RESEARCH ARTICLE

Bridging between material properties of proteins and the underlying molecular interactions

Guang Song[1,2]*

1 Department of Computer Science, Iowa State University, Ames, IA, United States of America, **2** Program of Bioinformatics and Computational Biology, Iowa State University, Ames, IA, United States of America

* gsong@iastate.edu

Abstract

In this work, we develop a novel method that bridges between material properties of proteins, particularly the modulus of elasticity, and the underlying molecular interactions. To this end, we employ both an all-atom normal mode analysis (NMA) model with the CHARMM force field and an elastic solid model for proteins and protein interfaces. And the "bridge" between the two models is a common physical property predictable by both models: the magnitude of thermal vibrations. This connection allows one to calibrate the Young's moduli of proteins and protein interface regions. We find that the Young's moduli of proteins are in the range of a few Gpa to 10 Gpa, while the Young's moduli of the interface regions are several times smaller. The work is significant as it represents the first attempt to systematically compute the elastic moduli of proteins from molecular interactions.

1 Introduction

Solid is one of the fundamental states of matter. Atoms in solids are packed tightly and kept together by physical interactions, such as ionic bonds (as in sodium chloride), covalent bonds (as in diamond), metalic bonds (as in metals), hydrogen bonds (as in ice), or van der Waals interactions (as in organic compounds) [1]. Solids held together mostly by van der Waals interactions are called van der Waals solids.

The type of force that keeps a solid together determines its material properties, such as elasticity. Solids held together by metalic, ionic, or covalent bonds appear stiff and have a much higher elastic moduli than solids held together mostly by non-bonded interactions such as proteins [2].

The elasticity of proteins has been studied in three primary ways, namely, protein fibers or fibrils [3], protein crystals [4], and protein capsids [5].

Among protein fibers, some exhibit high extensibility and have a very low Young's modulus, on the order of only 1–10 MPa. These fibers usually contain unstructured regions or high mobility motifs, undergo α-helix to β-strand transitions, or even unfold under external forces, and consequently they appear very soft (1-10 Mpa) [3]. Other protein fibers are significantly different as they "form highly regular, nearly crystalline arrangement of monomer units

without regions that can extend, change secondary structure or unfold" [3] and thus appear much stiffer. Experimental studies showed these fibers, such as actin [6], tubulin, collagen [2], and keratin [7], fibrin [3], β-lactoglobulin fibrils [8], had an elastic modulus in the range of a few Gpa. The Young's moduli of these fiber proteins were obtained through direct measurement of force under stretching, mostly using atomic force spectroscopy (AFM) [8, 9].

The mechanical properties of proteins were studied also using protein crystals, mostly the crystals of hen egg-white lysozyme (HEWL), a model protein. Table 1 summarizes the techniques used, proteins studied, and results obtained. It lists also results from the compressibility studies done in solvent [10] and on a crystal structure solved at high pressure (1,000 bar) [11]. The results show a large variation, from less than 1 Gpa to over 10 Gpa. Compressibility study by Gavish et al. [10] showed that under most conditions, the adiabatic compressibilities of lysozyme, hemoglobin, and myoglobin in solution were on the order of 10-20% of the compressibility of water, or 5-7 $\times 10^{-11}$ m^2/N. Consequently, the bulk moduli of these proteins, which are the inverse of their compressibilities, are over 10 Gpa. (The bulk modulus or volume modulus should not be confused with Young's modulus or tensile modulus, though they tend to have similar values for most solids.) Kundrot and Richards [11] found that the isothermal compressibility of lysozyme proteins was 4.7 10^{-3} kbar^{-1}, or 20 Gpa in bulk modulus. On the other hand, the Young's moduli obtained from protein crystals using techniques such as vibration, indentation, ultrasound, or Brillouin scattering are much smaller, from 0.3 Gpa to 5.5 Gpa (see Table 1). How to understand this discrepancy? It is possible that the larger values obtained from compressibility studies represent the elastic modulus of proteins themselves, while the elastic moduli of protein crystals probably represent the moduli of both proteins and protein interface regions, as well as other noncrystalline elements in the crystal cell such as intracrystalline liquid [4]. This would reconcile the seemingly conflicting results on elastic moduli of proteins in the literature.

The Young's moduli of proteins measured experimentally as reviewed above depend on a number of factors, including conditions of the samples (such as pH in solvent [10] or water content in crystals: moistened or dried [15], etc.), and the frequency range at which a measurement was carried out [17], which varies from static to kilohertz [12–14], to ultrasonic in the megahertz range [16, 17] and Brillouin light scattering method in the gigahertz range [18]. The elastic modulus of a protein is also temperature dependent. Morozov and Gevorkian analyzed that the mechanical properties of protein crystals at different temperatures and found that proteins became significantly more rigid below glass transition temperature, when the surface layer of proteins and their bound water became immobilized [4]. In this work, we also will look into the influence of temperature on a protein's elastic modulus.

The elastic modulus of proteins was studied also through protein capsids by nano-indentation using atomic force microscopy (AFM). Nano-indentation allows one to measure the stiffness of viral capsids. The elastic modulus of the capsid material can then be deduced from the

Table 1. A summary of studies on the elasticity of proteins using protein crystals (top half), or compressibility measurements (bottom half).

Techniques	Proteins	Young's modulus	References
vibrating reed	lysozyme	0.3-1.5 Gpa	[12–14]
indentation	lysozyme	0.49-4.2 Gpa	[15]
ultrasound	lysozyme	4.87-5.5 Gpa	[16, 17]
brillouin scattering	lysozyme	5.5 Gpa	[18]
		bulk modulus	
ultrasound in solution	lysozyme, Mb, Hb	>10 Gpa	[10]
X-ray at high pressure	lysozyme	>20 Gpa	[11]

https://doi.org/10.1371/journal.pone.0247147.t001

measured stiffness. However, in order to do so, in most cases a thin-shell model was assumed and thin-shell elasticity theory applied. The limitation of the thin shell model is that it neglects the molecular structure of the capsid and assumes a homogeneous material property throughout, as well as an idealized spherical geometry and uniform thickness [5]. Thick shell models combined with finite element analysis were also tried [19] and the effect of non-uniform geometry was investigated [20]. These work all represent a top-down approach since experimental determined stiffness values were used in all cases to fit the underlying elastic moduli of capsids. A review article in 2012 by Mateu [5] summarized the nano-indentation results of a dozen viral capsids: the stiffness of the viral capsids and the derived Young's moduli. The review showed the Young's moduli of capsid proteins varied over a span of more than one order of magnitude, from 0.14 GPa for CCMV to 1-3 GPa for MVM and Φ29 [5]. The range of variation in Young's moduli seen in capsids is consistent with what is seen in protein crystals (Table 1, top half) but is much lower overall than what is obtained from compressibility studies (Table 1, bottom half), A possible explanation is that elastic moduli thus obtained represent the average moduli of both proteins and protein interface regions, while the elastic moduli obtained from compressibility studies represent those of individual proteins.

The Young's moduli of globular and membrane proteins were also estimated theoretically or computationally [2, 6, 21, 22]. By approximating proteins as solid materials with sheets that interacted through van der Waals interactions, Howard estimated that the Young's modulus of proteins to be around 4 Gpa (see Appendix 3.1 of [2]). In the case of F-actin, Bathe applied axial stretching of his elastic solid model for actin to match experimental stretching stiffness data [6] and found that the effective Young's modulus of actin was 2.7 Gpa [21].

Contribution of this work. In this work, we develop a novel bottom-up approach for computing the elastic moduli of proteins from the underlying molecular interactions. Our approach utilizes both normal mode analysis (NMA) [23–25], a well established technique for studying the fluctuating dynamics of macromolecules, and elastic solid models (ESM) [26] developed more recently that treat macromolecules as elastic solids with material properties such as Young's modulus. Since both of these models can predict the magnitude of thermal vibrations of macromolecules, this commonality is used as a bridge to link material properties modeled in ESM [26] with molecular interactions used in NMA. To the best of our knowledge, This work represents the first attempt to determine the Young's moduli of proteins and protein interface regions separately and systematically from all-atom molecular interactions. We find the Young's modulus of proteins can be as high as 10 Gpa, while the Young's modulus of protein interface regions is several times smaller. Our work reconciles the high modulus values found through compressibility studies and low values found in protein crystals or capsids. The large span of variations of elastic moduli at interface regions provides an explanation also for the similar extent of variations seen in protein crystals and protein capsids.

2 Methods

2.1 σESM, a molecular surface-based elastic solid model

Recently, we presented a novel elastic solid model called αESM[26] for macromolecules based on alpha shape [27]. The model has a parameter alpha which was chosen empirically.

Here we present a new extension of αESM whose major improvement over αESM is that the alpha value is now tuned individually for each protein so that the final elastic solid model of the protein has the same volume as its solvent excluded volume, such as that computed by the MSMS algorithm [28]. Another significant difference of the present ESM model from αESM is that in representing the solid, not only the atomic coordinates of the protein are used, which is the case for αESM, but also selected points on the protein's molecular surface [21].

```
 1 function [K,M,shp] = sigmaESM(xyz, verts, faces, volume, mass, E)
 2 % simgaESM: ESM based on molecular surface
 3 % Author: Guang Song
 4 % K, M: stiffness and mass matrices of the elastic solid
 5 % xyz: atomic coordinates of the input structure
 6 % verts, faces, volume: vertices and faces of MSMS surface and its enclosed volume
 7 % mass: mass of atoms of the input structure
 8 % E: Young's modulus
 9 if nargin<6 E = 1.0; end
10 init_alpha = 3.6; % 2.1 atom radius + 1.5 (probe radius)
11 shp0 = alphaShape(xyz, init_alpha);
12 [pface] = boundaryFacets(shp0);
13 m = size(pface,1); % m: # of faces on the initial alpha shape
14 % simplify the molecular surface to have m faces
15 [rF, rV] = reducepatch(faces, verts, m);
16 xyz = [xyz; rV]; % append reduced surface vertices rV to xyz
17 % binary search for alpha so that the alpha shape has the given volume
18 alpha = findAlpha(xyz, volume);
19 [K, M, shp] = alphaESM(xyz, alpha, mass, E);
```

Fig 1. The algorithmic flow of *σESM* as expressed in a MATLAB script. A complete copy of the above MATLAB script is available at https://github.com/gsongISU/sigmaESMrelease. The scripts for computing stiffness matrix and mass matrix used in *alphaESM* are available at MATLAB file exchange (https://www.mathworks.com/matlabcentral/fileexchange/27826-fast-fem-assembly-nodal-elements), kindly contributed by Anjam and Valdman [30].

https://doi.org/10.1371/journal.pone.0247147.g001

Consequently, the final solid model has a similar surface to the molecular surface of the protein. In this aspect the model is similar to Bathe's elastic solid model [21]. We name this new variant of ESM *σESM*, since *σ* is a symbol commonly used for surface area or surface tension [29].

The procedural flow of *σESM* is given in Fig 1. The script takes as input *xyz*, which contains the coordinates of all the heavy atoms of a given protein structure that is available in its PDB file, and *verts*, *faces*, and *volume*, all three of which are outputs of the MSMS software [28] and represent respectively the vertices and faces of the solvent excluded surface mesh and its enclosed volume (line 1). The program first computes an initial alpha shape of the protein alone (line 11), which is used to set the number of faces that should be on the surface mesh (lines 12-13). It then simplifies the surface mesh produced by MSMS using MATLAB script *reducepatch* (line 15). The vertices on the reduced molecular surface are then appended to the coordinates of the atoms in the protein (line 16). Lastly, an alpha value is selected (line 18) so that the final alpha shape has the same volume as the the volume computed from MSMS [28]. Once the alpha value is determined, the script used for *αESM*[26] can be adopted to compute the stiffness matrix **K** and mass matrix **M** (line 19), from which eigenmodes and eigen frequencies are computed (Fig 1 in Ref. [26]).

Fig 2 shows the *αESM* model of a protein (pdb-id: 1aqb) in red. The molecular surface, or solvent excluded surface, computed from the MSMS software [28] is shown in blue as a triangular surface mesh. The total volume of the *αESM* solid model (shown in red) is 16,600 Å^3, while the solvent excluded volume is 23,720 Å^3. Fig 2(B) shows the *σESM* model in cyan, which has an extra layer over the *αESM* model and resembles closely the molecular surface in
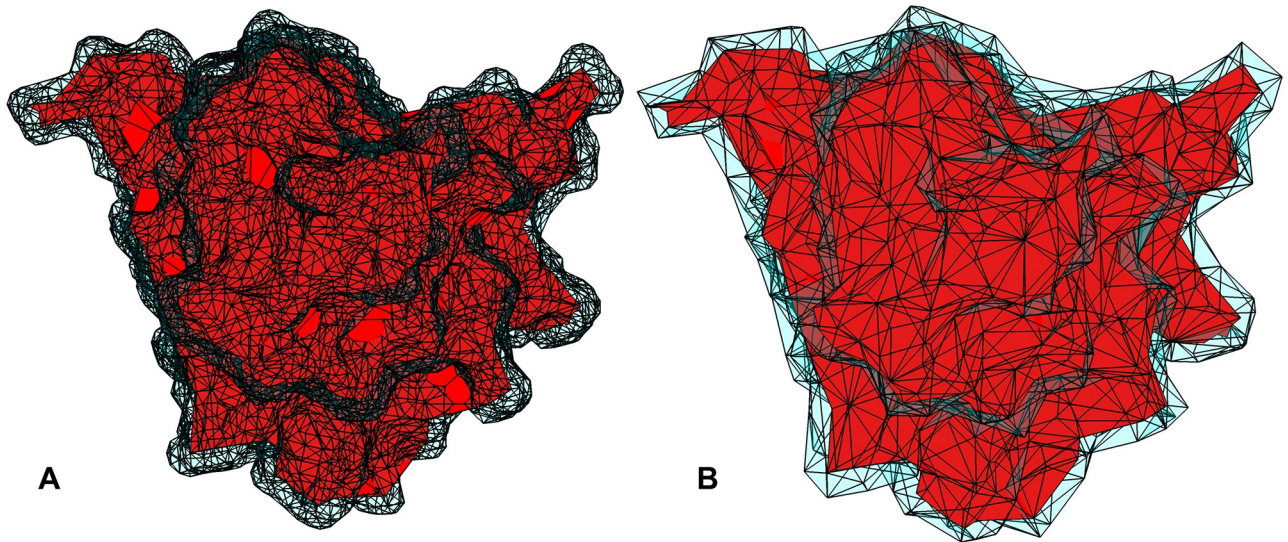
Fig 2. The *αESM* model [26] (in red) using an alpha value of 3.6 Å. The solvent excluded surface computed from MSMS [28] is shown in blue as a triangle mesh. (B) The *σESM* model: the difference between it and *αESM* (in red) is shown in cyan.

Fig 2(A). Specifically, the volume of *σESM* model is 23,770 Å^3, nearly the same as the protein's solvent excluded volume.

In summary, *σESM* is a more accurate representation than *αESM* in describing both molecular surface and and its enclosing volume.

The magnitude of thermal vibrations by *σESM*. Given a protein structure and its molecular surface mesh (which can be generated for instance by the MSMS software [28]), an application of the *σESM* script given in Fig 1 produces a stiffness (**K**) and a mass (**M**) matrix. The (global) stiffness matrix **K** is a 3*M*-by-3*M* matrix, where *M* is the number of nodes in the model, built with the element stiffness matrices of all the elements (tetrahedral in the current work). Each element stiffness matrix is a 12x12 matrix that can be written as:

$$\mathbf{k} = [\mathbf{B}]^T [\mathbf{D}][\mathbf{B}] V, \tag{1}$$

where **B** and **D** are standard matrices available in textbooks on finite element method [31] that depend solely on the coordinates of the four nodes of the tetrahedral element, and *V* is the volume of the tetrahedron. The unit of **K** is kcal/mol/Å^2. For the mass matrix **M** used in finite element method, one has the option of a consistent mass matrix or a lump-sum mass matrix [31]. In this work, the lump-sum mass matrix is used, which is the same as the mass matrix used in NMA and is diagonal. (The consistent mass matrix, on the other hand, is not a diagonal matrix.) According to Ref. [26], the mean-square thermal vibrations of atom *i* by ESM is:

$$\langle \mathbf{\Delta R}_i^2 \rangle_{\text{ESM}} = \frac{k_B T}{E} trace\left(\left[\mathbf{M}^{-\frac{1}{2}}\mathbf{K}_m^{-1}\mathbf{M}^{-\frac{1}{2}}\right]_{ii}\right), \tag{2}$$

where the subscript *ii* represents the ith 3-by-3 diagonal block and $\mathbf{K}_m$ is the mass-weighted stiffness matrix, i.e., $\mathbf{K}_m = \mathbf{M}^{-1/2} \mathbf{K} \mathbf{M}^{-1/2}$. *E* is the Young's modulus of the protein solid.

The magnitude of thermal vibrations of a whole protein is defined as the *mass-weighted average* of $\langle \mathbf{\Delta R}_i^2 \rangle_{\text{ESM}}$ over all the nodes [32], i.e.,

$$\langle \mathbf{\Delta R}^2 \rangle_{\text{ESM}} = \frac{\sum_i m_i \langle \mathbf{\Delta R}_i^2 \rangle_{\text{ESM}}}{\sum_i m_i} = \frac{k_B T trace(\mathbf{K}_m^{-1})}{E \sum_i m_i}, \tag{3}$$

where i is the index of the nodes that include both protein atoms and added molecular surface nodes. m_i are the mass of node i. Each surface node is given a mass of 1 atomic unit.

2.2 The sbNMA model

The sbNMA model developed by Na and Song [33] is an all-atom NMA model based on the CHARMM force field [34]. It was designed to maintain the high accuracy of classical NMA (cNMA) using all-atom force fields. By circumventing the cumbersome step of energy minimization, it can be applied directly to experimental structures. Extensive studies have demonstrated its accuracy [32, 33, 35–37]. The entire sbNMA code is publicly available at https://github.com/htna/sbNMA-Matlab. Given a protein structure, an application of the sbNMA code (particularly the sbNMA_PSF.m script available at the above link) produces a Hessian matrix **H** and a mass matrix **M** that is diagonal (the sbNMA_PSF script actually produces the diagonal of the mass matrix as a vector, which can be easily converted to a mass matrix when needed). The Hessian matrix **H** is a 3N-by-3N matrix containing the second partial derivatives of the potential energy function with respect to the (x,y,z) coordinates of the N atoms in a protein. The unit of the Hessian matrix is kcal/mol/Å^2.

Once the Hessian matrix **H** and the mass matrix **M** are obtained, the mean square fluctuation of atom i is [33]:

$$\langle \Delta \mathbf{R}_i^2 \rangle_{\text{NMA}} = k_B T trace\left(\left[\mathbf{M}^{-\frac{1}{2}} \mathbf{H}_m^{-1} \mathbf{M}^{-\frac{1}{2}} \right]_{ii} \right),$$

(4)

Thus,

$$\langle \Delta \mathbf{R}^2 \rangle_{\text{NMA}} = \frac{\sum_i m_i \langle \Delta \mathbf{R}_i^2 \rangle_{\text{NMA}}}{\sum_i m_i} = \frac{k_B T trace(\mathbf{H}_m^{-1})}{\sum_i m_i}$$

(5)

As will be shown later, the magnitude of thermal vibrations as computed from sbNMA in Eq (5) can be used to calibrate the Young's moduli of proteins.

2.3 B-factors, static disorder, and glass transition

The Debye–Waller factor (DWF), or B-factor as it is called in protein X-ray crystallography, is a factor used to describe for each atom the degree to which electron density spreads out. The spread around the mean position of each atom i is generally modeled as a Gaussian function and the magnitude of the spread is characterized by $\langle u_i^2 \rangle$, the mean-square displacement from the mean position. The isotropic Debye-Waller factor, or B-factor, is related to $\langle u_i^2 \rangle$ as:

$$\langle u_i^2 \rangle = \frac{3}{8\pi^2} B_i.$$

(6)

The magnitude of mean-square displacement of a whole protein is defined as the *mass-weighted* averages of $\langle u_i^2 \rangle$ over all the atoms [32], i.e.,

$$\langle u^2 \rangle = \frac{\sum_i m_i \langle u_i^2 \rangle}{\sum_i m_i},$$

(7)

where i is the atom index.

2.4 Determine the Young's moduli of proteins based on molecular interactions

sbNMA is an accurate NMA model for predicting the magnitude of thermal vibrations. However, like all NMA models and elastic network models, one major limitation of sbNMA is that it assumes harmonicity. Thus the predicted magnitude by sbNMA represents only a *lower bound* of the actual magnitude of fluctuations of proteins that may contain also anharmonic motions, especially at above T_g, the temperature of glass transition [38].

The magnitude of thermal vibrations can be obtained also experimentally. X-ray crystallography that is widely used in protein structure determination produces such magnitude information in atomic displacement parameters that are commonly known as B-factors. Since B-factors contain also a significant amount of static disorder [32], the magnitude of mean-square displacement obtained from B-factors thus represents an *upper bound* of the actual magnitude of fluctuations of proteins.

A possible third alternative for obtaining the magnitude of thermal fluctuations of proteins is molecular dynamics (MD) simulation [39, 40], which is able to take into account anharmonic motions and is unaffected by static disorder. However, it suffers from the problem of insufficient sampling of the conformation space. Consequently, the actual magnitude of thermal fluctuations is difficult to obtain using MD.

The Young's modulus represents a material property of a solid. By representing a protein using both a molecular model (as in NMA) and an elastic solid model as in *σESM*, we can link material properties with molecular interactions. Particularly, we can deduce a protein's Young's modulus by requiring the magnitude of thermal vibrations predicted by *σESM* in Eq (3) to be the same as the magnitude obtained from sbNMA in Eq (5) or the magnitude of mean-square displacement from B-factors in Eq (7). Since the magnitudes obtained from sbNMA and B-factors represent the lower and upper bounds of the actual magnitude respectively, the Young's moduli thus deduced represent the upper and lower bounds of the actual Young's moduli. Specifically, by requiring $\langle \Delta \mathbf{R}^2 \rangle_{\mathrm{ESM}} = \langle \Delta \mathbf{R}^2 \rangle_{\mathrm{NMA}}$ (Eqs (3) and (5)), we have:

$$E_{\mathrm{upper}} = \frac{trace(\mathbf{K}_m^{-1})/m_{tot}^{\mathrm{ESM}}}{trace(\mathbf{H}_m^{-1})/m_{tot}^{\mathrm{NMA}}}, \tag{8}$$

where m_{tot}^{ESM} and m_{tot}^{NMA} are the total masses in *σESM* and NMA, respectively. And similarly, by requiring $\langle \Delta \mathbf{R}^2 \rangle_{\mathrm{ESM}} = \langle u^2 \rangle$ (Eqs (3) and (7)), we have:

$$E_{\mathrm{lower}} = \frac{trace(\mathbf{K}_m^{-1})/m_{tot}^{\mathrm{ESM}}}{(\sum_i m_i \langle u_i^2 \rangle)/m_{tot}^{Bfac}}, \tag{9}$$

where m_{tot}^{Bfac} is the total mass of all the atoms with B-factors, which usually are the heavy atoms.

2.5 Quaternary structures based on crystal contacts

For each protein chain p in the dataset, we determine the all the other chains in its unit cell based on the space group and symmetry matrices given in its PDB file. Additionally we find all the neighboring cells to the current cell in all directions: altogether 27-1 = 26 neighboring cells are considered. Of all the chains in these cells, we identify the one that is the closest to p and denote this chain as q. p and q form a crystal contact and a quaternary structure. (In the case when the selected q has a very small contact surface with p, a different chain q is used instead.) The quaternary structures are saved in PDB files as xxxx2.pdb, where xxxx is the pdb-id, and are available at https://github.com/gsongISU/sigmaESMrelease, under folder *pdbDataset*. In

the following, we explain how to compute the Young's modulus of the interface region of a quaternary structure.

2.6 *σESM* of the quaternary structures

σESM is readily applicable to the aforementioned quaternary structures in the same fashion as to regular tertiary structures, and as a result, an elastic solid model is produced for each quaternary structure (see Fig 3 of structure 2pwa2.pdb). The nodes in the model consists of heavy atoms of two protein chains (regions in red) in the quaternary structure and the surface nodes (in cyan) given by the MSMS software [28]. The elements in the solid model are tetrahedra generated by alpha shape [27] (see Fig 1). Given the elastic solid model of a quaternary structure, the interface region, which is in blue in Fig 3, is identified as follows.

First, all the tetrahedral elements formed by atoms from chain *p* or surface nodes but not from chain *q* are identified. Let all the nodes of these elements be *nodes*1. *nodes*1 consists of all the atoms of chain *p* and some surface nodes, but no atoms from chain *q*. Similarly, compute *nodes*2 for the second chain *q*. Next, remove any common surface nodes shared by *nodes*1 and *nodes*2 from them. As a result, the intersection of *nodes*1 and *nodes*2 is an empty set. Finally, identify all the tetrahedral elements formed solely by *nodes*1 or *nodes*2. The remaining elements are identified as the interface region (in blue).

2.7 Computing the Young's modulus of the interface region

Next, we compute the Young's modulus of the interface region of the quaternary structure. This allows us to assess how "soft" or "firm" the region is. Again, our method to determine the material property of the interface region is to link it with the underlying molecular interactions using *σESM* and sbNMA, by requiring the two models to give the same magnitude of thermal vibrations.

Fig 3(B) shows the molecular structure model of the same quaternary structure after protonation using psfgen from VMD [42]. We apply sbNMA to compute the magnitude of thermal vibration using Eq (5). To focus on the interface region and to speed up the computations, we set the two proteins as rigid, i.e., the internal degrees of freedom within each protein chain is left out, by employing a projection matrix as was done in the RTB model [43]. The magnitude of thermal vibration of the whole structure (Fig 3(B)) is thus dictated solely by inter-protein interactions. Similarly, when applying *σESM* to the solid model in Fig 3(A), the two protein regions, which are in red, are set as rigid and only the interface region (in blue) is flexible. By requiring the two models to give the same magnitude of thermal vibration, we can determine the Young's modulus of the interface region, similar to what is done in Eq (8). The entire program for computing the Young's modulus of the interface region is given at https://github.com/gsongISU/sigmaESMrelease. The program uses *σESM* and sbNMA and a matrix projection module. All of them are available at the above website.

It is worth noting that only non-bonded interactions (possibly also disulfide bonds) are present at the interface region. Therefore it is expected that the Young's modulus of the interface region should be smaller than that of a protein itself, since the latter contains also bonded interactions. Also, quaternary structures constructed from crystal contacts as described above, especially after protonation, may contain steric clashing in the interface region. For atom pairs that are too close to each other (i.e., closer than their equilibrium separation distance which is the sum of their van der Waals radii), we use their equilibrium separation distances as their separation distances when computing the potential energy at the interface.
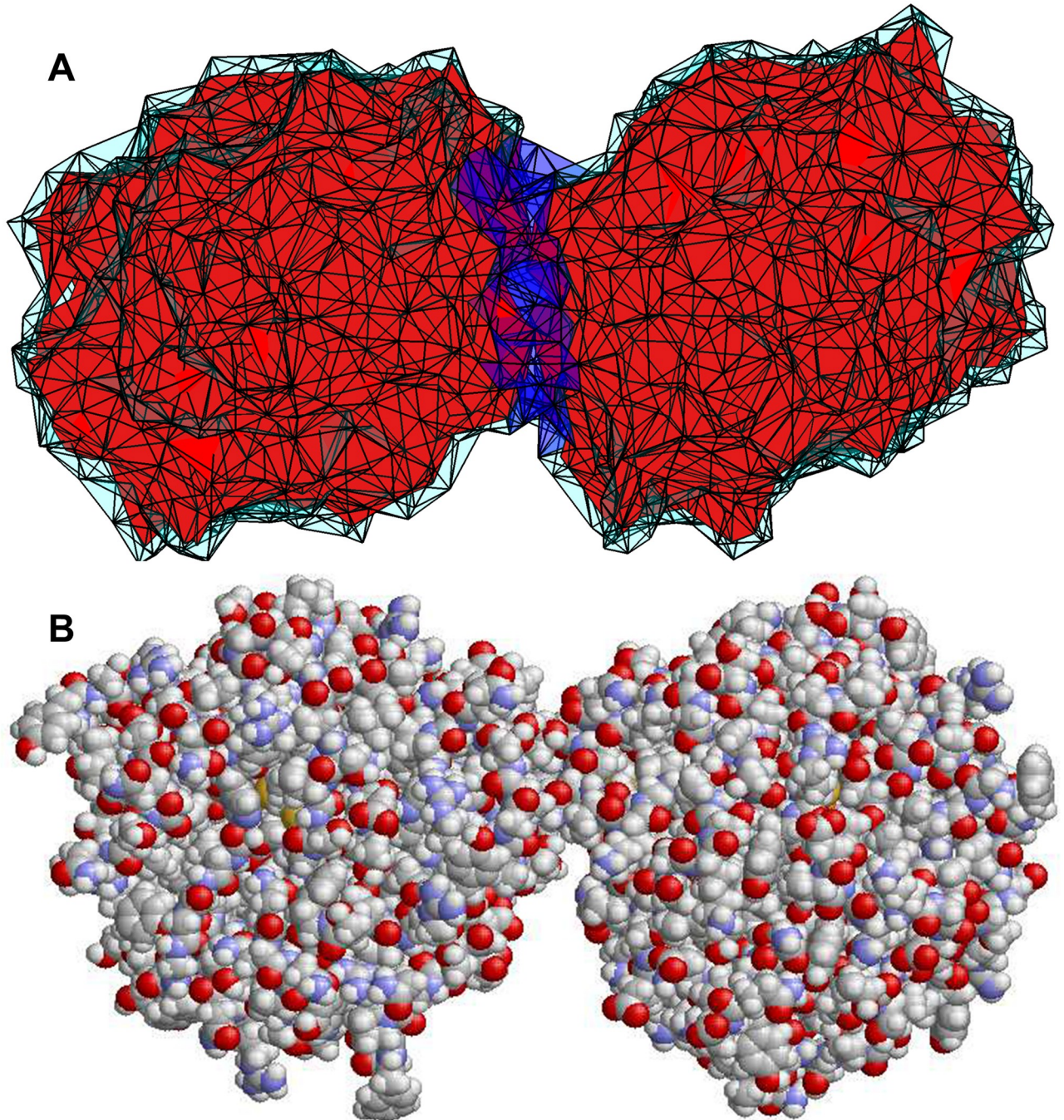
Fig 3. Models of a quaternary structure. (A) An elastic solid model of a quaternary structure (2pwa2.pdb). The pdb quaternary structure is available at https://github.com/gsongISU/sigmaESMrelease, under folder *pdbDataset*. (B) the all-atomic molecular model of the same structure. The image is generated by Rasmol [41].

https://doi.org/10.1371/journal.pone.0247147.g003

3 Results

3.1 The choice of protein dataset

In our recent work [32], we have shown that the percentage of thermal vibrations in B-factors is small in most proteins but is significantly large (50% or higher) for structures determined at

both high resolution (1.2 Å or higher) and high temperature (that is, room temperature). Unfortunately such structures are rare in the PDB, as most structures in the PDB [44] are determined at cryogenic temperature to reduce the radiation damage [45].

Based on this premise, we select a group of 18 structures determined at high resolution (1.1 Å or high) and at room temperature. This set of structures are taken from a list of 1522 structures used in our previous work [32] that was originally generated by PDB_SELECT [46, 47].

3.2 Upper bound and lower bound of thermal fluctuations in proteins

There are a few universal properties of globular proteins that are intriguing. First, it was established that the vibrational spectra of globular proteins, once properly normalized, follow a universal curve [35, 48]. It was shown later that this universality in vibrational spectrum held true also for protein capsids [36]. Second, it was shown in one of our recent work that a direct corollary of the universality in vibrational spectrum among globular proteins is that the magnitudes of their thermal vibrations are nearly universal, having a narrow distribution that peaks at 0.093 Å^2 at 100 K [32]. In obtaining this value, a harmonic potential is assumed. The harmonicity assumption is reasonable for thermal vibrations at or below 100 K. At higher temperature, particularly at above the glass transition temperature of proteins which is about 200-220 K, proteins undergo "glass transition" and become much more flexible [38]. As a result, the magnitude of mean-square displacements grows super-linearly. Therefore, the magnitude of thermal vibrations of globular proteins at 300 K is at least three times as high as 0.093 Å^2, or 0.279 Å^2. In other words, the magnitude of thermal vibrations computed from sbNMA (which assumes harmonicity) represents a lower bound of the actual magnitude at room temperature.

On the other hand, the magnitudes of thermal fluctuations of globular proteins are available in crystallographic B-factors, though in which unfortunately a large amount of static disorder co-resides. Because of static disorder, B-factors do not represent the actual magnitudes of thermal fluctuations but rather an upper bound.

The aforementioned upper and lower bounds define a range of the actual magnitudes of thermal vibrations of globular proteins. The range, though not as ideal as a single definite value, gives us a good sense of the magnitude.

Fig 4 shows the upper and lower bounds of the magnitudes of thermal fluctuations of the 18 structures in the dataset. For three out of the 18 structures, the lower bound obtained from the sbNMA computation is slightly higher than the upper bound obtained from B-factors. This could be due to the imperfection of sbNMA or uncertainties in B-factors, or both. However, for most proteins, the over trend is that the upper bound is about 1.6-1.7 times higher than the lower bound. The mean lower and upper bounds of the proteins shown in Fig 4 are 0.30 and 0.51 Å^2, respectively.

3.3 Glass transition

It is known that proteins undergo glass transition at T_g, the glass transition temperature, which is about 200-220 K for proteins [38]. At above T_g, proteins become much more flexible. Past work indicated that "the glass transition in hydrated samples is located in the surface layer of proteins and related to the (im)mobilization of the protein groups and strongly bound water." [4]. As a result, the magnitude of mean-square displacements grows super-linearly with temperature.

Fig 5 presents the magnitudes of thermal vibrations as computed from sbNMA [32, 33] of the 18 proteins in the dataset, as well as their mean-square displacements from crystallographic B-factors. The blue line represents the linear growth of mean-square displacements as a
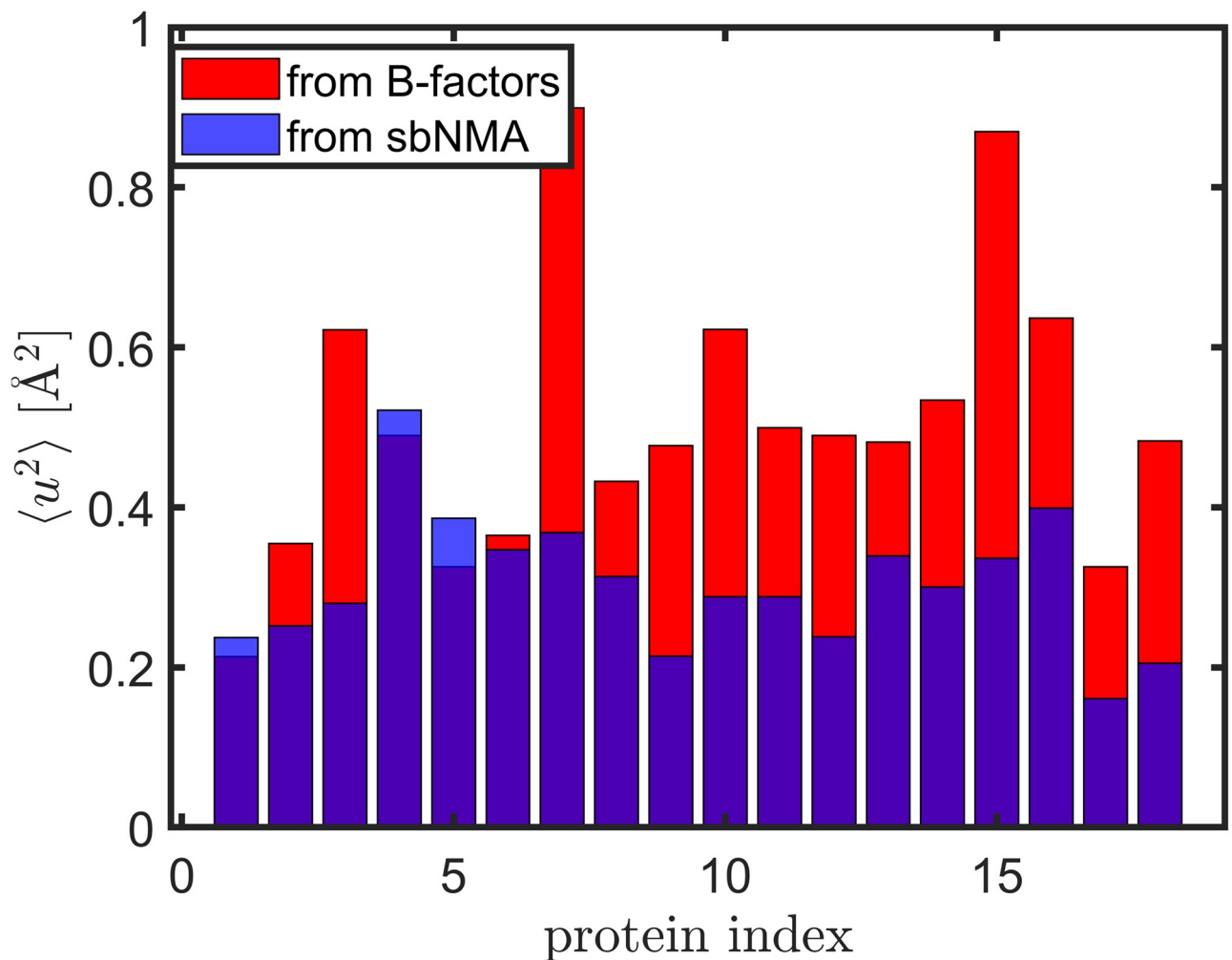
Fig 4. The magnitudes of thermal fluctuations obtained from B-factors (in red) and from sbNMA computations (in blue), which form the upper and lower bounds of the magnitudes of thermal fluctuations of the 18 structures in the dataset, listed by their sizes in the ascending order. For three out of the 18 structures, the lower bound is slightly higher than the upper bound.

function of temperature should there be no glass transition. The red line in Fig 5 represents a fitting to $\langle u^2 \rangle$ (of B-factors). A clear picture of glass transition emerges.

3.4 The range of Young's modulus

Fig 6 shows the range of Young's moduli computed from σESM and calibrated with sbNMA or B-factors. The median Young's modulus when fitting to B-factors is 6.03 Gpa, and 10.6 Gpa when fitting to sbNMA. sbNMA does not consider glass transition but only harmonic motions. Thus, the Young's moduli calibrated with sbNMA can be thought of as the Young's moduli of proteins at temperatures below T_g, while the Young's moduli fit to B-factors can be thought of as Young's moduli at the room temperature.

3.5 Estimating the Young's modulus of the interface region using sbNMA and CHARMM force field

The interface region is kept together mostly by non-bonded electrostatic and van der Waals interactions. Electrostatic interactions are difficult to include in normal mode analysis as they
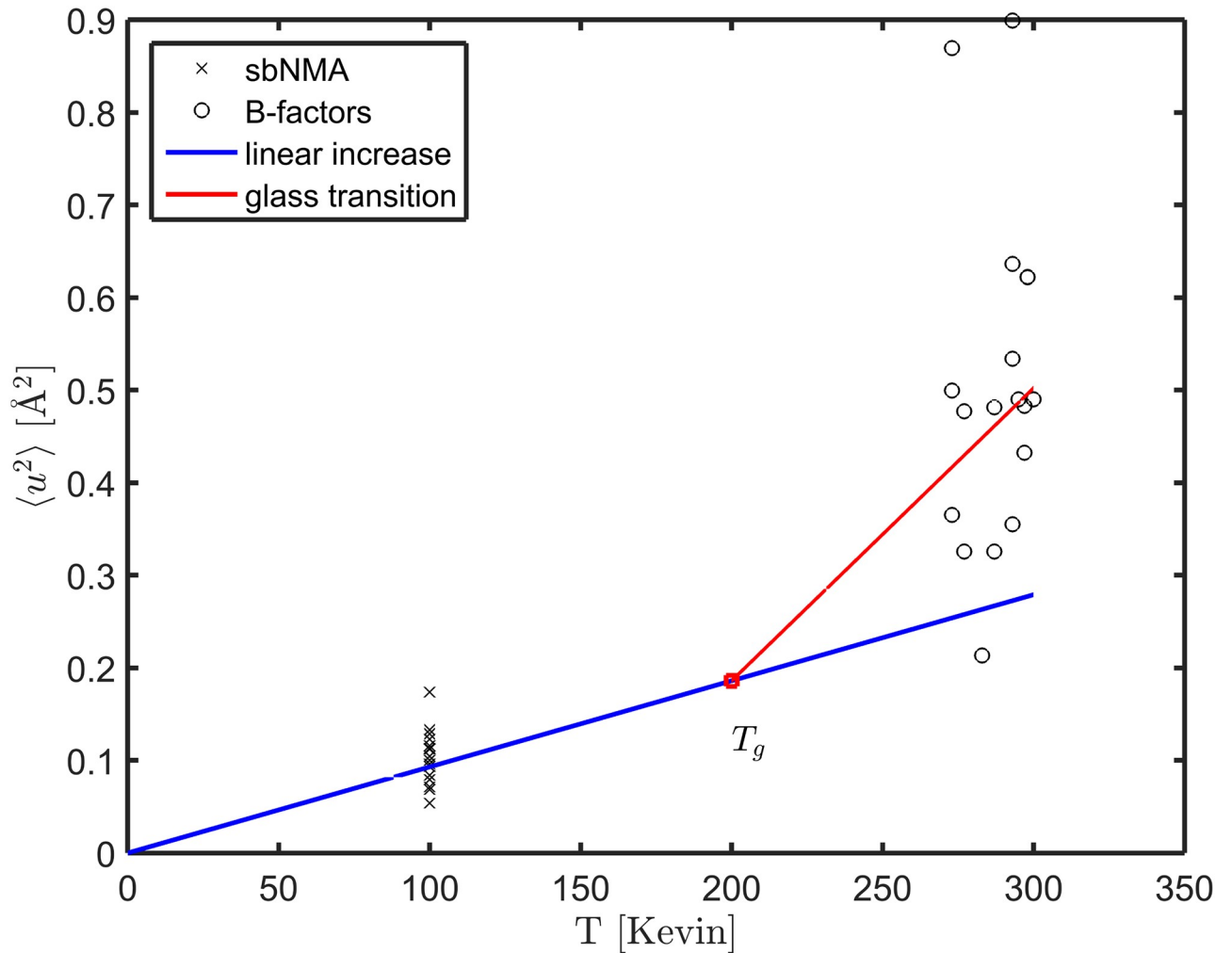
Fig 5. The magnitude of thermal fluctuations as a function of temperature. T_g denotes the temperature of glass transition that is 200-220K [38]. The magnitudes computed from sbNMA at 100K are marked by black crosses, while the magnitudes obtained from B-factors by black circles. The blue line represents the universal trend of the magnitude of harmonic vibrations by sbNMA with CHARMM force field, having a nearly universal value around 0.093 Å^2 at 100 K for all proteins [32], and 0.279 Å^2 at 300 K. At above T_g, anharmonic motions kick in and the magnitude of total thermal fluctuations increases superlinearly, as marked by the red line that is fitted to the black circles.

https://doi.org/10.1371/journal.pone.0247147.g005

introduce negative spring constants [33]. Our past studies showed that van der Waals interactions provided a greater contribution than the electrostatic interactions (Fig 1 in Ref. [33]) and a model without electrostatic interactions could still maintain most of the accuracy of the classical NMA [33]. For these reasons, our current model for the interactions at the interface includes only the van der Waals interactions. It is planned that a future release of the model should improve this deficiency. The van der Waals potential employed in the CHARMM force field is a 6-12 Leonard-Jones potential that takes the following form:

$$V_{\text{vdW}} = \epsilon \left(\left(\frac{r_0}{r}\right)^{12} - 2\left(\frac{r_0}{r}\right)^{6} \right). \tag{10}$$
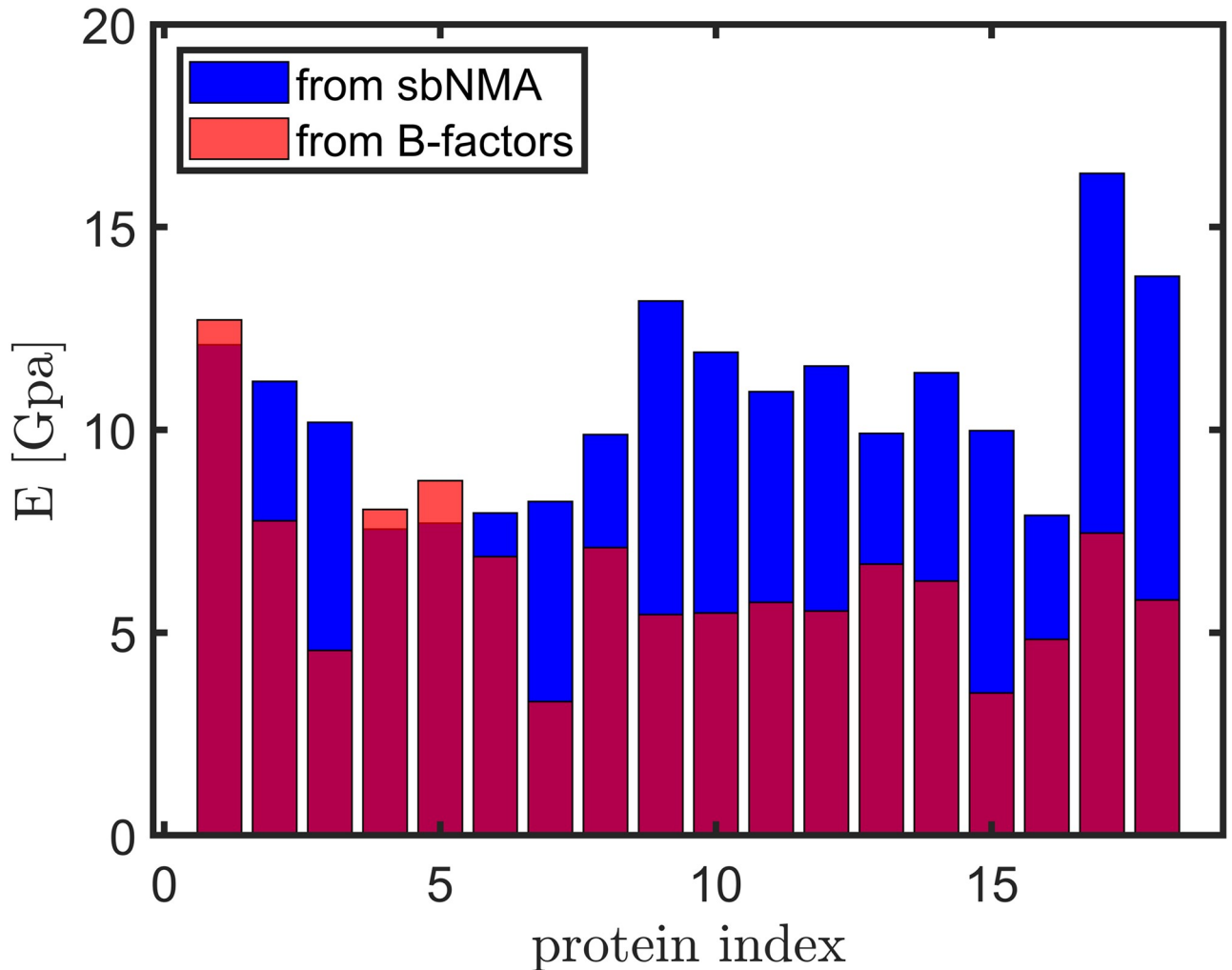
Fig 6. A bar plot of the Young's moduli (E) calibrated with sbNMA (in blue) and those with crystallographic B-factors (in red), for the 18 proteins in the dataset (listed by their sizes in the ascending order). The blue bars can be interpreted also the Young's moduli at temperatures below T_g, while the red bars Young's moduli at the room temperature.

https://doi.org/10.1371/journal.pone.0247147.g006

The spring constant due to the van der Waals potential is:

$$k_{\mathrm{vdW}} = \frac{\partial^2 V_{\mathrm{vdW}}}{\partial r^2} = \frac{12\epsilon}{r^2}\left(13\left(\frac{r_0}{r}\right)^{12} - 7\left(\frac{r_0}{r}\right)^6\right) = \frac{72\epsilon}{r_0^2}, \qquad (11)$$

the last step of which assumes $r = r_0$.

Fig 7 shows the histogram distribution of the van der Waals parameter ϵ of the 50 CHARMM atom types in par_all36_prot.prm and the histogram distribution of $k_{vdW}(r_0)$, which falls mostly in the range of 0.1 to 1 Kcal/mol/Å^2. It is thus clear that k_{vdW}, the spring constant due to van der Waals interactions, is much weaker than the spring constants for covalent bonds that are in the order of several hundreds Kcal/mol/Å^2.

Since the equilibrium distance between a pair of atoms are in the range of 2.5 to 4.5 Å, The Young's modulus of the interface region, E_{inter}, is estimated to be on the order of $\frac{k_{vdW}}{r_0}$, or 0.02 to 0.4 Kcal/mol/Å^3 (or 0.1 to 3 Gpa, the converting factor between the two units is 6.9). This is
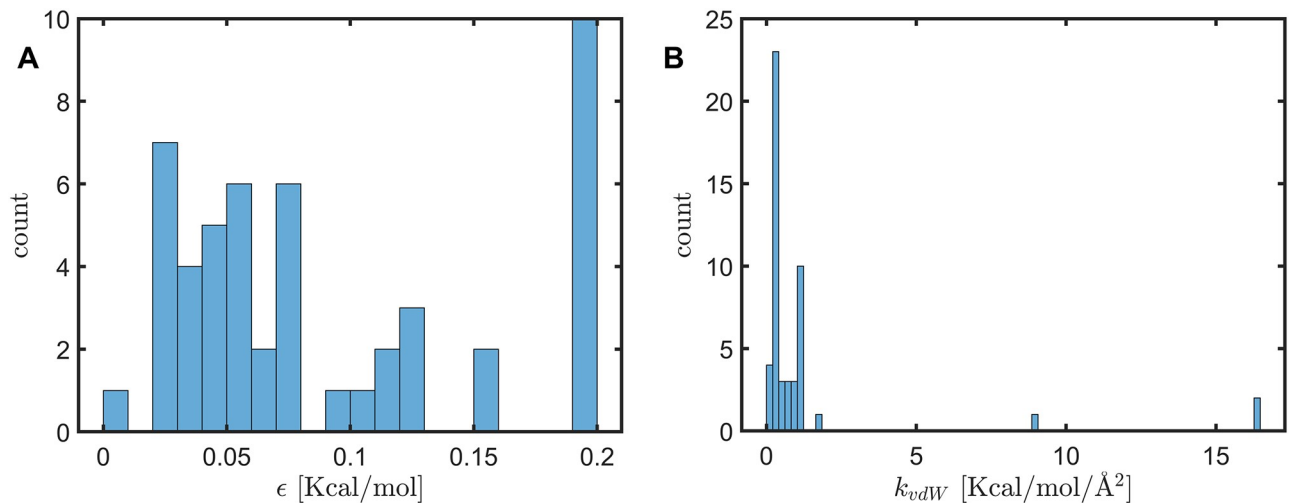
Fig 7. Histogram distributions of (A) the van der Waals parameter ϵ used in CHARMM and (B) of $k_{vdW}(r_0)$, the spring constant of van der Waal interactions.

2 to 3 orders smaller than that of covalent solids or ionic solids, which is expected since spring constants derived from van der Waals interactions are 2 to 3 orders smaller than those of covalent bonds or ionic bonds.

3.6 Young's moduli of the protein interface regions

Using the approach described in the Methods, we first generate quaternary structures of all the 18 proteins in the dataset. σESM (see Methods section for details) is then applied to compute the Young's moduli of the interface regions (the MATLAB scripts used for this computation are available at https://github.com/gsongISU/sigmaESMrelease). The results are given in Table 2. It is seen that while the Young's moduli of the proteins are a few Gpa, the Young's moduli of interface regions are about an order smaller. Such a significant difference in stiffness between proteins and protein interfaces must be due to the underlying chain connectivity of proteins and the associated bonded interactions. Proteins can be considered as van der Waals solids with "steel enforcement" [49]. Also listed in Table 2 are the buried surface areas of interface regions of the quaternary structures. The buried surface area is the difference between the total surface area of individual proteins and that of the complex, both of which are computed using MSMS [28].

The above 18 quaternary structures are artificially generated based on crystal contacts. An insightful reviewer commented that interfaces thus identified tend to be fortuitous and recommended that real protein-protein interfaces existing naturally should also be used and a comparison be made. For this reason, we selected another 18 high resolution structures (1.0 Å or higher) from the PDB that are in the form of homodimers in the asymmetric unit. Their pdb-ids and sizes are given in Table 3, as well as the buried surface areas and the Young's moduli of the interface regions.

To compare the Young's moduli of protein interfaces of the naturally existing dimers (Table 3) with those created by crystal contacts (Table 2), we show in Fig 8 a scatter plot between the buried surface areas and the Young's moduli of the interface regions, for both natural interfaces (red dots) and artificial interfaces created by crystal contacts (blue crosses). First, a significant correlation is seen between the buried surface areas and the Young's moduli (correlation coefficient: 0.77). Secondly, The natural interfaces (red dots) have a much higher

Table 2. The lower and upper bounds of Young's moduli of the 18 proteins and the Young's moduli of the interface regions of the corresponding quaternary structures artificially generated based on crystal contacts.

pdb-id	residue #	buried area [$Å^2$]	E_{inter}	E_{intra} (lower bound)	E_{intra} (upper bound)
1p9g	41	75	0.09	12.71	12.10
1rb9	52	263	0.58	7.76	11.19
1iro	53	388	0.56	4.56	10.19
2igd	61	171	0.40	8.04	7.56
1aho	64	493	1.17	8.75	7.70
5tog	75	541	0.56	6.88	7.95
1ctj	89	352	0.96	3.29	8.23
1lwb	122	251	0.58	7.10	9.88
1c7k	132	81	0.36	5.45	13.17
3r87	132	837	0.83	5.48	11.91
6gz8	133	293	1.13	5.75	10.95
5vg0	142	181	0.23	5.53	11.57
4b9g	146	264	1.02	6.69	9.91
4ann	176	527	0.57	6.27	11.41
4qa8	210	401	0.91	3.51	9.99
6h40	220	183	0.57	4.83	7.89
2pwa	279	203	0.44	7.45	16.32
6gy5	285	103	0.15	5.80	13.79
median	132	263	0.57	6.03	10.57

The lower bounds are obtained by fitting σESM results to B-factors, while the higher bounds are obtained by fitting the σESM results to sbNMA. Also listed in the table are the sizes of the 18 proteins and the buried surface areas of the interface regions. The unit for all Young's moduli is Gpa.

Table 3. The buried surface areas and the Young's moduli (E_{inter}) of the natural interfaces found in a list of 18 high resolution structures from the PDB that are in the form of homodimers in the asymmetric unit.

pdb-id	residue #	buried surface area [$Å^2$]	E_{inter} [Gpa]
4ynh	58	1,964	2.06
3rq9	84	900	1.60
4nds	94	1,279	2.26
2nmz	99	2,381	3.12
4unu	109	597	0.53
6j64	115	2,695	2.69
2xr4	116	1,987	1.89
4egu	118	2,256	2.02
2gud	121	2,997	3.50
4axo	137	2,615	2.09
5nld	138	1,058	1.15
5idb	142	1,390	1.89
2wyt	153	729	1.71
4a7v	153	704	1.50
5sy4	195	1,802	2.14
6rk0	214	5,313	3.77
3noq	229	3,327	2.17
4ypo	325	7,924	2.50
median	129	1,976	2.08

The buried surface areas and the Young's moduli of these natural interfaces are distinctly higher than those of the artificial interfaces seen in Table 2.
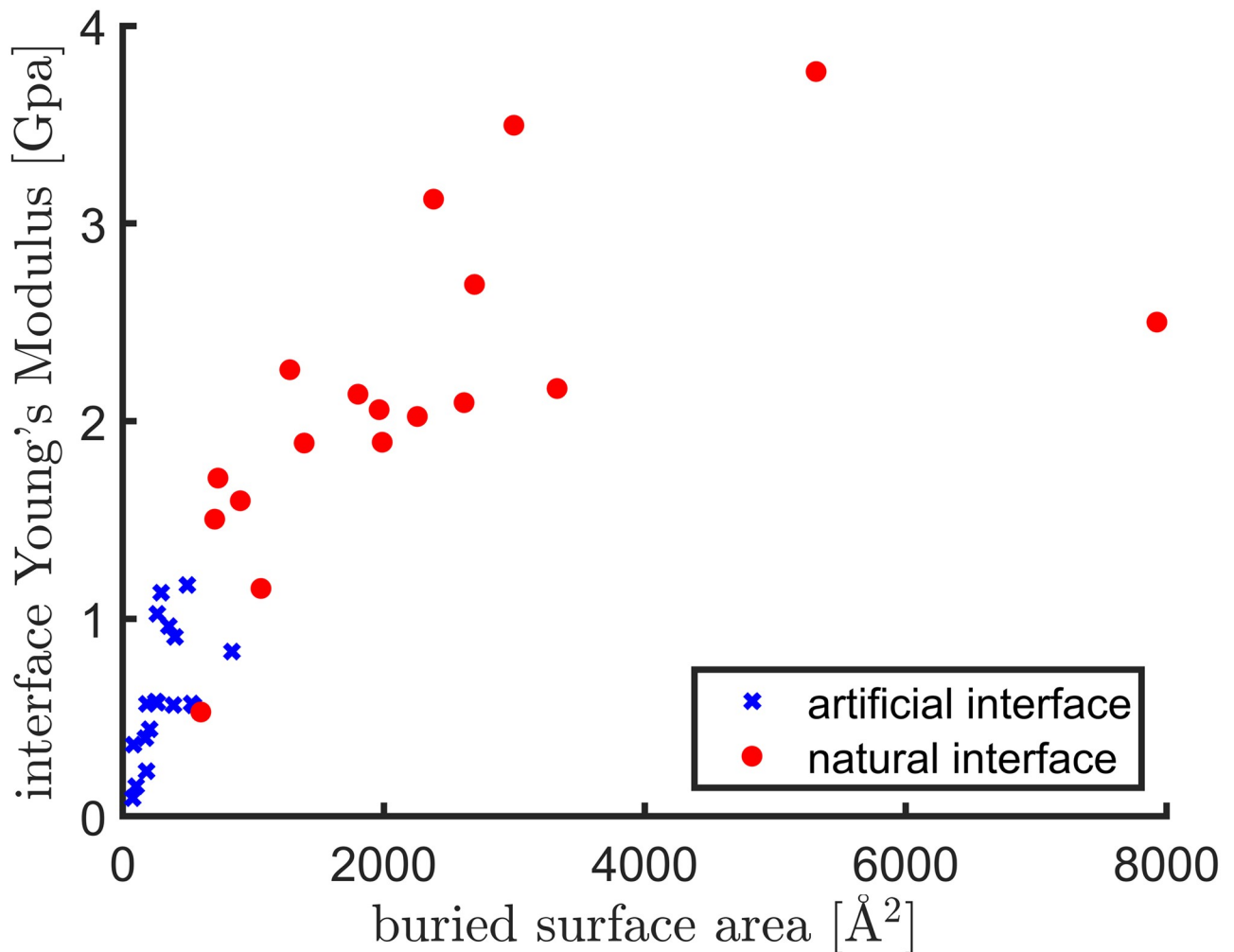
Fig 8. A scatter plot between the buried surface areas and the Young's moduli of the interface regions, for both artificial (blue crosses, data points from Table 2) and natural (red solid circles, data points from Table 3) interfaces.

https://doi.org/10.1371/journal.pone.0247147.g008

buried surface areas, though the two sets of proteins have similar sizes, and are much stiffer, having a median Young's modulus of about 2.0 Gpa that is about 4 times as high as that of artificial interfaces. Thirdly, even for the natural interfaces, the Young's moduli are still a few times lower than those of proteins. Since a clear distinction is seen between artificial interfaces and natural interfaces in Fig 8, a potential application of the present study is to use the interface's Young's modulus proposed in this work to assess the quality of a predicted protein interface.

4 Discussion

In this work, we have studied the elastic moduli of proteins and protein interfaces using a bottom-up approach. The material properties of solids are determined by the underlying physical interactions. For proteins and protein interfaces in particular, their material properties such as elasticity are dictated by the underlying molecular interactions.

The innovation of this work is the development of a novel method that can bridge between material properties at the macroscopic level and molecular interactions at the microscopic

level. To achieve this, we employ both an all-atom NMA model (here sbNMA [33]) and an elastic solid model (σESM) for proteins and protein interfaces. And the "bridge" between the two models is a common physical property predictable by both models: the magnitude of thermal vibrations (see Eq (8)), i.e.,

$$\langle \Delta \mathbf{R}^2 \rangle_{\text{ESM}} = \langle \Delta \mathbf{R}^2 \rangle_{\text{NMA}} \tag{12}$$

This connection allows us to determine the Young's moduli of proteins and protein interfaces using molecular interactions.

A similar bridge can be built between σESM and crystallographic B-factors (Eq 9) as well. A significant benefit of doing both is that the two constraints are complementary to each other and together they provide both an upper and lower bound for the Young's moluli of proteins.

Atoms at the interface region interact primarily through non-bonded interactions. One key realization is that the spring constant due to the van der Waals interaction, or k_{vdW}, is small and is mostly in the range of 0.1-1 kcal/mol/Å^2 (Fig 7). Since the separation distances r_0 between pairs of atoms at the interface region should be about the sum of their van der Waals radii, one may estimate the order of Young's modulus at the interface region by simply taking the ratio of the two, which comes to be about 0.1 to 3 Gpa. Our actual computations are in agreement with this estimation. Covalent solids such as diamond have a Young's modulus that is hundreds of Gpa or even over a thousand Gpa, which is 2-3 orders higher than that of proteins or protein interfaces. The difference has a simple physical explanation: the spring constants for covalent bonds (as in CHARMM force field) are 2-3 orders higher than the van der Waals spring constants while the separation distances between interacting atoms are similar. In a nutshell, the low Young's modulus at protein interface is directly due to the weak non-bonded interactions.

Another interesting finding is that, compared with protein interfaces, the Young's modulus of proteins is several times higher. After all, proteins are chains of amino acids linked together by covalent bonds. The covalent chain serves as a "steel enforcement" [49] that stiffens a protein solid. The abundance of secondary structures in proteins provides additional enforcement through hydrogen bonds. As a result, the internal of a protein is several times stiffer than protein interface.

As part of future work, we plan to apply the method to study the elastic properties of protein capsids. The stiffness of viral capsids has been measured for a large number of viruses using AFM technique [5]. The stiffness varies greatly among the viral capsids, with that of some capsids being an order of magnitude higher than that of others [5]. The physical cause of this large degree of variations in stiffness is not known. As mentioned in the Introduction, existing work, using with either the thin shell or thick shell models, all represents a top-down approach that uses experimental determined stiffness values to fit the underlying elastic moduli of capsids and were not able to provide a physical explanation for the observed stiffness. Our present approach has the potential to take into the account not only the structural details of protein capsids (instead of a shell with uniform spherical geometry and thickness), but also the striking difference in elastic properties between protein subunits and protein interface regions, and consequently may be used to *predict* capsid stiffness and provide an explanation for the observed differences in stiffness among viral capsids.

Author Contributions

Conceptualization: Guang Song.

Data curation: Guang Song.

Formal analysis: Guang Song.

Investigation: Guang Song.

Methodology: Guang Song.

Resources: Guang Song.

Software: Guang Song.

Supervision: Guang Song.

Validation: Guang Song.

Writing – original draft: Guang Song.

Writing – review & editing: Guang Song.

References

1. Grodzicki M. The concept of the chemical bond in solids. In Maksić Z., editor, *Theoretical Models of Chemical Bonding, Part 2: The Concept of the Chemical Bond*, pages 417–452. Springer-Verlag, 1990.

2. Howard J. *Mechanics of motor proteins and the cytoskeleton*. Sinauer Associates Publishers, Sunderland, Mass., 2001.

3. Guthold M., Liu W., Sparks E. A., Jawerth L. M., Peng L., Falvo M. A comparison of the mechanical and structural properties of fibrin fibers with other protein fibers. *Cell Biochem Biophys*, 49(3):165–81, 2007. https://doi.org/10.1007/s12013-007-9001-4 PMID: 17952642

4. Morozov V. N. and Gevorkian S. G. Low-temperature glass-transition in proteins. *Biopolymers*, 24 (9):1785–1799, 1985. https://doi.org/10.1002/bip.360240909

5. Mateu M. G. Mechanical properties of viruses analyzed by atomic force microscopy: a virological perspective. *Virus Res*, 168(1-2):1–22, 2012. https://doi.org/10.1016/j.virusres.2012.06.008 PMID: 22705418

6. Kojima H., Ishijima A., and Yanagida T. Direct measurement of stiffness of single actin-filaments with and without tropomyosin by in-vitro nanomanipulation. *Proceedings of the National Academy of Sciences of the United States of America*, 91(26):12962–12966, 1994. https://doi.org/10.1073/pnas.91.26.12962 PMID: 7809155

7. Fraser R. D. and Macrae T. P. Molecular structure and mechanical properties of keratins. *Symp Soc Exp Biol*, 34:211–46, 1980. PMID: 6166998

8. Gilbert J., Charnley M., Cheng C., Reynolds N. P., and Jones O. G. Quantifying young's moduli of protein fibrils and particles with bimodal force spectroscopy. *Biointerphases*, 12(4):041001, 2017. https://doi.org/10.1116/1.4996447 PMID: 29050479

9. Guthold M., Liu W., Stephens B., Lord S. T., Hantgan R. R., Erie D. A. Visualization and mechanical manipulations of individual fibrin fibers suggest that fiber cross section has fractal dimension 1.3. *Biophys J*, 87(6):4226–36, 2004. https://doi.org/10.1529/biophysj.104.042333 PMID: 15465869

10. Gavish B., Gratton E., and Hardy C. J. Adiabatic compressibility of globular proteins. *Proc Natl Acad Sci U S A*, 80(3):750–4, 1983. https://doi.org/10.1073/pnas.80.3.750 PMID: 6572366

11. Kundrot C. E. and Richards F. M. Crystal-structure of hen egg-white lysozyme at a hydrostatic-pressure of 1000 atmospheres. *Journal of Molecular Biology*, 193(1):157–170, 1987. https://doi.org/10.1016/0022-2836(87)90634-6 PMID: 3586017

12. Morozov V. N. and Morozova T. Y. Viscoelastic properties of protein crystals—triclinic crystals of hen egg-white lysozyme in different conditions. *Biopolymers*, 20(3):451–467, 1981. https://doi.org/10.1002/bip.1981.360200304 PMID: 7213938

13. Morozov V. N., Morozova T. Y., Kachalova G. S., and Myachin E. T. Interpretation of water desorption isotherms of lysozyme. *International Journal of Biological Macromolecules*, 10(6):329–336, 1988. https://doi.org/10.1016/0141-8130(88)90024-4

14. Zenchenko T. A., Pozharskii E. V., and Morozov V. N. A magnetic micromethod to measure young's modulus of protein crystals and other polymer materials. *J Biochem Biophys Methods*, 33(3):207–15, 1996. https://doi.org/10.1016/S0165-022X(96)00027-9 PMID: 9029264

15. Tait S., White E. T., and Litster J. D. Mechanical characterization of protein crystals. *Particle & Particle Systems Characterization*, 25(3):266–276, 2008. https://doi.org/10.1002/ppsc.200701112

16. Koizumi H., Tachibana M., and Kojima K. Elastic constants in tetragonal hen egg-white crystals containing large amount of water. *Phys Rev E Stat Nonlin Soft Matter Phys*, 79(6 Pt 1):061917, 2009. https://doi.org/10.1103/PhysRevE.79.061917 PMID: 19658534

17. Kitajima N., Tsukashima S., Fujii D., Tachibana M., Koizumi H., Wako K. Elastic constants in orthorhombic hen egg-white lysozyme crystals. *Phys Rev E Stat Nonlin Soft Matter Phys*, 89(1):012714, 2014. https://doi.org/10.1103/PhysRevE.89.012714 PMID: 24580264

18. Speziale S., Jiang F., Caylor C. L., Kriminski S., Zha C. S., Thorne R. E. Sound velocity and elasticity of tetragonal lysozyme crystals by brillouin spectroscopy. *Biophys J*, 85(5):3202–13, 2003. https://doi.org/10.1016/S0006-3495(03)74738-9 PMID: 14581220

19. Michel J. P., Ivanovska I. L., Gibbons M. M., Klug W. S., Knobler C. M., Wuite G. J. Nanoindentation studies of full and empty viral capsids and the effects of capsid protein mutations on elasticity and strength. *Proc Natl Acad Sci U S A*, 103(16):6184–9, 2006. https://doi.org/10.1073/pnas.0601744103 PMID: 16606825

20. Gibbons M. M. and Klug W. S. Influence of nonuniform geometry on nanoindentation of viral capsids. *Biophysical Journal*, 95(8):3640–3649, 2008. https://doi.org/10.1529/biophysj.108.136176 PMID: 18621831

21. Bathe M. A finite element framework for computation of protein normal modes and mechanical response. *Proteins*, 70(4):1595–609, 2008. https://doi.org/10.1002/prot.21708 PMID: 17975833

22. Lessen H. J., Fleming P. J., Fleming K. G., and Sodt A. J. Building blocks of the outer membrane: Calculating a general elastic energy model for beta-barrel membrane proteins. *J Chem Theory Comput*, 14 (8):4487–4497, 2018. https://doi.org/10.1021/acs.jctc.8b00377 PMID: 29979594

23. Go N., Noguti T., and Nishikawa T. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc. Natl. Acad. Sci. USA*, 80(12):3696–3700, June 1983. https://doi.org/10.1073/pnas.80.12.3696 PMID: 6574507

24. Levitt M., Sander C., and Stern P. S. The normal modes of a protein: Native bovine pancreatic trypsin inhibitor. *Int. J. Quant. Chem.*, 10:181–199, 1983.

25. Brooks B. and Karplus M. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. USA*, 80(21):6571–6575, November 1983. https://doi.org/10.1073/pnas.80.21.6571 PMID: 6579545

26. Song G. Shape-preserving elastic solid models of macromolecules. *Plos Computational Biology*, 16(5), 2020. https://doi.org/10.1371/journal.pcbi.1007855 PMID: 32407309

27. Edelsbrunner H. and Mucke E. P. Three-dimensional alpha shapes. *ACM Transactions on Graphics*, 13(1):43–72, 1994. https://doi.org/10.1145/174462.156635

28. Sanner M. F., Olson A. J., and Spehner J. C. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers*, 38(3):305–20, 1996. https://doi.org/10.1002/(SICI)1097-0282(199603)38:3%3C305::AID-BIP4%3E3.0.CO;2-Y PMID: 8906967

29. Inczédy J., Lengyel T., Ure A. M., and Gelencsér A. *Compendium of analytical nomenclature: definitive rules 1997*. Blackwell Science, Oxford, 3 edition, 1998.

30. Rahman T. and Valdman J. Fast matlab assembly of fem matrices in 2d and 3d: Nodal elements. *Applied Mathematics and Computation*, 219(13):7151–7158, 2013. https://doi.org/10.1016/j.amc.2011.08.043

31. Logan D. L. *A first course in the finite element method*. Thomson, Toronto, Canada, 4th edition, 2007.

32. Na H., Hinsen K., and Song G. The amounts of thermal vibrations and static disorder in protein x-ray crystallographic b-factors. *Authorea*, 2020.

33. Na H. and Song G. Bridging between normal mode analysis and elastic network models. *Proteins*, 82:2157–2168, September 2014. https://doi.org/10.1002/prot.24571 PMID: 24692201

34. MacKerell A. D., Bashford D., Bellott, Dunbrack R. L., Evanseck J. D., Field M. J. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, 102(18):3586–3616, April 1998. https://doi.org/10.1021/jp973084f PMID: 24889800

35. Na H., Song G., and ben Avraham D. Universality of vibrational spectra of globular proteins. *Phys Biol*, 13(1):016008, 2016. https://doi.org/10.1088/1478-3975/13/1/016008 PMID: 26907186

36. Na H. and Song G. Fast normal mode computations of capsid dynamics inspired by resonance. *Phys Biol*, 15(4):046003, 2018. https://doi.org/10.1088/1478-3975/aab813 PMID: 29557348

37. Na H. and Song G. Conventional nma as a better standard for evaluating elastic network models. *Proteins*, 83(2):259–67, 2015. https://doi.org/10.1002/prot.24735 PMID: 25471641

38. Ringe D. and Petsko G. A. The'glass transition' in protein dynamics: what it is, why it occurs, and how to exploit it. *Biophys Chem*, 105(2-3):667–80, 2003. https://doi.org/10.1016/S0301-4622(03)00096-6 PMID: 14499926

39. Fuglebakk E., Reuter N., and Hinsen K. Evaluation of protein elastic network models based on an analysis of collective motions. *J Chem Theory Comput*, 9(12):5618–28, 2013. https://doi.org/10.1021/ct400399x PMID: 26592296

40. Burden C. J. and Oakley A. J. Anisotropic atomic motions in high-resolution protein crystallography molecular dynamics simulations. *Phys Biol*, 4(2):79–90, 2007. https://doi.org/10.1088/1478-3975/4/2/002 PMID: 17664653

41. Sayle R. A. and Milnerwhite E. J. Rasmol—biomolecular graphics for all. *Trends in Biochemical Sciences*, 20(9):374–376, 1995. https://doi.org/10.1016/S0968-0004(00)89080-5 PMID: 7482707

42. Humphrey W., Dalke A., and Schulten K. VMD—Visual Molecular Dynamics. *J Mol. Graphics*, 14:33–38, 1996. https://doi.org/10.1016/0263-7855(96)00018-5 PMID: 8744570

43. Tama F., Gadea F.X., Marques O., and Sanejouand Y.H. Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins*, 41(1):1–7, 2000. https://doi.org/10.1002/1097-0134(20001001)41:1%3C1::AID-PROT10%3E3.0.CO;2-P PMID: 10944387

44. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 2000; 28(1):235–42. https://doi.org/10.1093/nar/28.1.235 PMID: 10592235

45. Rodgers D. W. Cryocrystallography. *Structure*, 2(12):1135–1140, Dec 1994. https://doi.org/10.1016/S0969-2126(94)00116-2 PMID: 7704524

46. Hooft R. W. W., Sander C., and Vriend G. Verification of protein structures: Side-chain planarity. *Journal of Applied Crystallography*, 29:714–716, 1996. https://doi.org/10.1107/S0021889896008631

47. Touw W. G., Baakman C., Black J., te Beek T. A. H., Krieger E., Joosten R. P. A series of pdb-related databanks for everyday needs. *Nucleic Acids Research*, 43(D1):D364–D368, 2015. https://doi.org/10.1093/nar/gku1028 PMID: 25352545

48. ben-Avraham D. Vibrational normal-mode spectrum of globular proteins. *Phys. Rev. B*, 47(21):14559–60, 1993. https://doi.org/10.1103/PhysRevB.47.14559 PMID: 10005814

49. K. Hinsen. The stuff that proteins are made of: physical properties of folded peptide chains, 2017. A research presentation, available at https://www.youtube.com/watch?v=RbOu1Vjp3Mc.