Journal of Molecular Graphics and Modelling 57 (2015) 106–113

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

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

Local elastic constants of LacI and implications for allostery

Andre A.S.T. Ribeiro, Vanessa Ortiz*

Columbia University, Dept. Chemical Engineering, United States

ARTICLE INFO

Article history:
Accepted 29 January 2015
Available online 7 February 2015

MSC:
00-01
99-00

Keywords:
Allostery
Mechanical properties
Elastic network model
Stiffness
Lac repressor protein

ABSTRACT

Allostery connects subtle changes in a protein's potential energy surface, induced by perturbations like ligand-binding, to significant changes in its function. Understanding this phenomenon and predicting its occurrence are major goals of current research in biophysics and molecular biology. In this paper we introduce a novel approach for studying complex structural transformations such as those typical for allostery. We show that the calculation and analysis of atomic elastic constants of a known allosterically regulated protein, *lac* repressor, highlights regions that are particularly prone to suffer structural deformation and are experimentally linked to allosteric function. The calculations are based on a high resolution, all-atom description of the protein. We also show that, for the present system, modifying the description of the system from an all-atom forcefield to an elastic network model yields qualitatively different results, indicating the importance of adequately describing the local environment surrounding the different parts of the protein.

1. Introduction

Cellular metabolism is characterized by a myriad of biochemical processes that need to be properly regulated. Perturbations of the cellular environment, such as changes in concentrations of specific substrates, may (de)activate some metabolic pathway(s). A thorough understanding of metabolism regulation has thus been a long-standing goal of biological research [1]. The study of feedback inhibition of enzymes provided a major breakthrough in the area, with the finding that enzymes may have two binding sites that are spatially distant but, nevertheless, influence each other. The situation in which an enzyme is inhibited by a molecule sterically different from its substrate was termed "allosteric inhibition" [2]. Subsequent rationalization of experimental evidence led to the development of two important phenomenological models: the concerted model of Monod et al. [3] and the sequential model of Koshland et al. [4]. Models based on statistical mechanics were later developed [5–7] and, more recently, the availability of computational simulation methods has shifted the focus to a mechanistic interpretation of allostery, in which changes in protein conformational equilibrium or dynamics are caused by the intricate balance of inter-atomic interactions that is perturbed upon effector binding [8–11].

Allosteric enzymes are increasingly acknowledged as potential targets for drug design due to the possibility of increased selectivity [12,13]. Furthermore, there is evidence that a vast number of proteins may exhibit allostery [10,14], thus making methodologies that may be able to identify allosteric sites and predict allosteric pathways extremely useful.

Experimental procedures are obviously the ultimate source of information regarding any molecular system. Nevertheless, computational methods have come a long way in establishing themselves as important tools to investigate the behavior of such systems. Statistical mechanics relates the microscopic states of a given system to its macroscopic properties. Therefore, in principle, an accurate energetic description of a protein is sufficient for the complete understanding of its biological function. On the microscopic level, protein energetics is characterized by a balance between different inter-atomic interactions [15], with small perturbations at specific sites potentially leading to major changes in conformational distributions. Therefore, a thorough characterization of allostery requires understanding of two aspects: (1) how signals propagate through the protein structure, and (2) which regions of the protein are likely to suffer structural deformations as a response to the applied perturbation.

On the first aspect, we have recently introduced a new energy-based network analysis method [16], which allows characterization of signaling pathways in proteins. This method assumes that signals travel more efficiently through residues that have strong interatomic interactions, and was able to correctly identify important residues for allosteric signal propagation in the allosteric enzyme imidazole glycerol phosphate synthase. In the work presented here,

* Corresponding author at: 809 Mudd Bldg., 500W 120th Street, New York, NY 10027, United States. Tel.: +1 212 854 7260; fax: +1 212 854 3054.
E-mail address: vortiz@columbia.edu (V. Ortiz).

http://dx.doi.org/10.1016/j.jmgm.2015.01.013
1093-3263/

we attempt to address the second aspect, which deals with connecting the energetic signals propagating through the molecule, to the resulting conformational changes associated with transitions between the active and inactive states of the molecule.

Several computational methods have been used to predict structural transformations [17]. These range from atomic-level molecular dynamics (MD) simulations [18] to more approximate methods such as normal mode analysis (NMA) of simplified potentials like elastic network models (ENM) [19]. Atomistic MD can provide the level of detail that characterization of some allosteric systems requires. However, the small integration steps coupled to the complexity of energy evaluation for a system containing thousands of atoms limit the time lengths that can be simulated. Even though insights into allostery have been gained by careful analysis of relatively short MD simulations [20–23], for the most part atomistic MD is still unable to access the time lengths (>μs) characteristic of many allosteric transitions [24,25]. Computationally less expensive methods will always involve some kind of approximation when compared to atomistic MD simulations. Common approaches include reducing the dimensionality of the molecule by using a coarse-grained description such as an elastic network model [26], and approximating its dynamical behavior in some way, with normal mode analysis (NMA) being a common example [27]. The disadvantages of the mentioned approaches are the neglect of local interactions, which may be important for conformational changes, and the use of a single, static structure. We believe the field could benefit from a computational method that can predict structural transformations in biomolecules, in a computationally efficient manner but without sacrificing the accuracy that comes from employing a high-resolution molecular description.

Our approach involves a method for determining which regions of a protein are likely to suffer structural changes upon small perturbations. The elastic modulus of a material provides information about its resistance to small applied deformations (i.e., stiffness). Materials with a high elastic modulus are stiffer, and oppose deformation more strongly. At the microscopic level, elastic constants can provide a stiffness map of a given system, identifying regions that are particularly prone to deformation [28]. Here we employ a method developed by Riggleman and co-workers [29] to calculate the local elastic constants of the lactose repressor protein (a known allosteric protein), and show that this quantity can predict the structural deformations associated with its allosteric transition from the active to inactive state.

2. Materials and methods

2.1. Local elastic constants

2.1.1. Calculation and implementation

A routine for calculation of local elastic constants was implemented in the GROMACS simulation package [30], following the method described in Ref. [29]. The elastic constant is interpreted as a change in atomic stress divided by a local strain. The protocol thus involves minimizing the energy of a system, applying a small strain along the x direction, and minimizing the energy of the deformed configuration. Local stresses are calculated using the method of Zhou [31], with the atomic stress tensor ($\boldsymbol{\sigma}$) of atom i being defined as

$$\boldsymbol{\sigma}_i = \frac{1}{2V_i}\sum_{i\neq j}\mathbf{r}_{ij}\otimes\mathbf{f}_{ij}, \tag{1}$$

where $\mathbf{r}_{ij}$ and $\mathbf{f}_{ij}$ are the position and force vectors between atoms i and j, respectively. The Voronoi volume of atom i (V_i) was calculated by implementing a routine that follows the method described by Gerstein and co-workers [32].

The local strain tensor of atom i ($\boldsymbol{\epsilon}_i$) is calculated using the method of Falk and Langer. [33] These authors suggested that the local strain be calculated as to minimize the non-affine character of the deformation. The analysis thus compares the position vectors $\mathbf{r}_i$ and $\mathbf{r}'_i$ before and after the deformation. The displacement D is then defined as

$$D_i = \sum_n\sum_k\left(r_n^{k\prime} - r_i^{k\prime} - \sum_j(\delta_{kj}+\epsilon_i^{kj})(r_n^j - r_i^j)\right)^2, \tag{2}$$

where the indices k and j denote spatial coordinates and the index n runs over atoms within the chosen cut-off value of atom i. The tensor $\boldsymbol{\epsilon}_i$ that minimizes Eq. (2) is defined as the local strain on atom i. [33]

Once local stresses and strains are determined, the atomic elastic constant for extension in the x direction ($C_{xx,i}$) can be calculated as

$$C_{xx,i} = \frac{\Delta\sigma_{xx,i}}{\epsilon_{xx,i}} \tag{3}$$

The necessary routines for local stress and strain calculation have been implemented in version 4.5.5 of the GROMACS simulation package. Local stresses are calculated inside the regular routines for energy/force calculation. A custom tool *g_elastic*, responsible for analyzing output generated by the *mdrun* tool and calculating the elastic constants, has been created.

2.1.2. FIRE minimization

To perform energy minimizations of complex biological systems, an efficient algorithm is needed. We have used the fast inertial relaxation engine (FIRE) [34] algorithm.

The FIRE algorithm [34] uses the following equation of motion to minimize the energy of the system:

$$\dot{\mathbf{v}}(t) = \mathbf{F}(t)/m - \gamma(t)|\mathbf{v}(t)|(\hat{\mathbf{v}}(t) - \hat{\mathbf{F}}(t)). \tag{4}$$

Our basic protocol involved minimizing the system until a maximum force of 10^{-2} kJ mol^{-1} nm^{-1} was reached. This typically took $\sim 5\times10^4$ minimization steps.

The FIRE algorithm is not implemented in the official GROMACS release. We have implemented it within version 4.5.5 of the code, using a standard *leap-frog* algorithm to integrate the equations of motion. FIRE minimizations can be run in parallel with the standard GROMACS domain decomposition technique.

We have compiled GROMACS with double-precision arithmetics in order to reach the chosen convergence criterion. The minimizations were run without constraints.

2.1.3. Solvent minimization

The main purpose of the present paper is to calculate elastic constants of a bio-molecular system. We also believe that a correct description of the local environment surrounding the protein is essential for obtaining an accurate account of complex biological phenomena. We have therefore calculated elastic constants of a protein described by an all-atom forcefield, and solvated with explicit water molecules. The minimization with the FIRE algorithm thus involves using Eq. (4) for all the atoms of the systems, i.e., solute and solvent molecules.

2.2. Systems studied

2.2.1. Lactose repressor protein

We have calculated local elastic constants for the lactose repressor protein (LacI), a known allosteric protein. The crystal structure of LacI [35] bound to an anti-inducer ligand (PDB entry 1EFA) was used as starting point for the calculations. The anti-inducer and DNA molecules present in the PDB file were removed, resulting in

a starting structure comprising one dimer (chains A and B of the PDB file). The system was described with the AMBER03 forcefield [36]. Long-range electrostatic interactions were treated with the particle mesh Ewald (PME) method, with a cut-off of 12 Å, while van der Waals interactions were switched off between 10 and 12 Å.

To calculate the elastic constants, the protein was solvated with TIP3P water molecules, and its energy minimized using the *steepest descent* algorithm. This was followed by 200 ps of MD (T = 300 K, P = 1 bar). The resulting configuration was then minimized using the FIRE algorithm, until a maximum force of 10^{-2} kJ mol^{-1} nm^{-1} was reached. A strain of 10^{-5} was then applied along the x direction, and the resulting configuration was minimized again with the FIRE algorithm. Analysis of the two minimized configurations allows the calculation of local elastic constants.

In order to have statistically meaningful results and to take molecular anisotropy into account, elastic constants for extension in different directions were calculated for LacI. To that end, we have used 290 vectors taken from a spherically symmetric distribution. For each calculation, the unsolvated protein structure was rotated to align the x axis with the chosen vector, and the protocol outlined above was then applied, resulting in 290 sets of elastic constants.

The elastic constants reported in this paper refer to heavy (non-hydrogen) atoms. This means that the strain is calculated from the displacement of the heavy atoms, the stress is calculated by summing the local stresses of the heavy atom with those of its bonded hydrogen neighbors, and the Voronoi volume V_i is calculated using only the positions of the heavy atoms, with the space occupied by hydrogens being distributed among neighboring heavy atoms. Voronoi volumes for different atom types of LacI are shown in Table S1, Supporting Information and are in accordance with previous results [32].

2.2.2. Trehalose

To ensure that the algorithm for the calculation of local elastic constants was correctly implemented, one of the systems analyzed in the original reference [29] (trehalose) was studied here as well. As in the case of LacI, we have used the AMBER03 forcefield to describe the system. We have also used the same electrostatic treatment with PME and switched van der Waals interactions. Following a short minimization, we performed a 2-ns MD simulation at P = 1 bar and T = 300 K. We took five frames from the last 200 ps as starting points for FIRE minimization and straining.

2.3. Analysis of elastic constants

2.3.1. Stress–strain curves

We have constructed stress–strain curves for both LacI and trehalose by applying 1000 successive strain steps of 10^{-5} to reach a total strain of 0.01. The stresses reported are the total stresses for the simulation boxes, i.e., $\sigma_{xx} = \sum_i \sigma_{xx,i}/V$, with the sum running over all atoms i and including both solvent and solute atoms. The Voronoi volumes V_i were not included in the local stresses, instead, the total stress was divided by the box volume V.

2.3.2. Stiffness evaluation

As outlined earlier, elastic constants for extension in different directions were calculated for LacI. Stiffness maps were generated by considering the entire set of elastic constant values. We defined the stiffness of an atom as the median of the distribution of its elastic constant values.

Even though our calculations were done at the atomic level, characteristic of the employed forcefield, we have analyzed the results by focusing on mechanical properties of groups of atoms, involving one or more amino-acid residues. We have defined the elastic constant of a group of atoms for extension along a given direction as the average of the corresponding elastic constants of its constituent atoms. The stiffness of a group of atoms was defined in an equivalent manner to the atomic case, i.e., the median of the distribution of elastic constant values.

2.3.3. Community analysis

We have performed a network analysis in order to divide our system in communities (regions). Our analysis is based on the construction of a weighted graph, in which each node represents one of the residues. Edges between nodes are drawn if the corresponding residues have any heavy atoms that are within 3 Åof each other.

We want to group together residues that are spatially close and have similar stiffness values. We have therefore taken the difference in stiffness values between two residues as a measure of inter-node "distance". In particular, we use the following definition:

$$d_{ij} = \exp(|s_i - s_j|/10), \tag{5}$$

where d_{ij} is the inter-node distance and s_i is the calculated stiffness of residue (node) i. As stated in the previous section, the stiffness of a group of atoms is defined as the median of the distribution of elastic constants of the group.

We have used the algorithm proposed by Newman [37] to find the most reasonable community division. The algorithm relies heavily on the idea of modularity, defined as

$$Q = \sum_{\mu=1}^{\eta} (e_{\mu\mu} - a_\mu^2), \tag{6}$$

where η is the number of communities, $e_{\mu\mu}$ is the fraction of edges linking nodes that belong to community μ and a_μ is the fraction of edges that are attached to vertices in community μ, i.e,

$$a_\mu = \sum_{\nu=1}^{\eta} e_{\mu\nu}. \tag{7}$$

The algorithm starts by assigning each node to a different community. The result of merging two different communities can readily be characterized by the corresponding value of ΔQ. The algorithm then performs $\eta - 1$ steps, joining the communities that maximize ΔQ at each step. The community division obtained at the step characterized by the maximal value of Q is taken as the optimal community structure.

The original algorithm was used to study unweighted graphs, however, as proposed by the authors, the generalization to weighted networks can be trivially done by taking the initial values of the elements e_{ij} to be proportional to the strength of the edges. In particular, we took $e_{ij} \propto 1/d_{ij}$.

2.3.4. Error analysis

To improve the statistics and provide error estimates, we have performed additional calculations for LacI. As stated earlier, our original set of calculations involved minimizing the energy of the system until a maximum force of 10^{-2} kJ mol^{-1} nm^{-1} was reached. We took the minimized unstrained configurations as starting point for further minimizations, and performed three additional sets of calculations, characterized by convergence criteria of 8×10^{-3}, 6×10^{-3} and 4×10^{-3} kJ mol^{-1} nm^{-1}. These additional minimizations were followed by straining and further minimization, thus generating additional elastic constant values for each of the straining directions and, consequently, three additional sets of stiffness values.

2.3.5. Elastic network model

In order to evaluate the importance of a detailed description of the protein system, we have calculated the elastic constants of LacI

described by an elastic network model (ENM). The ENM representation of biomolecules relies on a single parameter potential energy function: [19]

$$V = \sum_{a,b} \frac{C}{2} (|\mathbf{r}_{a,b}| - |\mathbf{r}^0_{a,b}|)^2, \quad (8)$$

where $\mathbf{r}_{a,b} = \mathbf{r}_a - \mathbf{r}_b$ and the zero superscript denotes the initial configuration. The sum runs over all pairs of atoms (a, b) within a chosen distance cut-off value, and the parameter C is chosen such that the model reproduces the cumulative density of modes obtained from an atomistic model. For our calculations, we used a distance cut-off of 4 Å, and C = 2 kJ/(mole-Å^2), values typically used with these models.

In addition, we calculated the elastic constants for LacI using an anisotropic network model (ANM), a more broadly used ENM that involves applying Eq. (8) to a coarse-grained description of the molecule. For the ANM, the coordinates of the α-carbons were used to define the positions of the network nodes, and values of 12 Åand 2 kJ/(mole-Å^2) were used for the cutoff and C parameters, respectively.

For both ENM and ANM models, we used the set of unstrained minimized configurations as initial configurations, applied a strain along the x direction ($\epsilon_{xx} = 10^{-5}$), and used the resulting configurations to calculate the elastic constants. We calculated elastic constants for extensions in different directions as outlined above.

3. Results and discussion

3.1. Significance of elastic constants of biological molecules in aqueous media

A given material can respond to deformations in different manners. More specifically, small deformations usually lead to elastic behavior, in which the system returns to the original state if the perturbation is removed, while larger deformations can lead to plastic behavior. The magnitude of external forces at which the transition between elastic and plastic behavior occurs depends on the chemical nature of the material.

In order to gauge how the concept of elastic deformations holds for a solvated protein, a stress–strain curve was constructed for LacI by successively deforming the system until a total strain of 0.01 was reached (1000 successive strains $\times$ 10^{-5}/strain). For this particular system, a total strain of 0.01 represented atom displacements of up to 1.5 Å. We have also performed the same calculations for a simulation box containing trehalose molecules, a small glass-forming system that was studied in Ref. [29].

The stress–strain curves (Fig. 1) show that, although the protein exhibits several plastic events (discontinuous stress drops) along the curve, the slope remains approximately constant. In addition, following a plastic event, a series of calculations were done in which a negative strain (compression) was applied until the system was taken to its original volume. The red line in Fig. 1 shows the corresponding results, with the slope also remaining constant. Finally, for comparison, a stress–strain curve was built for the trehalose system (blue line, Fig. 1). Trehalose also shows a linear dependence of stress on applied strain but, as expected, its elastic modulus is higher, reflecting the increased stiffness of this material. We compared the slopes of the two stress–strain curves in two different regions of the plots: between steps 0 and 20 and steps 980 and 1000. For LacI, the values were 9.75 GPa and 9.69 GPa. For trehalose, the values were 36.36 GPa and 32.22 GPa.

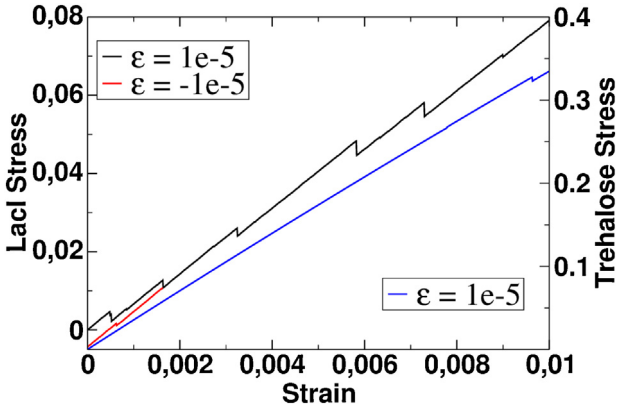

Fig. 1. Stress–strain curves for LacI (black and red) and trehalose (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

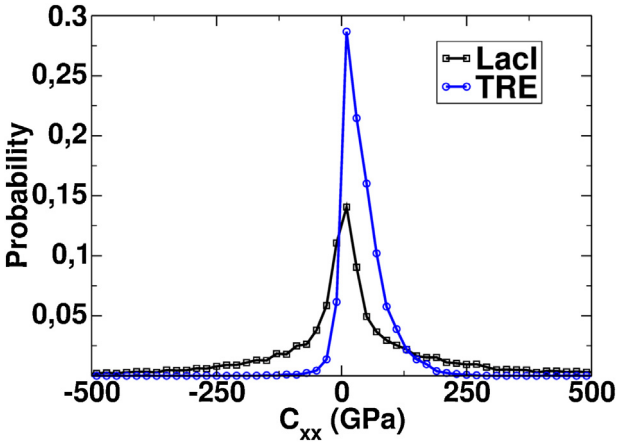

Fig. 2. Distributions of atomic elastic constants for LacI and trehalose.

3.2. Local frustration in biological molecules

The above results show that a macroscopic elastic constant of a solvated protein is well-defined. However, the purpose of the present paper is to calculate local elastic constants and investigate whether the stiffness map of a protein can highlight residues that are important for allosteric function. To analyze the elastic constants of both LacI and trehalose on the atomic level, we have calculated distributions of local elastic constants for both systems, with the results shown in Fig. 2. It can be seen that the distribution of elastic constant values is wider for LacI, with several atoms exhibiting negative elastic constants. This is indicative of an increased level of local frustration in the protein [15]. We have investigated this difference by recalculating the stress values without inclusion of specific energetic contributions. The results are included in Fig. S1 of Supporting Information, and show that the distributions of changes in local stress resulting from the non-bonded interactions are similar for the trehalose and protein systems. Inclusion of bonded interactions, particularly dihedral interactions (directly related to conformational flexibility), drive the distributions for trehalose and the protein away from each other.

3.3. Grouping of atoms reduces noise and highlights differences in stiffness for different secondary structure elements

Knowledge of the atomic elastic constants can be used to construct a stiffness map, highlighting atoms that are particularly prone to be deformed. However, the significant deviations in elastic

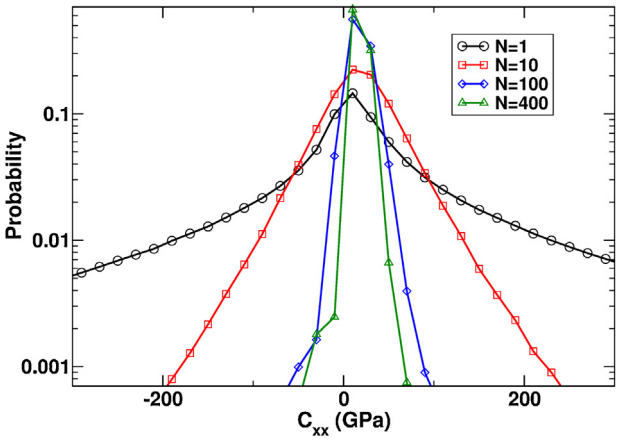

Fig. 3. Distributions of elastic constants for groups of atoms of different size.

constant values resulted in large statistical errors for a stiffness map constructed at the atomic level. To circumvent this problem, we have explored an approach of joining atoms in groups and calculating the elastic constants of the group as the average of the elastic constants of its constituent atoms. Fig. 3 shows how the width of the distributions of elastic constants dramatically decreases as the number of atoms included in the group is increased, indicating a possible degree of anti-correlation between elastic constants of neighboring atoms. Indeed, calculation of the average correlation coefficient between any atom and its bonded neighbors yields a value of −0.45, indicating a moderate anti-correlation.

This analysis was also applied to atoms belonging to specific types of secondary structure (α-helix vs. β-sheet). Table S2, Supporting Information shows that, while the widths of the distributions decrease for larger groups in both α-helices and β-sheets, α-helices showed lower elastic constants than β-sheets. This suggests that α-helices might be more prone to deformation and propagation of structural changes, in accordance with recent experimental results [38].

While Fig. 3 presents a convincing argument for wanting to group atoms, a decision still needs to be made about how to define the groups, and the atoms that belong to each one. A straightforward approach would be to use secondary structure classifications, however, there is no reason to believe that this structural division would also correspond to the most reasonable mechanical division of the system. In particular, we want to group neighboring residues that have similar stiffness. This can be done with help from network theory (see Section 2). When applied to the local elastic constants of LacI, our network analysis yielded an optimal community division with 23 distinct communities, ranging in size from 11 residues (80 atoms) to 52 residues (393 atoms). The communities are colored in Fig. 4A according to their respective stiffness values. Tables S3–S5, Supporting Information provide the calculated values, along with the corresponding errors. Figs. S2–S4, which highlight the communities in the different protein sub-domains, show that the division tends to group together residues that belong to the same secondary structure unit, even though this was not enforced.

3.4. Stiffness map of LacI reveals regions important for allosteric function

LacI is part of the so-called *lac* operon, a genetic switch that regulates the metabolism of lactose in bacteria [39]. LacI binds to the DNA that encodes for the proteins involved in lactose metabolism, blocking the transcription of genes that coordinate lactose utilization. Upon binding of an *inducer* molecule, LacI undergoes a conformational change that decreases its affinity to

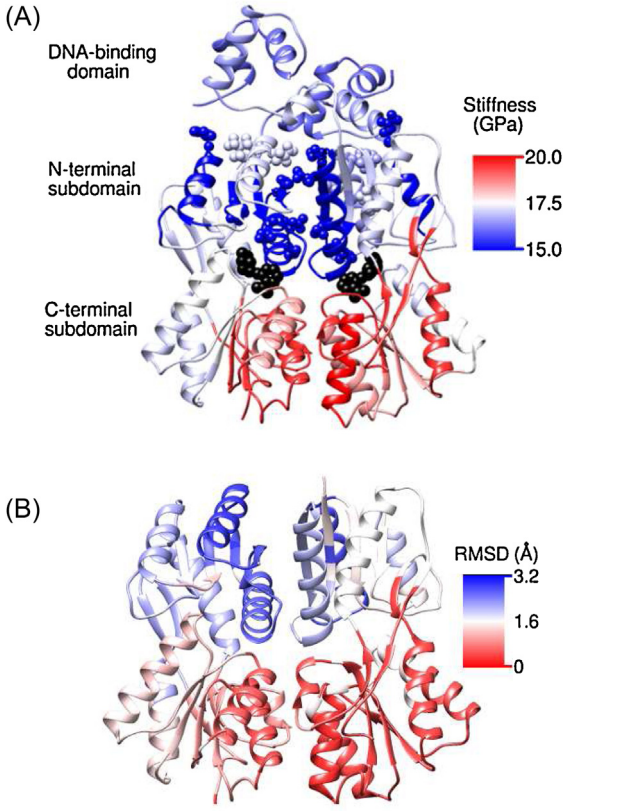

Fig. 4. (A)Stiffness map for LacI repressor. (B) Root mean squared deviation between PDB structures 1EFA and 2P9H. In stiffness map, residues are grouped in communities and colored according to the calculated stiffness, with blue (red) representing regions of low (high) stiffness. The DNA-binding and core domains are shown. The ligand molecules are represented as black spheres to highlight the binding pocket, but were removed from the structure for the calculation of the elastic constants. Residues that have been recently highlighted to be important in LacI's allosteric function [41] are shown as spheres. For the RMSD map, RMSD values were calculated after superimposing C_α atoms of the C-terminal domains of both structures.

DNA, allowing gene expression and activation of lactose metabolic pathways [35,40]. This protein is thus a prime example of an allosterically regulated system.

Fig. 4A shows the structure of the LacI dimer. Each monomer is composed of a DNA-binding domain (residues 1–61), a core domain (residues 62–330) and a tetramerization domain (331–360, not shown). The core domain of each monomer can be further divided in N-terminal and C-terminal sub-domains, with binding of inducer molecules (black spheres in Fig. 4A) occurring at the interface between these two sub-domains, and the allosteric signal being propagated through the N-terminal sub-domain to the DNA-binding domain [35].

The stiffness map of a protein can identify the structural changes that are most likely to occur as a result of a small perturbation such as binding of a ligand. In the case of LacI, the stiffness map clearly differentiates between the C-terminal and N-terminal subdomains.

The C-terminal subdomains are stiffer, while the interface between the N-terminal subdomains of the two monomers appears as the most flexible region. This region sits right in between the effector binding site (ligand represented as black spheres in Fig. 4A) and the DNA-binding domain, so it would be reasonable to infer that, in LacI, binding of the effector creates a local structural perturbation that propagates through the molecule via the low-stiffness regions, until reaching the DNA-binding domain. This is consistent with previous reports, [35] which show that the interface between the N-terminal subdomains are the regions through which the allosteric signal propagates, while the C-subdomains remain

stationary during the allosteric transition, anchoring the mobile N-terminal subdomains.

The use of a well-studied allosteric molecule such as LacI to test this methodology, allows us to perform additional checks, since LacI is one of the rare cases for which structures are available for both the active and inactive states. The structural changes that are caused by binding of inducer molecules can be highlighted by comparing the experimentally determined structures of LacI bound to anti-inducer and inducer molecules. This analysis is shown in Fig. 4B, where a root-mean-squared comparison between PDB structures 1EFA[35] and 2P9H[42] shows the regions that deform the most during the transition. Unfortunately, the coordinates of the DNA-binding domain residues are not determined in 2P9H, which is thus omitted. It can be seen that the stiffness map agrees with the experimental results in highlighting the interface between the N-terminal subdomains as the region most prone to suffer structural changes.

The importance of the interface between the N-terminal subdomains in the allosteric function of LacI is further highlighted by theoretical and experimental studies, with mutation of several residues in this region suppressing allosteric function [43,40,44]. Of particular interest is a recent report in which the authors used directed evolution in combination with saturation mutagenesis to identify mutations that re-establish allosteric function in LacI mutant D88A [41]. The mutated residues that were found to re-establish allosteric function are shown as spheres in Fig. 4A, and are mostly within the communities that have the lowest stiffness values (blue). Finally, we analyzed the trajectory obtained from a 100-ns NPT MD simulation of LacI at 300 K and 1 bar, to calculate the root-mean-squared-fluctuations (RMSF) for each residue. The RMSF map is shown in Fig. S5 of Supporting Information, and indicates that, without the application of a uniform strain and analysis of resulting structural changes, the N- and C-terminal subdomains are not discriminated.

3.5. Importance of local interactions

The most accurate approach for understanding complex biological phenomena such as allostery is to perform long molecular dynamics simulations, provided that the employed forcefield yields a good representation of the system in the chosen thermodynamic conditions. However, in most cases, such task is extremely expensive in practice, demanding significant computational power. For this reason, alternative, less expensive approaches have been extensively used in the past. The method introduced here provides one significant advantage over other methods: the description of the system is done at the atomic level, with a forcefield normally used in MD simulations. Furthermore, the effect of solvent molecules is explicitly considered. The use of an all-atom forcefield with explicit solvent molecules certainly increases the computational cost when compared to simpler models, therefore, it is appropriate to investigate whether such level of sophistication is indeed necessary.

The elastic network model (ENM) description of molecular systems introduced by Tirion [19] relies on a very simplistic approach to describe inter-atomic interactions. A single parameter energy equation is used, with harmonic potentials describing the interactions between all pairs of atoms that are within a cut-off radius. This method thus relies on the assumption that accurate modeling of local interactions is not essential for the description of slow, large-scale oscillations of the molecule. The fact that this assumption can be approximately true in several cases, along with the obvious computational simplicity of energy evaluation, has led to this approach being widely used in molecular modeling of bio-systems. A number of methods based on ENM or other simplified models of proteins have been recently proposed with the specific goal of studying structural changes triggered by allostery and/or ligand binding [45,46,27,47–51]. Similarly, an approach for calculating protein rigidity based on graph theory has been previously proposed,[52] however, incorporation of a detailed forcefield description with inclusion of explicit solvent molecules would not be feasible within this framework. Therefore, when applying these methods to the study of allostery and other processes that depend on perturbations of the bio-molecular systems, the assumption that local interactions are not relevant for determining the effects of such perturbations must also be made.

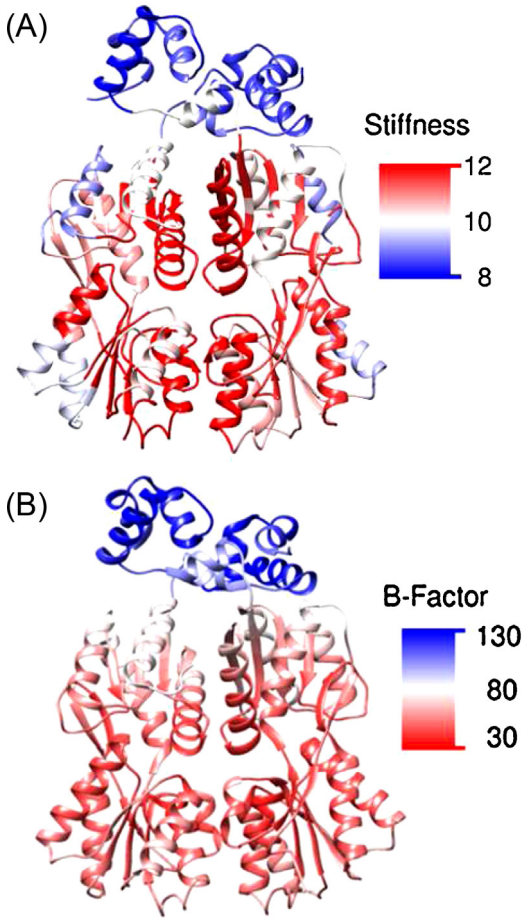

Fig. 5. (A) Stiffness (GPa) map for LacI represented by an elastic network model. (B) Experimental B-factors (Å^2) for LacI. Atomic B-factors were taken from PDB file 1EFA. Residues are grouped in the same communities as in Fig. 4A and colored according to the average value of the corresponding quantity.

We have calculated the elastic constants of LacI described by the ENM, with the corresponding stiffness map shown in Fig. 5A. It can be seen that the DNA-binding domains are highlighted as the least stiff regions. Furthermore, there is no significant difference between the N- and C-terminal sub-domains. In addition, we performed a normal-mode analysis (NMA) of LacI described by the ENM, using the program MMTK [53]. The low-frequency eigenvectors correspond to vibrations of the DNA-binding domain (not shown), in accordance with the stiffness map built using the ENM. It is important to note that the ENM description of this protein agrees with the experimentally determined B-factors (Fig. 5B). We have also calculated the LacI elastic constants based on the anisotropic network model (ANM), which is essentially a coarse grained version of the Tirion potential. The results, shown in Fig. S6, Supporting Information, are equivalent to the atomistic ENM.

The results above suggest that modeling local interactions in simple ways, such as by using an ENM description, may be sufficient for the determination of slow, large-scale vibrations of

biomolecules, but will most likely provide a poor account of the effects of local perturbations applied to the system. Therefore, if the allosteric behavior of a specific protein relies mainly on its large-scale motions, then an ENM type of description will be sufficient to characterize it. However, if the type of interactions that dominate the allosteric behavior of the protein are local, then we believe a method such as the one presented here would provide more reliable results.

Finally, we note that normal mode analysis (NMA) of proteins is not limited to elastic network models, with the employment of all-atom forcefields being a possibility. However, NMA is limited by the size of the system due to the necessity of diagonalizing the $3N \times 3N$ Hessian matrix, where N is the number of atoms. Applications of NMA thus usually disregard solvent effects or use an implicit solvent model to describe solute–solvent interactions [54–56]. For the method presented here, calculations using an implicit solvent without the employment of cutoffs (necessary for convergence of the minimization) were considerably slower than those using explicit solvent. Furthermore, NMA uses information on partial derivatives of the potential energy function with respect to the atomic coordinates. The Hessian matrix thus contains information about thousands of independent perturbations applied upon the system. The calculation of local elastic constants outlined here is based upon single perturbations of the entire system and the evaluation of how individual atoms respond to this perturbation. Combined with the possibility to include an explicit solvent representation in a straightforward manner, we believe that the calculation of local elastic constants has the potential to be a more effective approach to studying effects of perturbations of biological systems than normal mode analysis.

3.6. Computational efficiency

The present results were obtained at a fraction of the computational cost necessary for an informative MD simulation of the same system, while using the same kind of elaborate forcefield description. Specifically, we would have been able to run approximately 10 ns of MD had we used the same computational power over the same amount of time for the task.

It must also be noted that our minimization algorithm/protocol can be optimized for increased performance. The minimization efficiency seems to significantly depend on the values of some parameters, as opposed to the systems studied in the original reference [34]. We have not yet carried a systematic investigation of this dependence, but we expect that fine-tuning of the parameters for proteins can greatly increase performance. We also note that most of the computational effort is spent with solvent minimization. The employment of different convergence criteria for solute and solvent could further reduce the required computational effort.

4. Conclusions

Significant structural changes of biomolecules are frequently necessary for biological function. Computationally, there are a number of approaches that can be used to determine which regions of a protein are more prone to deformation, with normal mode analysis of elastic network models being a frequently used approach. In this paper, we have adapted a method, previously developed for the calculation of local elastic constants of materials in the melt, for applicability to proteins in water, and used it to analyze a known allosteric protein, the *lac* repressor.

We have shown that this kind of analysis is relevant to a solvated protein, since its macroscopic elastic constant is well-defined for the magnitudes of the deformations applied here. When comparing results for a solvated protein to those for a polymeric system in the melt, the main differences observed are in the magnitude of the macroscopic elastic constant (solvated proteins are softer than polymers in the melt), and in the width of the distribution of local elastic constants. This difference in the width of the distribution was shown to be a consequence of the increased level of local conformational frustration in the protein. In addition, analysis of different secondary structural elements in the protein results in larger average stiffness for residues in β-sheets, when compared to α-helical residues. The construction of a stiffness map for the lactose repressor protein, by analyzing its local elastic constants, is able to highlight specific regions of the protein as having low stiffness values. More importantly, these regions are exactly the ones that suffer significant structural deformation upon effector binding, even though some of these regions are not in close proximity to the effector binding pocket.

This method offers an accurate and fast way to identify structural changes associated with allosteric transitions, without the need to have structures available for the active and inactive states. It is accurate because it employs an all-atom description of the molecule and an explicit description for the solvent, and it is fast when compared to other methods that employ the same all-atom resolution. The use of all-atom resolution was shown here to be essential for the correct identification of regions that deform during the allosteric transition of LacI, as the employment of an elastic network model failed to reproduce the all-atom results. We highlight that the results presented here come from application of this method to analysis of a single protein. Therefore, further studies need to be performed to determine whether the present method can be successfully applied to other systems, and to make generalizations about the results obtained here.

Acknowledgments

The authors would like to acknowledge Dr. Robert Riggleman for kindly providing a source code that was used as a basis for the implementation of the necessary routines in the GROMACS package. This material is based upon work supported by the National Science Foundation under Grant No. 1125698.

Appendix A. Supplementary Data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmgm.2015.01.013.

References

[1] J.-P. Changeux, 50th Anniversary of the Word “Allosteric”, Protein Sci. 20 (7) (2011) 1119–1124, http://dx.doi.org/10.1002/pro.658.
[2] J. Monod, F. Jacob, General conclusions: teleonomic mechanisms in cellular metabolism, growth, and differentiation Cold Spring Harbor Symposia on Quantitative Biology, vol. 26, Cold Spring Harbor Laboratory Press, 1961, pp. 389–401.
[3] J. Monod, J. Wyman, J.-P. Changeux, On the nature of allosteric transitions: a plausible model, J. Mol. Biol. 12 (1965) 88–118.
[4] D.E. Koshland, G. Némethy, D. Filmer, Comparison of experimental binding data and theoretical models in proteins containing subunits, Biochemistry 5 (1) (1966) 365–385.
[5] A. Szabo, M. Karplus, A mathematical model for structure-function relations in hemoglobin, J. Mol. Biol. 72 (1) (1972) 163–197.
[6] A.W.-M. Lee, M. Karplus, Structure-specific model of hemoglobin cooperativity, Proc. Natl. Acad. Sci. U. S. A. 80 (December) (1983) 7055–7059.
[7] E.R. Henry, S. Bettati, J. Hofrichter, W.A. Eaton, A tertiary two-state allosteric model for hemoglobin, Biophys. Chem. 98 (1–2) (2002) 149–164.
[8] Q. Cui, M. Karplus, Allostery and cooperativity revisited, Protein Sci. 17 (2008) 1295–1307, http://dx.doi.org/10.1110/ps.03259908.coined.
[9] A. Cooper, D. Dryden, Allostery without conformational change, Eur. Biophys. J. 10 (1984) 3–10, 9.
[10] K. Gunasekaran, B. Ma, R. Nussinov, Is allostery an intrinsic property of all dynamic proteins? Proteins: Struct. Funct. Bioinf. 57 (3) (2004) 433–443, http://dx.doi.org/10.1002/prot.20232.

[11] R.J. Hawkins, T.C.B. McLeish, Coarse-grained model of entropic allostery, Phys. Rev. Lett. 93 (9) (2004) 098104, http://dx.doi.org/10.1103/PhysRevLett.93.098104.
[12] D.R. Groebe, Screening for positive allosteric modulators of biological targets., Drug Discov. Today 11 (13–14) (2006) 632–639, http://dx.doi.org/10.1016/j.drudis.2006.05.010.
[13] N.M. Goodey, S.J. Benkovic, Allosteric regulation and catalysis emerge via a common route, Nat. Chem. Biol. 4 (8) (2008) 474–482, http://dx.doi.org/10.1038/nchembio.98.
[14] J.E. Lindsley, J. Rutter, Whence cometh the allosterome? Proc. Natl. Acad. Sci. U. S. A. 103 (28) (2006) 10533–10535.
[15] J. Bryngelson, P. Wolynes, Spin glasses and the statistical mechanics of protein folding, Proc. Natl. Acad. Sci. U. S. A. 84 (November) (1987) 7524–7528.
[16] A.A.S.T. Ribeiro, V. Ortiz, Determination of signaling pathways in proteins through network theory: importance of the topology, J. Chem. Theory Comput. 10 (2014) 1762–1769.
[17] G. Collier, V. Ortiz, Emerging computational approaches for the study of protein allostery, Arch. Biochem. Biophys. 538 (2013) 6–15.
[18] M. Karplus, J.A. McCammon, Molecular dynamics simulations of biomolecules., Nat. Struct. Biol. 9 (9) (2002) 646–652, http://dx.doi.org/10.1038/nsb0902-646.
[19] M. Tirion, Large amplitude elastic motions in proteins from a single-parameter, atomic analysis, Phys. Rev. Lett. 77 (9) (1996) 1905–1908.
[20] M.S. Formaneck, L. Ma, Q. Cui, v Reconciling the "old" and "new" views of protein allostery: a molecular simulation study of chemotaxis Y protein (CheY)., Proteins 63 (4) (2006) 846–867, http://dx.doi.org/10.1002/prot.20893.
[21] P.M. Gasper, B. Fuglestad, E.A. Komives, P.R.L. Markwick, Allosteric networks in thrombin distinguish procoagulant vs. anticoagulant activities, Proc. Natl. Acad. Sci. U. S. A. 109 (52) (2012) 21216–21222, http://dx.doi.org/10.1073/pnas.1218414109/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1218414109.
[22] G. Manley, I. Rivalta, J.P. Loria, Solution NMR and computational methods for understanding protein allostery, J. Phys. Chem. B 117 (11) (2013) 3063–3073, http://dx.doi.org/10.1021/jp312576v.
[23] A.T. Vanwart, J. Eargle, Z. Luthey-Schulten, R.E. Amaro, Exploring residue component contributions to dynamical network models of allostery, J. Chem. Theory Comput. 8 (8) (2012) 2949–2961, http://dx.doi.org/10.1021/ct300377a.
[24] J.D. Durrant, J.A. McCammon, Molecular dynamics simulations and drug discovery, BMC Biol. 9 (2011) 71, http://dx.doi.org/10.1186/1741-7007-9-71.
[25] Z. Bu, D.J.E. Callaway, Proteins move! Protein dynamics and long-range allostery in cell signaling, Adv. Protein Chem. Struct. Biol. 83 (2011) 163–221, http://dx.doi.org/10.1016/B978-0-12-381262-9.00005-7.
[26] N. Kantarci-Carsibasi, T. Haliloglu, P. Doruker, Conformational transition pathways explored by Monte Carlo simulation integrated with collective modes, Biophys. J. 95 (12) (2008) 5862–5873, http://dx.doi.org/10.1529/biophysj.107.128447.
[27] D. Ming, M.E. Wall, Quantifying allosteric effects in proteins., Proteins: Struct. Funct. Bioinf. 59 (4) (2005) 697–707, http://dx.doi.org/10.1002/prot.20440.
[28] R.A. Riggleman, J.F. Douglas, J.J. de Pablo, Antiplasticization and the elastic properties of glass-forming polymer liquids, Soft Matter 6 (2) (2010) 292–304, http://dx.doi.org/10.1039/b915592a.
[29] R.A. Riggleman, J.J. de Pablo, Antiplasticization and local elastic constants in trehalose and glycerol mixtures, J. Chem. Phys. 128 (2008) 224504.
[30] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M.R. Shirts, J.C. Smith, P.M. Kasson, D. van der Spoel, B. Hess, E. Lindahl, GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit, Bioinformatics 29 (7) (2013) 845–854, http://dx.doi.org/10.1093/bioinformatics/btt055.
[31] M. Zhou, A new look at the atomic level virial stress: on continuum-molecular system equivalence, Proc. R. Soc. A 459 (2037) (2003) 2347–2392, http://dx.doi.org/10.1098/rspa.2003.1127.
[32] M. Gerstein, J. Tsai, M. Levitt, The volume of atoms on the protein surface: calculated from simulation, using Voronoi polyhedra, J. Mol. Biol. 249 (5) (1995) 955–966, http://dx.doi.org/10.1006/jmbi.1995.0351.
[33] M.L. Falk, J.S. Langer, Dynamics of viscoplastic deformation in amorphous solids, Phys. Rev. E 57 (6) (1998) 7192–7205, http://dx.doi.org/10.1103/PhysRevE.57.7192.
[34] E. Bitzek, P. Koskinen, F. Gähler, M. Moseler, P. Gumbsch, Structural relaxation made simple, Phys. Rev. Lett. 97 (17) (2006) 170201.
[35] C.E. Bell, M. Lewis, A closer view of the conformation of the Lac repressor bound to operator, Nat. Struct. Biol. 7 (3) (2000) 209–214, http://dx.doi.org/10.1038/73317.
[36] Y. Duan, C. Wu, S. Chowdhury, M. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, et al., A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations, J. Comput. Chem. 24 (16) (2003) 1999–2012.
[37] M.E.J. Newman, Fast algorithm for detecting community structure in networks, Phys. Rev. E 69 (6) (2004) 066133.
[38] S. Perticaroli, J.D. Nickels, G. Ehlers, H. O'Neill, Q. Zhang, A.P. Sokolov, Secondary structure and rigidity in model proteins, Soft Matter 9 (2013) 9548–9556, http://dx.doi.org/10.1039/C3SM50807B.
[39] L. Swint-Kruse, K.S. Matthews, Allostery in the LacI/GalR family: variations on a theme, Curr. Opin. Microbiol. 12 (2) (2009) 129–137, http://dx.doi.org/10.1016/j.mib.2009.01.009.
[40] M. Lewis, The lac repressor, C. R. Biol. 328 (6) (2005) 521–548, http://dx.doi.org/10.1016/j.crvi.2005.04.004.
[41] S. Meyer, R. Ramot, K. Kishore Inampudi, B. Luo, C. Lin, S. Amere, C.J. Wilson, Engineering alternate cooperative-communications in the lactose repressor protein scaffold, Protein Eng. Des. Sel. 26 (6) (2013) 433–443, http://dx.doi.org/10.1093/protein/gzt013.
[42] R. Daber, S. Stayrook, A. Rosenberg, M. Lewis, Structural analysis of lac repressor bound to allosteric effectors, J. Mol. Biol. 370 (4) (2007) 609–619, http://dx.doi.org/10.1016/j.jmb.2007.04.028 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2715899&tool=pmcentrez&rendertype=abstract
[43] T. Flynn, L. Swint-Kruse, Y. Kong, C. Booth, K. Matthews, J. Ma, Allosteric transition pathways in the lactose repressor protein core domains: asymmetric motions in a homodimer, Protein Sci. 12 (2003) 2523–2541, http://dx.doi.org/10.1110/ps.03188303.central.
[44] J. Suckow, P. Markiewicz, L.G. Kleina, J. Miller, B. Kisters-Woike, B. Müller-Hill, Genetic studies of the Lac repressor. XV: 4000 single amino acid substitutions and analysis of the resulting phenotypes on the basis of the protein structure, J. Mol. Biol. 261 (4) (1996) 509–523, http://dx.doi.org/10.1006/jmbi.1996.0479.
[45] Z.N. Gerek, S.B. Ozkan, Change in allosteric network affects binding affinities of PDZ domains: analysis through perturbation response scanning, PLoS Comput. Biol. 7 (10) (2011) e1002154, http://dx.doi.org/10.1371/journal.pcbi.1002154.
[46] D.U. Ferreiro, J.A. Hegler, E.A. Komives, P.G. Wolynes, On the role of frustration in the energy landscapes of allosteric proteins, Proc. Natl. Acad. Sci. U. S. A. 108 (9) (2011) 3499–3503, http://dx.doi.org/10.1073/pnas.1018980108.
[47] O.N.A. Demerdash, M.D. Daily, J.C. Mitchell, Structure-based predictive models for allosteric hot spots, PLoS Comput. Biol. 5 (10) (2009) e1000531, http://dx.doi.org/10.1371/journal.pcbi.1000531.
[48] A. Panjkovich, X. Daura, Exploiting protein flexibility to predict the location of allosteric sites, BMC Bioinf. 13 (2012) 273, http://dx.doi.org/10.1186/1471-2105-13-273.
[49] M. Ikeguchi, J. Ueno, M. Sato, A. Kidera, Protein structural change upon ligand binding: linear response theory, Phys. Rev. Lett. 94 (7) (2005) 078102, http://dx.doi.org/10.1103/PhysRevLett.94.078102.
[50] W. Zheng, B.R. Brooks, Normal-modes-based prediction of protein conformational changes guided by distance constraints, Biophys. J. 88 (5) (2005) 3109–3117, http://dx.doi.org/10.1529/biophysj.104.058453 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1305462&tool=pmcentrez&rendertype=abstract
[51] P. Weinkam, J. Pons, A. Sali, Structure-based model of allostery predicts coupling between distant sites, Proc. Natl. Acad. Sci. U. S. A. 109 (13) (2012) 4875–4880, http://dx.doi.org/10.1073/pnas.1116274109 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3324024&tool=pmcentrez&rendertype=abstract
[52] D.J. Jacobs, A.J. Rader, L.A. Kuhn, M.F. Thorpe, Protein flexibility predictions using graph theory, Proteins 44 (2) (2001) 150–165 http://www.ncbi.nlm.nih.gov/pubmed/11391777
[53] K. Hinsen, The molecular modeling toolkit: a new approach to molecular simulations, J. Comput. Chem. 21 (2) (2000) 79–85.
[54] I. Adamovic, S.M. Mijailovich, M. Karplus, The elastic properties of the structurally characterized myosin II S2 subdomain: a molecular dynamics and normal mode analysis, Biophys. J. 94 (10) (2008) 3779–3789, http://dx.doi.org/10.1529/biophysj.107.122028.
[55] T. Gaillard, E. Martin, E. San Sebastian, F.P. Cossío, X. Lopez, A. Dejaegere, R.H. Stote, Comparative normal mode analysis of LFA-1 integrin I-domains, J. Mol. Biol. 374 (1) (2007) 231–249, http://dx.doi.org/10.1016/j.jmb.2007.07.006.
[56] N. Reuter, K. Hinsen, J.-J. Lacapère, Transconformations of the SERCA1 Ca-ATPase: a normal mode study, Biophys. J. 85 (4) (2003) 2186–2197, http://dx.doi.org/10.1016/S0006-3495(03)74644-X.