Vol. 15, pp. 58–74 (2018)
doi: 10.2142/biophysico.15.0_58

Biophysics and Physicobiology

https://www.jstage.jst.go.jp/browse/biophysico/

Regular Article

Theoretical framework for analyzing structural compliance properties of proteins

Keisuke Arikawa

Department of Mechanical Engineering Kanagawa Institute of Technology, Atsugi, Kanagawa 243-0292, Japan

Received September 8, 2017; accepted December 27, 2017

We propose methods for directly analyzing structural compliance (SC) properties of elastic network models of proteins, and we also propose methods for extracting information about motion properties from the SC properties. The analysis of SC properties involves describing the relationships between the applied forces and the deformations. When decomposing the motion according to the magnitude of SC (SC mode decomposition), we can obtain information about the motion properties under the assumption that the lower SC mode motions or the softer motions occur easily. For practical applications, the methods are formulated in a general form. The parts where forces are applied and those where deformations are evaluated are separated from each other for enabling the analyses of allosteric interactions between the specified parts. The parts are specified not only by the points but also by the groups of points (the groups are treated as flexible bodies). In addition, we propose methods for quantitatively evaluating the properties based on the screw theory and the considerations of the algebraic structures of the basic equations expressing the SC properties. These methods enable quantitative discussions about the relationships between the SC mode motions and the motions estimated from two different conformations; they also help identify the key parts that play important roles for the motions by comparing the SC properties with those of partially constrained models. As application examples, lactoferrin and ATCase are analyzed. The results show that we can understand their motion properties through their lower SC mode motions or the softer motions.

Key words: protein motion, elastic network model, SC mode decomposition

Corresponding author: Keisuke Arikawa, Department of Mechanical Engineering Kanagawa Institute of Technology, 1030, Shimo-ogino, Atsugi, Kanagawa 243-0292, Japan.
e-mail: arikawa@me.kanagawa-it.ac.jp

All organisms are continuously exposed to various forces. At the tissue level, external forces such as touch apply pressure on the skin. Bones support weight, lungs undergo periodic stretching, and blood vessels experience the force of blood flows. At the cell level, each cell experiences forces from the neighboring cells and extra-cellular matrices. Intensive studies in mechanobiology have revealed that forces play important roles in life not only at the tissue or the cell level but also at the molecular level; these studies have also clarified the functions of many proteins (e.g., Integrin, Cadherin, Tarin, MscL) related to the response to mechanical forces [1–6]. Forces play important roles not only for the proteins related to the response to the mechanical forces. Studies based on normal mode analysis (NMA) clarified that the functional protein motions can be approximated by combining lower normal mode motions [7–22]. In general,

◀ ***Significance*** ▶

We propose methods for directly analyzing structural compliance (SC) properties of the elastic network models of proteins, methods for extracting information about the motion properties from the SC properties, and methods for quantitatively evaluating the motion properties. Under the assumption that the softer motions occur easily, we can obtain information about the motion properties, including the allosteric interactions between the specified parts, by decomposing the motion according to the magnitude of SC (SC mode decomposition). As application examples, we have analyzed lactoferrin and ATCase. The results show that we can understand their motion properties using their softer motions.

natural frequencies of soft objects are low and the flexibility of objects depends on the directions of the applied forces. Moreover, Ikeguchi *et al.* showed that the conformation changes of proteins when bound to ligands can be well predicted from their ligand-free forms using linear response theory (LRT) [23]. In their method, the proteins' responses to perturbations are assumed to be proportional to the fluctuations of their ligand-free forms, and the perturbations are expressed as forces between the ligands and proteins. From these facts, we can understand that motion properties of proteins have a close relationship with the responses to the forces.

The force response properties of an object can be characterized by the structural compliance properties of the object. For a particle connected to a wall by a linear spring, the structural compliance relates the displacement of the particle to the force applied on the particle. In this system, the structural compliance corresponds to the inverse of the spring constant of the spring; in this sense, it expresses the flexibility of the spring. When a force is applied to a 3D object, the structural compliance relates the deformation to the applied force. The structural compliance is generally represented as a matrix that connects the deformation vector and the force vector. The flexibility depends on the direction of the applied force. The structural compliance matrix contains information of the structural compliance properties, which mainly describe the directional dependence of the flexibility. Therefore, if we can directly analyze the structural compliance properties of proteins from their 3D structural data or PDB data [24],we will be able to obtain information about motion properties. The objective of this research is to formulate the method and understand the motion properties of proteins through their structural compliance properties.

As a protein model, we focus on the elastic network model (ENM). Tirion compared the simulation results of NMA calculated by using a complex semiempirical potential field with those using a simple potential field constructed by linear springs; she showed that the results for the lower normal mode motions were similar [9]. Tama *et al*. investigated 20 proteins whose 3D structural data of open and close forms are archived in the PDB. They compared the simulation results of NMA using C_α based ENM with the motions estimated from the difference between the open and close forms. They revealed that information on the nature of the conformation change of a protein is often contained in a single lower normal mode motion of its open form [12]. Various types of ENMs that are different at the coarse grain level and in the setting of the spring constant have been proposed. The results of the studies based on NMA by using ENM show that the essential information about functional motion is preserved in ENM in spite of the simplicity of the model. Owing to the reduction of computational load, it becomes possible to comprehend the essential motion properties of large proteins and even comprehensively analyze the proteins stored in PDB [11–15,17,20–22]. Therefore, we expect that it will be possible to obtain information about motion properties by analyzing structural compliance properties of ENMs with less computational cost.

Related to the NMA by using ENM, it is know that we can analyze ENM behavior when an external force is applied by adding external force terms to the formulation of NMA. By using the methods, the structural mechanisms that enable the allosteric effect were investigated [21], and roles of forces were investigated from the view point of mechanobiology [25]. In contrast to the methods, we directly analyze the structural compliance properties of ENMs and extract information about motion properties without calculating the normal modes. An ENM is a type of system with multiple degrees of freedom (DOF). The methods used in robotics are extremely useful for analyzing such systems. In particular, when the ENM is constructed by using dihedral angles, the model can be treated as an assembly of robotic arms; the dihedral angles of ENM correspond to the joint angles of robotic arms. The following examples show that various techniques in robotics have been applied for the analyses of protein motions [26–37]. The method for solving the inverse kinematics problem (i.e., the problem of calculating joint displacements from the specified hand configurations) is applicable for solving the loop closure problems in proteins (i.e., the problem of moving the localized part of the proteins without affecting the surrounding parts) [26,27,36]. Statics and path planning of robotic arms are applicable for describing the folding process of proteins [30]. The methods used for robot kinematics are effective for generating the trajectory between two conformations of proteins [35].

We have, so far, derived the basic formula for directly analyzing structural compliance properties of the ENM of proteins and extracting the information of motion properties by making the best use of the methods in robotics [37]. For more practical applications, we formulate the method in a more general form and present some methods for quantitatively evaluating the properties. In the general formulation, the forces are assumed to apply to the points and the groups of points (the groups are not treated as rigid but as flexible bodies). The methods for quantitatively evaluating the properties are formulated based on the screw theory and the considerations of the algebraic structures of the basic equations expressing structural compliance properties. As the application examples of the methods, we will show the results of the analyses of lactoferrin and aspartate transcarbamoylase.

Methods

Protein Model

Figure 1 illustrates the approximate protein model employed in this research. The model is main-chain-based ENM. The conformation of the model is expressed by the dihedral angles around the N-C_α bond axes (ϕ) and the C_α-C bond axes (ψ) (The side chain structures are not modeled in

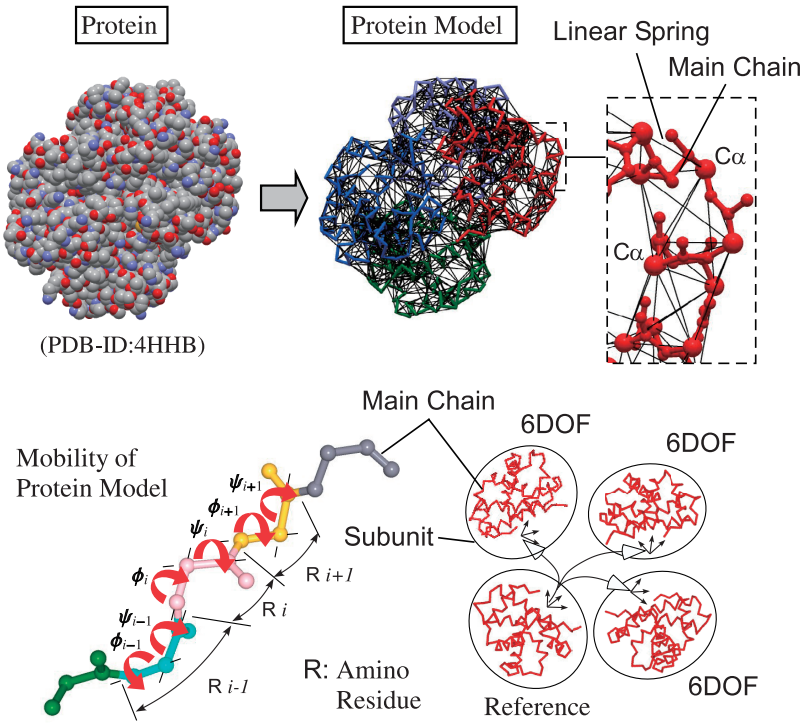

Figure 1 Protein model. The model is a main-chain-based ENM. The conformation of the model is expressed by the dihedral angles around the N-C_α bond axes (ϕ) and the C_α-C bond axes (ψ). For multisubunit proteins, the relative position and orientation (6DOF) between the main chains are added. Linear springs are placed between C_α whose distances are less than the threshold value. Their natural lengths are set as the distance between C_α in the PDB data.

this protein model). For multisubunit proteins, the relative position and orientation (6DOF) between the main chains are added, where the reference subunit can be specified arbitrarily. Linear springs are placed between the C_α whose distances are less than the threshold value or the cutoff value L_{th}. Their natural lengths are set as the distance between C_α in the PDB data.

The variables that are required to represent the conformation of the model are called conformation variables. Let n be the number of conformation variables (i.e., the DOF of the model), and $\boldsymbol{\theta} = (\theta_1 \cdots \theta_n)^T$ be the vector assembling all the conformation variables. The change of the vector $\Delta\boldsymbol{\theta}$ expresses the internal motion of the model. The exemplified protein model shown in Figure 1 is created by using the PDB data of hemoglobin (PDB-ID:4HHB), where L_{th} is set as 8 Å. The number of conformation variable is n = 574, and the number of springs is 2,817.

Definition of Structural Compliance

The basic concept of the proposed method is simple. We consider the deformations of the protein model under applied forces. Even under forces of the same magnitude, the deformation magnitude will depend on the force direction. The protein model is softer for the force directions which cause larger deformation magnitude. The basic concept of the method is that protein motions occur easily in the softer directions. Mathematically, the deformation vector $\Delta\boldsymbol{X}$ and the force vector $\boldsymbol{F}$ are related through $\Delta\boldsymbol{X} = \boldsymbol{CF}$. The matrix $\boldsymbol{C}$ is called the structural compliance matrix. The magnitude of structural compliance is given by $|\Delta\boldsymbol{X}|/|\boldsymbol{F}|$; note that the larger magnitude of structural compliance means that the protein model is softer for the applied force direction. By analyzing the properties of $\boldsymbol{C}$, we can obtain the motion modes called the SC mode. The lower SC mode motions align with the directions of larger magnitude of structural compliance (softer motions). From the decomposed SC mode motions, we can extract the motion properties of proteins. To improve the practicality of the method we define the structural compliance in a more general form and constrain the forces and the protein model (as shown below). Through these modifications, we can obtain much information related to the protein motions.

The structural compliance for this study is more general than normal. Normally, the structural compliance of an object is defined based on the relationship between the applied forces and the deformation of the part where the forces are applied. In contrast, as shown in Figure 2, we separate the parts where the deformations are evaluated (Deformation Part in Fig. 2) from the parts where forces are applied (Application Part in Fig. 2). The separation enables the analyses of the motion interactions between the parts distant from one another.

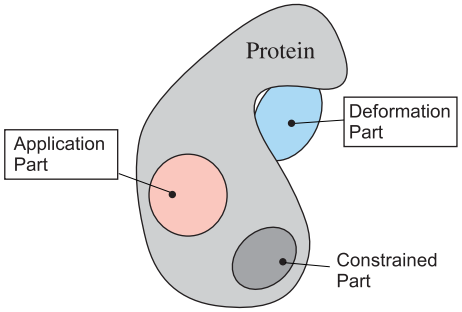

Figure 2 Definition of structural compliance. Normally, the structural compliance of an object is defined based on the relationships between the applied forces and the deformation of the part where the forces are applied. In contrast, in the general definition of the structural compliance, the parts where the deformations are evaluated (labeled as Deformation Part) are separated from the parts where forces are applied (labeled as Application Part). We assumed that no part of the protein model is fixed to the external environment. To realize this condition in statics, the applied forces are restricted to those in static equilibrium. In addition, the deformation of the specified parts in the protein model are constrained as required (labeled as Constrained Part).

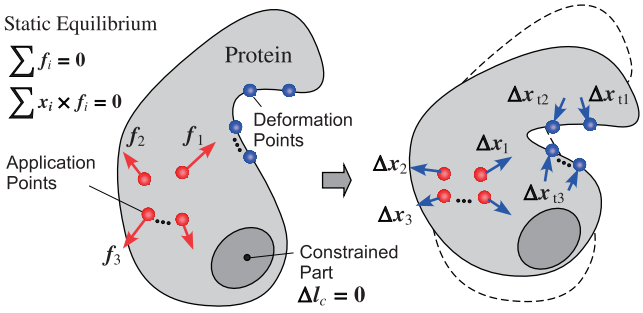

Figure 3 Calculation of structural compliance. External forces $\boldsymbol{f}_1, \cdots, \boldsymbol{f}_{n_f}$ are assumed to be applied at the specified n_f application points and the points displaced $\Delta\boldsymbol{x}_1, \cdots, \Delta\boldsymbol{x}_{n_f}$, where the forces are in static equilibrium ($\sum \boldsymbol{f}_i = \boldsymbol{0}$ and $\sum \boldsymbol{x}_i \times \boldsymbol{f}_i = \boldsymbol{0}$ are established). By the forces application, it is assumed that the specified n_t deformation points displaced $\Delta\boldsymbol{x}_{t1}, \cdots, \Delta\boldsymbol{x}_{tn_t}$. In the deformation, the vector $\boldsymbol{l}_c$ assembling the parameters constrained in the constrained parts (e.g., the distances between the specified points in the model) is kept unchanged (i.e., $\Delta\boldsymbol{l}_c = \boldsymbol{0}$).

We assume that no part of the protein model is fixed to the external environment. To realize this condition in statics, the applied forces are restricted to those in static equilibrium. In addition, we constrain the deformation of the parts in the protein model as required (Constrained Part in Fig. 2). By this constraint, we can express the constraints by ligand bindings, disulfide bonds, and so on. Moreover, the constraint to the model is effective for identifying the key parts that govern the motion properties. If the motion property changes greatly when a part is constrained, it can be inferred that the constrained part is the key part. It should be noted that the no part of the protein model including the constrained part is fixed to the external environment.

Structural Compliance Analysis

In this section, we summarize the basic formula for calculating the structural compliance of the protein model. As described earlier, we derived the formula by making the best use of the methods used in robotics. The details of the derivation of the formula are shown in our previous paper [37].

As shown in Figure 3, we assume that the external forces $\boldsymbol{f}_1, \cdots, \boldsymbol{f}_{n_f}$ are applied at n_f application points, and the points displaced $\Delta\boldsymbol{x}_1, \cdots, \Delta\boldsymbol{x}_{n_f}$. By assembling these vectors, we define the force vector $\boldsymbol{F}$ as $\boldsymbol{F} = (\boldsymbol{f}_1{}^T \cdots \boldsymbol{f}_{n_f}{}^T)^T$ and the displacement vector $\Delta\boldsymbol{X}$ as $\Delta\boldsymbol{X} = (\Delta\boldsymbol{x}_1{}^T \cdots \Delta\boldsymbol{x}_{n_f}{}^T)^T$. The work W_f done by the external forces is expressed as $W_f = \boldsymbol{F}^T\Delta\boldsymbol{X}/2 = \boldsymbol{F}^T\boldsymbol{J}_f\Delta\boldsymbol{\theta}/2$, where $\boldsymbol{J}_f$ is the Jacobian matrix that relates the displacements of the application points $\Delta\boldsymbol{X}$ and the changes of the conformation variables $\Delta\boldsymbol{\theta}$ ($\Delta\boldsymbol{X} = \boldsymbol{J}_f\Delta\boldsymbol{\theta}$). Let $\Delta\boldsymbol{l}_s$ be the vector assembling all the spring deflections in this deformation. The energy E_s stored in the springs is expressed as $E_s = \Delta\boldsymbol{l}_s^T\boldsymbol{K}\Delta\boldsymbol{l}_s/2 = \Delta\boldsymbol{\theta}^T\boldsymbol{J}_s{}^T\boldsymbol{K}\boldsymbol{J}_s\Delta\boldsymbol{\theta}/2$, where $\boldsymbol{K}$ is the diagonal matrix that contains spring constants as its diagonal elements, and $\boldsymbol{J}_s$ is the Jacobian matrix that relates the spring deflections $\Delta\boldsymbol{l}_s$ and $\Delta\boldsymbol{\theta}$ ($\Delta\boldsymbol{l}_s = \boldsymbol{J}_s\Delta\boldsymbol{\theta}$). The relationship $W_f = E_s$ gives the following equation.

$$\boldsymbol{F}^T\boldsymbol{J}_f\Delta\boldsymbol{\theta} = \Delta\boldsymbol{\theta}^T\boldsymbol{J}_s{}^T\boldsymbol{K}\boldsymbol{J}_s\Delta\boldsymbol{\theta} \tag{1}$$

Solving this equation about $\Delta\boldsymbol{\theta}$, we obtain Eq. (2), which expresses the deformation of the model or the changes in the conformation variables $\Delta\boldsymbol{\theta}$ when the force $\boldsymbol{F}$ is applied.

$$\Delta\boldsymbol{\theta} = (\boldsymbol{J}_s{}^T\boldsymbol{K}\boldsymbol{J}_s)^{-1}\boldsymbol{J}_f{}^T\boldsymbol{F} = \boldsymbol{D}\boldsymbol{F} \tag{2}$$

where $\boldsymbol{D}$ is defined as $\boldsymbol{D} = (\boldsymbol{J}_s{}^T\boldsymbol{K}\boldsymbol{J}_s)^{-1}\boldsymbol{J}_f{}^T$.

As explained in the previous section, we assume that the external forces are in static equilibrium ($\sum \boldsymbol{f}_i = \boldsymbol{0}$ and $\sum \boldsymbol{x}_i \times \boldsymbol{f}_i = \boldsymbol{0}$ are established). The equilibrium condition is linear about $\boldsymbol{F}$; therefore, it can be expressed in the form $\boldsymbol{B}\boldsymbol{F} = \boldsymbol{0}$. Using $\boldsymbol{B}^\perp$ (the columns form the orthonormal basis of the kernel space of $\boldsymbol{B}$), all $\boldsymbol{F}$ satisfying the condition are expressed as:

$$\boldsymbol{F} = \boldsymbol{B}^\perp\boldsymbol{F}_b \tag{3}$$

where $\boldsymbol{F}_b$ is an arbitrary vector. We regard $\boldsymbol{F}_b$ as the force vector instead of $\boldsymbol{F}$. $\boldsymbol{F}^T\Delta\boldsymbol{X} = \boldsymbol{F}_b{}^T\boldsymbol{B}^{\perp T}\Delta\boldsymbol{X}$ has the dimension of work. Therefore, the displacement $\Delta\boldsymbol{X}_b$ compatible with $\boldsymbol{F}_b$ should be defined by Eq. (4) so that $\boldsymbol{F}^T\Delta\boldsymbol{X}$ equals $\boldsymbol{F}_b{}^T\Delta\boldsymbol{X}_b$.

$$\Delta\boldsymbol{X}_b = \boldsymbol{B}^{\perp T}\Delta\boldsymbol{X} \tag{4}$$

We regard $\Delta\boldsymbol{X}_b$ as the displacement vector of the application points instead of $\Delta\boldsymbol{X}$.

Besides expressing static equilibrium, the matrix $\boldsymbol{B}$ plays another important role. The displacements expressed as $\Delta\boldsymbol{X} = \boldsymbol{B}^T\boldsymbol{w}$ (where $\boldsymbol{w}$ is an arbitrarily vector) are mapped to zero by $\boldsymbol{B}^{\perp T}$ ($\Delta\boldsymbol{X}_b = \boldsymbol{B}^{\perp T}\boldsymbol{B}^T\boldsymbol{w} = \boldsymbol{0}$). Thus, regardless of $\boldsymbol{F}$ and $\boldsymbol{F}_b$, the work done by the displacements expressed as $\Delta\boldsymbol{X} = \boldsymbol{B}^T\boldsymbol{w}$ becomes zero. Considering the fact that the work done by the forces in static equilibrium is zero when the application points are displaced rigidly (i.e., displaced without changing mutual distances), we can conclude that the

displacements expressed as $\Delta \boldsymbol{X} = \boldsymbol{B}^T \boldsymbol{w}$ are a type of rigid displacement of the application points. Therefore, the displacements $\Delta \boldsymbol{X}_b$ compatible with $\boldsymbol{F}_b$ express the type of relative displacements between the application points. Equation (4) implies that the linear mapping $\boldsymbol{B}^{\perp T}$ has the effect of eliminating the rigid displacements contained in $\Delta \boldsymbol{X}$.

Moreover, according to the need (see Fig. 2), we constrain the deformation of the parts in the protein model (no part is fixed to the external environment). Let $\boldsymbol{l}_c$ be the vector assembling the parameters that must be constrained (e.g., distances between the specified points in the model), and $\boldsymbol{J}_c$ be the Jacobian matrix that relates $\Delta \boldsymbol{l}_c$ with the changes of the conformation variables $\Delta \boldsymbol{\theta}$ ($\Delta \boldsymbol{l}_c = \boldsymbol{J}_c \Delta \boldsymbol{\theta}$). The conformation variables $\boldsymbol{\theta}$ must be changed without changing $\boldsymbol{l}_c$.The set of $\Delta \boldsymbol{\theta}$ that maintains $\Delta \boldsymbol{l}_c = \boldsymbol{0}$ is expressed as:

$$\Delta \boldsymbol{\theta} = \boldsymbol{J}_c^{\perp} \Delta \boldsymbol{\theta}_c \tag{5}$$

where $\boldsymbol{J}_c^{\perp}$ is the matrix whose columns are the orthonormal basis of the kernel space of $\boldsymbol{J}_c$, and $\Delta \boldsymbol{\theta}_c$ is an arbitrary vector.

Substituting Eq. (3) (the condition of static equilibrium of the forces) and Eq. (5) (the constraint to the protein model) into Eq. (2), we obtain the following equations.

$$\Delta \boldsymbol{\theta}_c = (\boldsymbol{J}_s^{*T} \boldsymbol{K} \boldsymbol{J}_s^*)^{-1} \boldsymbol{J}_f^{*T} \boldsymbol{F} = \boldsymbol{D}^* \boldsymbol{F} = \boldsymbol{D}^* \boldsymbol{B}^{\perp} \boldsymbol{F}_b \tag{6}$$

$$\Delta \boldsymbol{\theta} = \boldsymbol{J}_c^{\perp} \Delta \boldsymbol{\theta}_c = \boldsymbol{J}_c^{\perp} \boldsymbol{D}^* \boldsymbol{F} = \boldsymbol{J}_c^{\perp} \boldsymbol{D}^* \boldsymbol{B}^{\perp} \boldsymbol{F}_b \tag{7}$$

where $\boldsymbol{J}_f^* = \boldsymbol{J}_f \boldsymbol{J}_c^{\perp}$, $\boldsymbol{J}_s^* = \boldsymbol{J}_s \boldsymbol{J}_c^{\perp}$, and $\boldsymbol{D}^* = (\boldsymbol{J}_s^{*T} \boldsymbol{K} \boldsymbol{J}_s^*)^{-1} \boldsymbol{J}_f^{*T}$. It is possible to calculate the changes in the conformation variables $\Delta \boldsymbol{\theta}$ using Eq. (7) when the balanced external forces expressed by $\boldsymbol{F}_b$ are applied to the partially constrained model. In addition, we often encounter the problem in calculating the inverse of $\boldsymbol{J}_s^{*T} \boldsymbol{K} \boldsymbol{J}_s^*$ because of the ill condition of the matrix. The details about the method to deal with the problem are given in [37].

In the general definition of the structural compliance in this study, the deformation of the part specified separately from the application points must be evaluated (see Fig. 2). As shown in Figure 3, we specify the part by specifying n_t deformation points; the volume surrounded by the points forms the part. Let $\Delta \boldsymbol{x}_{t1}, \cdots, \Delta \boldsymbol{x}_{t\, n_t}$ be the displacement vectors of the deformation points, and $\Delta \boldsymbol{X}_t = (\Delta \boldsymbol{x}_{t1}{}^T \cdots \Delta \boldsymbol{x}_{t\, n_t}{}^T)^T$ be the vector assembling them. Defining the Jacobian matrix $\boldsymbol{J}_t$ that relates $\Delta \boldsymbol{X}_t$ and $\Delta \boldsymbol{\theta}$, we can express $\Delta \boldsymbol{X}_t$ as:

$$\Delta \boldsymbol{X}_t = \boldsymbol{J}_t \Delta \boldsymbol{\theta} = \boldsymbol{J}_t \boldsymbol{J}_c^{\perp} \Delta \boldsymbol{\theta}_c = \boldsymbol{J}_t^* \Delta \boldsymbol{\theta}_c \tag{8}$$

where $\boldsymbol{J}_t^* = \boldsymbol{J}_t \boldsymbol{J}_c^{\perp}$ (see Eq. (5)). $\Delta \boldsymbol{X}_t$ contains the rigid displacement of the deformation points. Based on the effect of the matrix $\boldsymbol{B}$ explained in Eq. (4), the matrix $\boldsymbol{B}_t$, which expresses the condition of the static equilibrium when forces are virtually applied at the specified n_t deformation points, can be used for eliminating the rigid displacement contained in $\Delta \boldsymbol{X}_t$. The deformation of the deformation part $\Delta \boldsymbol{X}_{bt}$ is expressed as:

$$\Delta \boldsymbol{X}_{bt} = \boldsymbol{B}_t^{\perp T} \Delta \boldsymbol{X}_t = \boldsymbol{B}_t^{\perp T} \boldsymbol{J}_t^* \Delta \boldsymbol{\theta}_c \tag{9}$$

Combining Eq. (9) with Eq. (6), we obtain the following equation.

$$\Delta \boldsymbol{X}_{bt} = \boldsymbol{B}_t^{\perp T} \boldsymbol{J}_t^* \Delta \boldsymbol{\theta}_c = \boldsymbol{B}_t^{\perp T} \boldsymbol{J}_t^* \boldsymbol{D}^* \boldsymbol{B}^{\perp} \boldsymbol{F}_b = \boldsymbol{C}_{bt} \boldsymbol{F}_b \tag{10}$$

where $\boldsymbol{C}_{bt} = \boldsymbol{B}_t^{\perp T} \boldsymbol{J}_t^* \boldsymbol{D}^* \boldsymbol{B}^{\perp}$ is the structural compliance matrix. Equation (10) is the basic equation for directly calculating the structural compliance when application and deformation parts are specified by points.

SC Mode Decomposition

As illustrated in Figure 4 (the application and deformation points are assumed to be the same in this illustration), the structural compliance will differ depending on the force directions. By applying singular value decomposition (SVD) to the structural compliance matrix $\boldsymbol{C}_{bt}$, we can decompose the motion according to the magnitude of structural compliance. Let $\boldsymbol{C}_{bt} = \boldsymbol{U} \boldsymbol{S} \boldsymbol{V}^T$ be the SVD of $\boldsymbol{C}_{bt}$, where $\boldsymbol{U}$ and $\boldsymbol{V}$ are the orthonormal matrices and $\boldsymbol{S}$ is a diagonal matrix that has singular values σ_i for the diagonal elements in decreasing order ($\sigma_1 \geq \sigma_2 \geq \cdots$). When $\boldsymbol{F}_b = [\boldsymbol{V}]_i$ (the i th column of $\boldsymbol{V}$, $|[\boldsymbol{V}]_i| = 1$) is substituted in Eq. (10), the magnitude of deformation $|\Delta \boldsymbol{X}_{bt}|$ is expressed as $|\Delta \boldsymbol{X}_{bt}| = |\boldsymbol{C}_{bt} [\boldsymbol{V}]_i| = \sigma_i$. In other words, when the value of i decreases, the structural compliance increases. To calculate the changes of the conformation variables $\Delta \boldsymbol{\theta}_{mode\ i}$ under forces applied in the direction $[\boldsymbol{V}]_i$, we substitute $\boldsymbol{F}_b = \alpha [\boldsymbol{V}]_i$ (α is a scalar value) into Eq. (7). $\Delta \boldsymbol{\theta}_{mode\ i}$ is expressed as:

$$\Delta \boldsymbol{\theta}_{mode\ i} = \alpha \boldsymbol{J}_c^{\perp} \boldsymbol{D}^* \boldsymbol{B}^{\perp} [\boldsymbol{V}]_i \tag{11}$$

We call this decomposition of motion according to the magnitude of the structural compliance as the SC mode decomposition, and $\Delta \boldsymbol{\theta}_{mode\ i}$ in Eq. (11) expresses the i th SC mode motion. In the calculation of SC mode motions, we have to specify the application and deformation points (both can be specified to the same points); however, we do not have to specify the force directions. The force directions are calcu-

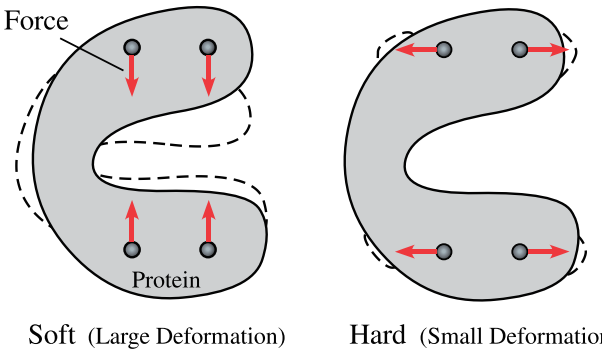

Figure 4 Force direction and structural compliance. The structural compliance will differ depending on the directions of the forces (the application and deformation points are assumed to be the same in this illustration). Although the magnitudes of the forces applied to the protein models are the same for both the cases, the deformation of the left case is larger than that of the right case.

lated by the SVD of the structural compliance matrix $\boldsymbol{C}_{bt}$ ($[\boldsymbol{V}]_i$ in Eq. (11) corresponds to the force directions).

As described earlier, the studies based on NMA clarified that the functional protein motions can be approximated by the combinations of lower normal mode motions. Combining this with the fact that the natural frequencies of soft objects are low in general, we can infer that the lower SC mode motions or the softer motions will occur easily.

SC Mode, Normal Mode, and Principal Component

The normal mode is based on vibration theory, and the lower normal mode motions express the motions associated with the lower natural frequencies. In contrast, the SC mode is based on structural mechanics, and the lower SC mode motions align with the higher structural compliance (the softer motions). As mentioned above, combining this with the fact that the natural frequencies of soft objects are low in general, we can understand that normal mode and SC mode motions are interrelated, but are not the same. When calculating the SC mode, we explicitly specify the application and deformation parts in the protein model. This procedure is not needed in NMA (which calculates the normal modes of the whole structure), so the calculation of the SC mode might appear more troublesome than that of the normal mode. However, by implementing this procedure, we can obtain direct information about the motion properties of the parts, even of small regions of the localized parts, by specifying the parts as application or deformation parts (we can also specify the parts for each residue). Moreover, by separately specifying the application and deformation parts, we can obtain direct information about the motion interaction among the different parts.

We mention the relationships between the SC mode decomposition and the principal component analysis (PCA). Given an ensemble of conformational data of a protein, PCA is effective to extract the collective motions contained in the conformations [19,20,38,39]. The principal components are calculated from the diagonalization of the covariance matrix constructed from the assembly of the displacement vectors of the points in the protein. In NMA, we focus on the change in energy of various conformations around the equilibrium state. In this sense, NMA can be understood from the viewpoint of PCA [20]. The structural compliance matrix $\boldsymbol{C}_{bt}$ does not directly express the covariance matrix; however, we can find the relationships between the SC mode decomposition and PCA as follows. The i th column of the structural compliance matrix $\boldsymbol{C}_{bt}$ expresses the deformation vector of the deformation part when the i th element of the force vector is one and the other elements are zero. Namely, the structural compliance matrix $\boldsymbol{C}_{bt}$ can be regarded as the assembly of the deformation vectors of the deformation part achieved by various forces of equal magnitude acting on the application part. We understand that $\boldsymbol{C}_{bt}\boldsymbol{C}_{bt}^{T}$ expresses the covariance matrix of these deformation vectors. In the SC mode decomposition, we applied SVD to $\boldsymbol{C}_{bt}$ ($= \boldsymbol{USV}^T$). Using the SVD of $\boldsymbol{C}_{bt}$, the covariance matrix can be rewritten as $\boldsymbol{C}_{bt}\boldsymbol{C}_{bt}^{T} = \boldsymbol{USV}^T\boldsymbol{VSU}^T = \boldsymbol{US}^2\boldsymbol{U}^T$. Thus, the covariance matrix is expressed as the products of the orthonormal matrix $\boldsymbol{U}$ and the diagonal matrix $\boldsymbol{S}^2$. According to the theory of PCA, $[\boldsymbol{U}]_i$ (the i th column of $\boldsymbol{U}$) corresponds to the direction of the i th principal component. In the calculation of the i th SC mode motion, $\alpha[\boldsymbol{V}]_i$ is substituted in $\boldsymbol{F}_b$ (force vector acting on the application part) in Eq. (7). If $\alpha[\boldsymbol{V}]_i$ is substituted in $\boldsymbol{F}_b$ in Eq. (10), we find that the deformation vector of the deformation part $\Delta\boldsymbol{X}_{bt}$ is expressed as $\Delta\boldsymbol{X}_{bt} = \alpha\boldsymbol{C}_{bt}[\boldsymbol{V}]_i = \alpha\boldsymbol{USV}^T[\boldsymbol{V}]_i = \alpha\sigma_i[\boldsymbol{U}]_i$ ($\Delta\boldsymbol{X}_{bt}$ directs in the direction of $[\boldsymbol{U}]_i$). Therefore, we can understand that $\Delta\boldsymbol{X}_{bt}$ caused by the i th SC mode motion directs in the direction of the i th principal component of the deformation vectors of the deformation part achieved by various forces of equal magnitude acting on the application part.

Structural Compliance Analysis Focusing on Flexible Groups

To calculate the structural compliance shown above, we assumed to apply forces to the points (i.e., application points) and evaluate the deformations at the points (i.e., deformation points). For the practical application of the analysis, it will be convenient if forces were not only applied to the points but also to the groups of points (e.g., secondary structures, domains, and subunits). Similarly, it will be convenient if deformation were not only evaluated between the points but also between the groups of points. If groups are treated as rigid bodies, the computational cost will be dramatically reduced because of the reduction of conformation variables (all of the conformation variables in the groups are fixed) [15,40]. However, the specification of groups has to be done carefully. If the specified groups include key parts, the structural compliance properties will completely change. For avoiding this problem, in the formulation given below, we do not treat the groups as rigid bodies but as flexible bodies.

As shown in Figure 5, in the structural compliance analysis focusing on the flexible group motions, we specify n_{fg} groups where forces and moments are applied (Application Groups in Fig. 5); we also specify n_{tg} groups whose relative motions are evaluated (Deformation Groups in Fig. 5) in the protein models. It should be noted that both the forces and the moments must be applied. In addition, according to the need, we specify the parts whose deformations are constrained (Constrained Part in Fig. 5). Similar to the cases that focus on the points (see Fig. 3), no part is fixed to the external environment.

We assume that the forces and moments are applied at the centroids of C_α in flexible groups. Let $\boldsymbol{x}_{ci}$ be the position vector of the centroid of the group i; $\boldsymbol{f}_{ci}$ and $\boldsymbol{\tau}_{ci}$ be the force and moment vectors applied at the centroid, respectively; $\Delta\boldsymbol{x}_{ci}$ be the translational displacement vector of the centroid, and $\Delta\boldsymbol{a}_{ci}$ be the angular displacement vector around the centroid. If we define the force vector $\boldsymbol{F}$ as $\boldsymbol{F} = (\boldsymbol{f}_{c1}{}^T\boldsymbol{\tau}_{c1}{}^T\boldsymbol{f}_{c2}{}^T\boldsymbol{\tau}_{c2}{}^T\cdots)^T$

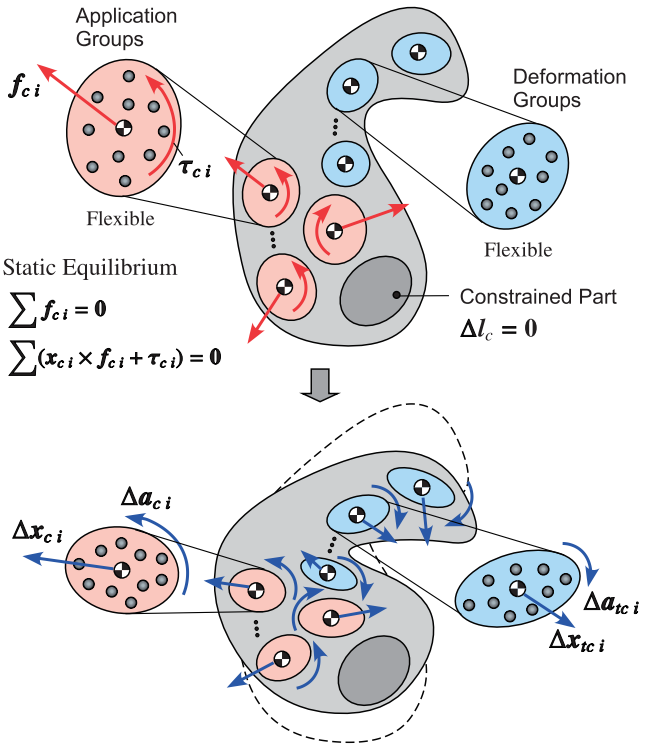

Figure 5 Calculation of structural compliance focusing on flexible groups. External forces $\boldsymbol{f}_{ci}$ and moments $\boldsymbol{\tau}_{ci}$ are assumed to be applied at the centroids of the specified application groups and the groups displaced $\Delta\boldsymbol{x}_{ci}$ in translation and $\Delta\boldsymbol{a}_{ci}$ in orientation, where the forces and moments are in static equilibrium ($\sum\boldsymbol{f}_{ci} = \boldsymbol{0}$ and $\sum(\boldsymbol{x}_{ci} \times \boldsymbol{f}_{ci} + \boldsymbol{\tau}_{ci}) = \boldsymbol{0}$ are established). By the forces and moments application, it is assumed that the specified deformation groups displaced $\Delta\boldsymbol{x}_{tci}$ in translation and $\Delta\boldsymbol{a}_{tc\,i}$ in orientation. It should be noted that, both the application and deformation groups are treated not as rigid but as flexible bodies. In the deformation, the vector $\boldsymbol{l}_c$ assembling the parameters that must be constrained in the constrained parts is kept unchanged (i.e., $\Delta\boldsymbol{l}_c = \boldsymbol{0}$).

and define the displacement vector $\Delta\boldsymbol{X}$ as $\Delta\boldsymbol{X} = (\Delta\boldsymbol{x}_{c1}{}^T\Delta\boldsymbol{a}_{c1}{}^T\Delta\boldsymbol{x}_{c2}{}^T\Delta\boldsymbol{a}_{c2}{}^T\cdots)^T$ so that $\boldsymbol{F}^T\,\Delta\boldsymbol{X}$ has the work dimension, then Eq. (2) is still valid for the case focusing on flexible group motions. However, the calculation of the Jacobian matrix $\boldsymbol{J}_f$ is not so straightforward because of the flexibility of the groups themselves.

We must define the Jacobian matrix $\boldsymbol{J}_f$ so that it expresses the relationship between the displacements of the flexible groups $\Delta\boldsymbol{X}$ and the changes of the conformation variables $\Delta\boldsymbol{\theta}$ ($\Delta\boldsymbol{X} = \boldsymbol{J}_f\Delta\boldsymbol{\theta}$). $\boldsymbol{J}_f$ can be divided into blocks $\boldsymbol{J}_{xc\,i}$ and $\boldsymbol{J}_{ac\,i}$ as shown below.

$$\begin{pmatrix} \Delta\boldsymbol{X}_{(6n_{fg}\times 1)} \\ \Delta\boldsymbol{x}_{c1} \\ \Delta\boldsymbol{a}_{c1} \\ \Delta\boldsymbol{x}_{c2} \\ \Delta\boldsymbol{a}_{c2} \\ \vdots \end{pmatrix} = \begin{pmatrix} \boldsymbol{J}_{f(6n_{fg}\times n)} \\ \boldsymbol{J}_{xc1} \\ \boldsymbol{J}_{ac1} \\ \boldsymbol{J}_{xc2} \\ \boldsymbol{J}_{ac2} \\ \vdots \end{pmatrix} \Delta\boldsymbol{\theta}_{(n\times 1)} \tag{12}$$

Here, $\boldsymbol{J}_{xc\,i}$ is the Jacobian matrix that relates $\Delta\boldsymbol{x}_{c\,i}$ and $\Delta\boldsymbol{\theta}$ ($\Delta\boldsymbol{x}_{c\,i} = \boldsymbol{J}_{xc\,i}\Delta\boldsymbol{\theta}$), and $\boldsymbol{J}_{ac\,i}$ is the Jacobian matrix that relates $\Delta\boldsymbol{a}_{c\,i}$ and $\Delta\boldsymbol{\theta}$ ($\Delta\boldsymbol{a}_{c\,i} = \boldsymbol{J}_{ac\,i}\Delta\boldsymbol{\theta}$). Figure 6 illustrates the motion of the flexible group i. Let $n_{ca\,i}$ be the number of C_α in the group i, $\Delta\boldsymbol{p}_{ij}$ be the displacement vector of the j th C_α in the group i, $\Delta\boldsymbol{P}_i$ be the vector assembling $\Delta\boldsymbol{p}_{ij}$ or $\Delta\boldsymbol{P}_i = (\Delta\boldsymbol{p}_{i1}{}^T\cdots\Delta\boldsymbol{p}_{in_{cai}}{}^T)^T$, and $\boldsymbol{J}_{p\,i}$ be the Jacobian matrix that relates $\Delta\boldsymbol{P}_i$ and $\Delta\boldsymbol{\theta}$ ($\Delta\boldsymbol{P}_i = \boldsymbol{J}_{p\,i}\Delta\boldsymbol{\theta}$).

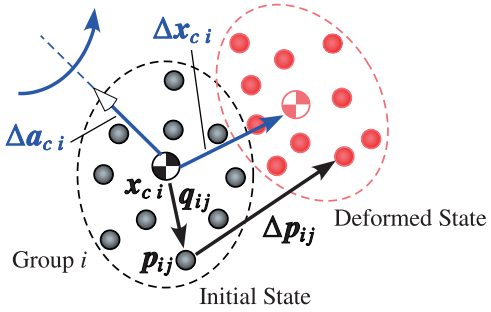

Figure 6 Translational and angular displacement of a flexible group. Here, group i moved with the deformation. It is necessary to calculate the translational displacement vector $\Delta\boldsymbol{x}_{c\,i}$ and the angular displacement vector $\Delta\boldsymbol{a}_{c\,i}$, which approximate the motion of the group i. We can calculate $\Delta\boldsymbol{x}_{c\,i}$ as the displacement vector of the centroid of the group i; however, the calculation of $\Delta\boldsymbol{a}_{c\,i}$ is not so straightforward. $\Delta\boldsymbol{p}_{ij}$ shows the displacement vector of the j th C_α in the group i. If the group i is perfectly rigid, $\Delta\boldsymbol{p}_{ij}$ is expressed as $\Delta\boldsymbol{p}_{ij} = \Delta\boldsymbol{x}_{c\,i} + \Delta\boldsymbol{a}_{c\,i} \times \boldsymbol{q}_{ij}$ ($\boldsymbol{q}_{ij}$ is the vector directed from the centroid to the j th C_α). The group is not perfectly rigid; therefore, we cannot determine $\Delta\boldsymbol{a}_{c\,i}$ such that this relationship holds strictly for every C_α. $\Delta\boldsymbol{a}_{c\,i}$ is determined as the least square error solution.

First, we consider the calculation of $\boldsymbol{J}_{xc\,i}$ in Eq. (12). The translational displacement vector of the centroid $\Delta\boldsymbol{x}_{c\,i}$ can be expressed as:

$$\Delta\boldsymbol{x}_{c\,i} = \frac{1}{n_{ca\,i}}\sum_j \Delta\boldsymbol{p}_{ij} = \frac{1}{n_{ca\,i}}\boldsymbol{E}_i'\Delta\boldsymbol{P}_i = \frac{1}{n_{ca\,i}}\boldsymbol{E}_i'\boldsymbol{J}_{p\,i}\Delta\boldsymbol{\theta} \tag{13}$$

where $\boldsymbol{E}_i' = \{\boldsymbol{E}\cdots\boldsymbol{E}\}$ is $3\times 3n_{ca\,i}$ matrix containing 3×3 identity matrix $\boldsymbol{E}$ as blocks. Comparing Eqs. (12) and (13), we obtain:

$$\boldsymbol{J}_{xc\,i} = \frac{1}{n_{ca\,i}}\boldsymbol{E}_i'\boldsymbol{J}_{p\,i} \tag{14}$$

Then, we consider the calculation of $\boldsymbol{J}_{ac\,i}$ in Eq. (12). If the group i is perfectly rigid, $\Delta\boldsymbol{p}_{ij}$ is expressed as:

$$\Delta\boldsymbol{p}_{ij} = \Delta\boldsymbol{x}_{c\,i} + \Delta\boldsymbol{a}_{c\,i} \times \boldsymbol{q}_{ij} \tag{15}$$

Here, $\boldsymbol{q}_{ij} = \boldsymbol{p}_{ij} - \boldsymbol{x}_{c\,i}$. $\Delta\boldsymbol{p}_{ij}$ and $\Delta\boldsymbol{x}_{c\,i}$ are already expressed by $\Delta\boldsymbol{\theta}$ using the Jacobian matrices $\boldsymbol{J}_{p\,i}$ and $\boldsymbol{J}_{xc\,i}$, respectively. The group is not perfectly rigid; therefore, it is impossible to determine $\Delta\boldsymbol{a}_{c\,i}$ such that Eq. (15) holds strictly for every C_α. Thus, we consider the least square error solution. Equation (15) can be rewritten as $\boldsymbol{q}_{ij} \times \Delta\boldsymbol{a}_{c\,i} = \Delta\boldsymbol{x}_{c\,i} - \Delta\boldsymbol{p}_{ij}$. By stacking this equation for all C_α in the group i, we obtain the following equation.

$$\begin{pmatrix} \tilde{\boldsymbol{q}}_{i1} \\ \vdots \\ \tilde{\boldsymbol{q}}_{i n_{cai}} \end{pmatrix} \Delta \boldsymbol{a}_{ci} = \begin{pmatrix} \Delta \boldsymbol{x}_{ci} - \Delta \boldsymbol{p}_{i1} \\ \vdots \\ \Delta \boldsymbol{x}_{ci} - \Delta \boldsymbol{p}_{i n_{cai}} \end{pmatrix} \tag{16}$$

where $\tilde{\boldsymbol{q}}_{ij}$ is a 3×3 skew symmetric matrix expressing the operation $\boldsymbol{q}_{ij}\times$. By defining $\tilde{\boldsymbol{Q}}_i = (\tilde{\boldsymbol{q}}_{i1}^T \cdots \tilde{\boldsymbol{q}}_{i n_{cai}}^T)^T$ and $\boldsymbol{J}_{xci}^r = (\boldsymbol{J}_{xci}^T \cdots \boldsymbol{J}_{xci}^T)^T$, we obtain the following equation.

$$\tilde{\boldsymbol{Q}}_i \Delta \boldsymbol{a}_{ci} = (\boldsymbol{J}_{xci}^r - \boldsymbol{J}_{pi}) \Delta \boldsymbol{\theta} \tag{17}$$

Here, Eq. (17) is a linear equation containing an unknown variable $\Delta \boldsymbol{a}_{ci}$. The least square error solution is expressed as:

$$\Delta \boldsymbol{a}_{ci} = \tilde{\boldsymbol{Q}}_i^{\#} (\boldsymbol{J}_{xci}^r - \boldsymbol{J}_{pi}) \Delta \boldsymbol{\theta} \tag{18}$$

where $\tilde{\boldsymbol{Q}}_i^{\#}$ is the pseudo inverse of $\tilde{\boldsymbol{Q}}_i$. Comparing Eqs. (12) and (18) we obtain:

$$\boldsymbol{J}_{aci} = \tilde{\boldsymbol{Q}}_i^{\#} (\boldsymbol{J}_{xci}^r - \boldsymbol{J}_{pi}) \tag{19}$$

Substituting Eqs. (14) and (19) into Eq. (12), we obtain $\boldsymbol{J}_f$ for the flexible group motions. It becomes possible to calculate the deformation when forces and moments are applied to the flexible groups by using Eq. (2).

When forces are applied to the points, the forces are assumed to be in static equilibrium. In a similar manner, forces and moments are assumed to be in static equilibrium. This condition is expressed by $\sum \boldsymbol{f}_{ci} = \boldsymbol{0}$ and $\sum (\boldsymbol{x}_{ci} \times \boldsymbol{f}_{ci} + \boldsymbol{\tau}_{ci}) = \boldsymbol{0}$. These equations are linear about $\boldsymbol{F}$; therefore, the condition can be expressed in the form $\boldsymbol{BF} = \boldsymbol{0}$. All the forces and moments in static equilibrium are expressed by Eq. (3). We regard $\boldsymbol{F}_b$ as force vector instead of $\boldsymbol{F}$. The displacement $\Delta \boldsymbol{X}_b$, which is compatible with $\boldsymbol{F}_b$, is expressed by Eq. (4). $\Delta \boldsymbol{X}_b$ expresses the relative displacement between the flexible groups, and we can eliminate the rigid motions contained in $\Delta \boldsymbol{X}$ by Eq. (4). Moreover, the effect of the constraint on the protein model is expressed by Eq. (5). As a result, the deformation of the protein model is expressed in the form of Eq. (7).

The structural compliance of the flexible group motions can be expressed in the form shown in Eq. (10). In this case, the Jacobian matrix $\boldsymbol{J}_t$ must be defined so that it expresses the relationship between the displacements of the deformation groups $\Delta \boldsymbol{X}_t = (\Delta \boldsymbol{x}_{tc1}{}^T \Delta \boldsymbol{a}_{tc1}{}^T \Delta \boldsymbol{x}_{tc2}{}^T \Delta \boldsymbol{a}_{tc2}{}^T \cdots)^T$ and the changes of conformation variables $\Delta \boldsymbol{\theta}$, where $\Delta \boldsymbol{x}_{tci}$ is the translational displacement vector, and $\Delta \boldsymbol{a}_{tci}$ is the angular displacement vector of the deformation group i (see Fig. 5). The Jacobian matrix $\boldsymbol{J}_t$ can be calculated in a manner similar to the calculation used for $\boldsymbol{J}_f$ for the flexible group motion. Moreover, in this case, $\boldsymbol{B}_t$ expresses the condition of the static equilibrium when forces and moments are virtually applied at specified n_{tg} deformation groups ($\boldsymbol{B}_t$ is used to extract the relative motion from $\Delta \boldsymbol{X}_t$).

SC Mode Decomposition Focusing on Flexible Groups

When decomposing motion according to the magnitude of structural compliance (i.e., SC mode decomposition), we should be careful about the mixed dimensions of the elements in the vectors. For evaluating the magnitude of structural compliance, we need to define the magnitudes of $\boldsymbol{F}_b$ and $\Delta \boldsymbol{X}_{bt}$. In flexible group motions, it is not appropriate to directly take the magnitude of $\boldsymbol{F}_b$ because $\boldsymbol{F} = \boldsymbol{B}^{\perp} \boldsymbol{F}_b$ contains elements with different dimensions (force and moment). For the same reason, directly taking the magnitude of $\Delta \boldsymbol{X}_{bt} = \boldsymbol{B}_t^{\perp T} \Delta \boldsymbol{X}_t$ is also not appropriate.

To deal with this problem of mixed dimensions, we define a new vector $\boldsymbol{F}'$ containing the same dimensional elements, and we assume that $\boldsymbol{F}'$ is related to $\boldsymbol{F}$ through a weighting matrix $\boldsymbol{W}$.

$$\boldsymbol{F} = \boldsymbol{W} \boldsymbol{F}' \tag{20}$$

Then the static equilibrium condition $\boldsymbol{BF} = \boldsymbol{0}$ can be rewritten as $\boldsymbol{B}'\boldsymbol{F}' = \boldsymbol{0}$, where $\boldsymbol{B}' = \boldsymbol{BW}$. All the $\boldsymbol{F}'$ and $\boldsymbol{F}$ satisfying the condition are expressed as:

$$\boldsymbol{F}' = \boldsymbol{B}'^{\perp} \boldsymbol{F}'_b \tag{21}$$

$$\boldsymbol{F} = \boldsymbol{W} \boldsymbol{B}'^{\perp} \boldsymbol{F}'_b \tag{22}$$

where $\boldsymbol{F}'_b$ is the arbitrary vector. We regard $\boldsymbol{F}'_b$ as the force vector instead of $\boldsymbol{F}_b$. Based on the considerations similar to Eq. (3), the displacement $\Delta \boldsymbol{X}'_b$ compatible to the force $\boldsymbol{F}'_b$ (see Eq. (4)) is expressed as:

$$\Delta \boldsymbol{X}'_b = (\boldsymbol{W} \boldsymbol{B}'^{\perp})^T \Delta \boldsymbol{X} = \boldsymbol{B}'^{\perp T} \boldsymbol{W}^T \Delta \boldsymbol{X} \tag{23}$$

$\Delta \boldsymbol{X}'_b$ expresses the relative displacement between the application groups. Moreover, in a similar manner, the relative displacement between the deformation groups $\Delta \boldsymbol{X}'_{bt}$ is expressed as:

$$\Delta \boldsymbol{X}'_{bt} = \boldsymbol{B}_t'^{\perp T} \boldsymbol{W}_t^T \Delta \boldsymbol{X}_t \tag{24}$$

where $\boldsymbol{B}'_t = \boldsymbol{B}_t \boldsymbol{W}_t$ and $\boldsymbol{W}_t$ is the weighting matrix for the deformation groups. Combining Eqs. (6), (8), (22) and (24), we can express the relationship between $\Delta \boldsymbol{X}'_{bt}$ and $\boldsymbol{F}'_b$ as follows.

$$\Delta \boldsymbol{X}'_{bt} = \boldsymbol{B}_t'^{\perp T} \boldsymbol{W}_t^T \boldsymbol{J}_t^* \boldsymbol{D}^* \boldsymbol{W} \boldsymbol{B}'^{\perp} \boldsymbol{F}'_b = \boldsymbol{C}'_{bt} \boldsymbol{F}'_b \tag{25}$$

where $\boldsymbol{C}'_{bt} = \boldsymbol{B}_t'^{\perp T} \boldsymbol{W}_t^T \boldsymbol{J}_t^* \boldsymbol{D}^* \boldsymbol{W} \boldsymbol{B}'^{\perp}$. In the SC mode decomposition, it becomes possible to avoid the problem of mixed dimension by using $\boldsymbol{C}'_{bt}$ instead of $\boldsymbol{C}_{bt}$ as a compliance matrix. From the SVD of the compliance matrix $\boldsymbol{C}'_{bt} = \boldsymbol{USV}^T$, the direction of the force vector corresponding to the i th SC mode motion is obtained as $[\boldsymbol{V}]_i$. Combining $\boldsymbol{F}'_b = \alpha [\boldsymbol{V}]_i$ (α is a scalar value), Eqs. (7) and (22), we can express the i th SC mode motion focusing on the flexible group motions as:

$$\Delta \boldsymbol{\theta}_{mode\,i} = \alpha \boldsymbol{J}_c^{\perp} \boldsymbol{D}^* \boldsymbol{W} \boldsymbol{B}'^{\perp} [\boldsymbol{V}]_i \tag{26}$$

In calculating the SC mode motions, it should be noted that we have to specify the application and deformation groups (both can be specified to the same); however, we do not have to specify $\boldsymbol{F}'_b$ or its direction.

For example, as shown in Figure 7, we may define $\boldsymbol{F}'$ and $\boldsymbol{W}$ by using a couple. In Figure 7, $\boldsymbol{\tau}$ is a moment vector, and $\boldsymbol{u}_\tau$ is the unit vector directed to $\boldsymbol{\tau}$. The moment can be replaced by a couple consisting of two forces of magnitude

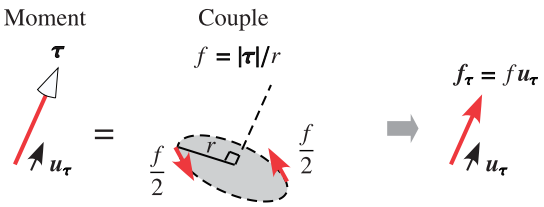

Figure 7 Weighting of the moment vector by using a couple. For example, the weighting of the moment vector can be defined through a couple. $\boldsymbol{\tau}$ is a moment vector, and $\boldsymbol{u}_\tau$ is the unit vector directed to $\boldsymbol{\tau}$. The moment can be replaced by a couple consisting of two forces of magnitude $f/2$, $f = |\boldsymbol{\tau}|/r$, where r is the perpendicular distance between the force vector and the moment vector. We define the force dimensional vector $\boldsymbol{f}_\tau = f\boldsymbol{u}_\tau$ whose magnitude and direction are expressed as f and $\boldsymbol{u}_\tau$, respectively. By using $\boldsymbol{f}_\tau$ instead of $\boldsymbol{\tau}$, it becomes possible to avoid the problem of mixed dimension.

$f/2$, $f = |\boldsymbol{\tau}|/r$, where r is the perpendicular distance between the force vector and the moment vector. We define the force dimensional vector $\boldsymbol{f}_\tau = f\boldsymbol{u}_\tau$ whose magnitude and direction are expressed as f and $\boldsymbol{u}_\tau$, respectively. Based on this definition, we define $\boldsymbol{f}_{c\tau i}$ corresponding to the moment vectors $\boldsymbol{\tau}_{ci}$ acting on the application group i (see Fig. 5). Here, we may use the average radius $r_{avi} = (\sum_j |\boldsymbol{q}_{ij}|)/n_{cai}$ (see Fig. 6) for the perpendicular distance r. Then, $\boldsymbol{F}'$ can be defined as $\boldsymbol{F}' = (\boldsymbol{f}_{c1}{}^T \boldsymbol{f}_{c\tau 1}{}^T \boldsymbol{f}_{c2}{}^T \boldsymbol{f}_{c\tau 2}{}^T \cdots)^T$. The corresponding weighting matrix $\boldsymbol{W}$ can be defined by the next diagonal matrix.

$$\boldsymbol{W} = diag(\boldsymbol{E}, r_{av1}\boldsymbol{E}, \boldsymbol{E}, r_{av2}\boldsymbol{E}, \cdots) \tag{27}$$

Here, $\boldsymbol{E}$ is the 3 × 3 identity matrix. Similarly, the weighting matrix for the deformation group $\boldsymbol{W}_t$ can also be defined.

Screw Approximation of Relative Motion Between Flexible Groups

When the change of conformation variables are obtained, it is easy to graphically express the motion. For example, we can easily draw the main chain structure of each SC mode motion by using $\Delta\boldsymbol{\theta}_{mode\,i}$ in Eq. (11) or (26). The value α in these equations should be modulated so that the motions are not large because the analyses are formulated based on instantaneous kinematics. For more quantitatively comprehending the characteristic of motions, it is effective to identify the screw approximating the relative motion between two flexible groups specified in the protein model [16,41–44]. An arbitrary relative motion between two rigid bodies can be expressed as a rotation around a unique axis and a translation along the same axis [45]. The axis is called the screw axis and the ratio between the translation and the rotation is called the pitch. After identifying the screw (i.e., the axis and the pitch) that approximates the relative motion between two specified flexible groups (see Fig. 8), we can regard the groups as virtually connected by the screw joint (The flexibility of the groups makes it impossible to identify the screw that strictly expresses the motions of all the points in the groups). For example, if the pitch of the identified screw is small enough and the axis passes through some residues, we

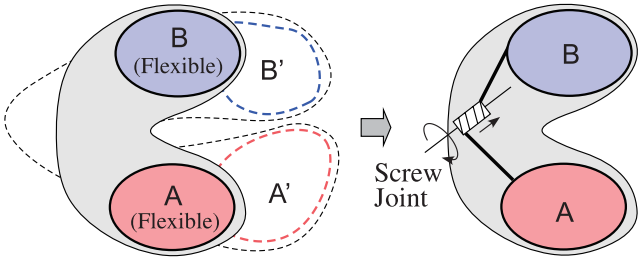

Figure 8 Screw approximation of relative motion between flexible groups. An arbitrary relative motion between two rigid bodies can be expressed as a rotation around a unique axis and a translation along the same axis. The axis is called the screw axis and the ratio between the translation and the rotation is called the pitch. If the screw (i.e., the axis and the pitch) approximating the relative motion between two flexible groups A and B is identified, it becomes possible to regard the groups as virtually connected by the screw joint (The flexibility of the groups makes it impossible to identify the screw that strictly expresses the motions of all the points in the groups).

can understand that the motion is the hinge-bending motion and the residues are hinge residues. For the proteins whose 3D structural data for different conformations are stored in PDB, it is possible to estimate the real protein motion, which we call the PDB motion, and to identify the screw between the specified groups.

For finite motions, such as PDB motions, the screws between the specified groups can be derived by applying the alignment algorithm for the two different conformations [43,44]. However, the formulations for analyzing structural compliance properties described above are based on instantaneous kinematics; therefore, the method for finite motion is not appropriate. We present a method for identifying the screws approximating the instantaneous motions expressed by $\Delta\boldsymbol{\theta}$. Here, $\Delta\boldsymbol{\theta}$ is treated as an infinitesimal change of conformation variables. As shown in Figure 9, we assume that two flexible groups A and B are specified in the protein model (the groups can be specified independently from the application and the deformation groups in Fig. 5). Let $\boldsymbol{x}_{ci}$ (i = A or B) be the positions of the centroids of the specified flexible groups, and let $\Delta\boldsymbol{x}_{ci}$ and $\Delta\boldsymbol{a}_{ci}$ be the instantaneous translational and angular displacements approximating the motion of the groups, respectively. Here, we can express $\Delta\boldsymbol{x}_{ci}$ and $\Delta\boldsymbol{a}_{ci}$ as functions of $\Delta\boldsymbol{\theta}$ in a manner similar to the formulation in the structural compliance analysis focusing on flexible groups (see Eqs. (13) and (18)). The approximated instantaneous translational and angular displacement of the flexible group B with respect to A ($\Delta\boldsymbol{x}_{cAB}$ and $\Delta\boldsymbol{a}_{cAB}$, respectively) are expressed as:

$$\Delta\boldsymbol{x}_{cAB} = (\Delta\boldsymbol{x}_{cB} - \Delta\boldsymbol{x}_{cA}) - \Delta\boldsymbol{a}_{cA} \times (\boldsymbol{x}_{cB} - \boldsymbol{x}_{cA}) \tag{28}$$

$$\Delta\boldsymbol{a}_{cAB} = \Delta\boldsymbol{a}_{cB} - \Delta\boldsymbol{a}_{cA} \tag{29}$$

According to the screw theory, the screw parameters that approximate the instantaneous relative motions between the flexible groups A and B, a point on the axis $\boldsymbol{x}_{0\,scw}$, the

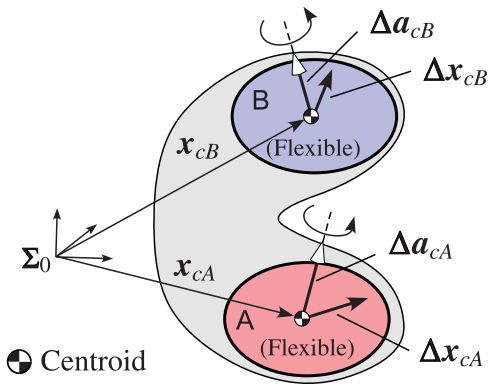

Figure 9 Calculation of the screw approximating instantaneous relative motion between flexible groups. The instantaneous translational and angular displacements approximating the motion of flexible groups ($\Delta \boldsymbol{x}_{ci}$ and $\Delta \boldsymbol{a}_{ci}$, respectively) can be expressed as a function of infinitesimal changes of conformation variables $\Delta\boldsymbol{\theta}$ in a manner similar to the formulation in the structural compliance analysis focusing on flexible groups. From $\Delta \boldsymbol{x}_{ci}$ and $\Delta \boldsymbol{a}_{ci}$, the screw approximating the instantaneous relative motion between the flexible groups can be calculated based on the screw theory.

direction of the axis $\boldsymbol{u}_{scw}$, and the pitch p_{scw} (the translation along the axis occurred in one radian rotation around the axis) are expressed as follows [45].

$$\boldsymbol{x}_{0\,scw} = \boldsymbol{x}_{cB} + \frac{\Delta \boldsymbol{a}_{cAB} \times \Delta \boldsymbol{x}_{cAB}}{|\Delta \boldsymbol{a}_{cAB}|^2} \tag{30}$$

$$\boldsymbol{u}_{scw} = \Delta \boldsymbol{a}_{cAB} \tag{31}$$

$$p_{scw} = \frac{\Delta \boldsymbol{a}_{cAB} \cdot \Delta \boldsymbol{x}_{cAB}}{|\Delta \boldsymbol{a}_{cAB}|^2} \tag{32}$$

When we consider the instantaneous screws of the SC mode motions, these screw parameters do not depend on α in Eqs. (11) and (26).

SC Mode Expansion

We formulated methods for calculating the structural compliance matrix that relates the deformation of the deformation part and the force acting on the application part, and for calculating the SC mode motions from the matrix (SC mode decomposition) for both the cases in which the deformation and the application parts are expressed by using the points and the flexible groups. The essential part in the formulation is the derivation of the deformation of the deformation part from the force applied to the application part. It is assumed that the deformation and application parts are specified (these can be specified to the same part) in the formulation; however, the values need not be specified for the deformation and force. In this section, in contrast, we consider the situation that the value of deformation of the deformation part, or the reference deformation, is given. We formulate a method for approximating the given reference deformation by the selected mode motions in the SC mode motions and for evaluating the contribution of each mode in the approximation. We term this analysis as SC mode expansion. One of the most practical application of SC mode expansion is the comparison of the SC mode motions and PDB motions. This comparison can be made by giving the reference deformation from the PDB data of different conformations. Here, in the following formulation, we use the superscript "(′)" for the symbols (e.g., $\boldsymbol{X}_{bt}^{(\prime)}$) for convenience. This means that the equations are valid for both the cases that the parts are specified by using points and flexible groups (e.g., $\Delta \boldsymbol{X}_{bt}$ and $\Delta \boldsymbol{X}'_{bt}$).

As explained for Eqs. (11) and (26), the column vectors of $\boldsymbol{V}$ (in the SVD of the structural compliance matrix $\boldsymbol{C}_{bt}^{(\prime)} = \boldsymbol{USV}^T$) correspond to the force vectors of SC mode motions. Let $\boldsymbol{V}_M$ be the matrix whose columns are selected from the columns of $\boldsymbol{V}$. The force vectors spanned by the selected column vectors can be expressed as $\boldsymbol{F}_b^{(\prime)} = \boldsymbol{V}_M \boldsymbol{f}_M$, where the elements of the vector $\boldsymbol{f}_M$ express the coefficients for the column vectors. Substituting this into Eq. (10) or (25), we obtain:

$$\Delta \boldsymbol{X}_{bt}^{(\prime)} = \boldsymbol{C}_{bt}^{(\prime)} \boldsymbol{V}_M \boldsymbol{f}_M \tag{33}$$

When $\Delta \boldsymbol{X}_{bt}^{(\prime)}$ is given as the reference deformation, $\boldsymbol{f}_M$ that approximate the given $\Delta \boldsymbol{X}_{bt}^{(\prime)}$ can be calculated by using the following equation.

$$\boldsymbol{f}_M = (\boldsymbol{C}_{bt}^{(\prime)} \boldsymbol{V}_M)^{\#} \Delta \boldsymbol{X}_{bt}^{(\prime)} \tag{34}$$

The force vector corresponding to the j th selected SC mode can be expressed as $\boldsymbol{F}_b^{(\prime)} = f_{Mj}[\boldsymbol{V}_M]_j$ (i.e., the product of j th element of $\boldsymbol{f}_M$ and j th column of $\boldsymbol{V}_M$). The deformation caused by this force is expressed as:

$$\Delta \boldsymbol{X}_{btj}^{(\prime)} = f_{Mj} \boldsymbol{C}_{bt}^{(\prime)} [\boldsymbol{V}_M]_j \tag{35}$$

We use $|\Delta \boldsymbol{X}_{btj}^{(\prime)}|$ as the measure of intensity for evaluating the contribution of the selected j th SC mode needed to approximate the given reference deformation, and we call $|\Delta \boldsymbol{X}_{btj}^{(\prime)}|$ the mode intensity of the j th mode.

As mentioned above, in one of the most practical applications of SC mode expansion, the reference deformations are given from the PDB motions or from the difference between the two PDB data of different conformations. Where, PDB motions are finite; however, SC mode motions expressed by Eqs. (11) and (26) are instantaneous. We use the directions of the PDB motions as the reference deformation. When specifying the relative displacement between the points, $\Delta \boldsymbol{X}_t$ (the displacements of the specified points) can be obtained from the difference of the positions between the two PDB data of different conformations. The deformation $\Delta \boldsymbol{X}_{bt}$ can be expressed as $\Delta \boldsymbol{X}_{bt} = \boldsymbol{B}_t^{\perp T} \Delta \boldsymbol{X}_t$ (see Eq. (9)). When specifying the relative displacement between the flexible groups, $\Delta \boldsymbol{X}_t$ can be obtained by using the alignment algorithm such as the Kabsch method [46,47]. The relative displacement with the weight for translation and rotation $\Delta \boldsymbol{X}'_{bt}$ is expressed as $\Delta \boldsymbol{X}'_{bt} = \boldsymbol{B}'^{\perp T}_t \boldsymbol{W}_t^T \Delta \boldsymbol{X}_t$ (see Eq. (24)). We use the direction of $\Delta \boldsymbol{X}_{bt}^{(\prime)}$ (or $\Delta \boldsymbol{X}_{bt}^{(\prime)} / |\Delta \boldsymbol{X}_{bt}^{(\prime)}|$) calculated from two PDB data of different conformations as the reference deformation.

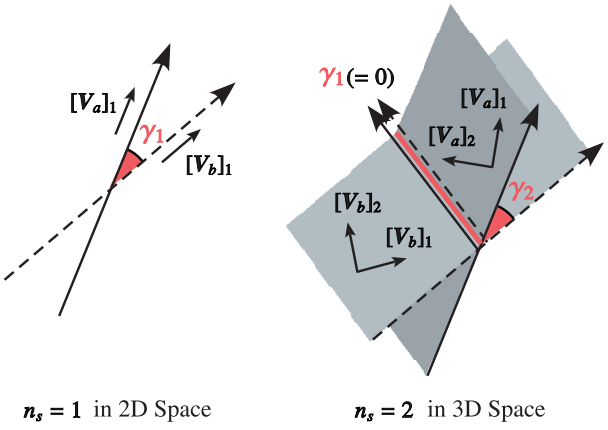

Figure 10 Principal angles between subspaces. For quantitatively evaluating the difference between the subspaces spanned by $[\boldsymbol{V}_a]_{1,2,\cdots,n_s}$ and $[\boldsymbol{V}_b]_{1,2,\cdots,n_s}$ (the force vectors corresponding to the first n_s SC mode motions of two different states), we focus on the principal angles between them. The figures illustrate the examples of principal angles. For the case of $\dim([\boldsymbol{V}]_i) = 2$, $n_s = 1$, one principal angle γ_1 is defined (left figure). For the case of $\dim([\boldsymbol{V}]_i) = 3$, $n_s = 2$, two principal angles γ_1 (= 0) and γ_2 are defined (right figure).

Comparison of Structural Compliance Properties

If the structural compliance property largely changes when a part in the protein model is constrained to be made rigid, we can infer that the constrained part plays an important role. Moreover, it is interesting to examine how the structural compliance properties change by ligand binding. For these applications, we show a method for quantitatively comparing the structural compliance properties.

The vector $[\boldsymbol{V}]_i$ in Eqs. (11) and (26) expresses the direction of the force vector corresponding to the i th SC mode motion. For comparing the structural compliance properties, we focus on the subspace spanned by $[\boldsymbol{V}]_{1,2,\cdots,n_s}$, where n_s is a number smaller than the dimension of $[\boldsymbol{V}]_i$. Let $[\boldsymbol{V}_a]_i$ and $[\boldsymbol{V}_b]_i$ be the force direction vectors corresponding to the i th SC mode motions of a protein model in the different states; for example, one state does not include the constrained part whereas the other state does. Between the subspaces spanned by $[\boldsymbol{V}_a]_{1,2,\cdots,n_s}$ and $[\boldsymbol{V}_b]_{1,2,\cdots,n_s}$, n_s principal angles $\gamma_{1,2,\cdots,n_s}$ can be defined. Figure 10 illustrates the examples of the principal angles when $\dim([\boldsymbol{V}]_i) = 2$, $n_s = 1$ and $\dim([\boldsymbol{V}]_i) = 3$, $n_s = 2$. To evaluate the difference between the subspaces, we define the index Γ expressed by the following equation.

$$\Gamma = \frac{1}{n_s}\sum \gamma_i \tag{36}$$

Here, Γ is the average of the principal angles, which takes the values from 0 to $\pi/2$. The larger the value is, the larger the difference of the structural compliance property is.

Results and Discussion

Conditions of Analyses

We provide two application examples by using the PDB data of lactoferrin and aspartate transcarbamoylase (ATCase). All the parameters for creating protein models and calculating the compliance matrix are the same as those in the application examples reported in our previous study [37]. The protein models are created by setting the threshold distance or the cutoff distance L_{th} for spanning springs as 8 Å and setting the spring constants as the inverse of the distance between the C_α in the PDB data. Moreover, as the parameters specified in the analyses focus on the flexible group motions, the weighting matrices $\boldsymbol{W}$ and $\boldsymbol{W}_t$ in Eq. (25) are defined by using Eq. (27). To calculate the screw approximating the relative motion between the specified flexible groups, we use the positions of C_α. For the alignment algorithm required to express the PDB motions, the Kabsch method is applied for the positions of C_α in the specified flexible groups.

The proposed methods were implemented in an original computer program coded in the C++ language. Singular valued decompositions and matrix multiplications of large matrices were performed in Intel Math Kernel Library (Intel Corporation). The results were graphically expressed in the viewer software Molfeat (FiatLux Corporation).

Lactoferrin

Figure 11 illustrates the 3D structure of lactoferrin. It consists of 691 amino residues. As shown in the conformation PDB-ID:1LFH, it has a large cavity between the N1 domain (residues 1–91, 252–333) and the N2 domain (residues 92–251). Comparing conformations PDB-ID:1LFH and 1LFG, we find that lactoferrin opens and closes the cavity. We reported that first SC mode motion of the protein model created using PDB-ID:1LFH (the distances between the C_α in the residues forming disulfide bonds were constrained) agreed with the PDB motion [37]. In the analysis, both the application and deformation points were specified for C_α in the seven residues related to the ligand binding.

We analyzed the protein model of lactoferrin in more detail focusing on the structural compliance properties between the N1 and N2 domains. The two application groups were specified for all the C_α in the N1 and N2 domains, and the two deformation groups were specified to be the same as the application groups. Calculating the structural compliance matrix $\boldsymbol{C}'_{bt}$ (Eq. (25)) and applying SVD to $\boldsymbol{C}'_{bt}$, we obtained six SC mode motions (Eq. (26)). Figure 12 illustrates the first three SC mode motions. In this figure, the screws approximating the SC mode motions (see Fig. 9) and those approximating the PDB motions (between the conformations of PDB-ID:1LFH and 1LFG) are shown; here the pitch value is expressed as the translation along the screw axis in one rotation around the axis (Å/rev). The cavity closing motion is observed in the first SC mode motion.

Using SC mode expansion, the intensities of six SC mode

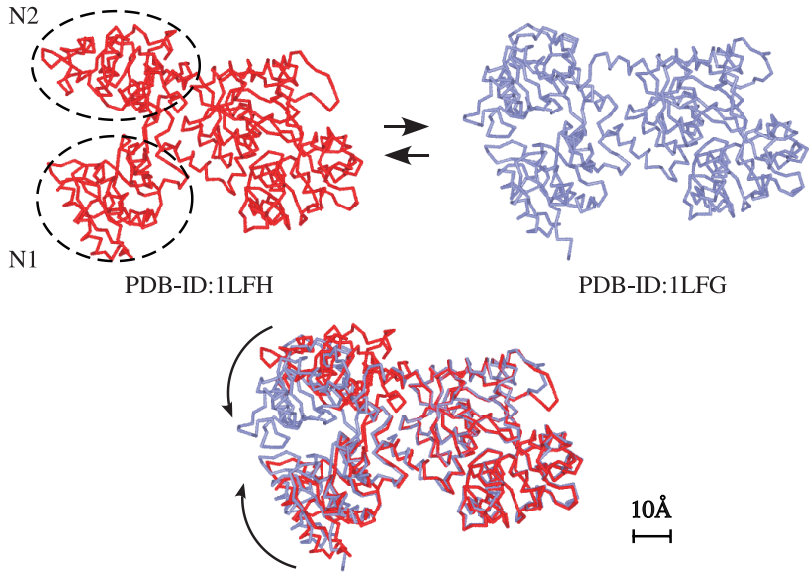

Figure 11 Structure and motion of lactoferrin. Lactoferrin consists of 691 amino residues. As shown in the conformation PDB-ID:1LFH, lactoferrin has a large cavity between the N1 and the N2 domains. Comparing conformations PDB-ID:1LFH and 1LFG, we find that lactoferrin opens and closes the cavity.

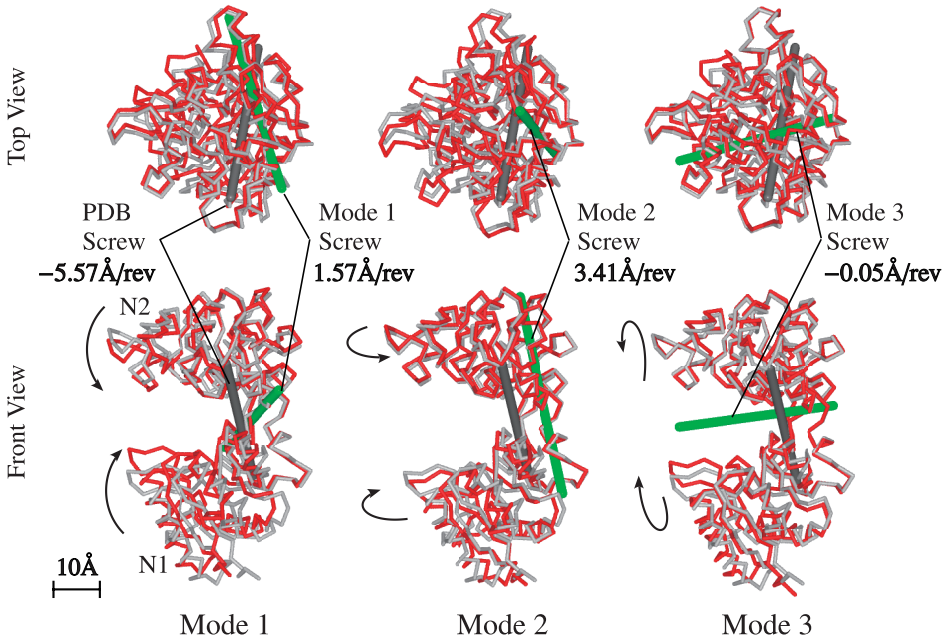

Figure 12 SC mode motions of lactoferrin. The motions were obtained by analyzing the protein model created from PDB-ID:1LFH. Among the six SC mode motions, the first three motions are shown (only the N1 and N2 domains are shown). In the analysis, the two application groups were specified for all the C_α in the N1 and N2 domains, and the two deformation groups were specified to be the same as the application groups (see Fig. 5). In each illustration, the screws approximatingthe (instantaneous) relative motion between the N1 and N2 domains calculated from the SC mode motions and the PDB motion (between the conformations of PDB-ID:1LFH and 1LFG) are shown. The cavity closing motion is observed in the first SC mode motion.

motions (the magnitude of $\Delta \boldsymbol{X}'_{btj}$ for each j in Eq. (35)) were calculated for approximating the reference deformation. The result is shown in Figure 13. The reference deformation ($\Delta \boldsymbol{X}'_{bt}$ in Eq. (34)) expresses the direction of the relative displacement between the deformation groups calculated from the difference between PDB-ID:1LFH and 1LFG (direction of the PDB motion). We can observe that contribution of the first SC mode motion is the highest and that the direction of

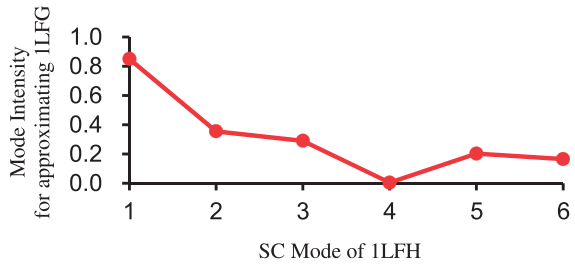

Figure 13 SC mode expansion of lactoferrin. The graph expresses the intensities of SC mode motions, or the magnitude of $\Delta \boldsymbol{X}'_{bt\,j}$ for each j in Eq. (35), calculated from PDB-ID:1LFH for approximating the motion direction from 1LFH to 1LFG (PDB motion). It can be observed that the intensity of the first SC mode motion is the highest and the direction of the PDB motion is approximated by the lower mode motions.

the PDB motion is approximated by the lower mode motions. Here, note that the SC mode motions were calculated from the model created using PDB-ID:1LFH; however, the pattern of the mode intensity shown in Figure 13 depends not only on PDB-ID:1LFH but also on 1LFG.

Next, we calculated the index Γ (Eq. (36)) when the mutual distances between the C_α in the part in the protein model were constrained. The part was specified by the sphere whose radius was 8 Å, and the calculation of Γ was repeated by scanning the center for all C_α. Figure 14 shows the result of this analysis when $n_s = 1$, 2, and 3. The values of the index Γ corresponding to the center of the constrained spheres are represented by the graphs and the shade mapped to the main chain structure (the lighter color expresses a larger value of Γ). It is known that the hinge axis between the N1 and N2 domains passes through 91 and 251 [42]; therefore, these residues play important roles in the internal motion of lactoferrin. In Figure 14, we can observe that the index Γ indicates large values when the parts around the C_α near the real hinge residues 91 and 251 are constrained.

In the analyses based on SC mode motion, we must correctly specify the application and deformation parts depending on the analysis objectives. For example, to understand the motion properties between the N1 and N2 domains from the structural compliance properties in the above analyses, we specified the two application groups for all the C_α in the N1 and N2 domains, and specified the two deformation groups as identical to the application groups. If the application and deformation groups are assigned to regions of the N1 domain only, the cavity closing motion will not appear.

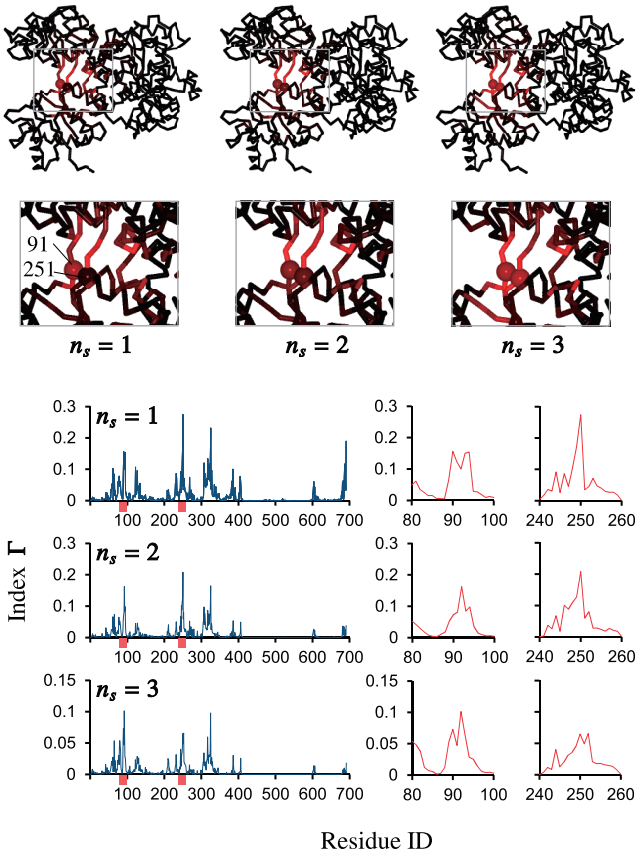

Figure 14 Constrained part scanning of lactoferrin. The changes of structural compliance properties are shown when mutual distances between the C_α in the localized part are constrained. For evaluating the changes, the index Γ (Eq. (36)) was used. The constrained part was specified by the sphere whose radius was 8 Å, and the calculation of Γ was repeated by scanning the center for all C_α. The results are shown for the case n_s = 1, 2, and 3 (see Fig. 10). The values of the index Γ corresponding to the center of the constrained spheres are represented by the graphs and the shade mapped to the main chain structure (the lighter color expresses a larger value of Γ). We can observe that the index Γ indicates large values when the parts around the C_α near the real hinge residues 91 and 251 are constrained.

Aspartate Transcarbamoylase (ATCase)

Figure 15 illustrates the 3D structure of ATCase. It consists of 2,778 amino residues and symmetrically arranged 12 subunits. Among the subunits, $C_{1\sim6}$ and $R_{1\sim6}$ are called the catalytic units and the regulatory units, respectively. The groups of subunits $C_{1,2,3}$ and $C_{4,5,6}$ are mutually connected by three limbs consisting of $R_{1,2}$, $R_{3,4}$, and $R_{5,6}$. During the transition between the R and T states (PDB-ID:1D09 and 1ZA1, respectively), $C_{1,2,3}$ and $C_{4,5,6}$ move relative to each other like a screw. It is known that the ligand binding to the regulatory units cause the screw-like motion between the two groups of catalytic units [48–50]. ATCase is a typical example of a protein that shows the allosterc effect.

For analyzing the motion interaction between the regulatory units and the catalytic units based on the structural compliance properties, we specified three application groups $R_{1,2}$, $R_{3,4}$, and $R_{5,6}$ and two deformation groups $C_{1,2,3}$ and $C_{4,5,6}$ to the protein model created from the PDB data of the R state (PDB-ID:1D09). We obtained 12 SC mode motions. Among the 12 SC mode motions, the last six mode motions corresponded to the motions of the regulatory units (i.e., the application groups) that do not affect the relative motion between the catalytic units (i.e., the deformation groups).

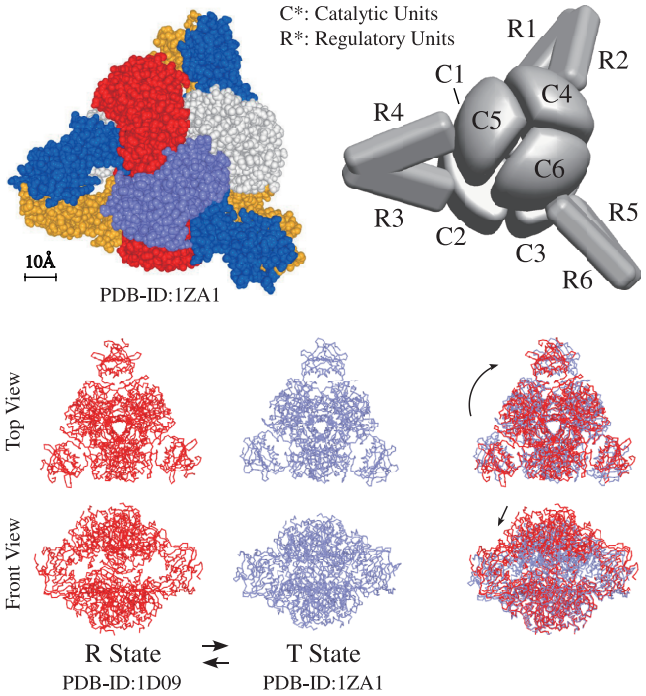

Figure 15 Structure and motion of aspartate transcarbamoylase (ATCase). ATCase consists of 2,778 amino residues and symmetrically arranged 12 subunits. Among the subunits, $C_{1\sim6}$ and $R_{1\sim6}$ are called the catalytic units and the regulatory units, respectively. During the transition between the R and T states (PDB-ID:1D09 and 1ZA1, respectively), $C_{1,2,3}$ and $C_{4,5,6}$ move relative to each other like a screw. The ligand binding to the regulatory units cause the screw-like motion. ATCase is a typical example of a protein that shows the allosteric effect.

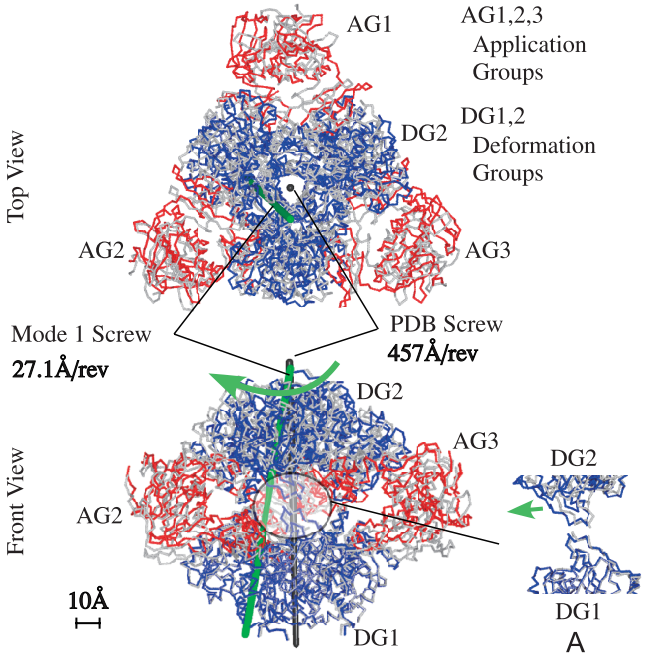

Figure 16 First SC mode motion of ATCase. The motion was obtained by analyzing the protein model created from PDB-ID:1D09. For analyzing the motion interaction between the regulatory units and the catalytic units based on the structural compliance properties, we specified three application groups $R_{1,2}$(AG1), $R_{3,4}$(AG2), and $R_{5,6}$(AG3) and two deformation groups $C_{1,2,3}$(DG1) and $C_{4,5,6}$(DG2) to the protein model. The screws approximating the (instantaneous) relative motion between the catalytic units $C_{1,2,3}$ and $C_{4,5,6}$ calculated from the first SC mode motion and the PDB motion (between the conformations of PDB-ID:1D09 and 1ZA1) are shown. We can observe the screw-like motion between the catalytic units $C_{1,2,3}$ and $C_{4,5,6}$ in the first SC mode motion. The screw axis is near that of the PDB motion; however, the pitch is much smaller (27.1 and 457 Å/rev). The structural compliance for this motion is high (soft), as shown in A, because the collisions between the convex parts of the catalytic units are avoided.

Figure 16 illustrates the first SC mode motion. In the illustration, the screws approximating the relative motion between the catalytic units $C_{1,2,3}$ and $C_{4,5,6}$ calculated from the first SC mode motion and the PDB motion (between PDB-ID:1D09 and 1ZA1) are shown. We can observe the screw-like motion between the catalytic units in the first SC mode motion. The screw axis is near that of the PDB motion; however, the pitch is much smaller than that of the PDB motion (27.1 and 457 Å/rev). Figure 17 shows the result of SC mode expansion when approximating the direction of the PDB motion (from PDB-ID:1D09 to 1ZA1) by SC mode motions. We can observe that the intensity of the first mode motion is the second highest and that of the sixth mode motion is the highest (the magnitudes of $\Delta \boldsymbol{X}'_{bt\,1}$ and $\Delta \boldsymbol{X}'_{bt\,6}$ in Eq. (35), respectively). In other words, the sixth SC mode motion (the hardest motion, except the motions that do not affect the relative motion between the specified deformation groups) is the one required the most to approximate the direction of the PDB motion. As shown in Figure 18, in the sixth SC mode motion, the upper and lower catalytic units are mutually compressed.

This result might appear to contradict the assumption that the lower SC mode motions will occur easily. However, it should be remembered that the formulations related to SC mode decomposition are based on instantaneous kinematics, whereas PDB motions are finite. In the conformation of the R state shown in Figure 15, the upper and lower catalytic units are close to each other. In real protein motion, during the first stage of the conformation change from the R state to the T state, it can be considered that motion like the sixth SC mode motion hardly occurs because of the collision among the convex parts of the catalytic units (see Fig. 18A). After the rotational motion or the screw motion of the small pitch, like the first SC mode motion, the convex parts in the catalytic units separate from each other (see Fig. 16A), and it becomes possible for the units to approach each other more closely. Therefore, we can infer that motion like the first SC mode motion actually occurs at the first stage during the conformation change.

As this example shows, we can understand the motion properties related to the allosteric interaction to some extent from the structural compliance properties. At the same time, the results indicate the limitation of the current method for

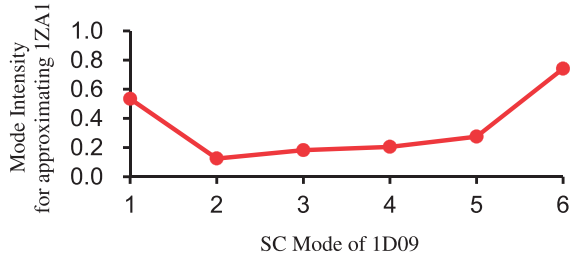

Figure 17 SC mode expansion of ATCase. The graph expresses the intensities of SC mode motions, or the magnitude of $\Delta \boldsymbol{X}'_{bt\,j}$ for each j in Eq. (35), calculated from PDB-ID:1D09 for approximating the motion direction from 1D09to 1ZA1 (PDB motion). It can be observed that the intensity of the first mode motion is the second highest and that of the sixth mode motion is the highest.

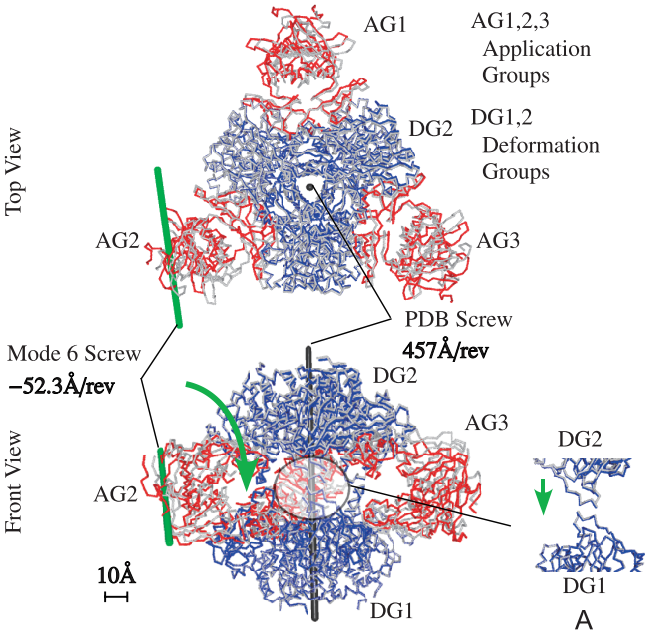

Figure 18 Sixth SC mode motion of ATCase. The motion was obtained by analyzing the protein model created from PDB-ID:1D09 under the same condition in Figure 16. It can be observed that the upper and lower catalytic units (DG1 and 2) are mutually compressed. The structural compliance for this motion is low (hard), as shown in A, because of the collisions among the convex parts of the catalytic units.

obtaining information about nonlinear large conformation changes.

Conclusion

We have formulated the methods for directly analyzing structural compliance properties of the ENM of proteins and for extracting the motion properties from the properties in a general form. When decomposing the motion according to the magnitude of structural compliance between the specified parts (SC mode decomposition), we can obtain information about the motion properties under the assumption that the lower SC mode motions or the softer motions occur easily. Moreover, for quantitative discussions, we have formulated the methods for calculating screws approximating the instantaneous relative motions between specified flexible groups, methods for approximating the PDB motions by the combinations of the SC mode motions (SC mode expansion), and methods for evaluating the changes in the properties by using principal angles (index Γ). For application examples, we analyzed lactoferrin and ATCase. The results showed that we could understand their motion properties including the allosteric interactions through their lower SC mode motions or the softer motions. The results also showed the limitations of the methods used to obtain information about nonlinear large conformation changes.

Although within limitations, by applying the presented theoretical framework for analyzing the structural compliance properties, we can expect to obtain information related to protein motions such as the conformation changes, the structures that enable allosteric effects, the effects of ligand bindings, and the key parts that govern the motion properties. In this study, the ENM focusing on the dihedral angles of the main chains was employed as a protein model. We can apply the methods not only to this type of ENM but also to other types of ENMs such as all-atom ENM and Cartesian-coordinate-based ENM by switching the calculations of the Jacobian matrices. In the future, we will include examples of the analyses for different types of proteins based on the theoretical framework. In addition, the SC mode and normal mode motions could be quantitatively compared through the SC mode expansion. This comparison is an interesting future prospect.

Acknowledgment

This research was supported by the Grant-in-Aid for Scientific Research by the Ministry of Education, Culture, Sports, Science and Technology, Japan.

Conflicts of Interest

Keisuke Arikawa declares that he has no conflict of interest.

Author Contribution

Keisuke Arikawa conceived and formulated the methods, developed the computer program, carried out the simulations, discussed the simulation results, and wrote the manuscript.

References

[1] Jacobs, C. R., Huang, H. & Kwon, R. Y. *Introduction to Cell Mechanics and Mechanobiology* (Garland Science, New York, 2012).
[2] Wang, J. H. C. & Thampatty, B. P. An introductory review of cell mechanobiology. *Biomech. Model. Mechanobiol.* **5**, 1–16 (2006).

[3] Orr, A. W., Helmke, B. P., Blackman, B. R. & Schwartz, M. A. Mechanisms of Mechanotransduction. *Dev. Cell* **10**, 11–20 (2006).
[4] Hayakawa, K., Tatsumi, H. & Sokabe, M. Actin stress fibers transmit and focus force to activate mechanosensitive channels. *J. Cell Sci.* **121**, 496–503 (2008).
[5] Hirata, H., Ttsumi, H. & Sokabe, M. Mechanical forces facilitate actin polymerization at focal adhesions in a zyxin-dependent manner. *J. Cell Sci.* **121**, 2795–2804 (2008).
[6] DuFort, C. C., Paszek, M. J. & Weaver, V. M. Balancing forces: architectural control of mechanotransduction. *Nat. Rev. Mol. Cell Biol.* **12**, 308–319 (2011).
[7] Gō, N. A theorem on amplitudes of thermal atomic fluctuations in large molecules assuming specific conformations calculated by normal mode analysis. *Biophys. Chem.* **35**, 105–112 (1990).
[8] Kidera, A. & Gō, N. Refinement of protein dynamic structure: normal mode refinement. *Proc. Natl. Acad. Sci. USA* **87**, 3718–3722 (1990).
[9] Tirion, M. M. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* **77**, 1905–1908 (1996).
[10] Thomas, A., Field, M. J. & Perahia, D. Analysis of the low-frequency normal modes of the R state of aspartate transcarbamylase and a comparison with the T state modes. *J. Mol. Biol.* **261**, 490–506 (1996).
[11] Tama, F., Gadea, F. X., Marques, O. & Sanejouand, Y. H. Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins* **41**, 1–7 (2000).
[12] Tama, F. & Sanejouand, Y.-H. Conformational change of proteins arising from normal mode calculations. *Protein Eng.* **14**, 1–6 (2001).
[13] Atilgan, A. R., Durell, S. R., Jernigan, R. L., Demirel, M. C., Keskin, O. & Bahar, I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* **80**, 505–515 (2001).
[14] Krebs, W. G., Alexandrov, V., Wilson, C. A., Echols, N., Yu, H. & Gerstein, M. Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic. *Proteins* **48**, 682–695 (2002).
[15] Schuyler, A. D. & Chirikjian, G. S. Normal mode analysis of proteins: a comparison of rigid cluster modes with C_α coarse graining. *J. Mol. Graph. Model.* **22**, 183–193 (2004).
[16] Wako, H., Kato, M. & Endo, S. ProMode: a database of normal mode analyses on protein molecules with a full-atom model. *Bioinformatics* **20**, 2035–2043 (2004).
[17] Bahar, I. & Rader, A. J. Coarse-grained normal mode analysis in structural biology. *Curr. Opin. Struct. Biol.* **15**, 586–592 (2005).
[18] Petrone, P. & Pande, V. S. Can conformational change be described by only a few normal modes? *Biophys. J.* **90**, 1583–1593 (2006).
[19] Hayward, S. & de Groot, B. L. Normal Modes and Essential Dynamics. in *Molecular Modeling of Proteins.* (Kukol, A. ed.) pp. 89–106 (Humana Press, New York, 2008).
[20] Bahar, I., Lezon, T. R., Bakan, A. & Shrivastava, H. Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins. *Chem. Rev.* **110**, 1463–1497 (2010).
[21] Togashi, Y. Screening for mechanical responses of proteins using coarse-grained elastic network models. *NOLTA* **7**, 190–201 (2016).
[22] Taguchi, J. & Kitao, A. Dynamic profile analysis to characterize dynamic-driven allosteric sites in enzymes. *Biophys. Physicobiol.* **13**, 117–126 (2016).
[23] Ikeguchi, M., Ueno, J., Sato, M. & Kidera, A. Protein structural change upon ligand binding: linear response theory. *Phys. Rev. Lett.* **94**, 078102 (2005).
[24] Protein Data Bank. Available at http://www.rcsb.org
[25] Chen, J., Xie, Z. & Wu, Y. Sudy of protein structural deformations under external mechanical perturbations by a coarse-grained simulation method. *Biomech. Model. Mechanobiol.* **15**, 317–329 (2016).
[26] Canutescu, A. A. & Dunbrack Jr., R. L. Cyclic coordinate descent: a robotics algorithm for protein loop closure. *Protein Sci.* **12**, 963–972 (2003).
[27] Cahill, S., Cahill, M. & Cahill, K. On the kinematics of protein folding. *J. Comput. Chem.* **24**, 1364–1370 (2003).
[28] Kazerounian, K. Is design of new drugs a challenge for kinematics? in *Advances in Robot Kinematics.* (Lenarčič, J. & Thomas, F. eds.) pp. 134–144 (Springer, Netherlands, 2002).
[29] Kazerounian, K. From mechanisms and robotics to protein conformation and drug design. *J. Mech. Des. N.Y.* **126**, 40–45 (2004).
[30] Kazerounian, K., Latif, K. & Alvarado, C. Protofold: a successive kinetostatic compliance method for protein conformation prediction. *J. Mech. Des. N.Y.* **127**, 712–717 (2004).
[31] Subramanian, R. & Kazerounian, K. Improved molecular model of a peptide unit for proteins. *J. Mech. Des. N.Y.* **129**, 1130–1136 (2007).
[32] Kazerounian, K. Protein Molecules: Evolution's Design for Kinematic Machines. in *21st Century Kinematics.* (McCarthy, J. M. ed.) pp. 217–244 (Springer, London, 2012).
[33] Sharma, G., Badescu, M., Dubey, A., Mavroidis, C., Tomassone, S. M. & Yarmush, M. L. Kinematics and workspace analysis of protein based nano-actuators. *J. Mech. Des. N.Y.* **127**, 718–727 (2005).
[34] Chirikjian, G. S., Kazerounian, K. & Mavroidis, C. Analysis and design of protein based nanodevices: challenges and opportunities in mechanical design. *J. Mech. Des. N.Y.* **127**, 695–698 (2005).
[35] Diez, M., Petuya, V., Martínez-Cruz, L. A. & Hernández, A. A biokinematic approach for the computational simulation of proteins molecular mechanism. *Mech. Mach. Theory* **46**, 1854–1868 (2011).
[36] Gipson, B., Hsu, D., Kavraki, L. E. & Latombe, J-C. Computational models of protein kinematics and dynamics: beyond simulation. *Annu. Rev. Anal. Chem.* **5**, 273–291 (2012).
[37] Arikawa, K. Structural compliance analysis and internal motion properties of proteins from a robot kinematics perspective: formulation of basic equations. *J. Mech. Robot.* **8**, 021028 (2016).
[38] Amadei, A., Linssen, A. B. M. & Berendsen, H. J. C. Essential Dynamics of Proteins. *Proteins* **17**, 412–425 (1993).
[39] Yang, L., Eyal, E., Bahar, I. & Kitao, A. Principal component analysis of native ensembles of biomolecular strucutres (PCA_NEST): insights into functional dynamics. *Bioinformatics* **25**, 606–614 (2009).
[40] Kim, M. K., Jernigan, R. L. & Chirikjian, G. S. Rigid-cluster models of conformational transition in macromolecular machines and assemblies. *Biophys. J.* **89**, 43–55 (2005).
[41] Subbiah, S. *Protein Motions* (R. G. Landes Company, Texas, 1996).
[42] Gerstein, M., Anderson, B. F., Norris, G. E., Baker, E. N., Lesk, A. M. & Chothia, C. Domain closure in lactoferrin: two hinges produce a see-saw motion between alternative close-packed interfaces. *J. Mol. Biol.* **234**, 357–372 (1993).
[43] Hayward, S., Kitao, A. & Berendsen, H. J. C. Model-free methods of analyzing domain motions in proteins from simulation: a comparison of normal mode analysis and molecular dynamics simulation of lysozyme. *Proteins* **27**, 425–437 (1997).
[44] Hayward, S. & Berendsen, H. J. C. Systematic analysis of

domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. *Proteins* **30**, 144–154 (1998).
[45] Tsai, L. W. *Robot Analysis* (John Wiley & Sons, New Jersey, 1999).
[46] Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Cryst.* **32**, 922–923 (1976).
[47] Kabsch, W. A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Cryst.* **34**, 827–828 (1978).
[48] Petsko, G. A. & Ringe, D. *Protein Structure and Function* (New Science Press, London, 2004).
[49] Whitford, D. *Proteins: Structure and Function* (John Wiley & Sons, West Sussex, 2005).
[50] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. & Walter P. *Molecular Biology of the Cell (5th edition)* (Garland Science, New York, 2008).