B-factors and average displacements due to the application of random forces on the external protein surface, with PDB water molecules included.

Figure 9 shows the correlation coefficients of the wpfANM of the infrared fluorescent protein (PDB: 5AJG), and how it depends on the adopted shrink factor for the surface. From the calculations, we obtain a correlation with the mode-based fluctuations  $\rho_{W,FL}$  equal to 0.84, a correlation with the displacements from the application of forces to the entire structure  $\rho_{W,FR,ALL}$  equal to 0.85, a correlation with the displacements from the application of forces only to the water molecules  $\rho_{W,FR,WAT}$  equal to 0.63, and a correlation with the displacements due to the perturbations only on the surface  $\rho_{W,FR,EXT}$  that varies with the shrink factor and reaches a maximum of 0.83 for the optimal shrink factor of 0.3. In this case, it is remarkable that by applying random perturbations on only 31% of the nodes (corresponding to a shrink factor of 0.3), we obtain a high correlation with the experimental data (0.83 Pearson coefficient). It should be noted that this correlation is found by comparing the experimental B-factors of all protein residues (both on the surface and within the core) to the computed average displacements due to the application of perturbations only on the external part of the elastic network. Thus, it follows that even perturbing a small portion of the protein surface (31%) allows us to predict fairly accurately the fluctuations of the entire protein. This result has its origin in the strong coupling throughout the elastic network model: since the ENM is a highly cooperative model, the perturbation of only a small part of the structure can indeed generate fluctuations over the entire protein. This arises from the specific features of the three-dimensional protein structure and all of the internal connections, which are both built into the ENM.



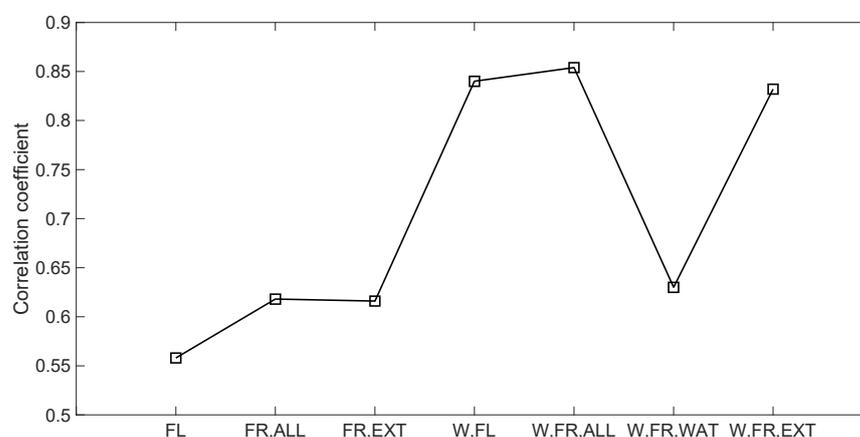
**Figure 9.** Experimental B-factors vs. mode-based fluctuations and average displacements due to random forces for the infrared fluorescent protein (PDB: 5AJG), with PDB water molecules included. The continuous blue line refers to  $\rho_{W,FL}$ , the dashed orange line to  $\rho_{W,FR,ALL}$ , the dotted red line to  $\rho_{W,FR,EXT}$ , and the dashed-dotted green line to  $\rho_{W,FR,WAT}$ . The correlations from the application of forces only on the protein's external surface, i.e.,  $\rho_{W,FR,EXT}$ , depends on the selected shrink factor, in the range 0–1. For each shrink factor, the values reported close to the marker represent the fraction of external nodes out of the total of 533 nodes (301 C $^{\alpha}$  atoms + 252 water molecules) of the network.

Figures S6–S10 in the Supplementary Material show similar results as those reported in Figure 9, but for different exponents  $p$  in the water-augmented ENM, namely,  $p = 1, 2, 4, 6$  and  $12$ . Despite some differences in the numerical values of the correlation coefficients, similar conclusions are reached as for those in Figure 9.

### 3.3. Comparison between pfANM and wpfANM Results

In the previous sections, it has been shown that perturbing the protein structure with random forces, either on the entire structure or on the surface, generally leads to a fairly accurate prediction of the protein fluctuations and flexibility. In a large number of cases, it has also been found that the agreement with the experimental B-factors was improved

with respect to considering the traditional fluctuations of the usual elastic network. As an example, the results shown in the previous sections for the infrared fluorescent protein (PDB: 5AJG) are reported together in Figure 10 in terms of correlation coefficients with the experimental B-factors. Several observations can be made regarding this figure. First, it is evident that the traditionally employed internal protein fluctuations provide the lowest correlation with the experimental data, with a correlation of about 55% ( $\rho_{FL}$ ). Conversely, applying random perturbations on the same elastic network leads to a 5% gain in the correlations with the B-factors ( $\rho_{FR,ALL}$  and  $\rho_{FR,EXT}$ ). Furthermore, adding the PDB water molecules to the elastic network further improves the correlation with the experimental data. In this case, considering the internal protein fluctuations of the water-enriched elastic network or applying random forces to it yields correlation coefficients of about 85% ( $\rho_{W,FL}$ ,  $\rho_{W,FR,ALL}$  and  $\rho_{W,FR,EXT}$ ). Also, applying random forces only on these water molecules, i.e., not perturbing the protein molecule directly but only the water molecules attached to the network (see Figure 5a), leads to a correlation of about 60% ( $\rho_{W,FR,WAT}$ ), which is still higher than the correlation obtained with the traditional mode-based internal protein fluctuations of the ENM ( $\rho_{FL} = 55\%$ ).



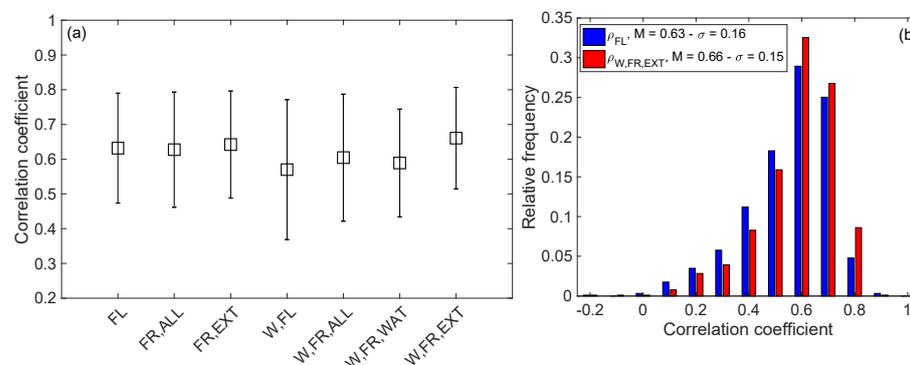
**Figure 10.** Infrared fluorescent protein (PDB: 5AJG): correlation coefficients obtained from the seven types of analyses, as reported in Table 1 and in the previous sections.

Figures S11–S15 in the Supplementary Material show similar outcomes, but are obtained by changing the exponent  $p$  of the ENM. As can be seen, similar conclusions are drawn. In all cases, it is found that perturbing the protein with random forces provides an enhancement of the correlation with experimental B-factors, rather than considering the traditional ENM with only intra-protein interactions. Moreover, the inclusion of waters in the ENM leads to further improvements in the correlation, with some gains in the Pearson coefficient being as high as 30–35%.

The example shown in Figure 10 obviously refers to one single case, but these results were found for a decent amount of protein structures. For other protein structures, the addition of water molecules to the protein network led to results which were quite similar to the ones obtained in the classical way, i.e., calculating the mode-based fluctuations of the traditional no-water elastic network.

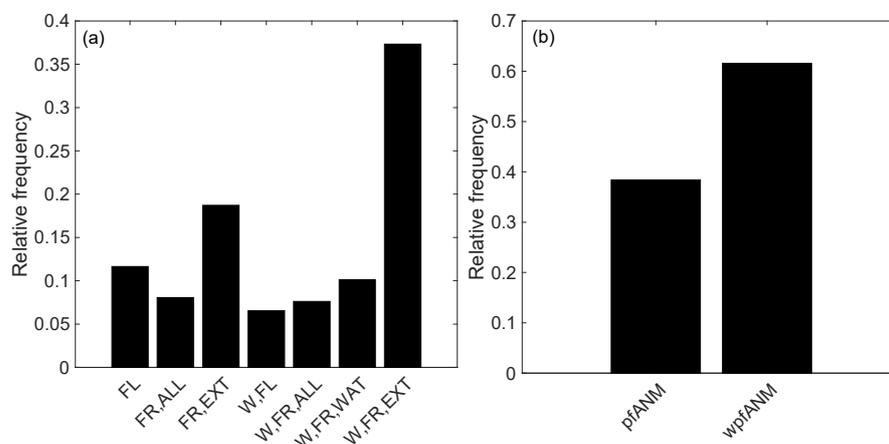
Figure 11a reports the median values and standard deviations of the seven correlation coefficients obtained for the dataset of 919 single-chain protein structures, as reported in the keys of Figures 1c and 6d. As can be observed, the median values lie in the range 0.60–0.65, and present a similar standard deviation (15–20%). However, the distribution with the highest median value and the lowest standard deviation was found for the analysis involving the application of random perturbation on the external surface of the wpfANM, i.e.,  $\rho_{W,FR,EXT}$ . A direct comparison between the statistical distribution of  $\rho_{W,FR,EXT}$  and the one related to the traditional mode-based internal protein fluctuations of the no-water pfANM, i.e.,  $\rho_{W,FL}$ , is reported in Figure 11b. From the direct comparison of the two

distributions, it is clear that applying random forces on the surface of the wpfANM leads to an overall, yet slight, improvement of the correlation with the experimental data.



**Figure 11.** (a) Correlation coefficients obtained from the seven types of analyses (median values and standard deviations of the statistical distributions) for the 919 single-chain protein structures; (b) direct comparison between the statistical distributions of  $\rho_{FL}$  and  $\rho_{W,FR,EXT}$ .

Based on the correlation coefficients obtained for our entire dataset of protein structures, the analysis among the seven performed ones (Table 1) that gave the highest correlations with the experimental data was noted. Figure 12a shows the relative number of such occurrences for each type of analysis. It was found that, out of the 919 investigated protein structures, the mode-based fluctuations of the pfANM provided the highest correlation coefficient in 107 cases (11.6%), the application of random forces on the entire pfANM in 74 cases (8.1%), the application of forces on the external surface of the pfANM for 172 cases (18.7%), the mode-based fluctuations of the wpfANM in 60 cases (6.5%), the application of forces on the entire wpfANM in 70 cases (7.62%), the application of forces only on the water molecules of the wpfANM in 93 cases (10.1%), while the application of forces on the external surface of the wpfANM were provided in 343 cases (37.3%). As can be seen and has already been discussed concerning Figure 11, the application of random forces on the surface of the water-enriched protein network is statistically the best performing, although not the only one, with regards to better predicting protein fluctuations and flexibility in terms of experimental B-factors. Moreover, looking cumulatively at the analyses *FR,EXT* and *W,FR,EXT*, in 515 cases (56.0%), the application of random forces on the external surface of the protein network yields the highest correlation coefficients, thus confirming that perturbing the external protein surface can induce a response in good agreement with the experimental B-factors.



**Figure 12.** (a) Relative frequency of the highest correlation coefficient for each type of analysis for the dataset of 919 single-chain protein structures; (b) relative frequency of the highest correlation coefficient for the pfANM or the wpfANM.

Figure 12b shows the occurrence of the highest correlation coefficients for the two models, i.e., the pfANM vs. wpfANM. From the outcomes, it was obtained that in 353 cases (38.41%) the highest correlation with the B-factors was obtained with the pfANM, whereas in the remaining 566 cases (61.59%) the wpfANM was allowed to reach the highest correlation with the experimental data. This confirms that considering the PDB water molecules might actually enhance the prediction of the protein fluctuations and therefore the numerical evaluation of the experimental B-factors.

#### 4. Conclusions

Research carried out in the last decades has shown that protein fluctuations and flexibility, as measured by the experimental B-factors, mainly arise from the internal protein motions and inherent dynamics. The dynamics are known to originate from the tertiary structure, as recognized within the fundamental sequence-structure-dynamics-function paradigm of protein action. Therefore, it should be more appropriate to say that protein fluctuations and flexibility arise from the protein structure and can be mediated by its dynamics. As a matter of fact, in a recent work [62], we have shown that the overall protein flexibility, as measured by the experimental B-factors, can also be retrieved by applying pairwise static forces to the protein ENM and measuring the amount of compliance against these external pulling forces.

In this paper we have proposed an additional viewpoint as regards protein fluctuations and flexibility. We applied random static forces throughout the protein elastic network and evaluated the response of the network via the computation of average nodal displacements. From the comparison of these average displacements against the experimental B-factors, we found that the protein flexibility, and therefore its fluctuations, can indeed be elucidated with such a procedure. Also, we found that if these perturbations are applied on the protein surface, and if crystallized water molecules are also inserted into the model, higher correlations with experimental data can often be found. This suggests that protein fluctuations can also be seen as (fully or partly) the response of the protein structure to external forces, which might be induced by the continuous collisions of the solvent and other solute molecules around the protein structure.

It is important to mention that the goal of the analysis presented here was to propose a new methodology, and therefore a new perspective, to understand protein fluctuations. However, no additional work has been carried out yet as regards the optimization of the elastic model upon which the random perturbations are applied. This might eventually improve the correlation with the experimental data and is our next goal. As described in the text, the pfANM has been used for the standard ENM, whereas a wpfANM has been generated in order to account for the presence of crystallized water, where water molecules have simply been added as additional nodes to the network. We plan to optimize the elastic model by including different spring constants for the  $C^\alpha$ - $C^\alpha$  connections,  $C^\alpha$ -water connections and water-water connections, which should simulate more realistically the different atomic interactions (residue-residue, residue-water, water-water).

Attention has been paid to the external surface of the network. Thus, we are also planning to use a different version of ENM, where the contact topology is not generated by using the traditional cut-off limit, but using alpha-shapes associated with Delaunay tessellations. A recent work from Koehl et al. [36] showed that such a procedure is able to generate elastic models with enhanced agreement with experimental data. Applying external forces on such optimized models, and adding water molecules as well, might enhance the correlation with experimental B-factors, allowing for the better explanation of fluctuations, and therefore the way a protein moves and functions.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/app12052344/s1>, Figures S1–S5: Experimental B-factors vs. mode-based fluctuations and average displacements due to random forces for the infrared fluorescent protein (PDB: 5AJG), with  $p = 1, 2, 4, 6$  and  $12$  for the decay of spring constants in the ENM, Figures S6–S10: Experimental B-factors vs. mode-based fluctuations and average displacements due to random forces for the

infrared fluorescent protein (PDB: 5AJG), with PDB water molecules included, with  $p = 1, 2, 4, 6$  and  $12$  for the decay of spring constants in the ENM, Figures S11–S15: Infrared fluorescent protein (PDB: 5AJG): correlation coefficients obtained from the seven types of analyses, as reported in Table 1 and in the previous sections. Case with  $p = 1, 2, 4, 6$  and  $12$  for the decay of spring constants in the ENM.

**Author Contributions:** Conceptualization, D.S., P.M.K. and R.L.J.; methodology, D.S. and P.M.K.; software, D.S.; validation, D.S., P.M.K. and R.L.J.; formal analysis, D.S. and P.M.K.; investigation, D.S. and P.M.K.; resources, R.L.J.; data curation, D.S.; writing—original draft preparation, D.S.; writing—review and editing, D.S., P.M.K. and R.L.J.; visualization, D.S., P.M.K. and R.L.J.; supervision, R.L.J.; project administration, R.L.J.; funding acquisition, R.L.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** We gratefully acknowledge the support from NIH grant R01GM127701.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Additional data is available upon request from the authors.

**Acknowledgments:** We thank Research IT @Iowa State University for helping with some aspects of the computing.

**Conflicts of Interest:** The authors declare that they have no conflict of interest.

## References

1. Debye, P. Interferenz von Röntgenstrahlen und Wärmebewegung. *Ann. Phys.* **1913**, *348*, 49–92. [[CrossRef](#)]
2. Trueblood, K.N.; Bürgi, H.-B.; Burzlaff, H.; Dunitz, J.D.; Gramaccioli, C.M.; Schulz, H.H.; Shmueli, U.; Abrahams, S.C. Atomic Displacement Parameter Nomenclature. Report of a Subcommittee on Atomic Displacement Parameter Nomenclature. *Acta Crystallogr. Sect. A* **1996**, *52*, 770–781. [[CrossRef](#)]
3. Sherwood, D.; Cooper, J. *Crystals, X-rays and Proteins: Comprehensive Protein Crystallography*; OUP Oxford: Oxford, UK, 2010.
4. Na, H.; Hinsen, K.; Song, G. The Amounts of Thermal Vibrations and Static Disorder in Protein X-ray Crystallographic B-factors. *Proteins Struct. Funct. Bioinform.* **2021**, *89*, 1442–1457. [[CrossRef](#)] [[PubMed](#)]
5. Karplus, P.A.; Schulz, G.E. Prediction of chain flexibility in proteins—A tool for the selection of peptide antigens. *Naturwissenschaften* **1985**, *72*, 212–213. [[CrossRef](#)]
6. Schlessinger, A.; Rost, B. Protein flexibility and rigidity predicted from sequence. *Proteins Struct. Funct. Bioinform.* **2005**, *61*, 115–126. [[CrossRef](#)]
7. Yuan, Z.; Zhao, J.; Wang, Z.-X. Flexibility analysis of enzyme active sites by crystallographic temperature factors. *Protein Eng. Des. Sel.* **2003**, *16*, 109–114. [[CrossRef](#)]
8. Radivojac, P.; Obtadovic, Z.; Smith, D.K.; Zhu, G.; Vucetic, S.; Brown, C.J.; David Lawson, J.; Keith Dunker, A. Protein flexibility and intrinsic disorder. *Protein Sci.* **2004**, *13*, 71–80. [[CrossRef](#)]
9. Kuczera, K.; Kuriyan, J.; Karplus, M. Temperature dependence of the structure and dynamics of myoglobin. A simulation approach. *J. Mol. Biol.* **1990**, *213*, 351–373. [[CrossRef](#)]
10. Teilum, K.; Olsen, J.G.; Kragelund, B.B. Functional aspects of protein flexibility. *Cell. Mol. Life Sci.* **2009**, *66*, 2231. [[CrossRef](#)]
11. Huber, R.; Bennett, W.S., Jr. Functional significance of flexibility in proteins. *Biopolymers* **1983**, *22*, 261–279. [[CrossRef](#)]
12. Bahar, I.; Jernigan, R.L.; Dill, K.A. *Protein Actions: Principles & Modeling*; Garland Science: New York, NY, USA, 2017.
13. Karplus, M.; Kuriyan, J. Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6679–6685. [[CrossRef](#)] [[PubMed](#)]
14. McCammon, J.A.; Gelin, B.R.; Karplus, M. Dynamics of folded proteins. *Nature* **1977**, *267*, 585–590. [[CrossRef](#)] [[PubMed](#)]
15. Hospital, A.; Goñi, J.R.; Orozco, M.; Gelpí, J.L. Molecular dynamics simulations: Advances and applications. *Adv. Appl. Bioinform. Chem.* **2015**, *8*, 37–47. [[PubMed](#)]
16. Meinhold, L.; Smith, J.C. Fluctuations and Correlations in Crystalline Protein Dynamics: A Simulation Analysis of Staphylococcal Nuclease. *Biophys. J.* **2005**, *88*, 2554–2563. [[CrossRef](#)]
17. Pang, Y.-P. Use of multiple picosecond high-mass molecular dynamics simulations to predict crystallographic B-factors of folded globular proteins. *Heliyon* **2016**, *2*, e00161. [[CrossRef](#)]
18. Go, N.; Noguti, T.; Nishikawa, T. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 3696–3700. [[CrossRef](#)]
19. Brooks, B.; Karplus, M. Harmonic dynamics of proteins: Normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 6571–6575. [[CrossRef](#)]
20. Levitt, M.; Sander, C.; Stern, P.S. Protein normal-mode dynamics: Trypsin inhibitor, crambin, ribonuclease and lysozyme. *J. Mol. Biol.* **1985**, *181*, 423–447. [[CrossRef](#)]
21. Ben-Avraham, D. Vibrational normal-mode spectrum of globular proteins. *Phys. Rev. B* **1993**, *47*, 14559–14560. [[CrossRef](#)]

22. Dykeman, E.C.; Sankey, O.F. Normal mode analysis and applications in biological physics. *J. Phys. Condens. Matter* **2010**, *22*, 423202. [[CrossRef](#)]
23. Bahar, I.; Cui, Q. *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems*; Chapman & Hall: London, UK, 2006.
24. Dehouck, Y.; Bastolla, U. Why are large conformational changes well described by harmonic normal modes? *Biophys. J.* **2021**, *120*, 5343–5354. [[CrossRef](#)] [[PubMed](#)]
25. Tirion, M.M. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* **1996**, *77*, 1905–1908. [[CrossRef](#)] [[PubMed](#)]
26. Haliloglu, T.; Bahar, I.; Erman, B. Gaussian dynamics of folded proteins. *Phys. Rev. Lett.* **1997**, *79*, 3090–3093. [[CrossRef](#)]
27. Bahar, I.; Atilgan, A.R.; Erman, B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold. Des.* **1997**, *2*, 173–181. [[CrossRef](#)]
28. Rader, A.J.; Chennubhotla, C.; Yang, L.-W.; Bahar, I. The Gaussian Network Model: Theory and Applications. In *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems*; Cui, Q., Bahar, I., Eds.; Chapman & Hall: London, UK, 2006; pp. 41–64.
29. Micheletti, C.; Carloni, P.; Maritan, A. Accurate and Efficient Description of Protein Vibrational Dynamics: Comparing Molecular Dynamics and Gaussian Models. *Proteins Struct. Funct. Genet.* **2004**, *55*, 635–645. [[CrossRef](#)] [[PubMed](#)]
30. Bahar, I.; Erman, B.; Jernigan, R.L.; Atilgan, A.R.; Covell, D.G. Collective motions in HIV-1 reverse transcriptase: Examination of flexibility and enzyme function. *J. Mol. Biol.* **1999**, *285*, 1023–1037. [[CrossRef](#)] [[PubMed](#)]
31. Bahar, I.; Jernigan, R.L. Cooperative fluctuations and subunit communication in tryptophan synthase. *Biochemistry* **1999**, *38*, 3478–3490. [[CrossRef](#)]
32. Atilgan, A.R.; Durell, S.R.; Jernigan, R.L.; Demirel, M.C.; Keskin, O.; Bahar, I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* **2001**, *80*, 505–515. [[CrossRef](#)]
33. Yang, L.; Song, G.; Jernigan, R.L. Protein elastic network models and the ranges of cooperativity. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 12347–12352. [[CrossRef](#)]
34. Eyal, E.; Yang, L.W.; Bahar, I. Anisotropic network model: Systematic evaluation and a new web interface. *Bioinformatics* **2006**, *22*, 2619–2627. [[CrossRef](#)]
35. Kim, M.H.; Lee, B.H.; Kim, M.K. Robust elastic network model: A general modeling for precise understanding of protein dynamics. *J. Struct. Biol.* **2015**, *190*, 338–347. [[CrossRef](#)] [[PubMed](#)]
36. Koehl, P.; Orland, H.; Delarue, M. Parameterizing elastic network models to capture the dynamics of proteins. *J. Comput. Chem.* **2021**, *42*, 1643–1661. [[CrossRef](#)] [[PubMed](#)]
37. Orellana, L.; Rueda, M.; Ferrer-Costa, C.; Lopez-Blanco, J.R.; Chacón, P.; Orozco, M. Approaching elastic network models to molecular dynamics flexibility. *J. Chem. Theory Comput.* **2010**, *6*, 2910–2923. [[CrossRef](#)] [[PubMed](#)]
38. Tama, F.; Sanejouand, Y.H. Conformational change of proteins arising from normal mode calculations. *Protein Eng.* **2001**, *14*, 1–6. [[CrossRef](#)] [[PubMed](#)]
39. Yang, L.; Song, G.; Jernigan, R.L. How well can we understand large-scale protein motions using normal modes of elastic network models? *Biophys. J.* **2007**, *93*, 920–929. [[CrossRef](#)]
40. Petrone, P.; Pande, V.S. Can conformational change be described by only a few normal modes? *Biophys. J.* **2006**, *90*, 1583–1593. [[CrossRef](#)] [[PubMed](#)]
41. Mahajan, S.; Sanejouand, Y.H. On the relationship between low-frequency normal modes and the large-scale conformational changes of proteins. *Arch. Biochem. Biophys.* **2015**, *567*, 59–65. [[CrossRef](#)]
42. Sanejouand, Y.-H. Normal-mode driven exploration of protein domain motions. *J. Comput. Chem.* **2021**, *42*, 2250–2257. [[CrossRef](#)]
43. Mahajan, S.; Sanejouand, Y.H. Jumping between protein conformers using normal modes. *J. Comput. Chem.* **2017**, *38*, 1622–1630. [[CrossRef](#)]
44. Zheng, W.; Doniach, S. A comparative study of motor-protein motions by using a simple elastic-network model. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 13253–13258. [[CrossRef](#)]
45. Zheng, W.; Brooks, B.R. Normal-modes-based prediction of protein conformational changes guided by distance constraints. *Biophys. J.* **2005**, *88*, 3109–3117. [[CrossRef](#)] [[PubMed](#)]
46. Dobbins, S.E.; Lesk, V.I.; Sternberg, M.J.E. Insights into protein flexibility: The relationship between normal modes and conformational change upon protein-protein docking. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 10390–10395. [[CrossRef](#)]
47. Tobi, D.; Bahar, I. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 18908–18913. [[CrossRef](#)] [[PubMed](#)]
48. Khade, P.M.; Scaramozzino, D.; Kumar, A.; Lacidogna, G.; Carpinteri, A.; Jernigan, R.L. hdANM: A new comprehensive dynamics model for protein hinges. *Biophys. J.* **2021**, *120*, 4955–4965. [[CrossRef](#)] [[PubMed](#)]
49. Kim, M.K.; Chirikjian, G.S.; Jernigan, R.L. Elastic models of conformational transitions in macromolecules. *J. Mol. Graph. Model.* **2002**, *21*, 151–160. [[CrossRef](#)]
50. Kim, M.K.; Jernigan, R.L.; Chirikjian, G.S. Efficient generation of feasible pathways for protein conformational transitions. *Biophys. J.* **2002**, *83*, 1620–1630. [[CrossRef](#)]
51. Kim, M.K.; Jernigan, R.L.; Chirikjian, G.S. Rigid-cluster models of conformational transitions in macromolecular machines and assemblies. *Biophys. J.* **2005**, *89*, 43–55. [[CrossRef](#)]

52. Maragakis, P.; Karplus, M. Large amplitude conformational change in proteins explored with a plastic network model: Adenylate kinase. *J. Mol. Biol.* **2005**, *352*, 807–822. [[CrossRef](#)]
53. Eom, K. Conformational Changes of Protein Analyzed Based on Structural Perturbation Method. *Multiscale Sci. Eng.* **2021**, *3*, 62–66. [[CrossRef](#)]
54. Orellana, L.; Yoluk, O.; Carrillo, O.; Orozco, M.; Lindahl, E. Prediction and validation of protein intermediate states from structurally rich ensembles and coarse-grained simulations. *Nat. Commun.* **2016**, *7*, 12575. [[CrossRef](#)]
55. Ikeguchi, M.; Ueno, J.; Sato, M.; Kidera, A. Protein structural change upon ligand binding: Linear response theory. *Phys. Rev. Lett.* **2005**, *94*, 078102. [[CrossRef](#)]
56. Atilgan, C.; Atilgan, A.R. Perturbation-response scanning reveals ligand entry-exit mechanisms of ferric binding protein. *PLoS Comput. Biol.* **2009**, *5*, e1000544. [[CrossRef](#)]
57. Atilgan, C.; Gerek, Z.N.; Ozkan, S.B.; Atilgan, A.R. Manipulation of conformational change in proteins by single-residue perturbations. *Biophys. J.* **2010**, *99*, 933–943. [[CrossRef](#)] [[PubMed](#)]
58. Gerek, Z.N.; Ozkan, S.B. Change in allosteric network affects binding affinities of PDZ domains: Analysis through perturbation response scanning. *PLoS Comput. Biol.* **2011**, *7*, e1002154. [[CrossRef](#)] [[PubMed](#)]
59. Liu, J.; Sankar, K.; Wang, Y.; Jia, K.; Jernigan, R.L. Directional Force Originating from ATP Hydrolysis Drives the GroEL Conformational Change. *Biophys. J.* **2017**, *112*, 1561–1570. [[CrossRef](#)] [[PubMed](#)]
60. Scaramozzino, D.; Piana, G.; Lacidogna, G.; Carpinteri, A. Low-Frequency Harmonic Perturbations Drive Protein Conformational Changes. *Int. J. Mol. Sci.* **2021**, *22*, 10501. [[CrossRef](#)]
61. Eyal, E.; Bahar, I. Toward a molecular understanding of the anisotropic response of proteins to external forces: Insights from elastic network models. *Biophys. J.* **2008**, *94*, 3424–3435. [[CrossRef](#)]
62. Scaramozzino, D.; Khade, P.M.; Jernigan, R.L.; Lacidogna, G.; Carpinteri, A. Structural Compliance: A New Metric for Protein Flexibility. *Proteins Struct. Funct. Bioinform.* **2020**, *88*, 1482–1492. [[CrossRef](#)]
63. Sen, T.Z.; Feng, Y.; Garcia, J.V.; Kloczkowski, A.; Jernigan, R.L. The extent of cooperativity of protein motions observed with elastic network models is similar for atomic and coarser-grained models. *J. Chem. Theory Comput.* **2006**, *2*, 696–704. [[CrossRef](#)]
64. Frey, M. Water structure associated with proteins and its role in crystallization. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **1994**, *50*, 663–666. [[CrossRef](#)]
65. Bhat, T.N.; Bentley, G.A.; Boulot, G.; Greene, M.I.; Tello, D.; Dall’Acqua, W.; Souchon, H.; Schwarz, F.P.; Mariuzza, R.A.; Poljak, R.J. Bound water molecules and conformational stabilization help mediate an antigen-antibody association. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 1089–1093. [[CrossRef](#)] [[PubMed](#)]
66. Hayward, S.; Kitao, A.; Hirata, F.; Go, N. Effect of solvent on collective motions in globular protein. *J. Mol. Biol.* **1993**, *234*, 1207–1217. [[CrossRef](#)] [[PubMed](#)]
67. Chandler, D. Interfaces and the driving force of hydrophobic assembly. *Nature* **2005**, *437*, 640–647. [[CrossRef](#)] [[PubMed](#)]
68. Nakasako, M. Large-scale networks of hydration water molecules around bovine  $\beta$ -trypsin revealed by cryogenic X-ray crystal structure analysis. *J. Mol. Biol.* **1999**, *289*, 547–564. [[CrossRef](#)] [[PubMed](#)]
69. Lins, L.; Thomas, A.; Brasseur, R. Analysis of accessible surface of residues in proteins. *Protein Sci.* **2003**, *12*, 1406–1417. [[CrossRef](#)]
70. Prabhu, N.; Sharp, K. Protein-solvent interactions. *Chem. Rev.* **2008**, *106*, 1616–1623. [[CrossRef](#)]
71. Brysbaert, G.; Blossey, R.; Lensink, M.F. The inclusion of water molecules in residue interaction networks identifies additional central residues. *Front. Mol. Biosci.* **2018**, *5*, 88. [[CrossRef](#)]
72. Horvath, I.; Jeszenoi, N.; Balint, M.; Paragi, G.; Hetenyi, C. A fragmenting protocol with explicit hydration for calculation of binding enthalpies of target-ligand complexes at a quantum mechanical level. *Int. J. Mol. Sci.* **2019**, *20*, 4384. [[CrossRef](#)]
73. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [[CrossRef](#)]
74. Eyal, E.; Lum, G.; Bahar, I. The anisotropic network model web server at 2015 (ANM 2.0). *Bioinformatics* **2015**, *31*, 1487–1489. [[CrossRef](#)]
75. Scaramozzino, D.; Lacidogna, G.; Piana, G.; Carpinteri, A. A finite-element-based coarse-grained model for global protein vibration. *Meccanica* **2019**, *54*, 1927–1940. [[CrossRef](#)]
76. Giordani, G.; Scaramozzino, D.; Iturrioz, I.; Lacidogna, G.; Carpinteri, A. Modal analysis of the lysozyme protein considering all-atom and coarse-grained finite element models. *Appl. Sci.* **2021**, *11*, 547. [[CrossRef](#)]
77. Khade, P.M.; Kumar, A.; Jernigan, R.L. Characterizing and Predicting Protein Hinges for Mechanistic Insight. *J. Mol. Biol.* **2020**, *432*, 508–522. [[CrossRef](#)] [[PubMed](#)]
78. Kurkcuoglu, O.; Jernigan, R.L.; Doruker, P. Mixed levels of coarse-graining of large proteins using elastic network model succeeds in extracting the slowest motions. *Polymer* **2004**, *45*, 649–657. [[CrossRef](#)]
79. Tsai, J.; Taylor, R.; Chothia, C.; Gerstein, M. The packing density in proteins: Standard radii and volumes. *J. Mol. Biol.* **1999**, *290*, 253–266. [[CrossRef](#)]