



# Computer Simulation of Protein Materials at Multiple Length Scales: From Single Proteins to Protein Assemblies

Kilho Eom<sup>1</sup>

Received: 7 August 2018 / Accepted: 12 September 2018 / Published online: 11 January 2019  
© Korean Multi-Scale Mechanics (KMSM) 2019

## Abstract

Computer simulation of protein materials and their dynamic or mechanical behavior is of high significance, as proteins perform their functions through their structural changes in response to a force (or stimulus). The computer simulation enables the detailed insight into the structure and behavior of proteins at atomistic resolution, which is inaccessible with experimental toolkits such as single-molecule experiments. With the advancement of computing resources, the computer simulation has recently played a vital role as a virtual microscopy in understanding and characterizing the structure and behaviors of protein materials at atomistic resolution. For examples, computer simulations allow for gaining insight into how some protein domains can exhibit the remarkable mechanical properties and functions. In this article, we would like to address the current state-of-arts of computer simulations that have been employed for studying the structure and properties (or behaviors) of proteins at multiple length scales ranging from single proteins to protein assemblies. Specifically, we summarize various computational modeling techniques, ranging from atomistic models to coarse-grained models and continuum models, applicable for modeling protein structures at multiple length scales from single proteins to protein assemblies. This paper discusses how such various computational modeling/simulation techniques can be employed for studying the structure and properties of protein materials at multiple length scales. This paper sheds light on computer simulations, which are able to unveil the hidden, complex mechanisms related to the structure and properties of protein materials at multiple length scales.

**Keywords** Computer simulation · Protein materials · Multiple length scales · Atomistic simulation · Coarse-grained modeling · Protein mechanics

## Introduction

It is of high importance to study the dynamic (or mechanical) behavior of a single protein molecule, as a protein molecule performs its function through its structural changes upon a chemical stimulus (e.g. ligand binding) and/or a force. For example, hemoglobin transports the oxygen when the oxygen is bound to the hemoglobin. The ability of hemoglobin to bind with oxygen is attributed to the existence of the hemoglobin's various conformations (structures), few of which may correspond to the conformation of oxygen-bound hemoglobin. In other words, a protein molecule can explore many conformations, some of which corresponds

to the ligand-bound state, because the energy landscape of a protein is very rugged in that there are multiple local minima (or equilibria), and that the energy barrier between these minima is comparable to the amount of thermal energy [1–3]. This unique dynamic feature of a protein molecule enables it to perform its function such as ligand binding or unbinding. Moreover, some protein domains are able to perform remarkable mechanical functions and to exhibit excellent mechanical properties. For instance, an immunoglobulin (Ig) domain in a muscle protein chain titin has a high resilience against a mechanical force in such a way that hydrogen bonds formed between the  $\beta$ -strands of an Ig domain are able to sustain a mechanical force [4–7]. In addition, the high resilience of a protein domain is also attributed to its unfolding behavior, that is, a force-driven unfolding behavior of a protein folded domain results in its energy dissipation, and consequently, the high mechanical resilience [8]. These two representative examples suggest the necessity of studying

✉ Kilho Eom  
kilhoem@skku.edu

<sup>1</sup> Biomechanics Laboratory, College of Sport Science, Sungkyunkwan University (SKKU), Suwon 16419, Republic of Korea

the dynamic (or mechanical) behavior of proteins in order to gain a fundamental insight into their functions.

Protein assemblies have recently received significant attention due to their important role in disease pathology [9–12] as well as their applicability in bionanotechnology [13–15]. As a representative example, amyloid fibrils have been highlighted due to their role in the pathogenesis of various diseases ranging from neurodegenerative diseases [16, 17] to type 2 diabetes [18] and cardiovascular diseases [19]. The amyloid fibrils are formed through self-assembly process, that is, amyloidogenic proteins (e.g. denatured proteins) are aggregated so as to form the fibril structure [10]. Once amyloid fibrils are formed by self-assembly process, they are quite rigid and structurally stable in a physiological condition, resulting in the formation of amyloid plaques. These fibrils have recently suggested to play a role in disease pathologies due to their remarkable mechanical properties. Specifically, the elastic modulus of amyloid fibrils is in a range of 1–10 GPa [20–22], which is a few orders of magnitude higher than the elastic modulus of cell membrane. The growth of an amyloid fibril on a cell membrane results in the mechanical disruption of cell membrane [23] due to the difference between elastic moduli of amyloid fibril and cell membrane [24]. In addition, a spider silk crystal, whose molecular structure is similar to that of amyloid fibril, possesses anomalous mechanical properties whose elastic modulus is in the order of 1–10 GPa [25]. The remarkable properties of these protein assemblies (e.g. amyloid fibril, and spider silk crystal) led researchers to study the underlying principles showing how the excellent mechanical properties of protein assemblies can be determined.

For last two decades, single-molecule experimental techniques [26] based on atomic force microscopy (AFM) [27–29] and/or optical tweezer [30] have played a central role in probing the properties and functions of protein molecules. As an example, AFM-based single-molecule pulling experiments have allowed for gaining a quantitative insight into the response of a protein domain to a mechanical force. Specifically, AFM-based pulling experiments are able to capture the force-driven unfolding event of a protein domain based on measuring the force–extension (equivalent to force–displacement) relationship of a protein domain during its unfolding process. Despite the ability of single-molecule experimental techniques to quantitatively probe the properties of protein molecules, they are very restrictive in that they are unable to provide how the molecular structure of a protein molecule is changed (e.g. deformed) due to a force. In addition, these experimental techniques cannot unveil the underlying principles showing how the properties and behaviors of protein materials are determined. Due to these limitations of single-molecule experimental techniques, computer simulations can play an alternative role in revealing the underlying mechanisms of the dynamic (or

mechanical) behaviors and properties of protein materials at atomistic resolution. In particular, computer simulations based on atomistic models are also able to provide a quantitative insight into the relationship between the molecular structure of protein materials and their properties [31, 32]. Furthermore, computer simulations are capable of quantitatively understanding how the molecular structure of protein materials (e.g. protein assemblies) can be determined. The computer simulations can be implemented based on different modeling techniques at different length scales, ranging from atomistic model to coarse-grained model and continuum model. Here, we note that this paper excludes the details of a continuum mechanics-based simulation, which has recently been applied for studying the dynamics of proteins [33, 34], while we briefly review the continuum model that is useful in analyzing the result of computer simulations based on atomistic or coarse-grained models.

In this work, we address the current state-of-arts of computational simulations of protein materials at multiple length scales, ranging from a protein domain to a protein assembly, for unveiling the hidden, complex mechanisms of how the remarkable properties (or functions) of protein materials are determined. In particular, we describe the current state-of-arts of computational simulations at multiple length scales based on atomistic model, coarse-grained model, and/or continuum modeling, which are applicable for understanding the structure–property–function relationship of protein materials at different length scales ranging from a single protein domain to protein assembly. In addition, we present the computer simulation-based analyses of single proteins such as their unfolding mechanics and/or conformational changes. Moreover, we introduced the computer simulation-based characterization of the structures and properties of protein assemblies such as amyloid fibrils and viral capsids. Finally, we conclude our paper with some remarks and outlooks for future works in computational modeling of protein materials.

## Computational Methodologies

This Section demonstrates the fundamental principles of computational methodologies that have been applied for modeling protein materials. Here, we present the computational simulation methodologies at different length scales, based on atomistic model or coarse-grained model.

### Atomistic Modeling/Simulation

#### Molecular Dynamics Simulations

The atomistic molecular dynamics (MD) simulation had been dated back to 1950s, when Alder and Wainwright [35]

first considered the atomistic modeling for studying the phase transition of hard-sphere gas system. From 1950s to 1960s, the MD simulations were extensively employed for studying the phase transition of gas and solution. Not until 1970s were the MD simulations employed for modeling biological molecules such as DNA, RNA, and protein. In 1977, Karplus and colleagues [36] had utilized the MD simulations for studying the conformational dynamics of bovine trypsin inhibitor (that is a very small protein molecule). In 1970s, Karplus [37], Levitt [38], and Warshel studied the dynamics of biological molecules (that is related to their function) by using the MD simulations, which led them to become a pioneer in the MD simulation-based study of biological molecules and their dynamics. The pioneering works of Karplus, Levitt, and Warshel allowed them to win a Nobel prize in Chemistry 2013 [39]. Since 2000s, the MD simulations have been extensively used for gaining a quantitative insight into the structure, dynamics, properties, and functions of biological molecules such as proteins, DNAs, and RNAs [40–43].

The fundamental principle of atomistic MD simulation is to numerically solve the Newton's equation of motion for a molecular system composed of atoms. Specifically, the position vector of each atom, which is a component of a molecular system, is numerically obtained as a function of time based on the numerical integration of the Newton's equation of motion with computing forces acting on each atom. Here, the forces acting on each atom originate from inter-atomic interaction, thermal energy, and interaction between the molecular system and an environment (e.g. water molecules) surrounding the molecular system. The equation of motion for the molecular system is written as [44]

$$m_i \partial_t^2 \mathbf{r}_i = -\nabla_i V(\mathbf{r}_1, \dots, \mathbf{r}_N) \quad (1)$$

where  $\mathbf{r}_i$  is the position vector of an  $i$ -th atom in a molecular system,  $m_i$  is the mass of the  $i$ -th atom,  $N$  is the total number of atoms for the molecular system,  $V$  is the potential energy of the molecular system, and a symbol  $\nabla_i$  represents a spatial gradient with respect to position vector  $\mathbf{r}_i$ . With a given initial position vector  $\mathbf{r}_i(0)$  for the molecular system, the time-dependent position vector  $\mathbf{r}_i(t)$  of the molecular system can be obtained by the numerical integration of the equation of motion given by Eq. (1). The details of numerical integration scheme are well described in ref [44]. The dynamic behavior of a biological molecule (e.g. protein) can be analyzed based on calculating the fluctuation matrix  $\mathbf{Q}$  that consists of block matrices  $\mathbf{Q}_{ij}$  defined as [45]

$$\mathbf{Q}_{ij} = \langle [\mathbf{r}_i(t) - \langle \mathbf{r}_i(t) \rangle] \otimes [\mathbf{r}_j(t) - \langle \mathbf{r}_j(t) \rangle] \rangle \quad (2)$$

where a symbol  $\otimes$  indicates a tensor product, and an angle bracket represents an ensemble average. The root mean-square fluctuation (RMSF) of an  $i$ -th atom can be measured as follows:

$$\begin{aligned} \xi_i &= \sqrt{\text{tr} \mathbf{Q}_{ii}} = \sqrt{\langle [\mathbf{r}_i(t) - \langle \mathbf{r}_i(t) \rangle] \otimes [\mathbf{r}_i(t) - \langle \mathbf{r}_i(t) \rangle] \rangle} \\ &= \sqrt{\langle [x_i(t) - \langle x_i(t) \rangle]^2 + [y_i(t) - \langle y_i(t) \rangle]^2 + [z_i(t) - \langle z_i(t) \rangle]^2 \rangle}. \end{aligned} \quad (3)$$

Here,  $x_i$ ,  $y_i$ , and  $z_i$  represent the components of a position vector  $\mathbf{r}_i(t)$  in the  $x$ -,  $y$ -, and  $z$ -directions, respectively. The RMSF of a protein is very useful in not only understanding the fluctuation motion of the protein but also validating the atomistic model by fitting the RMSF obtained from the MD simulation to that acquired from experiments based on nuclear magnetic resonance or X-ray crystallography. In addition, the correlated motion of a protein can be quantitatively described using a correlation factor (CF) defined as

$$c_{ij} = \text{tr} \mathbf{Q}_{ij} / \sqrt{\text{tr} \mathbf{Q}_{ii} \cdot \text{tr} \mathbf{Q}_{jj}}. \quad (4)$$

It should be noted that the value of CF is in a range of  $-1$  to  $1$ . The value of CF being  $+1$  indicates that two atoms  $i$  and  $j$  move collectively (i.e. together) in the same direction, while the value of CF being  $-1$  means that these two atoms move in the opposite direction, and the value of CF being zero implies that the motion of an atom  $i$  is not correlated with that of other atom  $j$ .

For a quantitatively analysis of the mechanical response of a protein to a force (as can be obtained from single-molecule pulling experiment), the potential field should be modified by including the strain energy stored in a force probe that is used to pull the protein.

$$U = V(\mathbf{r}_1, \dots, \mathbf{r}_N) + \frac{K}{2} (R_{ij} - vt)^2 \quad (5)$$

where  $K$  is the stiffness of a force probe,  $R_{ij}$  is a distance between two atoms, one of which is fixed and the other is pulled by a force probe,  $v$  is a pulling rate, and  $t$  is the time. The equation of motion for a protein that is pulled by a force probe is given by

$$m_i \partial_t^2 \mathbf{r}_i(t) = -\nabla_i U(\mathbf{r}_1, \dots, \mathbf{r}_N, t). \quad (6)$$

The numerical simulation of solving the Newton's equation of motion given by Eq. (6) for studying the mechanical response of a protein is referred to as steered molecular dynamics (SMD) simulations [46]. The SMD simulations enables the acquisition of the force–displacement relationship of a biomolecule (e.g. DNA, RNA, and proteins), and consequently, the extraction of the mechanical properties of a biomolecule.

### Normal Mode Analysis

Over the last two decades, normal mode analysis (NMA) has played a pivotal role in understanding the dynamic

behavior of a protein structure [47], as the NMA coupled with statistical mechanics theory has allowed for calculating the fluctuation matrix of the protein structure as long as its stiffness matrix is evaluated. For an insight into a relationship between the fluctuation behavior of a protein and its stiffness, we consider a simple molecular system composed of two identical atoms (i.e. diatomic molecule). The potential field of a diatomic molecule is approximated as harmonic potential given by  $V = (\gamma/2)u^2$ , where  $\gamma$  is a force constant, and  $u$  is a distance between two atoms. The kinetic energy of the diatomic molecule is given as  $L = (m/2)v^2$ , where  $m$  is the atomic mass, and  $v$  is the velocity defined as  $v = du/dt$ . The statistical mechanics theory provides a probability distribution,  $p(u, v)$ , to find a specific position and velocity ( $u, v$ ) on the phase space given as

$$p(u, v) = Z^{-1} \exp(-H/k_B T) = Z^{-1} \exp(-\{L + V\}/k_B T) \quad (7)$$

where  $k_B$  and  $T$  are the Boltzmann's constant and absolute temperature, respectively,  $H$  is the Hamiltonian of the diatomic molecule, and  $Z$  is the partition function defined as

$$Z = \iint_{-\infty}^{\infty} \exp[-H(u, v)/k_B T] dudv. \quad (8)$$

The ensemble averages of the distance and the mean-square distance are obtained as follows:

$$\langle u \rangle = Z^{-1} \iint_{-\infty}^{\infty} u \cdot \exp[-\{L(v) + V(u)\}/k_B T] dudv = 0 \quad (9.a)$$

$$\langle u^2 \rangle = Z^{-1} \left[ \int_{-\infty}^{\infty} \exp(-mv^2/2k_B T) dv \right] \cdot \left[ \int_{-\infty}^{\infty} u^2 \exp(-\gamma u^2/2k_B T) du \right] = \gamma^{-1} k_B T. \quad (9.b)$$

The statistical mechanics theory given by Eq. (9.b) suggests that the fluctuation behavior of a molecular system is dependent on its stiffness in such a way that the mean-square fluctuation of the system is inversely proportional to its stiffness.

Now, we consider a molecular system that is composed of  $N$  atoms. For the dynamic motion of a protein near its equilibrium state, its potential field can be approximated to a quasi-harmonic potential such as [48–50]

$$V = \frac{1}{2} \mathbf{r}^T \mathbf{K} \mathbf{r} \quad (10)$$

where  $\mathbf{K}$  is the stiffness matrix of a molecular system,  $\mathbf{r}$  is the atomic coordinates of the system, and a super-script  $T$  indicates a transpose of a vector. With a spectral

decomposition of the stiffness matrix  $\mathbf{K}$  such as  $\mathbf{K} = \mathbf{P}^T \Lambda \mathbf{P}$ , where  $\Lambda$  is a diagonal matrix consisting of eigenvalues  $\lambda_j$  for the stiffness matrix and  $\mathbf{P}$  is the modal matrix, the potential energy can be rewritten as

$$V = \frac{1}{2} \mathbf{y}^T \Lambda \mathbf{y} = \frac{1}{2} \sum_{j=1}^{3N-6} \lambda_j y_j^2. \quad (11)$$

Here,  $\mathbf{y}$  is a transformed coordinate defined as  $\mathbf{y} = \mathbf{P} \mathbf{r}$ . It should be noted that summation in Eq. (11) excludes six zero normal modes (i.e. rigid body modes). The statistical mechanics theory provides a probability function  $p(\mathbf{y})$  to find a specific state  $\mathbf{y}$  given by

$$p(\mathbf{y}) = Z^{-1} \exp \left[ -\frac{1}{2} (k_B T)^{-1} \sum_{j=1}^{3N-6} \lambda_j y_j^2 \right]. \quad (12)$$

From a probability function  $p(\mathbf{y})$  given by Eq. (12), the ensemble averages of quantities,  $y_i$  and  $y_i \cdot y_j$ , can be found as follows.

$$\langle y_i \rangle = \iint_{-\infty}^{\infty} y_i \cdot p(\mathbf{y}) d\mathbf{y} = 0 \quad (13.a)$$

$$\langle y_i \cdot y_j \rangle = \int_{-\infty}^{\infty} y_i \cdot y_j \cdot p(\mathbf{y}) d\mathbf{y} = \lambda_i^{-1} k_B T \cdot \delta_{ij} \quad (13.b)$$

where  $\delta_{ij}$  is a Kronecker delta defined as  $\delta_{ij} = 1$  if  $i = j$ , otherwise  $\delta_{ij} = 0$ . Here, we note that the repeated index in Eq. (13.b) does not imply the Einstein summation rule. With using a transformation  $\mathbf{r} = \mathbf{P}^T \mathbf{y}$ , the fluctuation matrix  $\mathbf{Q}$  of a molecular system can be written as

$$\mathbf{Q} = \sum_{j=1}^{3N-6} \lambda_j^{-1} k_B T \mathbf{P}_j \otimes \mathbf{P}_j = \mathbf{P}^T \Lambda^{-1} \mathbf{P} \quad (14)$$

where  $\mathbf{P}_j$  is the  $j$ -th column vector of a modal matrix  $\mathbf{P}$ , that is,  $\mathbf{P}_j$  is the  $j$ -th normal mode of the system.

## Coarse-Grained Modeling

### Principle of Coarse-Graining

For more than three decades, coarse-grained (CG) modeling techniques have been established and widely utilized for understanding the dynamic behavior of a large protein structures [51–53]. Though a full atomistic MD simulations have recently allowed for gaining insight into the dynamics of a small protein (consisting of less than few hundreds of residues) at millisecond time scales [54, 55], the full atomistic

MD simulations are still computationally restrictive in characterizing the dynamic behavior of a large protein structures at long timescale [56]. In order to overcome the computational limitation of atomistic MD simulations, CG modeling approaches have been widely developed in such a way that a group of atoms, which move collectively, are replaced by a single bead (particle), and that the motion of such beads (that represent the groups of atoms) is numerically predicted based on Newtonian equation of motion while the interaction potential energies between beads have to be properly defined.

The process to construct a CG model from atomistic model is as follows: As briefly demonstrated above, a group of atoms are replaced with a single bead as long as such atoms undergo a similar (or mostly collective) motion. Specifically, a group of atoms are selected based on a criterion given by Eq. (15)

$$\zeta^2 = \tau^{-1} \int_0^\tau \sum_{i,j \in \Omega} |\mathbf{r}_i(t) - \mathbf{r}_j(t)| < \varepsilon \tag{15}$$

where  $\Omega$  represents the set of atoms that are grouped, and  $\tau$  is a short simulation time scale (e.g.  $\sim 1$  ns). Here, all-atom MD simulation is used with a short timescale (e.g.  $\sim 1$  ns) in order to measure a criterion given by Eq. (15). Once the group of atoms are replaced with a single bead, its position vector can be obtained from a following equation.

$$\mathbf{R}_j(t) = \frac{\sum_{l=1}^{n_j} m_l \mathbf{r}_l(t)}{\sum_{l=1}^{n_j} m_l} \tag{16}$$

Here, subscript  $j$  indicates the  $j$ -th group of atoms,  $m_l$  and  $\mathbf{r}_l$  are the molecular weight and position vector of the  $l$ -th atom, respectively, and  $n_j$  is the number of atoms in the  $j$ -th group. It should be noted that the position vector of a bead (representing a group of atoms) is equivalent to the center of mass for the atoms that are grouped. Once the groups of atoms are replaced with CG beads, the effective potential field that can be prescribed to CG beads has to be defined. The effective potential field can be written as

$$V = \sum_{j=1}^{N-1} \frac{k_D}{2} (D_j - D_j^0)^2 + \sum_{j=1}^{N-2} \frac{k_\Omega}{2} (\Omega_j - \Omega_j^0)^2 + \sum_{j=1}^{N-3} \frac{k_\Phi}{2} (\Phi_j - \Phi_j^0)^2 + \sum_{j=1}^{N-1} \sum_{k=j+1}^N X_{jk}^{-1} A Q_j Q_k + \sum_{j=1}^{N-1} \sum_{k=j+1}^N 4E_0 \left[ (\Upsilon/X_{jk})^6 - (\Upsilon/X_{jk})^{12} \right] \tag{17}$$

where  $k_D$ ,  $k_\Omega$ , and  $k_\Phi$  are the force constants for strain energies corresponding to the stretching of a pseudo-covalent bond, bending of a pseudo-bond angle, and twisting of a pseudo-dihedral angle, respectively,  $N$  is the total number of CG beads,  $D_j$  is the distance of a pseudo-covalent bond such as  $D_j = |\mathbf{D}_j| = |\mathbf{R}_{j+1} - \mathbf{R}_j|$ ,  $\Omega_j$  is a pseudo-bond angle defined as  $\cos \Omega_j = \mathbf{D}_j \cdot \mathbf{D}_{j+1} / D_j D_{j+1}$ ,  $\Phi_j$  is a pseudo-dihedral angle

given as  $\cos \Phi_j = (\mathbf{D}_j \times \mathbf{D}_{j+1}) \cdot \mathbf{D}_{j+2} / (|\mathbf{D}_j \times \mathbf{D}_{j+1}| \cdot |\mathbf{D}_{j+2}|)$ ,  $Q_j$  is an effective charge acting on a  $j$ -th CG bead,  $X_{jk}$  is a distance between two CG beads  $j$  and  $k$  such as  $X_{jk} = |\mathbf{R}_j - \mathbf{R}_k|$ ,  $E_0$  is the energy scale of van der Waal’s interaction between CG beads, and  $\Upsilon$  is a length scale for van der Waal’s interaction. The major procedure of CG modeling is to determine the parameters of effective potential field given by Eq. (17). The effective potential field can be validated by comparing the results of CG model-based simulations with those of atomistic MD simulations.

Once CG model is constructed as described above, the dynamics of a CG model can be analyzed based on either NMA or MD-like simulation. In particular, the NMA can be applied to the CG model while the stiffness matrix of CG model is estimated based on the position vector of CG beads. Here, the stiffness matrix of CG model ( $\mathbf{K}$ ) is composed of (off-diagonal) block matrices  $\mathbf{K}_{ij}$  given by

$$\mathbf{K}_{ij} = -\nabla_i \nabla_j V(\mathbf{R}_1, \dots, \mathbf{R}_N). \tag{18}$$

Here, gradient  $\nabla_i$  is defined with respect to the coordinates of CG beads, such as  $\nabla_i = \partial / \partial \mathbf{R}_i$ . As a consequence, once the eigen-values and their corresponding normal modes are computed, the statistical mechanics theory given by Eq. (14) enables the quantitation of the fluctuation matrix  $\mathbf{Q}$  for the CG model.

On the other hand, with empirical potential field prescribed to the CG model given as Eq. (17), the Newtonian equation of motion for the CG model can be described as follows.

$$M_i \partial_t^2 \mathbf{R}_i = -\nabla_i V(\mathbf{R}_1, \dots, \mathbf{R}_N). \tag{19}$$

Here,  $M_i$  is the effective molecular weight of the  $i$ -th CG bead. The time-dependent trajectories of CG beads are obtained by numerically integrating the Newtonian equation of motion given by Eq. (19). This method is referred to as coarse-grained MD (CG-MD) simulations.

### Go-Like Model

In late 1970s, Go and coworkers [57–59] had developed a CG model in such a way that the alpha carbons of a protein structure are selected to represent the CG beads of the CG model, while the other atoms are neglected. The empirical potential field prescribed to CG beads is defined as follows.

$$V = \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left[ \left\{ \frac{k_1}{2} (d_{ij} - d_{ij}^0)^2 + \frac{k_2}{4} (d_{ij} - d_{ij}^0)^4 \right\} \delta_{j,i+1} + 4e_0 \left\{ (\sigma/d_{ij})^{12} - (\sigma/d_{ij})^6 \right\} (1 - \delta_{j,i+1}) \right] \tag{20}$$

where  $N$  is the total number of CG beads,  $d_{ij}$  is the distance between two CG beads  $i$  and  $j$ ,  $k_1$  and  $k_2$  are force constants for harmonic and quartic potential, respectively, for covalent bond,  $e_0$  and  $\sigma$  are the energy parameter and length scale, respectively, for non-bonded interaction, superscript 0 indicates the equilibrium state, and  $\delta_{i,j}$  is a Kronecker delta defined as  $\delta_{i,j} = 1$  if  $i = j$ ; otherwise,  $\delta_{i,j} = 0$ .

For recent decades, Go-like model has been extensively employed for studying the mechanical response of a protein molecule due to a force. For simulating the force-driven structural change of a protein molecule, the Langevin equation of motion was considered as follows [60].

$$m_i \partial_t^2 \mathbf{r}_i = -\nabla_i U(\mathbf{r}) - \eta \partial_t \mathbf{r}_i + \mathbf{f}_i(t). \quad (21)$$

Here,  $m_i$  and  $\mathbf{r}_i$  are the mass and position vector of an  $i$ -th CG bead, respectively,  $\eta$  is a friction coefficient,  $\mathbf{f}_i(t)$  is a Gaussian random force acting on the  $i$ -th CG bead, and  $U(\mathbf{r})$  is an effective potential field defined as  $U = V + V_L$  with  $V$  and  $V_L$  being the potential field of the CG protein structure given by Eq. (20) and a potential field for a force probe, respectively. It should be noted that  $\mathbf{f}_i(t)$  satisfies the fluctuation–dissipation theorem such as  $\langle \mathbf{f}_i(t) \cdot \mathbf{f}_i(\tau) \rangle = 2\eta k_B T \delta(t - \tau)$ , where  $k_B$  and  $T$  are the Boltzmann's constant and absolute temperature, respectively, an angle bracket  $\langle \rangle$  indicates an ensemble average (equivalent to the time average for the sufficient long-time simulation for an ergodic system), and  $\delta(t)$  is the Dirac delta function. In addition, we note that the potential field of a force probe is typically assumed to a harmonic potential given as  $V_L = (K/2)(d_{1,N} - vt)^2$ , where  $v$  is a pulling speed with which a protein is extended, and  $K$  is the stiffness of a force probe. By numerically solving the Langevin equation given by Eq. (21), one is able to obtain the dynamic trajectory of a coarse-grained protein structure in response to a force-driven extension with a given pulling speed, which results in gaining the force–extension (i.e. force–displacement) curve for a protein structure, and consequently, its physical properties (e.g. elastic modulus).

### Elastic Network Model

Though the Go-like model [57–59] is able to depict the dynamic and mechanical response of a protein structure, it is still computationally limited for studying the dynamic behavior of a large protein complex (or assembly), whose size is much larger than that of a single protein. In order to reduce a computational cost for simulating the dynamic behavior of large protein complex, Tirion [61] introduced an elastic network model (ENM), which assumes a harmonic potential for neighboring alpha carbon atoms. Bahar and colleagues [62] developed a Gaussian network model (GNM), in which a protein structure is described as a network of Gaussian chains that are widely used in rubber elasticity. Despite the inability of GNM to dictate the directionality of normal modes, GNM

was extensively utilized for studying the fluctuation dynamics and its correlated functions of large protein complexes. Atilgan et al. [63] introduced an anisotropic network model (ANM) which is able to depict the anisotropic fluctuation behavior of a protein structure. Here, we note that ANM is conceptually identical to ENM. The normal modes of ENM was found to be useful in predicting the conformational changes of a large protein complex due to a stimulus (such as ligand-binding).

As described above, the key concept of ENM modeling is to construct a harmonic potential field prescribed to alpha carbon atoms such as [61–65]

$$V = \frac{\gamma}{2} \sum_{i=1}^{N-1} \sum_{j=i+1}^N (d_{ij} - d_{ij}^0)^2 \cdot H(R_c - d_{ij}^0) \quad (22)$$

where  $\gamma$  is a force constant,  $R_c$  is a cut-off distance (usually given by  $R_c = 7\text{--}10 \text{ \AA}$ ), and  $H(x)$  is a Heaviside unit step function defined as  $H(x) = 1$  if  $x > 0$ ; otherwise,  $H(x) = 0$ .

In case of GNM, the Kirchhoff matrix of a protein structure is given by [62]

$$\Gamma_{ij} = \left. \frac{\partial^2 V}{\partial u_i \partial u_j} \right|_{\mathbf{r}=\mathbf{r}^0} = -\gamma H(R_c - d_{ij}^0) \cdot (1 - \delta_{ij}) + \delta_{ij} \sum_{k \neq i} \gamma H(R_c - d_{ik}^0). \quad (23)$$

Here,  $u_i$  is the fluctuation of the  $i$ -th alpha carbon atom (i.e. CG bead) defined as  $u_i = |\mathbf{r}_i - \mathbf{r}_i^0|$  with  $\mathbf{r}_i$  being the position vector of the  $i$ -th alpha carbon atom. Consequently, as described earlier, the fluctuation matrix of a protein structure can be obtained as follows.

$$\mathbf{Q}_{ij}^{GNM} = \left\langle (\mathbf{r}_i - \mathbf{r}_i^0) \cdot (\mathbf{r}_j - \mathbf{r}_j^0) \right\rangle = \sum_{k=2}^N \lambda^{-1} k_B T \mathbf{q}_i^k \mathbf{q}_j^k \quad (24)$$

where  $\lambda_k$  is the  $k$ -th eigen-value of the Kirchhoff matrix, and  $\mathbf{q}_i^k$  is the  $i$ -th component of the  $k$ -th normal mode of the Kirchhoff matrix. With obtaining the fluctuation matrix from GNM, one is able to easily evaluate the RMSF and CF with using Eqs. (3) and (4), which are useful in describing the fluctuation motion and correlated motion of a protein structure.

On the other hand, ANM (or ENM) assumes the anisotropic fluctuation behavior of a protein structure, while GNM presumes the isotropic fluctuation of the protein structure. The harmonic potential given by Eq. (22) can be rewritten as [63]

$$V = \frac{1}{2} \sum_{ij} (\mathbf{u}_i - \mathbf{u}_j)^T \mathbf{K}_{ij} (\mathbf{u}_i - \mathbf{u}_j) \quad (25)$$

where  $\mathbf{u}_i$  is the fluctuation vector of the  $i$ -th alpha carbon atom defined as  $\mathbf{u}_i = \mathbf{r}_i - \mathbf{r}_i^0$  with  $\mathbf{r}_i$  and superscript 0 being the position vector of the  $i$ -th alpha carbon atom and equilibrium state, respectively, and  $\mathbf{K}_{ij}$  is the off-diagonal ( $3 \times 3$ ) stiffness matrix given by

$$\mathbf{K}_{ij} = -\gamma \left| \mathbf{d}_{ij}^0 \right|^{-2} H \left( R_c - d_{ij}^0 \right) \cdot \mathbf{d}_{ij}^0 \otimes \mathbf{d}_{ij}^0 \tag{26}$$

Here,  $\mathbf{d}_{ij}$  is a distance vector (between two alpha carbon atoms  $i$  and  $j$ ) defined as  $\mathbf{d}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ , and a symbol  $\otimes$  indicates a tensor product. Once the off-diagonal stiffness matrices are estimated, it is straightforward to obtain the  $(3N \times 3N)$  stiffness matrix  $\mathbf{K}$  for the protein structure. As mentioned earlier, the fluctuation matrix of a protein structure can be acquired as follows.

$$\mathbf{Q} = \langle (\mathbf{r} - \mathbf{r}^0) \otimes (\mathbf{r} - \mathbf{r}^0) \rangle = \sum_{k=7}^{3N} \xi_k^{-1} k_B T \mathbf{v}^k \otimes \mathbf{v}^k \tag{27}$$

where  $\mathbf{r}$  is the position vector of all alpha carbon atoms, defined as  $\mathbf{r} = [\mathbf{r}_1, \dots, \mathbf{r}_N]$ ,  $\xi_k$  is the  $k$ -th eigen-value of the stiffness matrix  $\mathbf{K}$ , and  $\mathbf{v}^k$  is the  $k$ -th normal mode of the stiffness matrix  $\mathbf{K}$ . Here, it should be noted that in Eq. (27), six zero modes corresponding to the rigid body motion of a protein structure are excluded for computing the fluctuation matrix  $\mathbf{Q}$ .

### Continuum Model

Though computer simulations based on atomistic and/or CG models are able to provide the fundamental insight into the dynamic and mechanical behavior of protein structures at molecular level, a continuum model is useful in analyzing the results of atomistic MD simulations of a large protein structure (e.g. protein fibril). For example, while the mechanical response of a protein fibril can be obtained from all-atom MD simulations or CG-MD simulations, its mechanical properties such as elastic modulus and fracture stress can be extracted from a continuum model such as Euler–Bernoulli beam model. This Section briefly reviews the continuum models that can be employed for characterizing the mechanical properties of protein structures ranging from a single protein to protein assembly.

### Polymer Chain Statistics

Polymer chain model has been useful in characterizing the properties of one-dimensional biological structure such as fiber-like structure [66]. Specifically, polymer chain model allows one to measure the thermal fluctuation behavior of a biological structure resulting in the evaluation of the persistent length of the biological structure. Here, the persistent length is defined as  $l_p = D/k_B T$ , where  $D$  is the bending rigidity of the biological structure. Among polymer chain models, wormlike chain (WLC) model has been successfully able to capture the fluctuation behavior of a various biological molecules such as DNA and protein fibril. Statistical mechanics theory based on WLC model provides the mean-squared value of end-to-end distance ( $R$ ) in the form of [67]

$$\langle R^2 \rangle / l^2 = 4(l_p/l) [1 - 2(l_p/l) \{1 - \exp(-l/2l_p)\}] \tag{28}$$

in two-dimensional space

where  $l$  is the contour length of a biological structure. There are two ways to measure the persistent length of a biological molecule—the first approach is to place a biological molecule on the surface, and then its end-to-end distance is measured as a function of the contour length by atomic force microscopy (AFM) imaging [68]. The alternative approach is to fix the one end of a biological molecule, while its other end is labeled with a fluorescence dye, and then the fluctuation behavior of the free end of the biological molecule is estimated [69].

For last two decades, single-molecule force spectroscopy based on AFM or optical tweezer has played a vital role in characterizing the mechanical response and properties of protein materials. For example, the mechanical behavior of a muscle protein titin, which is responsible for the elasticity of a muscle fiber, has been analyzed based on a single-molecule force spectroscopy experiment [4, 70–73]. In addition, the mechanical response of a DNA molecule, which is associated with gene transcriptions, has been characterized by a single-molecule force experiment [30, 74–76]. It is shown that when a protein is extended (with a small amount of displacement before the folded domain of a protein is denatured), the mechanical response of the protein is well dictated by WLC model, which provides a relation between the force and displacement such as [77]

$$Fl_p/k_B T = (u/l) + \frac{1}{4}(1 - u/l)^{-1} - \frac{1}{4} \tag{29}$$

where  $F$  is the force acting on the molecule, and  $u$  is the displacement (i.e. end-to-end extension) of the molecule. WLC model is very useful in extracting the mechanical property (i.e. persistent length) of a biological molecule when it is stretched by a single-molecule force spectroscopy.

### Elastic Beam Model

Though the computer simulations based on atomistic or CG models are able to depict the mechanical response of a protein, a continuum elasticity theory is necessary in characterizing the mechanical properties of the protein. For example, the mechanical property of a one-dimensional structure such as protein fibril [78] (or nanoscale system [79]) can be measured by considering an elastic beam model such as Euler–Bernoulli beam model. In particular, the elastic beam model allows for evaluating the mechanical property (e.g. elastic modulus) of a protein fibril based on its vibrational characteristics (e.g. natural frequencies) or its mechanical deformation behavior (i.e.

force–displacement relation), which can be obtained from computer simulations based on atomistic or CG models.

The vibrational behavior of a protein fibril can be described by the equation of motion for Euler–Bernoulli beam model such as [80]

$$\rho A \partial_t^2 w(x, t) + \partial_x^2 [EI \partial_x^2 w(x, t)] = 0 \quad (30)$$

where  $\rho$ ,  $A$ ,  $E$  and  $I$  represent the density, cross-sectional area, elastic modulus, and cross-sectional moment of inertia for a protein fibril, respectively,  $w(x, t)$  is the transverse deflection of the fibril,  $x$  is a coordinate defined along the longitudinal direction of the fibril, and  $t$  is the time. The frequency of the fibril is written in the form of

$$\omega = (\lambda/L)^2 \sqrt{EI/\rho A} \quad (31)$$

Here,  $L$  is the length of the fibril, and  $\lambda$  is an eigenvalue that depends on the boundary condition. From Eq. (31), one is able to estimate the (bending) elastic modulus of a protein fibril as long as the frequency of the fibril is measured from computer simulations based on atomistic or CG models. The Euler–Bernoulli beam theory provides the relationship between the force and displacement for the fibril such as [81]

$$w(x_0) = \alpha(x_0) \cdot (FL^3/EI) \quad (32)$$

where  $x_0$  is the location at which a force  $F$  is applied, and  $\alpha(x_0)$  is a constant that is a function of  $x_0$  and a boundary condition. If the force–displacement relation of a protein fibril is measured based on the bending deformation of the fibril with using computer simulations based on atomistic models, it is straightforward to evaluate the bending elastic modulus of the fibril based on Eq. (32). In addition, for a recent decade, Timoshenko beam model has been highlighted in order to understand the role of the length of a protein fibril in its bending elastic property [69]. Timoshenko beam model provides the persistent length ( $l_p$ ) of a protein fibril in the form of [69, 82]

$$l_p = l_p^\infty \left[ 1 + (a/b) \left( c l_p^\infty k_B T / G_S A L^2 \right) \right]^{-1} \quad (33)$$

$l_p^\infty$  is the persistent length for a very long fibril,  $G_S$  is the shear modulus of the fibril, and parameters  $a$  and  $b$  are the constants that depend on the boundary condition, and a parameter  $c$  is a constant that is dependent on the cross-sectional shape of the fibril. Here, we note that the bending rigidity of a protein fibril is given by  $EI = l_p k_B T$ .

## Mechanical Characterization of a Single Protein

For last two decades, there are a lot of efforts that have been made to characterize the mechanical response of a biological molecule such as DNA and/or protein, as its mechanical

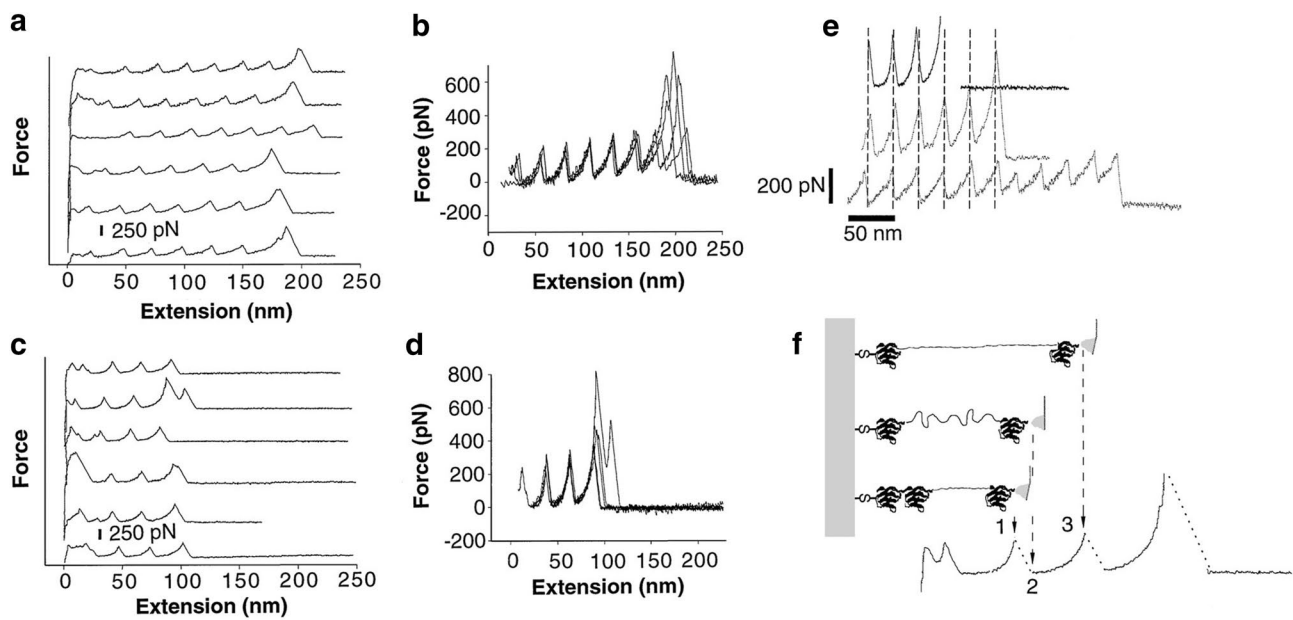
response has been associated with its biological functions. For instance, during the gene transcriptions, the unwinding of double-stranded DNA molecule is necessary. The underlying mechanisms of DNA unwinding have been revealed by a single-molecule force spectroscopy experiment [26, 30, 83] and/or computer simulations [84] based on atomistic or coarse-grained models. In addition, some proteins have recently been found to bear a mechanical loading, which endows such proteins to perform a mechanical function. As an example, protein domains found in a muscle protein titin, which is responsible for the passive elasticity of a muscle, have been found to resist a mechanical force even in the order of 100 pN, which is much larger than a thermal fluctuation force typically in the order of 1 pN (see Fig. 1) [4, 70–72, 85]. This implies that for gaining a fundamental insight into the design principles of mechanically strong proteins, the mechanical response of such proteins has to be characterized. Here, we review the mechanical response of a protein, particularly the mechanical unfolding of a protein domain and the conformational change of a protein, with using computer simulations based on atomistic or CG models.

## Mechanical Unfolding of Proteins

The study on the mechanical unfolding of proteins based on all-atom MD simulations or CG-MD simulations has been of high importance, as this study allows for not only fundamental insights into the mechanical stability, unfolding, and folding of protein molecules, but also validation of the physical theory such as reaction-rate theory.

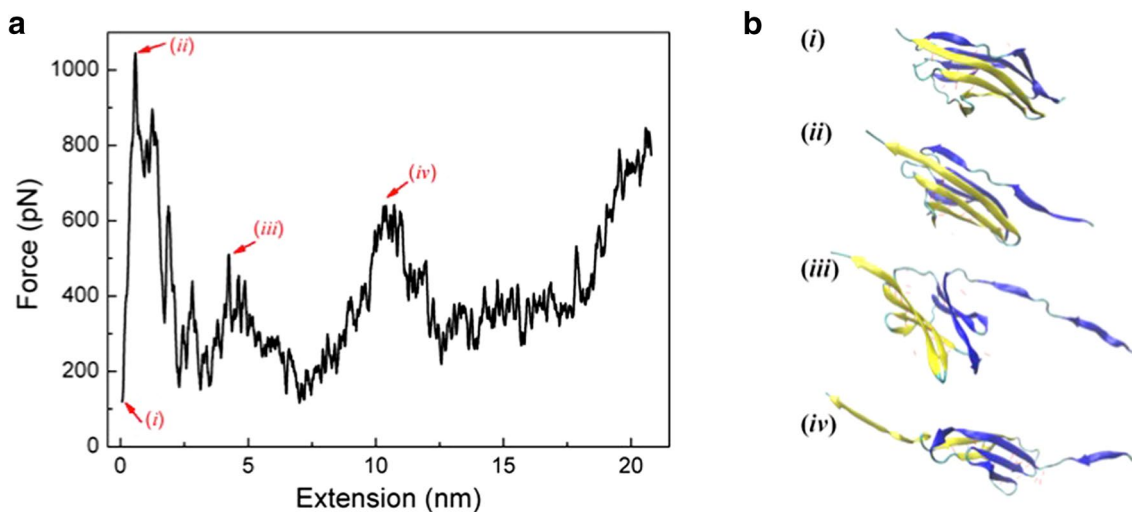
The single-molecule force spectroscopy has allowed for characterizing the mechanical unfolding behavior of proteins such as immunoglobulin (Ig)-like domains [4, 70–72, 85] and fibronectin3 (fn3) domains [6, 86, 87] found in a muscle protein titin. It is shown that the force–displacement curve of Ig-like domains resembles the sawtooth-like curve such that force peaks in the sawtooth-like force curve correspond to the unfolding forces, at which each Ig-like domain unfolds (Fig. 1). Before a protein domain is unfolded by a force, the force–displacement of the domain is well depicted by WLC model, i.e. Equation (29). For more than two decades, steered molecular dynamics (SMD) simulation [46] has allowed for unveiling the underlying mechanisms of protein unfolding process. In particular, though a single-molecule force spectroscopy experiment is able to actually measure the mechanical response of a protein (e.g. Ig-like domain), such an experiment does not provide any detailed insight into how some proteins exhibit remarkable mechanical properties. The underlying mechanisms of how a protein domain is able to resist a force can be unveiled by SMD simulations. Specifically, SMD simulations are able to show





**Fig. 1** Atomic force microscopy (AFM)-based single-molecule force spectroscopy experiments of a protein chain made of a few immunoglobulin (Ig) domains. **a** Experimental results of force-extension curves for a protein chain composed of 8 Ig domains. **b** Six independent single-molecule force spectroscopy experimental results are superimposed. **c** Force-extension curves for a protein chain consisting of 4 Ig domains. **d** Six independent experimental results of force-extension curves are superimposed. **e** Force-extension curves for

a protein chain made of 4 Ig domains (top) or 8 Ig domains (middle), and native titin (bottom). The distance between two neighboring force peaks is measured as  $\sim 25$  nm, which indicates that each force peak corresponds to the force, at which a single Ig domain is fully unfolded. **f** Schematic illustration showing the mechanical response of a protein chain made of Ig domains. Figure is adopted with permission from Ref. [4]. Copyright (1997) The American Association for the Advancement of Science



**Fig. 2** Steered molecular dynamics (SMD) simulation on the mechanical unfolding of immunoglobulin-1 (I1) domain. **a** Force-extension curve for I1 domain that is extended with a pulling rate of 0.001 nm/s. **(b)** The molecular structures of I1 domain at initial state

(i), first force peak (ii), and other force peaks (iii) and (iv). When I1 domain is stretched, hydrogen bonds (indicated by red-dashed lines) between  $\beta$  strands are ruptured so that I1 domain begins to be unfolded

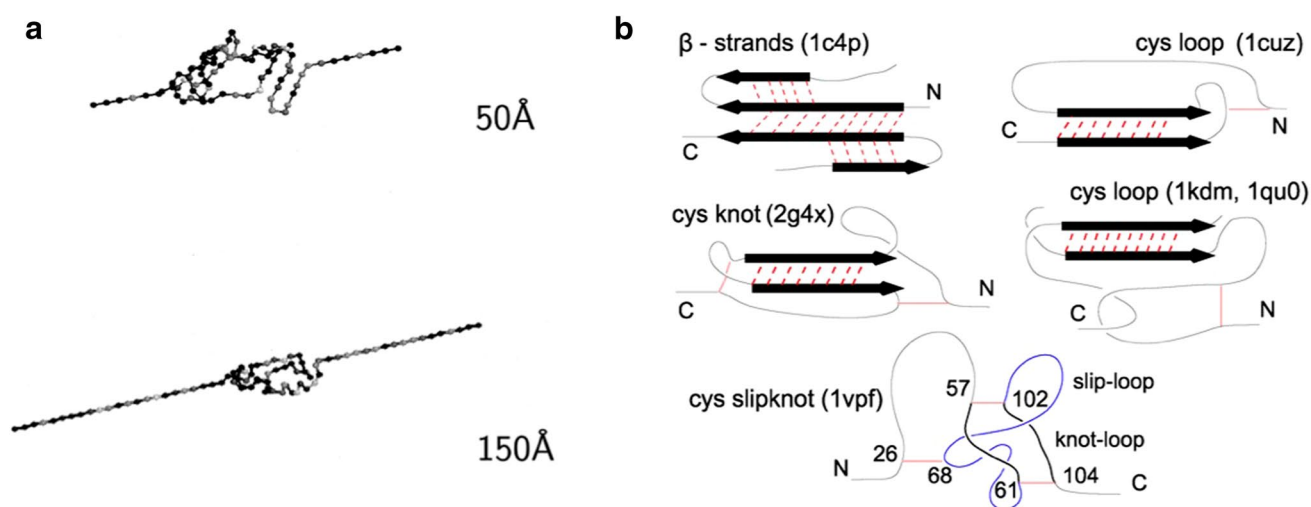
the unfolding pathway of a protein, that is, the time evolution of the protein structure due to a force (Fig. 2) [7, 88–90].

Although all-atom SMD simulations are useful in understanding the unfolding mechanisms of a protein, they are computationally restrictive in analyzing the unfolding

process of a large protein complex. In order to overcome this computational restriction, CG-MD simulations have been extensively employed, for more than a decade, for gaining insight into the unfolding mechanisms of a protein. Cieplak and coworkers [60, 91] have utilized the Go-like model for simulating the unfolding process of a protein molecule. They showed that the force–displacement response of a protein domain obtained from CG-MD simulations based on Go-like model is comparable to that acquired from a single-molecule force spectroscopy, and that Go-like model is able to provide the unfolding pathway of a protein domain (Fig. 3a). Due to the computational efficiency of Go-like model, it has been widely considered for studying how the thermal or mechanical stability of a protein domain is determined. It is shown that, from the results of Go-like model, the thermal or mechanical unfolding pathway of a protein domain is governed by its native topology [91]. Cieplak and colleagues [92] studied the mechanical unfolding responses of ~17,000 protein domains using CG-MD simulations based on Go-like model in order to gain a fundamental insight into the role of the topology of protein domains in their mechanical unfolding responses. Based on the simulations based on Go-like model with considering ~17,000 protein domains, Cieplak et al. [92] suggested the optimal topology of a protein structural motif, which maximizes the unfolding force of protein domain (Fig. 3b). In addition, they have also investigated the mechanical stability of a protein knot, which occupies < 1% of whole protein structures, using CG-MD simulations based on Go-like model [93]. They found that the presence of knot improves the stability of a protein domain in such a way that

the knot increases the life time of intermediate unfolding states during the unfolding process of protein domain. A previous study by Sulkowska et al. [94] reported that, by using Go-like model, the higher force applied to a knotted protein domain, the longer intermediate and metastable state, which is attributed to the high force-driven jamming of the knot resulting in the longer lifetime of intermediate unfolding states.

In recent years, Thirumalai and colleagues [95–97] developed a CG model, that is, self-organized polymer (SOP) model, whose potential field is conceptually similar to that of Go-like model. Specifically, SOP model considers the potential field that consists of the energies associated with covalent bonds and native contacts, whereas the mathematical form of the potential field of SOP is quite different from that of Go-like model. Thirumalai and coworkers [95] have shown that, by using CG-MD simulations based on SOP model, there are possible multiple unfolding pathways for a protein domain such that the alternative unfolding pathway of green fluorescence protein (GFP), which is inaccessible with conventional all-atom MD simulations, can be observed by SOP model. This is attributed to the fact that a pulling rate used in all-atom MD simulations is a few orders of magnitude larger than that employed in CG-MD simulations based on SOP model and/or single-molecule force spectroscopy experiments. The higher pulling rate used in all-atom MD simulations leads to the deterministic rupture of non-covalent bonds such as hydrogen bonds in a protein domain, while the lower pulling rate results in the stochastic rupture event of hydrogen bonds in the protein domain. In addition, a recent study by



**Fig. 3** **a** Coarse-grained molecular dynamics (CG-MD) simulations based on Go-like potential model for mechanical unfolding of titin immunoglobulin (Ig) domain. The snapshot of titin Ig domain that is extended with 5 nm (top) or 15 nm (bottom) was shown from the results of CG-MD simulations based on Go-like model. Figures are adopted with permission from Ref. [60]. Copyright (2002)

John Wiley and Sons. **b** Optimal structural motifs that are found in mechanically strong proteins are obtained from CG-MD unfolding simulations of > 17,000 protein structures. Top five structural motifs that are responsible for high mechanical strength. Figures are adopted from Ref. [92] under Creative Commons Attribution License

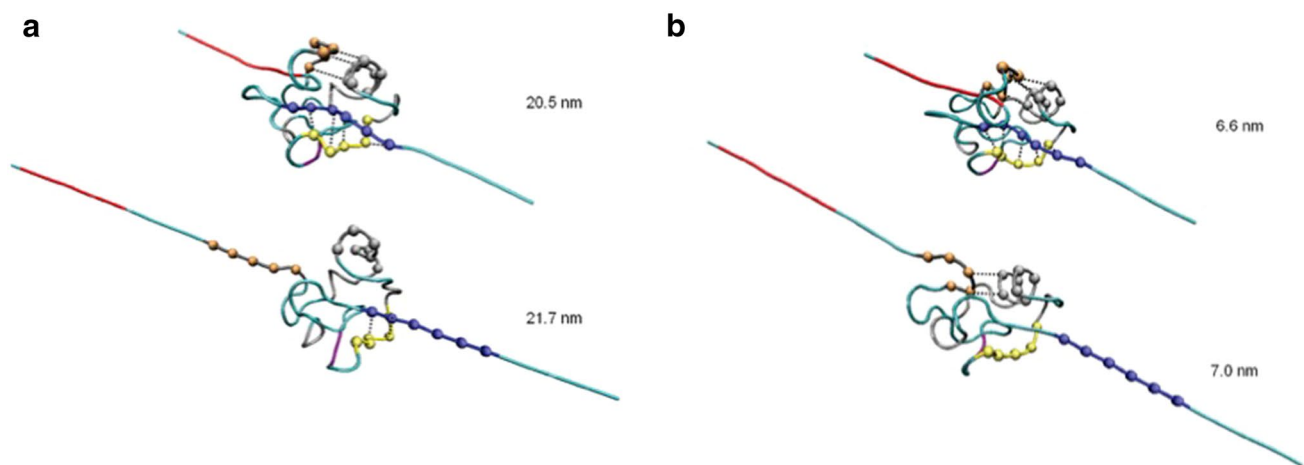
Thirumalai et al. [98] has interestingly shown the ability of SOP model-based simulations to unveil the refolding mechanisms of a protein domain under the quench of a mechanical force. Specifically, the refolding dynamics of force-driven unfolded protein under the significant decrease of a force applied to the protein from stretching force (with amount of  $\sim 100$  pN) to quenched force (around 10 pN) was studied with using CG-MD simulations based on SOP model. It is shown that refolding time for a unfolded protein is dependent on the amount of quenched force in such a way that the quench force allows the unfolded protein to explore the intermediate, metastable unfolded states during the refolding of the unfolded protein [98]. In addition, SOP model-based CG-MD simulations enabled Thirumalai and coworkers [96] to study the force-driven hopping of RNA hairpin between two intermediate states. The SOP model-based simulation results provide the force-dependent time evolution of end-to-end distance for RNA hairpin, which is comparable to that obtained from single-molecule force spectroscopy experiments. This SOP model-based simulations allowed for accurate description of the folding kinetics of RNA hairpin with dependence on its amino acid sequence.

Moreover, computer simulations such as all-atom MD simulations and CG-MD simulations based on Go-like model (or SOP model) are very useful in validating a physical theory, which has played an important role in interpreting the results of single-molecule force spectroscopy experiments. For instance, the kinetics of dissociation of protein complex such as antigen–antibody complex was experimentally observed by a single-molecule force spectroscopy [99], and this kinetics was successfully analyzed by a reaction-rate theory (ranging from Arrhenius law to Kramers' theory). In particular, Bell [100] revisited

the Arrhenius law [2] in order to theoretically obtain the relationship between the kinetic rate of molecular dissociation (i.e. bond rupture) and a force applied to induce the molecular dissociation event. The theory developed by Bell for molecular dissociation was verified by not only a single-molecule force spectroscopy but also all-atom MD simulation. In addition, the reaction rate theory of protein unfolding has recently been developed by considering the effect of force probe [101]. Previous study by Freund [101] and Arya et al. [102] revisited the Kramers' theory [103] with including the strain energy of a force probe (which is used to stretch a protein molecule) in the total potential energy for a force-driven protein unfolding, and reported that the stiffness of a force probe affects the kinetic rate of protein unfolding, and consequently, the force required to unfold the protein. This theoretical prediction was validated by CG-MD simulations based on Go-like model. Specifically, our previous study [104] employed CG-MD simulations based on Go-like model in order to verify the robustness of the theory developed by Freund [101] and Arya et al. [102], and found that the dependence of unfolding force on the stiffness of a force probe acquired from CG-MD simulations based on Go-like model is comparable to that predicted by the theory. In addition, our CG-MD simulations show the effect of the stiffness of a force probe in the unfolding pathway of a protein (Fig. 4).

### Conformational Change of Proteins

Proteins perform their biological functions through their conformational changes driven by a ligand (or inhibitor). For instance, the conformational change of protein kinases is responsible for their biological function such as signaling



**Fig. 4** Mechanical unfolding pathway for ubiquitin using **a** a soft force probe or **b** a stiff force probe. When a ubiquitin is extended by a soft loading device, the hydrogen bonds between  $\beta 1$  and  $\beta 5$  begin to be ruptured at an extension of 20.5 nm. However, for the stretching of

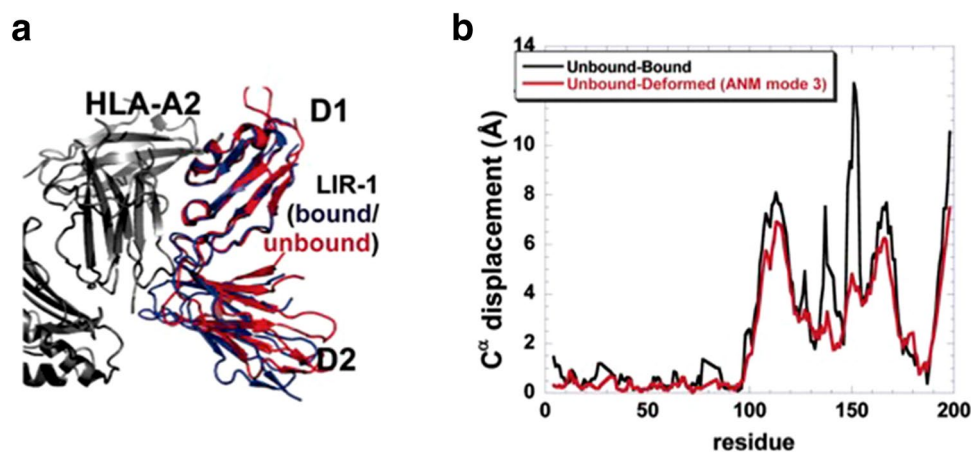
ubiquitin using a stiff loading device, the hydrogen bonds between  $\beta 1$  and  $\beta 5$  are fractured at an extension of 6.6 nm. Figures are adopted with permission from Ref. [104]. Copyright (2012) AIP Publishing

pathway (that is relevant for cellular function or dysfunction resulting in disease pathologies such as cancer) [105]. This highlights the importance of predicting (or analyzing) the conformational change of proteins (due to ligand-binding) for not only further understanding of their biological functions but also the effective design of novel drug molecules (that are able to regulate the conformational change of proteins) [106].

Though all-atom MD simulations have recently been extensively employed for simulating the long time-scale dynamic behavior of small proteins [54, 107], they are still unable to simulate the conformational change of proteins, as the time-scale relevant to the conformational change of proteins is still larger than that accessible with currently available all-atom MD simulations. For a recent decade, ENM has been widely utilized for computationally efficient analysis of the conformational transitions of proteins. One may ask why ENM is able to predict the conformational change of proteins despite the simplicity of ENM modeling. This is attributed to the important role of the native topology of proteins in their dynamics [108]. A previous study by Teeter et al. [48] reports that the conformational dynamics of proteins is very insensitive to the details of potential fields, which implies that the conformational dynamics of proteins is not affected by the details of inter-atomic interactions. In a similar spirit, Lu and Ma [109] showed that the conformational fluctuation behavior of protein is unaffected by perturbation of Hessian (stiffness) matrix as long as the native contacts of a protein are maintained. In addition, it has been reported that the thermal fluctuation of an unbound protein structure can render it to explore the conformation of a ligand-bound protein structure because of the feature

of free energy landscape for the protein structure in such a way that the free energy landscape has multiple metastable, intermediate conformations, some of which may correspond to the conformation of ligand-bound state [1]. These previous studies [1, 48, 109] validates the robustness of ENM modeling in analyzing the conformational dynamics of proteins such as their fluctuation and/or their conformational changes.

For a recent decade, the low-frequency normal modes of a protein predicted from ENM have been accepted to predict the conformational change of a protein. For example, Bahar and coworkers [110–112] have analyzed the conformational change of a protein based on the low-frequency normal modes obtained from ENM. In particular, they showed that the conformational change of a protein can be predicted based on low-frequency normal modes, with which the protein structure is perturbed (Fig. 5). As described in Ref. [110], the displacement of a protein corresponding to its conformational change is represented in the form of  $\mathbf{u}_i = \sum_{k=1}^n \alpha_k \mathbf{q}_i^k$ , where  $\mathbf{u}_i$  is the displacement of the  $i$ -th residue of the protein,  $\mathbf{q}_i^k$  is the  $k$ -th low-frequency normal mode vector for the  $i$ -th residue,  $\alpha_k$  is a fitting parameter, and  $n$  is the number of low-frequency normal modes that are utilized to describe the conformational change. They have also shown that the conformational change of protein kinases upon inhibitor binding can be also predicted using the low-frequency normal mode-based perturbation of the protein structure [111]. These previous studies by Bahar and coworkers [110, 111] have supported the hypothesis of population shift model, which suggests that the conformation of ligand-bound state can be accessible based on the conformation of an unbound



**Fig. 5** **a** The structure of leukocyte Ig-like receptor (LIR) in unbound state (red) or ligand-bound state (blue). The bound state is referred to as the complex formed with HLA-A2 (black). **b** The difference in the alpha carbon atoms of LIR between unbound and bound states (black). This difference can be inferred from the structural perturbation of LIR in unbound state along the 3rd lowest-frequency normal

modes obtained from ENM. This indicates that the low-frequency normal modes of a protein in its unbound state allows it to explore its ligand-bound structure. Figures are adopted with permission from Ref. [110]. Copyright (2005) National Academy of Sciences, U.S.A (color figure online)

protein due to the ruggedness of the free energy landscape of the unbound protein [1]. In a similar spirit, Karplus and colleagues [113] have predicted the conformational change of myosin V based on a projection of displacement vector corresponding to the conformation change into the normal mode space. They have provided that the conformational change of myosin V can be predicted with using few low-frequency normal modes. Furthermore, Wolynes and coworkers [114] had developed a theoretical model to predict the conformational change of proteins in such a way that the low-frequency normal modes of ENM were employed to generate the intermediate conformations of the protein during its conformational transition.

Moreover, Maragakis and Karplus [115] developed a plastic network model (PNM) in order to predict the conformational transition pathway of proteins. Specifically, PNM was developed in such a way that the harmonic potential field of intermediate state is obtained by mixing the harmonic potential fields (i.e. the potential fields of ENMs) corresponding to two end states, that are, unbound and ligand-bound states. The harmonic potential field of intermediate state is assumed to be in the form of  $E = (1/2)[E_1 + E_2 - ((E_1 - E_2)^2 + 4\epsilon^2)^{1/2}]$ , where  $E_1$  and  $E_2$  represent harmonic potential fields (i.e. the potentials of ENM) corresponding to two end states, and  $\epsilon$  is a parameter related to the degree of mixing two harmonic potential fields. The concept of PNM by mixing two potential fields was inspired by a quantum mechanical coupling of two potential energy surfaces. In a similar manner, Takagi et al. [116] suggested a dual Go-like model (DGM) by mixing two potential fields corresponding to two end conformations, while the potential field of a protein structure is assumed to be in the form of Go-like model, i.e. Equation (20), rather than the ENM-like harmonic potential. Furthermore, Hummer and colleagues [117] had developed a mixed elastic network model (MENM) to predict the conformational transition pathway of proteins by mixing two ENM harmonic potentials at two end states. The harmonic potential field of intermediate state is presumed to be written as  $E = -k_B T \ln[\exp((E_1 + \epsilon_1)/k_B T) + \exp((E_2 + \epsilon_2)/k_B T)]$ , where  $\epsilon_1$  and  $\epsilon_2$  are parameters that are related to the degree of mixing two potential energies.

In addition, Jernigan and coworkers [118] introduced an elastic network interpolation (ENI) to acquire the conformational transition pathways of proteins. In their ENI scheme, a displacement vector  $\mathbf{u}_i^k$  corresponding to a conformational transition at the  $k$ -th intermediate state is obtained by minimizing a cost function (i.e. strain energy for conformational transition) defined as  $E^k = (1/2) \sum \left[ \left\{ \left( \mathbf{r}_i^k + \mathbf{u}_i^k \right) - \left( \mathbf{r}_j^k + \mathbf{u}_j^k \right) \right\} - d_{ij}^k \right]^2$ , where  $\mathbf{r}_i^k$  is the position vector of  $i$ -th residue at  $k$ -th intermediate state. Here,  $d_{ij}^k$  is an interpolated inter-residue distance given by  $d_{ij}^k = \alpha_k d_{ij}^0 + (1 - \alpha_k) d_{ij}^F$ , where  $d_{ij}$  is a distance between two residues  $i$  and  $j$ , superscripts 0 and  $F$  indicate initial (unbound)

state and final (ligand-bound) state, respectively, and  $\alpha_k$  is a parameter to indicate the  $k$ -th intermediate state obtained by linear interpolation between two states. Moreover, they have extended ENI coupled to a rigid cluster model [119], in which the rigid domains of a protein are modeled as a rigid body, while the interaction between the rigid domains is described by elastic springs. A recent study by Kim et al. [120] has suggested a normal mode-guided elastic network interpolation (NGENI) in such a way that low-frequency normal modes are used in the linear interpolation for obtaining an intermediate conformation. In their work, it is shown that NGENI is more computationally efficient when compared with ENI for predicting the conformational transition pathways of large protein complexes. Kidera and colleagues [121] developed a theoretical model to understand the conformational change of protein structures based on a linear response theory with ENM modeling. In their work [121], a displacement vector  $\mathbf{u}_i$  for the conformational change is written as  $\mathbf{u}_i = (k_B T)^{-1} \sum_k \mathbf{q}_i^k \lambda_k \sum_j \mathbf{q}_j^k \cdot \mathbf{f}_j$ , where  $\mathbf{q}_i^k$  is a  $k$ -th normal mode vector for  $i$ -th residue,  $\lambda_k$  is the  $k$ -th eigenvalue of a covariance (compliance) matrix, which is a pseudo-inverse of the stiffness (Hessian) matrix of ENM, and  $\mathbf{f}_j$  is a force vector acting on  $j$ -th residue due to ligand-binding. In a similar spirit, Demirel and Lesk [122] employed ENM for evaluating a force associated with ligand-receptor binding. Specifically, a force related to ligand-receptor binding is obtained based on a linear response theory in such a way that the force  $\mathbf{f}_i$  is given by  $\mathbf{f}_i = k_B T \sum_l \mathbf{q}_i^l \lambda_l^{-1} \sum_j \mathbf{q}_j^l \cdot \mathbf{u}_j$ , where  $\mathbf{u}_j$  is the displacement vector of  $j$ -th residue (for a receptor) corresponding to the conformational change from unbound state to ligand-bound state.

In recent years, there are a few efforts that have been made to couple ENM-based structural perturbation scheme and all-atom molecular simulations for quantitatively understanding the conformational change of proteins. It should be noted that intermediate conformations predicted by the low-frequency normal modes of ENM may be sometimes physically unacceptable, because of the simplicity of ENM, which does not have any information for residue-specific interactions, side chain-dependent interactions, and so on. Bahar and coworkers [123] had coupled the ENM modeling and all-atom MD simulations for predicting the conformational transition of proteins. Specifically, the all-atom MD simulation was used to obtain an equilibrium structure for a protein, while ENM modeling was employed to introduce a structural perturbation of the protein. The iterative computations based on ENM normal mode-based perturbation followed by all-atom MD simulations are able to provide an insight into the conformational transition of proteins. In a similar spirit, Doruker et al. [124] had combined ENM normal mode-based structural perturbation scheme and all-atom Monte-Carlo (MC) simulations for studying the conformational transition of proteins. In their work, the structural perturbation of a protein based on normal modes of ENM

followed by all-atom MC simulation was iteratively implemented. These studies [123, 124] show the possibility of coupling two different scale modeling techniques for studying the conformational (structural) changes of proteins in such a way that a CG model (e.g. ENM) was used to induce the structural perturbation of proteins while all-atom MD simulation was considered to gain their equilibrium structures. In other words, two different scale modeling techniques can be coupled so as to develop a novel multi-scale simulation techniques, which can provide an insight into protein dynamics typically inaccessible with only all-atom MD simulations.

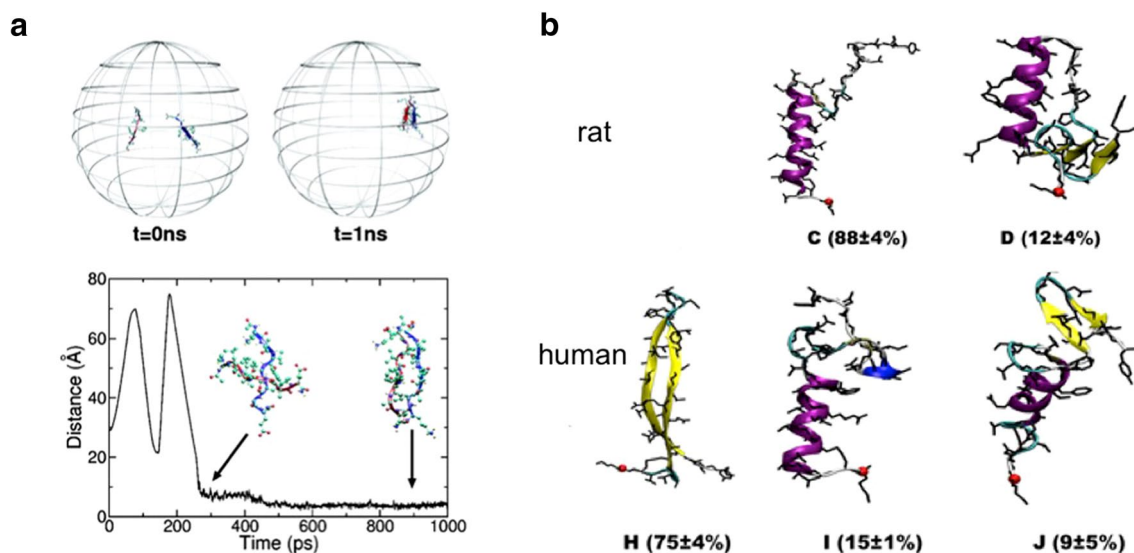
## Characterization of Protein Assemblies

Understanding the formation and properties of protein assemblies is of high importance for gaining fundamental insight into the biological function of these assemblies such as disease pathologies (or etiologies). For example, amyloid materials are formed by self-assembly process of protein aggregation resulting in the different structural forms of amyloids ranging from nanometer to micrometer (or even larger) scales [10, 20]. These amyloid materials have been known to play a crucial role in the pathogenesis of various diseases from neurodegenerative diseases [17] to type 2 diabetes [18] and cardiovascular diseases [19]. This suggests the necessity of unveiling how the amyloid materials are formed as an assembled structure driven by protein

aggregation. That is, it is very essential to characterize the structural feature and formation of amyloid materials for understanding the underlying mechanisms of disease pathologies driven by protein assembly. In addition, it has recently been suggested that the mechanical properties of protein assemblies are responsible for their biological functions. For instance, the formation of mechanically stiff amyloid fibrils on a mechanically soft cell membrane results in the mechanical disruption of the cell membrane [24]. This Section presents the structural, mechanical, and dynamic characterization of protein assemblies such as amyloid materials and viral capsids based on computer simulation techniques at different scales ranging from all-atom MD simulations to CG-MD simulations.

## Structural Characteristics of Amyloid Materials

For a recent decade, there are efforts that have been made to understand the formation and structures of amyloid materials based on computer simulations. A previous study by Karplus and coworkers [125] reported MD simulations of dimerization for understanding the underlying mechanisms of the formation of amyloid dimers, which were formed by self-assembled aggregation of two monomers (Fig. 6a). They showed the ability of all-atom MD simulations to probe the structural properties of amyloid dimer isoforms, while their work [125] is still computationally restricted in gaining insight into the kinetics of protein aggregation resulting in the formation of amyloid fibrils. A pioneering work



**Fig. 6** **a** Equilibrium molecular dynamics (MD) simulations for the formation of amyloid dimers. Top panel shows the snapshots of two monomers at time  $t=0$  and a dimer formed by aggregation of two monomers at time  $t=1$  ns. Figures are adopted with permission from Ref. [125]. Copyright (2004) National Academy of Sciences, U.S.A. **b** Replica exchange MD (REMD) simulation results show the domi-

nant structures of islet amyloid polypeptide (IAPP) monomers found in rat (top panel) or human (bottom panel). It is interestingly shown that rat IAPP is most likely to exist in the form of  $\alpha$  helix, while human IAPP is more likely to exist in the form of  $\beta$ -hairpin, which is responsible for aggregation of IAPP monomers. Figures are adopted from Ref. [132] under Creative Commons Attribution License

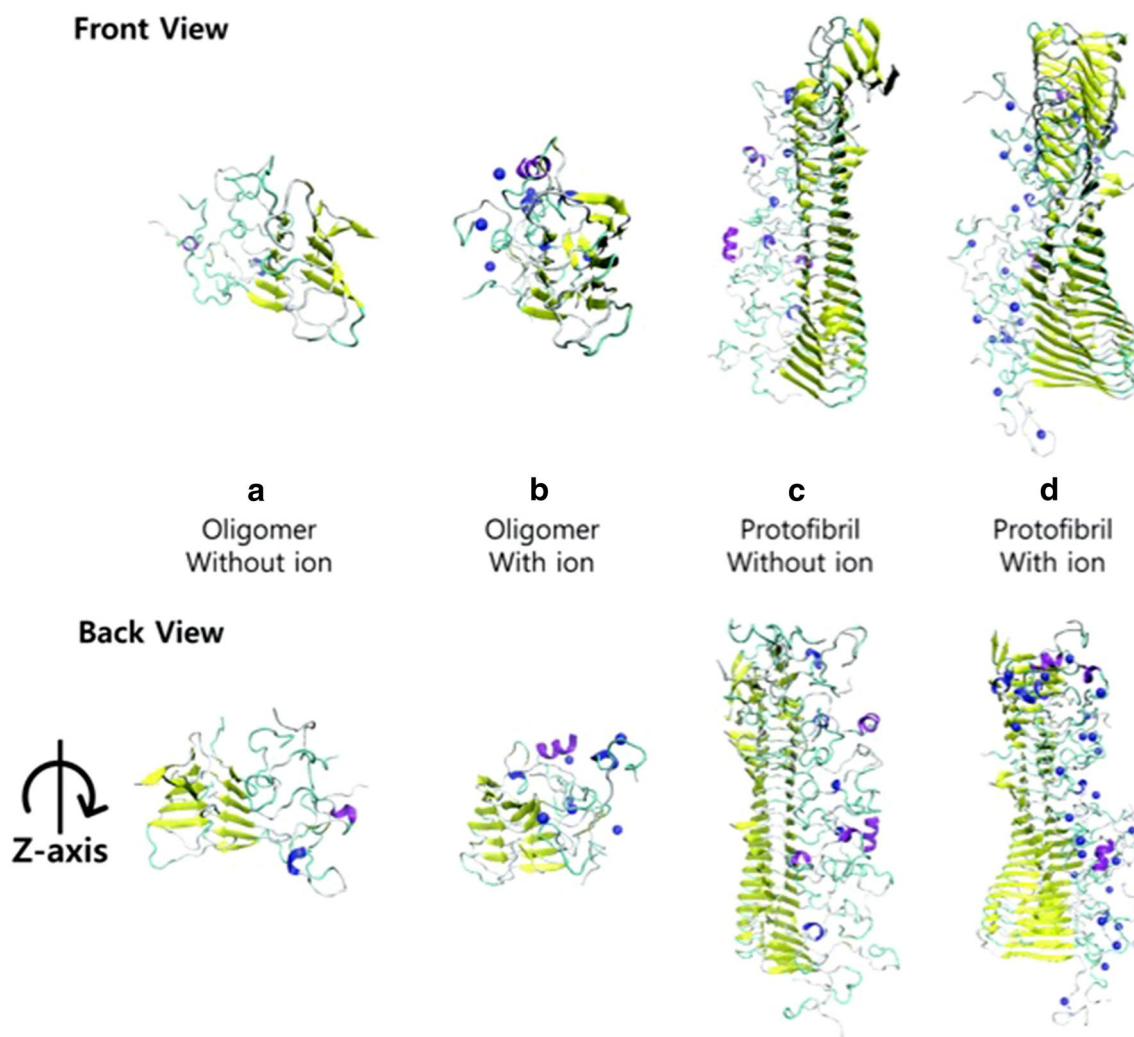
by DeMarco and Daggett [126] reported that all-atom MD simulations may allow for predicting the structural conversion of cellular prion protein (PrP<sup>C</sup>) into the scrapie isoform of prion protein (PrP<sup>Sc</sup>), which is responsible for the formation of prion fibrils involving in various transmissible spongiform encephalopathies. This pioneering work [126] is still limited in quantitative understanding of the kinetics of structural conversion into dimerization (and subsequently, fibrillization). In recent years, with the advancement of computational resources, the all-atom MD simulations have been able to probe the more detailed information of the formation of amyloid materials. For example, Chong and Ham [127] have studied the dimerization of Alzheimer- $\beta$  (A $\beta$ ) monomers using all-atom MD simulations. In their work [127], they are able to unveil the underlying mechanisms of the dimerization process such that a water-mediated hydration force plays a crucial role in the formation of A $\beta$  dimer. In addition, all-atom MD simulations with enhanced sampling methods such as replica exchange [128] or metadynamics [129, 130] have allowed for characterizing not only the formation of amyloid dimers but also the inherent structural feature of amyloid monomer, which may be correlated with self-assembly process of protein aggregation. For instance, a bias-exchanged metadynamics simulations have provided the pathway of the dimerization of islet amyloid polypeptide (IAPP) and the intermediate structural forms of IAPP dimers [131]. A recent study by Shea and coworkers [132] have shown the ability of replica exchange MD (REMD) simulations to probe the inherent structural properties of IAPP monomers with the effect of sequence mutations. In their work [132], it is found that human IAPP (hIAPP) monomer is able to exhibit more contents of  $\beta$ -sheets, which are responsible for protein aggregation, than rat IAPP (Fig. 6b). Their finding [132] is consistent with clinical observation [18] that though IAPP proteins are found in both human and rat, human suffers from type 2 diabetes due to aggregation of IAPP proteins, whereas the type 2 diabetes is not found in a case of rat. In addition, Chakraborty and Das [133] have recently employed REMD to probe the inherent structural properties of A $\beta$  monomer and the effect of inhibitors in the structural properties of A $\beta$  monomers. They found that an inhibitor, which is a six-residue A $\beta$  fragment (referred to as A $\beta$  hexapeptide), makes a critical impact on the secondary structural propensity of A $\beta$  monomers. In particular, A $\beta$  hexapeptide reduces the contents of  $\beta$ -sheets in an A $\beta$  monomer, which implies that A $\beta$  hexapeptide inhibits the aggregation of A $\beta$  monomers. A recent study by Shea and colleagues [134] have also considered REMD simulations for gaining insight into the role of osmolytes (e.g. urea or trimethylamine N-oxide, abbreviated as TMAO) in the structural propensity of tau fragment peptide. In their REMD simulations [134], the osmolytes induces the shift of the populations of secondary structures for tau fragment

peptide, which is consistent with experimental observation that the osmolytes inhibits the formation of amyloid fibrils but promotes the formation of helical amyloid oligomers. The previous studies reviewed here suggest the increasing role of all-atom MD simulations in revealing hidden, complex mechanisms that determine the structure and formation of amyloid materials at different length scales ranging from oligomeric structures to fibril structures.

## Mechanical Properties of Amyloid Materials

Mechanical characterization of amyloid materials is a priori requisite for gaining insight into the molecular mechanism of disease pathologies, which is attributed to the fact that once amyloid fibrils are formed by self-assembly process (i.e. protein aggregation), these fibrils are quite stable and are unlikely to be denatured (or ruptured) in physiological conditions so as to deposit onto an organ (e.g. brain, pancreas, etc.) resulting in the disease pathologies. In addition, the mechanical properties of amyloid fibrils are directly linked to the amyloid-driven pathologies. For example, a recent study by Zewail and coworkers [24] suggest that the formation of mechanically stiff amyloid fibrils (or protofibrils) on a mechanically soft cell membrane results in the disruption of the membrane. In addition, the fracture property (i.e. division rate) of prion fibrils are found to be correlated with prion infectivity, that is, cell-to-cell prion propagation [135].

For a recent decade, there are a lot of attempts to measure the mechanical stability and properties of amyloid fibrils. For instance, we studied the dependence of mechanical stability and properties of hIAPP amyloid fibrils on their structural feature, that is, the pattern of cross- $\beta$  structures made based on stacking of peptides, by using all-atom MD simulations [136]. We have shown that the amyloid fibrils formed based on antiparallel stacking of hIAPP peptides are quite stable when compared with the fibrils constructed based on other stacking patterns (e.g. parallel stacking of peptides), which highlights the importance of the molecular structure (i.e. cross- $\beta$  structure) of amyloid fibrils in their stability. In recent years, we have employed all-atom MD simulations in order to study how metal ions, which are found in patients suffering from neurodegenerative diseases, affect the formation and stability of A $\beta$  structures at different length scales, such as amyloid oligomers and fibrils [137]. We found that Zn<sup>2+</sup> ions improve the stability of A $\beta$  oligomers while these ions reduce the stability of A $\beta$  fibrils, which is consistent with clinical observation that high concentration of Zn<sup>2+</sup> ions promotes the formation of A $\beta$  oligomers that are known as toxic agent to functional cells (Fig. 7). In addition, a recent study by MacPhee et al. [138] reports the effect of protonation state in the stability of transthyretin (TTR) amyloid oligomers based on all-atom MD simulations. In their work [138], it is shown that the protonation state affects the



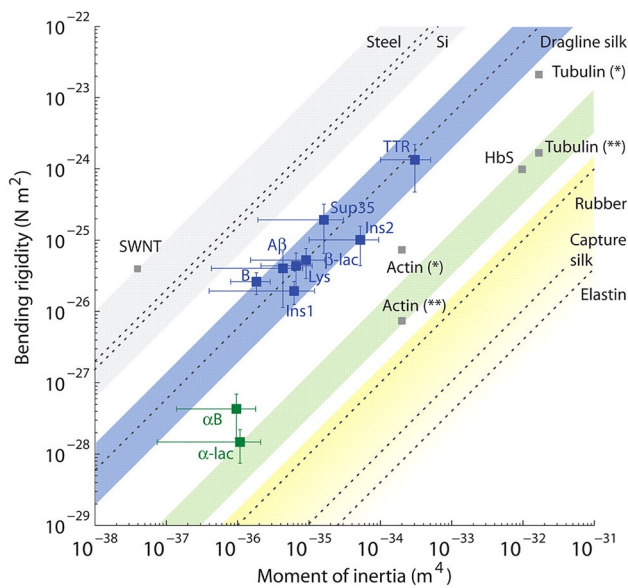
**Fig. 7** Effect of metal ions on the mechanical stability of amyloid aggregates at different length scales such as oligomer and fibril based on all-atom molecular dynamics (MD) simulations. It is shown that metal ions make an amyloid oligomer be an ordered structure, while

they distort the structure of amyloid fibril resulting in the decrease of its structural stability. Figures are adopted with permission from Ref. [137]. Copyright (2018) Royal Society of Chemistry

stability of amyloid oligomers so as to result in the different form (e.g. decamer, hexamer, etc.) of amyloid oligomers. It is found that the decamer is stable regardless of the protonation state of TTR, whereas the protonation results in the stable formation of the hexamer, and de-protonation favors the formation of higher order oligomers. These previous studies [136–138] show the ability of all-atom MD simulations to characterize the mechanical stability and formation of amyloid structures at different length scales ranging from oligomers to protofibrils. It should be noted that, however, the all-atom MD simulations are still computationally prohibited for simulating the stability of a very long amyloid fibril whose length scale is larger than 100 nm. Until now, the all-atom MD simulations is computationally able to characterize the amyloid protofibrils whose length scale is in the order of 10 nm.

Recently, there are many recent works for studying the mechanical properties of amyloid fibrils. A pioneering work by Knowles et al. [139] reports the mechanical properties of amyloid fibrils based on AFM experiments and ENM-based CG simulations. It is shown that the elastic modulus of amyloid fibrils is estimated in the order of 1 to 10 GPa, which is comparable to that of a mechanically strong protein material such as spider silk (Fig. 8). In their work [139], it should be noted that the persistent length of amyloid fibrils (equivalent to their bending rigidity) is computed based on AFM images of the fibrils coupled with polymer chain statistics, i.e. Equation (28). The remarkable elastic modulus of amyloid fibrils is attributed to interactions between neighboring  $\beta$ -sheets in the cross- $\beta$  structure of amyloid fibrils. This hypothesis was verified in a recent study by Buehler and coworkers [25], who showed

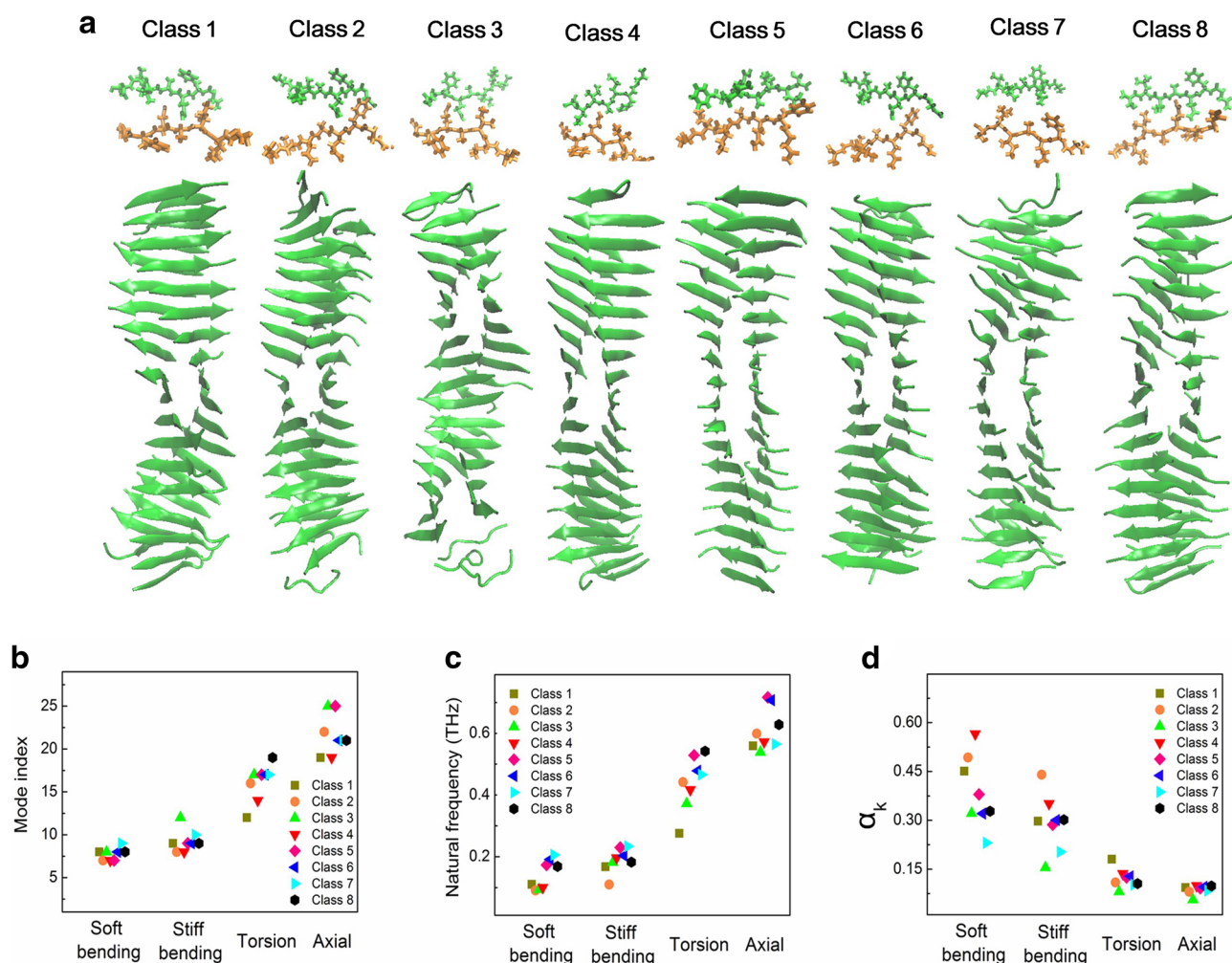




**Fig. 8** Mechanical properties of amyloid materials in comparison with those of other materials. Blue region shows the mechanical properties of amyloid fibrils, while green region indicates the mechanical properties of other protein materials. It is shown that the mechanical property (e.g. elastic modulus) of amyloid fibrils exceeds that of other protein materials. The mechanical properties of amyloid fibrils were measured by atomic force microscopy (AFM) imaging experiments coupled with polymer chain statistics. Figures are adopted with permission from Ref. [139]. Copyright (2007) The American Association for the Advancement of Science

that the geometric confinement of hydrogen bonds between  $\beta$ -sheets gives rise to the remarkable elastic properties of  $\beta$ -sheet-rich protein materials such as amyloid materials and spider silk crystals. As the all-atom MD simulations are computationally limited for simulating large protein structures such as amyloid fibrils with their length of  $> 10$  to  $> 100$  nm, CG model-based simulations have been taken into account for characterizing the mechanical properties of long amyloid fibrils. For example, Buehler and coworkers [140] have employed ENM-based NMA simulations for measuring the elastic properties of  $A\beta$  fibrils. They have shown that the low-frequency vibration modes of  $A\beta$  fibrils correspond to the bending, twisting, and stretching deformation modes, and that the elastic moduli of  $A\beta$  fibrils (with the twofold or threefold symmetries) are measured in a range of  $\sim 20$  to  $\sim 30$  GPa. Here, we note that the extraction of mechanical properties (e.g. elastic moduli) for amyloid fibrils is based on the Euler–Bernoulli beam theory, i.e. Equation (31), which relates the elastic moduli of the fibrils to their natural frequencies that were computed from ENM-based NMA simulations. It is found that the bending rigidity of  $A\beta$  fibrils (equivalent to their persistent length) is dependent on their length scales, which is ascribed to the length-dependent shear

effect on the bending deformation of the fibrils. Moreover, we studied the mechanical properties of hIAPP fibrils as a function of their structural features, i.e. stacking pattern of  $\beta$ -sheets, by using ENM-based simulations [141]. We found that the antiparallel stacking of  $\beta$ -sheets renders the hIAPP fibril to exhibit the higher bending rigidity, which highlights the role of  $\beta$ -sheet stacking patterns in not only the mechanical stability of amyloid fibrils [136] but also their mechanical properties [141]. We have also shown that the dependence of the bending rigidity of hIAPP fibrils on their length is well fitted to the Timoshenko beam theory, i.e. Equation (33), which sheds light on the shear effect in the bending properties of amyloid fibrils. In addition, we have investigated the mechanical properties of prion fibrils using ENM-based NMA simulations [142]. It is interestingly found that the length-dependent mechanical (or vibrational) properties of prion fibrils provide an insight into the critical size of prion fibrils, at which their infectivity is maximized, and that the helical structure of prion fibrils determines their elastic properties. Furthermore, we measured the vibrational and elastic properties of hIAPP fibrils (with their length of  $\sim 10$  nm) using all-atom MD simulations coupled with Euler–Bernoulli beam theory [136]. In particular, the natural frequencies of hIAPP fibrils for their vibrational modes are obtained based on all-atom MD simulations, while the elastic moduli of the fibrils are extracted based on Euler–Bernoulli beam theory that relates the natural frequencies to the elastic moduli of the fibrils. The natural frequencies of hIAPP fibrils with their length of  $\sim 10$  nm are critically dependent on their structural characteristics such as the stacking pattern (Fig. 9). We have shown that the antiparallel stacking of  $\beta$ -sheets results in the higher elastic moduli of hIAPP fibrils [136], which is consistent with our previous finding [141] based on ENM simulations. It is also found that low-frequency vibrational modes (corresponding to bending, torsional, and stretching deformation modes) mostly contribute to the thermal fluctuation of amyloid fibrils. Furthermore, a recent study by Na and colleagues [143] has reported that, using ENM-based NMA simulations, the low-frequency vibrational modes of triplet prion fibril are able to depict its conformational changes due to pH. This suggests that the low-frequency vibrational modes of amyloid fibrils play an important role in not only extracting their mechanical properties but also gaining insight into their conformational changes due to a chemical stimulus such as pH change. Recently, we have studied the mechanical properties of multi-stranded amyloid fibrils based on ENM-based simulations coupled with all-atom MD simulations [144]. In our recent study [144], all-atom MD simulations were employed to obtain the equilibrium structure of multi-stranded amyloid fibrils, while ENM-based simulations were used to extract the



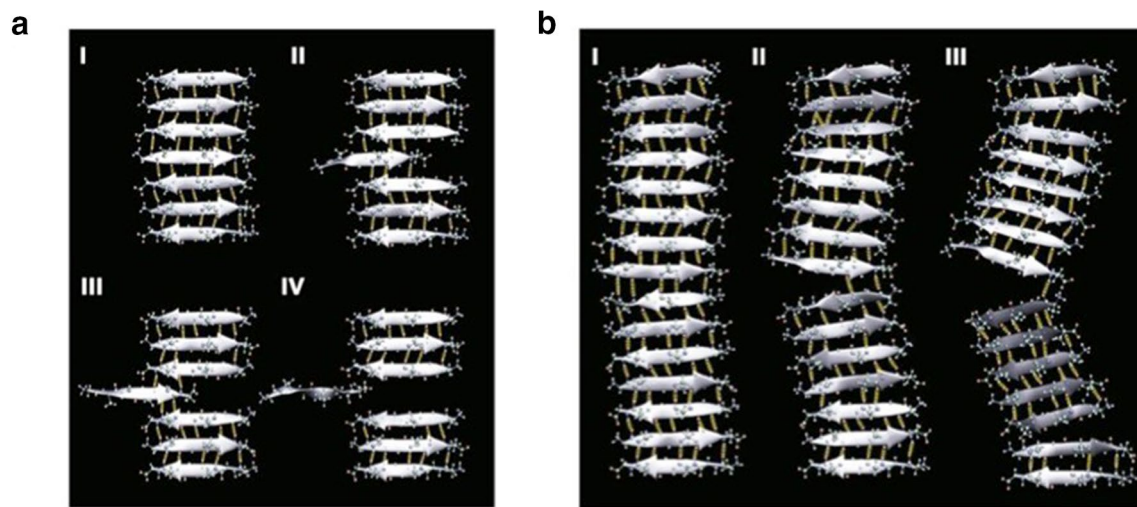
**Fig. 9** **a** Equilibrium molecular dynamics (MD) simulations of human islet amyloid polypeptide (hIAPP) protofibrils constructed based on eight different stacking patterns. MD simulations suggest that hIAPP protofibrils are mechanically stable as long as they are made based on antiparallel stacking of peptide chains. **b** Mode index for the bending, stretching, and torsional deformation modes of hIAPP protofibrils constructed based on 8 different structural

classes. **c** Natural frequencies corresponding to bending, torsional, and stretching deformation modes for hIAPP protofibrils are measured from MD simulations. **d** Contribution of bending, torsional, or stretching deformation modes to the thermal fluctuation motion of hIAPP protofibrils made based on 8 different structural classes. Figures are adopted from Ref. [136] under Creative Commons Attribution License

mechanical properties of the fibrils. We have shown that the dependence of persistent length (equivalent to bending rigidity) for multi-stranded amyloid fibrils is fitted to a scaling law of  $l_p \propto [(nq)^2 - \sin^2(nq)]^{1/2}$  [68], which was obtained from a helical model. It is also shown that a H31Y point mutation, where histidine residue locating in 31st residue from the N-terminus is changed into tyrosine residue, results in a change in not only the structure of multi-stranded amyloid fibrils but also their elastic moduli.

In recent years, all-atom SMD simulations have been extensively employed for studying not only the mechanical properties of amyloid fibrils but also their fracture mechanisms. Buehler and coworkers [25] have interestingly reported that the fracture mechanism of  $\beta$ -sheet crystals is

governed by their length scale in such a way that shear deformation dominates the force-driven fracture of short  $\beta$ -sheet crystals with their length of  $< 3$  nm, while the fracture of  $\beta$ -sheet crystals with their length of  $> 3$  nm is governed by bending deformation (Fig. 10). A recent study by Gräter and coworkers [145] have studied the mechanical deformation mechanisms and properties of  $\beta$ -sheet crystals, which are found in amyloid fibrils and/or spider silk, using all-atom SMD simulations. They have found that the direction of a force applied to the  $\beta$ -sheet crystals significantly affects not only their deformation (and fracture) mechanisms but also their elastic moduli. In particular,  $\beta$ -sheet crystal are able to effectively resist a force applied to the perpendicular direction of the  $\beta$ -strand, whereas it becomes mechanically weak when



**Fig. 10** Mechanical deformation mechanisms of  $\beta$ -sheet crystals with their length of **a**  $L=2.83$  nm or **b**  $L=6.56$  nm based on all-atom steered molecular dynamics (SMD) simulations. When a  $\beta$ -strand in the  $\beta$ -sheet crystal with its length of  $L=2.83$  nm is pulled out, the fracture of the crystal is dominated by shear deformation-driven fracture. On the other hand, for a  $\beta$ -sheet crystal with its length of

$L=6.56$  nm, the pulling of a  $\beta$ -strand in the crystal leads to the bending deformation-driven fracture of the  $\beta$ -sheet crystal. These SMD simulations suggest the role of the length scale of  $\beta$ -sheet crystals in their fracture mechanisms based on a competition between shear and bending deformations. Figures are adopted with permission from Ref. [25]. Copyright (2010) Springer Nature

a force is applied to the parallel direction of the  $\beta$ -strand. Recently, we have studied the mechanical deformation mechanisms and properties of hIAPP fibrils as a function of their length scales by using all-atom SMD simulations [146]. It is shown that the dependence of bending elastic property and fracture mechanism for amyloid fibrils on their length scales is attributed to the length-dependent competition between shear and bending deformation modes, and that the length-dependent elastic property of the fibrils obtained from all-atom SMD simulation is well fitted to Timoshenko beam theory. We found that this length-dependent fracture mechanism for the fibrils is attributed to the length-dependent rupture mechanisms of hydrogen bonds formed between  $\beta$ -sheets (in the cross- $\beta$  structure of the fibril). In particular, for a short fibril, several hydrogen bonds formed between  $\beta$ -sheets are able to cooperatively resist a force so that at a critical value of a force, these hydrogen bonds are simultaneously ruptured. On the other hand, for a long fibril, the hydrogen bonds between  $\beta$ -sheets are fractured one-by-one, which results in the mechanical fragility of the long fibril. In recent years, we have investigated the fracture mechanisms of a helical prion fibril using all-atom SMD simulations [147]. The elastic moduli of an infectious prion fibril formed based on a left-handed  $\beta$ -helix and a non-prion fibril constructed based on a right-handed  $\beta$ -helical structure are measured as  $\sim 18$  GPa and  $\sim 13$  GPa, respectively. However, it is interestingly found that the fracture toughness of an infectious prion fibril is smaller than that of non-prion fibril [147], which is consistent a recent finding [135] that infectious prion fibril is more fragile. A previous study by

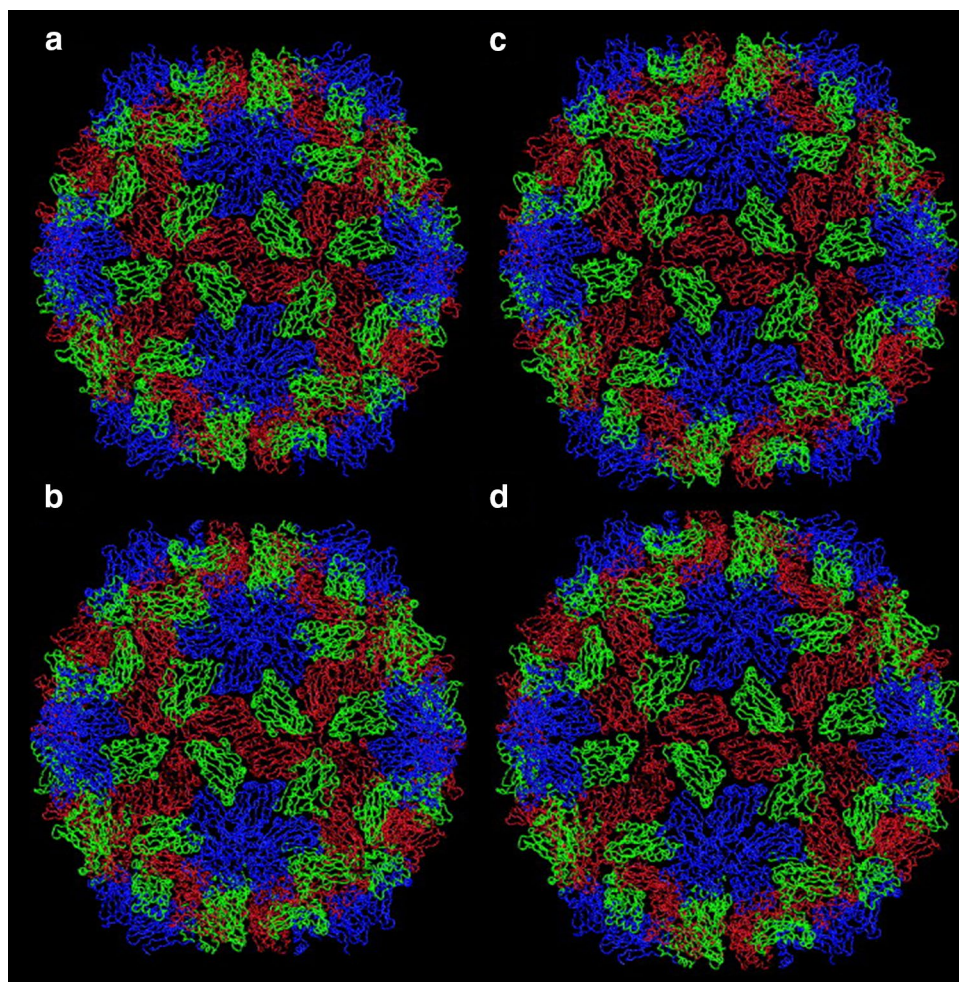
Buehler and colleagues [148] has interestingly investigated the relationship between the molecular structure of amyloid fibrils and their mechanical (e.g. elastic) properties. It should be noted that amyloid fibrils can be formed based on a few types of structural models such as stacked  $\beta$ -sheets,  $\beta$ -helix, and stacked  $\beta$ -helices. They found that amyloid fibrils formed based on stacked  $\beta$ -helices exhibit the high fracture strength, while the fibrils constructed based on  $\beta$ -helix possess the high elastic moduli. The mechanical properties of amyloid fibrils are highly correlated with their molecular structure (i.e. structure type), which determines the rupture mechanisms of hydrogen bonds within the fibril.

### Mechanical and Dynamic Properties of Viral Capsids

Viruses are nano-sized particles that are formed by self-assembly process. For a recent decade, there are efforts that have been conducted for characterizing the mechanical and dynamic properties of viral capsids [149], as these properties are closely related to the biological function of viruses such as gene packaging. Viral capsids are mechanically strong so as to resist disassembly, so that they are able to contain a gene information (i.e. DNA or RNA molecules inside the virus). Here, we review computational efforts that have been made, for a recent decade, to characterize the dynamic and mechanical properties of viral capsids using CG and/or atomistic simulations.

Tama and Brooks [150] had utilized ENM-based NMA simulations for characterizing the dynamic behavior of

**Fig. 11** Maturation dynamics of rice yellow mottle virus (RYMV) and southern bean mosaic virus (SMBV) was analyzed by elastic network model (ENM)-based normal mode analysis (NMA) simulations. The initial structures of **a** RYMV and **b** SMBV are shown. The structures of **c** RYMV and **d** SMBV in swollen state are obtained from ENM-based NMA simulations. Figures are adopted with permission from Ref. [150]. Copyright (2005) Elsevier



icosahedral viral capsids. They found that few low-frequency vibrational modes obtained from ENM are able to sufficiently depict the conformational changes of icosahedral viral capsids such as their shape change from spherical to faceted polyhedral. Their study [150] suggests that icosahedral viral capsids are able to adapt their shapes due to their flexibility (Fig. 11). A previous study by van Vlijmen and Karplus [151] reported the vibrational characteristics of icosahedral viruses based on all-atom NMA. In particular, the stiffness of viruses is computed based on the second derivative of the potential energy prescribed to the virus structure with respect to its atomic coordinates, and then this stiffness is used for NMA to find the natural frequencies of the virus and their corresponding vibrational modes. The low-frequency vibrational modes of poliovirus capsid are shown to be correlated with its shape adaptation. Rader et al. [152] studied the maturation dynamics of bacteriophage HK97 capsid using ENM-based NMA simulations. The authors found that the collective low-frequency vibrational motions of procapsid are highly correlated with its conformational change related to its maturation. This also confirms the role of flexibility for viral capsid in its shape adaptation. Kim

et al. [119] studied the maturation dynamics of HK97 virus capsid using rigid cluster ENI simulations. They showed the conformational transition pathways of HK97 virus capsid. These previous studies [119, 150–152] are still restricted for understanding the dynamic process of maturation (and its related to conformational change) for viral capsids, because the NMA-based studies do not consider the time-evolution dynamics of viral capsids but takes into account only their vibrational modes at equilibrium state. Arkhipov et al. [153] had developed a CG model that is applicable for studying the time-evolution dynamics of viral capsids. Their CG-MD simulations show that some viral capsids such as an empty satellite plant virus (SMTV) collapse rapidly within 500 ns, while other capsids such as brome mosaic virus (BMV) are highly flexible but very stable for a long time of  $> 1 \mu\text{s}$ . It is found that interlocking between the subunits of viral capsid plays a crucial role in its structural stability. Recently, Perilla and Schulten [154] have reported the mechanical and dynamic behavior of viral capsid (i.e. HIV-1 capsid) using all-atom MD simulations based on high-performance parallel supercomputing. In particular, the dynamic trajectories of 64 million atom-sized HIV-1 capsid for the timescale of

1  $\mu\text{s}$  were obtained from all-atom MD simulations. Based on this long-time all-atom MD simulations, the stability of HIV-1 capsid was studied by measuring the size of HIV-1 capsid as a function of time, and the low-frequency vibrational motions of HIV-1 capsid were also investigated.

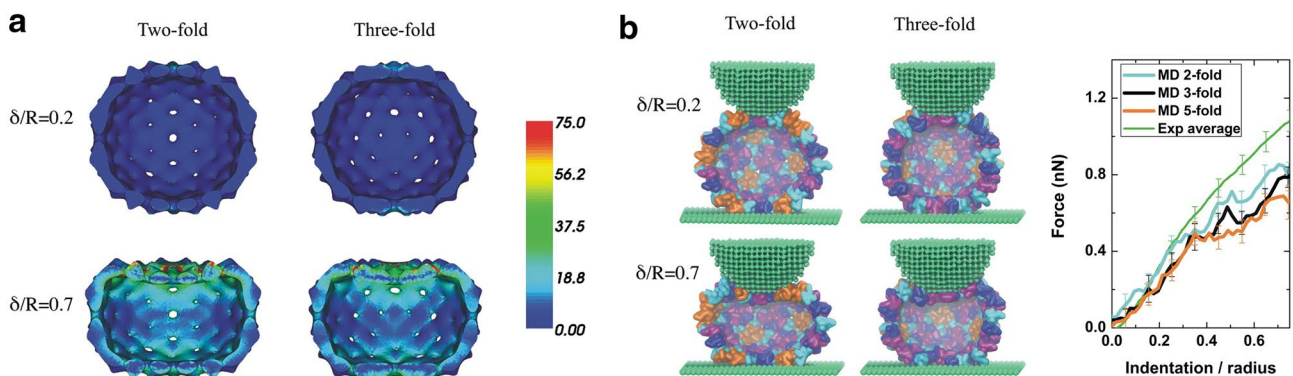
The mechanical properties of viral capsids have recently been probed by AFM indentation experiments [149]. The elastic moduli of viral capsids were found to be 0.1 to 1.8 GPa with dependence on the Föppl-von Kármán (FvK) number,  $\gamma$ , which is defined as  $\gamma = ER^2/D$ , where  $E$ ,  $R$ , and  $D$  are the Young's modulus, radius, and bending rigidity of the viral capsid. This FvK number, which quantifies the competition between bending energy and tension energy during the deformation of the viral capsid, determines the shape of the capsid, i.e. whether it exhibits spherical or icosahedral faceted shapes. In recent years, Widom and coworkers [155] have considered ENM-based NMA simulations coupled with continuum elasticity theory, i.e. Lamé equation (for sphere). It is shown that the low-frequency vibration modes of viral capsid, obtained from ENM-based NMA simulations, were comparable to those predicted by Lamé equation, which suggests the robustness of a continuum elastic model in characterizing the mechanical properties of viral capsids. The Young's modulus ( $E$ ) and Poisson's ratio ( $\nu$ ) of viral capsids were measured as  $\nu = 0.2 \sim 0.3$  and  $E/\gamma = 6.5 \sim 36.2 \text{ nm}^{-1}$ , where  $\gamma$  is a force constant used in ENM (here,  $\gamma = \sim 0.5 \text{ N/m}$ ), i.e.  $E = 3.2 \sim 18.1 \text{ GPa}$ . A previous study by Roos et al. [156] employed a continuum elastic model-based finite element simulations for characterizing the squeezing (i.e. indentation) mechanics of viral capsids. Specifically, the authors considered the finite deformation continuum hyperelasticity (with a neo-Hookean constitutive law) in finite element simulations to probe the indentation mechanism of viral capsids (Fig. 12). It is shown that the mechanical indentation behavior of viral capsids is critically

dependent on their shape that can be described by FvK number. For the capsids with their shape of  $FvK < 150$ , their indentation response is well dictated by a linear relation between indentation displacement and force. However, for the capsids with  $150 < FvK < 800$ , the indentation response becomes nonlinear, whereas the indentation behavior of capsids with  $Fv > 800$  is well depicted by a linear response as long as the indentation displacement becomes 40% of the radius of the capsid.

## Conclusion

In this paper, we introduce recent efforts that have been made, for a last decade, to characterize protein materials at multiple length scales ranging from single proteins to protein assemblies, based on various simulation techniques (at different length scales) from atomistic MD simulations to CG simulations.

First, we have shown that CG-MD simulations (using Go-like or SOP potential) have been extensively employed to understand the mechanical unfolding or refolding dynamics of a single protein molecule due to a force. Here, we note that though all-atom SMD simulations are able to provide a useful insight into protein unfolding (or refolding) mechanisms of a single protein, they are still computationally prohibited for directly interpreting the results of single-molecule force spectroscopy experiments. Specifically, the timescale accessible with all-atom SMD simulations is quite different from that used in a single-molecule force spectroscopy experiment such that a pulling rate used in SMD simulations is larger by six orders of magnitude than that considered in a single-molecule force spectroscopy experiment. This discrepancy between these timescales results in the extensive usage of CG-MD simulations for understanding



**Fig. 12** Indentation mechanics of Hepatitis B virus (HBV) capsid, whose icosahedral geometry is given as  $T=4$  with two- or three-fold symmetry axes. **a** Finite element simulation shows the mechanical deformation mechanisms of HBV capsids with their two- or three-fold symmetry axes during the indentation. **b** Coarse-grained molecu-

lar dynamics (CG-MD) simulation results for the indentation mechanisms of HBV capsids with their different symmetries (left), and the force–displacement curve of indented HBV capsids with two-, three-, or five-fold symmetry axes (right). Figures are adopted with permission from Ref. [156]. Copyright (2010) Elsevier

the unfolding (or refolding) mechanisms of single proteins. However, it should be noted that although CG-MD simulations are able to provide the unfolding dynamics of a single protein with a timescale comparable to that used in experiment, these simulations have to be cautiously used for interpreting the unfolding (or refolding) mechanisms of a single protein, because the CG models may not carry the all atomistic information of the protein. For example, a single point mutation (i.e. single amino acid substitution) may affect the structure, dynamic (e.g. unfolding and/or refolding), and properties of a single protein, which typically cannot be captured by CG models. Recently, with the improvement of computing resources as well as the development of enhanced sampling methods, all-atom MD (or SMD) simulations may be more extensively utilized for interpreting the results of a real single-molecule force spectroscopy experiment. For instance, metadynamics scheme has recently been developed and widely utilized to probe the dynamics of a protein with a timescale relevant to the experiment. Here, the principle of metadynamics is to apply a Gaussian penalty function to the potential field of a protein in order to speed up the reaction rate of a protein related to its structural (conformational) changes due to a force or other stimulus (e.g. ligand-binding) [129, 130]. Our recent studies report the unfolding mechanisms of ubiquitin [157] and/or prion protein [158] based on all-atom metadynamics simulations. In our recent studies [157, 158], it is shown that all-atom model-based metadynamics simulations allow for unveiling the unfolding mechanics of a protein with a timescale relevant to a single-molecule force spectroscopy experiment. Despite few studies [157, 158] of protein unfolding based on an enhanced sampling-based all-atom simulations (e.g. all-atom metadynamics), we believe that recent development of enhanced sampling techniques may enable all-atom MD simulations to be practically considered for interpreting the results of a single-molecule force spectroscopy experiment in near future.

For studying the conformational changes of a single protein molecule, we have discussed the extensive usage of CG models such as ENM. All-atom MD simulations have not been still applicable for revealing the underlying mechanisms of the conformational changes, as their available timescale is much shorter than that relevant to the conformational changes. This has led to the development and usage of CG models such as ENM for studying the conformational changes. The robustness of ENM in predicting the conformational changes is attributed to the fact that the conformational changes of a protein are correlated with its low-frequency vibrational modes. Nevertheless, ENM-based NMA simulations are unable to provide insight into the time-evolution dynamics of conformational changes for a protein. One possibility to overcome this restriction is the development of computational techniques based on coupling between CG model (i.e. ENM) and all-atom MD

simulations. As discussed in a paper by Bahar and coworkers [123], the low-frequency vibrational modes obtained from ENM were used to explore conformational spaces relevant to the conformational changes, while all-atom MD simulations were utilized to obtain the equilibrium dynamics of a protein during its conformational transitions. We anticipate that the sampling based on the normal modes (obtained from CG models) may be coupled to all-atom MD simulations for gaining a profitable knowledge on the conformational transitions of a protein.

For simulating large protein assemblies such as amyloid materials and viral capsids, we have shown that CG or continuum model-based simulations are able to provide the fruitful insight into the structure, dynamics, and properties of these assemblies. Despite a recent study by Perilla and Schulten [154] reporting the dynamics of viral capsids based on all-atom MD simulations, these atomistic simulations are still computationally unfavorable for predicting and characterizing the structure, dynamics, and properties of protein assemblies. We have presented many recent works on CG model (e.g. ENM) and/or continuum model-based simulations for characterizing the mechanical properties of protein assemblies such as amyloid materials and viral capsids, whose size is typically  $> 100$  nm. However, these simulations based on CG models (e.g. ENM) are unable to provide the time-evolution dynamics of mechanical behavior (e.g. fracture) for protein assemblies such as amyloid materials. Though continuum model-based finite element simulations may provide the time-evolution of mechanical deformation for protein assemblies, they require a priori knowledge on a constitutive law (e.g. parameters related to material properties) for protein assemblies. One of routes to improve the simulation techniques for characterizing the protein assemblies is to combine all-atom MD simulations and CG model-based simulations (or continuum model-based simulations such as finite element simulation). Specifically, as described in our recent work [144], all-atom MD simulations are considered to obtain the equilibrium structure of structural segment for protein materials (e.g. amyloid materials), while simulations based on CG model (e.g. ENM) are applied to a whole protein assembly, which was constructed based on the equilibrium structure of structural segment, in order to characterize the properties of this assembly. Another avenue to improve the simulation techniques is that all-atom MD simulations are taken into account to compute parameters (e.g. parameters for a constitutive law) that are required for a finite element simulation of protein assembly, while finite element simulations are performed to depict the time-evolution of mechanical deformation for protein assembly [159].

In conclusion, computer simulations based on various models at multiple length scales, ranging from atomistic model to CG and continuum models, have recently played invaluable role in predicting, characterizing, and revealing

the hidden, complex, underlying mechanisms of proteins' dynamics and mechanics, which govern their biological functions. These simulations may help to not only interpret the experimental results (e.g. results obtained from single-molecule experiments) but also provide new insights into proteins' dynamics and mechanics, which are still difficult to be observed in experiments due to their limited spatial and temporal resolutions. In addition, these computer simulations may also allow for providing an important design principles for developing a novel materials such as new-class protein materials [160–162] and nano-bio composite material [163–165] (i.e. material formed by coupling protein material and nanomaterial).

**Acknowledgements** This work was supported by the National Research Foundation of Korea (NRF) under Grant no. NRF-2015R1A2A2A04002453.

## References

- J. Monod, J. Wyman, J.P. Changeux, *J. Mol. Biol.* **12**, 88 (1965)
- H. Frauenfelder, S.G. Sligar, P.G. Wolynes, *Science* **254**, 1598 (1991)
- K. Arora, C.L. Brooks III, *Proc. Natl. Acad. Sci. USA* **104**, 18496 (2007)
- M. Rief, M. Gautel, F. Oesterhelt, J.M. Fernandez, H.E. Gaub, *Science* **276**, 1109 (1997)
- L. Tskhovrebova, J. Trinick, *J. Biol. Chem.* **279**, 46351 (2004)
- M. Rief, M. Gautel, A. Schemmel, H.E. Gaub, *Biophys. J.* **75**, 3008 (1998)
- H. Lu, B. Isralewitz, A. Krammer, V. Vogel, K. Schulten, *Biophys. J.* **75**, 662 (1998)
- B.L. Smith et al., *Nature* **399**, 761 (1999)
- G. Merlini, V. Bellotti, *N. Engl. J. Med.* **349**, 583 (2003)
- I. Cherny, E. Gazit, *Angew. Chem. Int. Ed.* **47**, 4062 (2008)
- I.W. Hamley, *Chem. Rev.* **112**, 5147 (2012)
- P.C. Ke et al., *Chem. Soc. Rev.* **46**, 6492 (2017)
- A.M. Kushner, Z. Guan, *Angew. Chem. Int. Ed.* **50**, 9026 (2011)
- M.I. Solar, M.J. Buehler, *Nat. Nanotechnol.* **7**, 417 (2012)
- A.R. Studart, *Adv. Mater.* **24**, 5024 (2012)
- F. Chiti, C.M. Dobson, *Annu. Rev. Biochem.* **75**, 333 (2006)
- M.B. Pepys, *Annu. Rev. Med.* **57**, 223 (2006)
- J.W.M. Hoppener, B. Ahren, C.J.M. Lips, *N. Engl. J. Med.* **343**, 411 (2000)
- C. Rapezzi et al., *Nat. Rev. Cardiol.* **7**, 398 (2010)
- T.P.J. Knowles, M.J. Buehler, *Nat. Nanotechnol.* **6**, 469 (2011)
- B. Choi, T. Kim, S.W. Lee, K. Eom, *J. Nanomater.* **2016**, 5873695 (2016)
- J. Adamcik, R. Mezzenga, *Curr. Opin. Colloid Interface Sci.* **17**, 369 (2012)
- M.F.M. Engel et al., *Proc. Natl. Acad. Sci. USA* **105**, 6033 (2008)
- A.W.P. Fitzpatrick, S.T. Park, A.H. Zewail, *Proc. Natl. Acad. Sci. USA* **110**, 10976 (2013)
- S. Keten, Z. Xu, B. Ihle, M.J. Buehler, *Nat. Mater.* **9**, 359 (2010)
- K.C. Neuman, T. Lionnet, J.F. Allemand, *Annu. Rev. Mater. Res.* **37**, 33 (2007)
- D.J. Muller, M. Krieg, D. Alsteens, Y.F. Dufrene, *Curr. Opin. Biotechnol.* **20**, 4 (2009)
- D.J. Muller, Y.F. Dufrene, *Nat. Nanotechnol.* **3**, 261 (2008)
- E.M. Puchner, H.E. Gaub, *Curr. Opin. Struct. Biol.* **19**, 605 (2009)
- C. Bustamante, Z. Bryant, S.B. Smith, *Nature* **421**, 423 (2003)
- M. Karplus, G.A. Petsko, *Nature* **347**, 631 (1990)
- M. Karplus, J.A. McCammon, *Nat. Struct. Mol. Biol.* **9**, 646 (2002)
- M. Bathe, *Proteins Struct. Funct. Bioinform.* **70**, 1595 (2008)
- G. Yun, J. Kim, D.-N. Kim, *J. Comput. Aided Mol. Des.* **31**, 609 (2017)
- B.J. Alder, T.E. Wainwright, *J. Chem. Phys.* **27**, 1208 (1957)
- J.A. McCammon, B.R. Gelin, M. Karplus, *Nature* **267**, 585 (1977)
- M. Karplus, *Annu. Rev. Biophys. Biomol. Struct.* **35**, 1 (2006)
- M. Levitt, *Nat. Struct. Biol.* **8**, 392 (2001)
- R. Van Noorden, *Nature* **502**, 280 (2013)
- J.A. McCammon, S. Harvey, *Dynamics of proteins and nucleic acids* (Cambridge University Press, Cambridge, 1987)
- S.W. Cranford, M.J. Buehler, *Biomaterialomics* (Springer, New York, 2012)
- J. Norberg, L. Nilsson, *Acc. Chem. Res.* **35**, 465 (2002)
- P. Auffinger, E. Westhof, *Curr. Opin. Struct. Biol.* **8**, 227 (1998)
- M.P. Allen, D.J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, Oxford, 1987)
- A. Amadei, A.B.M. Linssen, H.J.C. Berendsen, *Proteins Struct. Funct. Genet.* **17**, 412 (1993)
- M. Sotomayor, K. Schulten, *Science* **316**, 1144 (2007)
- Q. Cui, I. Bahar, *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems* (CRC Press, Boca Raton, 2005)
- M.M. Teeter, D.A. Case, *J. Phys. Chem.* **94**, 8091 (1990)
- B. Brooks, M. Karplus, *Proc. Natl. Acad. Sci. USA* **80**, 6571 (1983)
- S. Hayward, N. Go, *Annu. Rev. Phys. Chem.* **46**, 223 (1995)
- G.A. Voth, *Coarse-Graining of Condensed Phase and Biomolecular Systems* (CRC Press, Boca Raton, 2009)
- A.J. Rader, *Curr. Opin. Pharmacol.* **10**, 753 (2010)
- G.S. Ayton, W.G. Noid, G.A. Voth, *MRS Bull.* **32**, 929 (2007)
- D.E. Shaw et al., *Science* **330**, 341 (2010)
- H.S. Chung, S. Piana-Agostinetti, D.E. Shaw, W.A. Eaton, *Science* **349**, 1504 (2015)
- R.O. Dror, R.M. Dirks, J.P. Grossman, H. Xu, D.E. Shaw, *Annu. Rev. Biophys.* **41**, 429 (2012)
- Y. Suezaki, N. Go, *Int. J. Peptid. Protein Res.* **7**, 333 (1975)
- Y. Ueda, N. Go, *Int. J. Peptid. Protein Res.* **8**, 551 (1976)
- Y. Ueda, H. Taketomi, N. Go, *Biopolymers* **17**, 1531 (1978)
- M. Cieplak, T.X. Hoang, M.O. Robbins, *Proteins Struct. Funct. Genet.* **49**, 114 (2002)
- M.M. Tirion, *Phys. Rev. Lett.* **77**, 1905 (1996)
- T. Haliloglu, I. Bahar, B. Erman, *Phys. Rev. Lett.* **79**, 3090 (1997)
- A.R. Atilgan et al., *Biophys. J.* **80**, 505 (2001)
- K. Eom, G. Yoon, J.-I. Kim, S. Na, *J. Comput. Theor. Nanosci.* **7**, 1210 (2010)
- K. Eom, S.-C. Baek, J.-H. Ahn, S. Na, *J. Comput. Chem.* **28**, 1400 (2007)
- M. Fixman, J. Kovac, *J. Chem. Phys.* **58**, 1564 (1973)
- C. Rivetti, M. Guthold, C. Bustamante, *J. Mol. Biol.* **264**, 919 (1996)
- J. Adamcik et al., *Nat. Nanotechnol.* **5**, 423 (2010)
- F. Pampaloni et al., *Proc. Natl. Acad. Sci. USA* **103**, 10248 (2006)
- P.E. Marszalek et al., *Nature* **402**, 100 (1999)
- L. Tskhovrebova, J. Trinick, J.A. Sleep, R.M. Simmons, *Nature* **387**, 308 (1997)
- P.M. Williams et al., *Nature* **422**, 446 (2003)
- M.S. Kellermayer, S.B. Smith, H.L. Granzier, C. Bustamante, *Science* **276**, 1112 (1997)

74. J. Gore et al., *Nature* **442**, 836 (2006)
75. C. Baumann, S. Smith, V. Bloomfield, C. Bustamante, *Proc. Natl. Acad. Sci. USA* **94**, 6185 (1997)
76. C. Bustamante, J.F. Marko, E.D. Siggia, S. Smith, *Science* **265**, 1599 (1994)
77. J.F. Marko, E.D. Siggia, *Macromolecules* **28**, 8759 (1995)
78. A. Kis et al., *Phys. Rev. Lett.* **89**, 248101 (2002)
79. K. Eom, H.S. Park, D.S. Yoon, T. Kwon, *Phys. Rep.* **503**, 115 (2011)
80. L. Meirovitch, *Analytical Methods in Vibrations* (Macmillan, New York, 1967)
81. J.M. Gere, *Mechanics of Materials*, 6th edn. (Thomson Learning, Boston, 2003)
82. S.P. Timoshenko, *Philos. Mag.* **41**, 744 (1921)
83. J.F. Allemand, D. Bensimon, V. Croquette, *Curr. Opin. Struct. Biol.* **13**, 266 (2003)
84. W.K. Olson, V.B. Zhurkin, *Curr. Opin. Struct. Biol.* **10**, 286 (2000)
85. H. Li et al., *Nature* **418**, 998 (2002)
86. S.P. Ng, R.W. Rounsevell, A. Steward, C.D. Geierhaas, P.M. Williams, *J. Mol. Biol.* **350**, 776 (2005)
87. A. Krammer, H. Lu, B. Isralewitz, K. Schulten, V. Vogel, *Proc. Natl. Acad. Sci. USA* **96**, 1351 (1999)
88. M. Gao, M. Wilmanns, K. Schulten, *Biophys. J.* **83**, 3435 (2002)
89. H. Lu, K. Schulten, *Proteins* **35**, 453 (1999)
90. M. Gao, D. Craig, V. Vogel, K. Schulten, *J. Mol. Biol.* **323**, 939 (2002)
91. M. Cieplak, T.X. Hoang, M.O. Robbins, *Proteins Struct. Funct. Genet.* **49**, 104 (2002)
92. M. Sikora, J.I. Sulkowska, M. Cieplak, *PLoS Comput. Biol.* **5**, e1000547 (2009)
93. J.I. Sulkowska, P. Sulkowski, P. Szymczak, M. Cieplak, *J. Am. Chem. Soc.* **132**, 13954 (2010)
94. J.I. Sulkowska, P. Sulkowski, J.N. Onuchic, *Phys. Rev. Lett.* **103**, 268103 (2009)
95. M. Mickler et al., *Proc. Natl. Acad. Sci. USA* **104**, 20268 (2007)
96. C. Hyeon, G. Morrison, D. Thirumalai, *Proc. Natl. Acad. Sci. USA* **105**, 9604 (2008)
97. R.I. Dima, H. Joshi, *Proc. Natl. Acad. Sci. USA* **105**, 15743 (2008)
98. C. Hyeon, G. Morrison, D.L. Pincus, D. Thirumalai, *Proc. Natl. Acad. Sci. USA* **106**, 20288 (2009)
99. R. Merkel, P. Nassoy, A. Leung, K. Ritchie, E. Evans, *Nature* **397**, 50 (1999)
100. G.I. Bell, *Science* **200**, 618 (1978)
101. L.B. Freund, *Proc. Natl. Acad. Sci. USA* **106**, 8818 (2009)
102. A. Maitra, G. Arya, *Phys. Rev. Lett.* **104**, 108301 (2010)
103. P. Hanggi, P. Talkner, M. Borkovec, *Rev. Mod. Phys.* **62**, 251 (1990)
104. G. Yoon, S. Na, K. Eom, *J. Chem. Phys.* **137**, 025102 (2012)
105. R.G. Smock, L.M. Gierasch, *Science* **324**, 198 (2009)
106. T. Force, D.S. Krause, R.A. Van Etten, *Nat. Rev. Cancer* **7**, 332 (2007)
107. S. Piana, K. Lindorff-Larsen, D.E. Shaw, *Proc. Natl. Acad. Sci. USA* **110**, 5915 (2013)
108. E. Paci, M. Karplus, *Proc. Natl. Acad. Sci. USA* **97**, 6521 (2000)
109. M.Y. Lu, J.P. Ma, *Biophys. J.* **89**, 2395 (2005)
110. D. Tobi, I. Bahar, *Proc. Natl. Acad. Sci. USA* **102**, 18908 (2005)
111. A. Bakan, I. Bahar, *Proc. Natl. Acad. Sci. USA* **106**, 14349 (2009)
112. C. Chennubhotla, A.J. Rader, L.W. Yang, I. Bahar, *Phys. Biol.* **2**, S173 (2005)
113. M. Cecchini, A. Houdusse, M. Karplus, *PLoS Comput. Biol.* **4**, e1000129 (2008)
114. O. Miyashita, J.N. Onuchic, P.G. Wolynes, *Proc. Natl. Acad. Sci. USA* **100**, 12570 (2003)
115. P. Maragakis, M. Karplus, *J. Mol. Biol.* **352**, 807 (2005)
116. F. Takagi, M. Kikuchi, *Biophys. J.* **93**, 3820 (2007)
117. W. Zheng, B.R. Brooks, G. Hummer, *Proteins Struct. Funct. Bioinform.* **69**, 43 (2007)
118. M.K. Kim, W. Li, B.A. Shapiro, G.S. Chirikjian, *J. Biomol. Struct. Dyn.* **21**, 395 (2003)
119. M.K. Kim, R.L. Jernigan, G.S. Chirikjian, *Biophys. J.* **89**, 43 (2005)
120. B.H. Lee et al., *PLoS One* **12**, e0185658 (2017)
121. M. Ikeguchi, J. Ueno, M. Sato, A. Kidera, *Phys. Rev. Lett.* **94**, 078102 (2005)
122. M.C. Demirel, A.M. Lesk, *Phys. Rev. Lett.* **95**, 208106 (2005)
123. B. Isin, K. Schulten, E. Tajkhorshid, I. Bahar, *Biophys. J.* **95**, 789 (2008)
124. N. Kantarci-Carsibasi, T. Haliloglu, P. Doruker, *Biophys. J.* **95**, 5862 (2008)
125. W. Hwang, S. Zhang, R.D. Kamm, M. Karplus, *Proc. Natl. Acad. Sci. USA* **101**, 12916 (2004)
126. M. DeMarco, V. Daggett, *Proc. Natl. Acad. Sci. USA* **101**, 2293 (2004)
127. S.-H. Chong, S. Ham, *Proc. Natl. Acad. Sci. USA* **109**, 7636 (2012)
128. Y. Sugita, Y. Okamoto, *Chem. Phys. Lett.* **314**, 141 (1999)
129. B. Alessandro, B. Massimiliano, P. Michele, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **1**, 826 (2011)
130. L. Alessandro, L.G. Francesco, *Rep. Prog. Phys.* **71**, 126601 (2008)
131. L.E. Buchanan et al., *Proc. Natl. Acad. Sci. USA* **110**, 19285 (2013)
132. C. Wu, J.-E. Shea, *PLoS Comput. Biol.* **9**, e1003211 (2013)
133. S. Chakraborty, P. Das, *Sci. Rep.* **7**, 9941 (2017)
134. Z.A. Levine, L. Larini, N.E. LaPointe, S.C. Feinstein, J.-E. Shea, *Proc. Natl. Acad. Sci. USA* **112**, 2758 (2015)
135. M. Tanaka, S.R. Collins, B.H. Toyama, J.S. Weissman, *Nature* **442**, 585 (2006)
136. G. Yoon, M. Lee, J.I. Kim, S. Na, K. Eom, *PLoS ONE* **9**, e88502 (2014)
137. M. Lee, J.I. Kim, S. Na, K. Eom, *Phys. Chem. Chem. Phys.* **20**, 8951 (2018)
138. M. Porrini, U. Zachriae, P.E. Barran, C.E. MacPhee, *J. Phys. Chem. Lett.* **4**, 1233 (2013)
139. T.P. Knowles et al., *Science* **318**, 1900 (2007)
140. Z.P. Xu, R. Paparcone, M.J. Buehler, *Biophys. J.* **98**, 2053 (2010)
141. G. Yoon, J. Kwak, J.I. Kim, S. Na, K. Eom, *Adv. Funct. Mater.* **21**, 3454 (2011)
142. G. Yoon, Y.K. Kim, K. Eom, S. Na, *Appl. Phys. Lett.* **102**, 011914 (2013)
143. H. Choi et al., *RSC Adv.* **5**, 49263 (2015)
144. G. Yoon et al., *Phys. Biol.* **12**, 066021 (2015)
145. S. Xiao, S. Xiao, F. Grater, *Phys. Chem. Chem. Phys.* **15**, 8765 (2013)
146. B. Choi, G. Yoon, S.W. Lee, K. Eom, *Phys. Chem. Chem. Phys.* **17**, 1379 (2015)
147. B. Choi, T. Kim, E.S. Ahn, S.W. Lee, K. Eom, *Nanoscale Res. Lett.* **12**, 228 (2017)
148. M. Solar, M.J. Buehler, *Nanotechnology* **25**, 105703 (2014)
149. W.H. Roos, R. Bruinsma, G.J.L. Wuite, *Nat. Phys.* **6**, 733 (2010)
150. F. Tama, C.L. Brooks, *J. Mol. Biol.* **345**, 299 (2005)
151. H.W.T. van Vlijmen, M. Karplus, *J. Mol. Biol.* **350**, 528 (2005)
152. A.J. Rader, D.H. Vlad, I. Bahar, *Structure* **13**, 413 (2005)
153. A. Arkhipov, P.L. Freddolino, K. Schulten, *Structure* **14**, 1767 (2006)
154. J.R. Perilla, K. Schulten, *Nat. Commun.* **8**, 15959 (2017)
155. Z. Yang, I. Bahar, M. Widom, *Biophys. J.* **96**, 4438 (2009)



156. W.H. Roos et al., *Biophys. J.* **99**, 1175 (2010)
157. P. Cao, G. Yoon, W. Tao, K. Eom, H.S. Park, *Sci. Rep.* **5**, 8757 (2015)
158. W. Tao, G. Yoon, P. Cao, K. Eom, H.S. Park, *J. Chem. Phys.* **143**, 125101 (2015)
159. R. Phillips, M. Dittrich, K. Schulten, *Annu. Rev. Mater. Res.* **32**, 219 (2002)
160. R.V. Mannige et al., *Nature* **526**, 415 (2015)
161. Y. Mou, J.-Y. Yu, T.M. Wannier, C.-L. Guo, S.L. Mayo, *Nature* **525**, 230 (2015)
162. B. Dai et al., *Proc. Natl. Acad. Sci. USA* **112**, 2996 (2015)
163. G. Grigoryan et al., *Science* **332**, 1071 (2011)
164. K.-H. Kim et al., *Nat. Commun.* **7**, 11429 (2016)
165. Y.H. No et al., *J. Phys. Chem. Lett.* **8**, 3734 (2017)