



Review

Comparing the intrinsic dynamics of multiple protein structures using elastic network models[☆]



Edvin Fuglebakk, Sandhya P. Tiwari, Nathalie Reuter^{*}

Department of Molecular Biology, University of Bergen, Pb. 7803, N-5020 Bergen, Norway
 Computational Biology Unit, Department of Informatics, University of Bergen, Pb. 7803, N-5020 Bergen, Norway

ARTICLE INFO

Article history:

Received 30 June 2014

Received in revised form 15 September 2014

Accepted 16 September 2014

Available online 28 September 2014

Keywords:

Elastic network models

Protein dynamics

Intrinsic dynamics

Normal mode analysis

ABSTRACT

Background: Elastic network models (ENMs) are based on the simple idea that a protein can be described as a set of particles connected by springs, which can then be used to describe its intrinsic flexibility using, for example, normal mode analysis. Since the introduction of the first ENM by Monique Tirion in 1996, several variants using coarser protein models have been proposed and their reliability for the description of protein intrinsic dynamics has been widely demonstrated. Lately an increasing number of studies have focused on the meaning of slow dynamics for protein function and its potential conservation through evolution. This leads naturally to comparisons of the intrinsic dynamics of multiple protein structures with varying levels of similarity.

Scope of review: We describe computational strategies for calculating and comparing intrinsic dynamics of multiple proteins using elastic network models, as well as a selection of examples from the recent literature.

Major conclusions: The increasing interest for comparing dynamics across protein structures with various levels of similarity, has led to the establishment and validation of reliable computational strategies using ENMs. Comparing dynamics has been shown to be a viable way for gaining greater understanding for the mechanisms employed by proteins for their function. Choices of ENM parameters, structure alignment or similarity measures will likely influence the interpretation of the comparative analysis of protein motion.

General significance: Understanding the relation between protein function and dynamics is relevant to the fundamental understanding of protein structure–dynamics–function relationship. This article is part of a Special Issue entitled Recent developments of molecular dynamics.

© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

1. Introduction

Folded proteins are remarkably dense with heterogeneously distributed densities [1], which reflect the uneven distribution of interatomic forces within. Their response to thermal forces is expected to proceed by preferably deforming the least compact regions, while keeping the most compact ones rigid. Atoms tightly coupled on short time-scales are expected to remain tightly coupled on longer time-scales, especially between unfolding events. This suggests that estimates of the atomic density distribution within a folded protein can capture its collective degrees of freedom. It also motivates the extrapolation from the analysis of intrinsic properties of the structure to collective motions occurring on, for example, the millisecond time-scale. Estimates of the

atomic density distribution can also replace information about the exact chemistry involved in stabilising the fold, similar to how the elastic response of macroscopic materials can be calculated without atomic detail. While implicit in some of the early work on elastic network models, the relationship between the local packing density and the flexibility of proteins was explicitly treated by Halle [2].

Likewise, elastic network models (ENMs) are based on the simple idea that a protein can be described as a set of particles connected by springs. These can then be used to describe the protein's intrinsic flexibility using normal mode analysis (NMA). Monique Tirion pioneered the field in 1996, when she showed that a single-parameter potential could reproduce the slow dynamics obtained with a more complicated potential [3]. This simplification makes the potential insensitive to the details of the equilibrium structure, which has minimal energy by construction. Models from experimental structure determination can thus be used directly, without the costly energy minimisation associated with the use of chemical force fields. Tirion's model has later been further simplified, in particular by increasing its coarseness, as seen in ENMs of interacting residues rather than atoms [4,5]. ENMs provide a simple and interpretable description of the protein's collective motion, which can be conveniently coarse-grained and are computationally

Abbreviations: ENM, elastic network model; NMA, normal mode analysis; PCA, principal component analysis

[☆] This article is part of a Special Issue entitled Recent developments of molecular dynamics.

^{*} Corresponding author at: University of Bergen, Department of Molecular Biology, Pb. 7803, N-5020 Bergen, Norway. Tel.: +47 555 84040; fax: +47 555 89683.

E-mail addresses: Edvin.Fuglebakk@mbi.uib.no (E. Fuglebakk), Sandhya.Tiwari@mbi.uib.no (S.P. Tiwari), nathalie.reuter@mbi.uib.no (N. Reuter).

inexpensive. For these reasons, they rapidly have replaced molecular mechanics force fields that had been used for NMA of proteins earlier [6–10].

The robustness of NMA with ENMs for the description of slow collective motions in proteins can seem surprising, given its simple construction. The motivation outlined above for using ENMs involved some brave assumptions, and it was not necessarily clear beforehand that these assumptions were valid. In particular, the harmonic approximation used for investigating dynamics of large conformational changes and the absence of frictions such as those caused by the solvent. Yet, early studies comparing NMA and experimental structural data, or molecular dynamics simulations, did validate the use of NMA with coarse-grained models. Validation against detailed molecular mechanics force fields on large protein datasets has shown that even coarser models than the one suggested by Tirion still reproduce the slow dynamics obtained from molecular simulations (e.g. [11–14]). Furthermore, several studies have shown that in many cases, a few low-energy normal modes account for most of the structure difference between two conformational states [15–18]. Conformational changes can be described by just a few low-energy normal modes intimately linked to the structure, indicating that proteins systematically make use of these low-energy modes to achieve their function. The importance of these modes for protein function has naturally led to the question of the evolutionary conservation of their slow dynamics, analogous to the conservation between structure and sequence. Fig. 1 illustrates the relationship between the similarities in structural shape and intrinsic domain motion described by the low energy normal modes from the ENMs of two distantly related P-type ATPases.

Examples of comparative dynamics analysis include studying a set of proteins that represent various functional states of a given enzyme upon ligand-binding [19,20], evaluating the conservation of dynamics within a homologous protein family [21–27], or within a set of proteins that possess the same fold despite low sequence identity [28,29]. In a recent article, Cristian Micheletti comprehensively reviewed the use of dynamics as an aid for sequence and structure alignments of proteins [30]. It has been shown, when comparing structures of homologous proteins

and their intrinsic dynamics, that protein structures evolve along low-energy modes [14,31,32]. Furthermore, a number of studies have shown that low-energy modes are robust to sequence variations [14, 29,33–37]. The use of ENMs for comparative protein dynamics has the potential to teach us more about a wide range of topics. To name a few, these can include the effects of ligand or allosteric effector binding in an active or allosteric site, changes in oligomeric state, changes in sequence or structure through evolution, and the level of similarity in dynamics between functionally similar enzymes.

Together with the question of the evolutionary conservation of internal dynamics has come the need to reliably compare computed dynamics for a set of protein structures. Due to the scarcity of experimental data describing protein dynamics, molecular modelling at large is an attractive alternative that has earlier demonstrated its predictive power through numerous applications. ENMs are a model of choice for such studies, even if computing power has admittedly become more affordable than it was at the advent of ENMs and molecular dynamics simulations on microsecond time-scales are becoming increasingly accessible to the research community. The tractability and simplicity of ENMs are unparalleled by molecular mechanics force fields and ENMs defined with transferrable parameters can be easily applied to large numbers of protein structures in automated ways. Beyond the choice of the ENM and its parameterisation, comparing internal dynamics of several protein structures comes with a set of methodological choices, which are not obvious, but can significantly affect the outcome of the comparative dynamics analysis. After an introduction to the formalism of ENMs and their parameterisation, we focus on aspects that are directly relevant for comparative analysis of multiple protein structures, such as the similarity measures used to compare computed dynamics, the influence of the alignment methods and ways to include the influence of regions in the structures that are not similar in sequence or conserved into the comparison. Next, using selected examples, we describe how comparing protein intrinsic dynamics can be successfully used to understand conformational changes upon ligand binding, functional oligomerisation states and the overall role of intrinsic dynamics in protein function. Finally we list some of the most commonly used software and libraries for ENM calculations.

2. Elastic network models

2.1. Formalism

Since Tirion's contribution [3], further simplifications of the ENMs have been made. Tirion's model was an elastic network with a node for each atom and springs with uniform force-constants between all pairs of nodes closer than a distance-based cut-off. Upon realising that a good density estimate can be made even without atomic detail and that backbone motion can be largely decoupled from side-chain movement, Hinsen et al. [4] introduced a model with non-uniform distance dependent force-constants, connecting only C α atoms. Atilgan et al. [5] also applied Tirion's uniform force constant model at the C α granularity. This model is particularly convenient to visualise, and is illustrated in Fig. 2. Another popular density-based model has been the early Gaussian network model (GNM) [38]. While it obtains density estimates in a way that is similar to Atilgan et al., this model does not employ a Hookean potential. The interpretation of GNMs is therefore different from the ENMs.

Since the initial ENMs, many variants have been proposed. More detailed descriptions of the local backbone configurations have been investigated, such as parameters dependent on the secondary structure of the backbone [39,40], the reintroduction of chemical bond information or other kinds of residue specific interaction terms [41–43] as well as the modelling of side-chain locations [44]. On the other hand, simplifications to fewer coordinates have been proposed, both in terms of simpler coordinate systems [45,46] and less granular representations

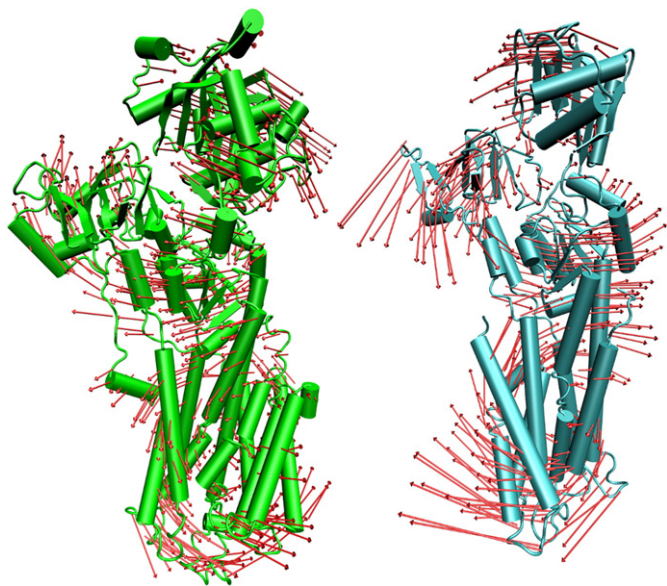


Fig. 1. Normal mode vectors from elastic network models of two distantly related proteins. The SR Calcium ATPase 1 (PDB ID: 1WPG [126], green) and the Copper-transporting P1B-type ATPase (PDB ID: 3RFU [127], cyan) have similar low frequency modes as illustrated here by the third lowest energy modes of each protein (red arrows). These vectors show the flexibility of the four domains of the proteins with respect to each other. This is an example where two structures with similar shapes yield comparable normal mode vectors from ENMs. The normal mode vector fields for these structures were computed using WEBnma [110] and the images were rendered in VMD [128].

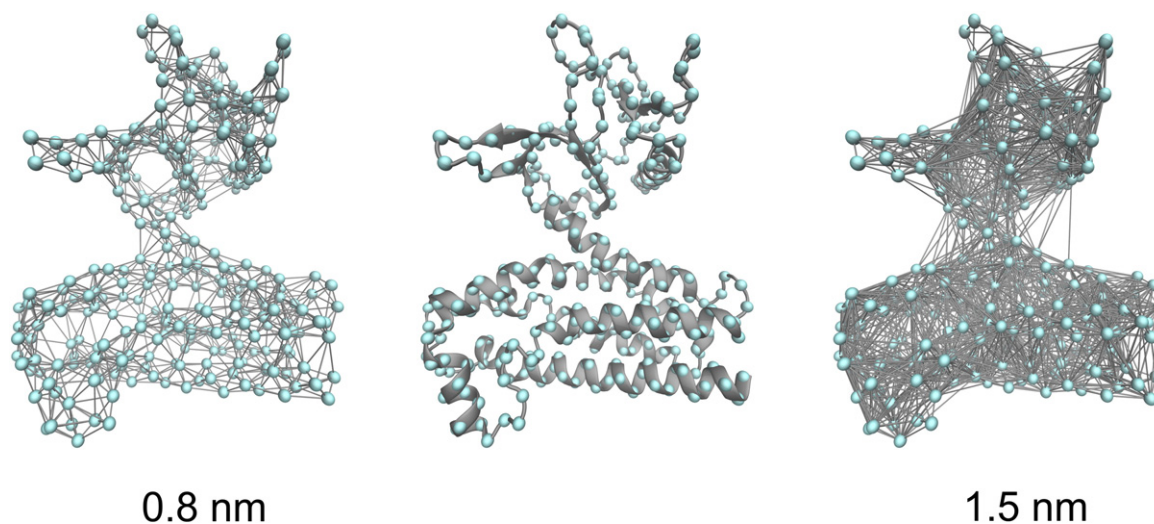


Fig. 2. Elastic network models with different cut-offs. Illustrates the effect on uniform force constants for elastic network models. The models are constructed for the triple functional domain protein (PDB ID: 1NTY [129]). Centre: The cartoon representation has C α atoms highlighted as cyan spheres. Left: A model that connects all residues with C α atoms within 0.8 nm of each other. Right: A model that connects all residues with C α atoms within 1.5 nm of each other. For clarity, connections between buried residues are not shown if they are more than 0.8 nm apart.

of proteins [47]. Despite all this variety, ENMs can be understood in terms of a single unifying formalism, the details of which are as follows.

The ENMs model the protein as a network of Hookean springs that connect all residues, which are typically represented by nodes located at the centre of their C α atom. Interactions between atoms are described by the pair potential for a given configuration of a protein:

$$V_{ij}(\mathbf{r}) = \frac{k_{ij}}{2} \left(\left\| \mathbf{r}_i - \mathbf{r}_j \right\| - \left\| \mathbf{r}_i^0 - \mathbf{r}_j^0 \right\| \right)^2 \quad (1)$$

where \mathbf{r}_i is the position of a residue i , in the configuration of the protein \mathbf{r} , the superscript 0 denotes the equilibrium conformation and k_{ij} is the force constant for the spring connecting residues i and j . Here, k_{ij} is typically determined by a scalar function of distance between connected nodes. Apart from the choice of granularity of the model, the function for determining k_{ij} is the most important difference between different ENMs. The potential energy of the entire network is the sum of this pair potential over all pairs:

$$V(\mathbf{r}) = \sum_{i=1}^N \sum_{j=i+1}^N V_{ij}(\mathbf{r}) \quad (2)$$

where N is the number of nodes in the network. Expanding this potential as a Taylor series around \mathbf{r}^0 reveals the following form of the potential for a configuration obtained by infinitesimal displacement from its equilibrium configuration:

$$V(\mathbf{r}) = \frac{1}{2} (\mathbf{r} - \mathbf{r}^0)^T \mathbf{H} (\mathbf{r} - \mathbf{r}^0) \quad (3)$$

with \mathbf{H} as the matrix of partial second order derivatives of the potential. With respect to Cartesian coordinates, this is a $3N \times 3N$ matrix. The elements of \mathbf{H} can be specified in terms of 3×3 submatrices corresponding to each pair of nodes:

$$\mathbf{H}_{ij} = \begin{cases} -\frac{k_{ij}}{\left\| \mathbf{r}_i^0 - \mathbf{r}_j^0 \right\|^2} (\mathbf{r}_i^0 - \mathbf{r}_j^0) (\mathbf{r}_i^0 - \mathbf{r}_j^0)^T, & i \neq j \\ \sum_{l \neq i} -\mathbf{H}_{il}, & i = j \end{cases} \quad (4)$$

Since \mathbf{H} is a symmetric matrix, the potential energy of a configuration \mathbf{r} can be written in terms of its eigendecomposition:

$$V(\mathbf{r}) = \sum_{m=1}^{3N} \lambda_m \left((\mathbf{r} - \mathbf{r}^0)^T \mathbf{v}_m \right)^2 \quad (5)$$

where \mathbf{v}_m represents the normalised eigenvectors and λ_m the corresponding eigenvalues of \mathbf{H} . These eigenvectors form an orthogonal basis for the configurational space of the protein, so that they each provide energetically independent contributions to the potential energy of \mathbf{r} . These independent modes of deformation are referred to as the normal modes of the network, and they describe motion intrinsic to the protein structure. Because of the coarseness of the model, eigenvalues and hence energies are not interpreted exactly, but the separation between eigenvalues are informative of the relative energetic cost of different structural deformations. Since rigid-body rotations and translations of the network are not restrained, the six modes corresponding to rigid-body motion in Cartesian coordinates will have zero energy. The modes describing rigid body displacements are referred to as trivial modes.

Since normal mode analysis has a long tradition in chemistry for analysing small vibrational molecules, the above formalism is often presented as an eigendecomposition of the mass-weighted Hessian. In that case, the elastic network is considered as a coupled harmonic oscillator and the eigenvalues are the squared frequencies of vibration along the corresponding modes. While the vibrational normal modes are a perfectly valid decomposition of motion, it is worth stressing that solvated proteins cannot in general be expected to be vibrational along their lower energy modes [48] and thus, this requires cautious interpretation of the oscillator model.

For equally normalised displacements, the quadratic dependence of energy on the spatial extent of deformations causes large local deformations to be more energetically expensive than collective motions that involve only small changes to each spring. Therefore low-energy modes are expected to be collective. By a similar reasoning, collective motions can be expected to have larger amplitudes, as local deformations are constrained by the stronger local interactions. In fact, for a harmonic potential, the displacements along low-energy normal modes are exactly the deviations along high-variance principal components. The Boltzmann distribution for the potential given in Eq. (3) is a multivariate Gaussian distribution with a covariance matrix proportional to the inverse of \mathbf{H} . Because of the zero energy associated with rigid movement of the protein,

this inverse is not defined, but the Moore–Penrose pseudo-inverse \mathbf{C} , can for many applications be regarded as a covariance matrix of internal deformations:

$$\mathbf{C} = \sum_{m=7}^{3N} \frac{1}{\lambda_m} \mathbf{v}_m \mathbf{v}_m^T \quad (6)$$

where the sum runs over the nontrivial modes. This implies that the eigenvectors \mathbf{v}_m can be regarded as the principal components of this covariance matrix \mathbf{C} , with variance $1/\lambda_m$. The covariance along each of the Cartesian coordinates of a pair of nodes i and j is proportional to C_{ij} , which denotes a 3×3 matrix. The trace of the submatrices C_{ii} is proportional to the mean squared thermal fluctuation of node i :

$$\langle \|\mathbf{r}_i - \mathbf{r}_i^0\|^2 \rangle \propto \text{tr}(\mathbf{C}_{ii}) = \sum_{m=7}^{3N} \frac{1}{\lambda_m} \mathbf{v}_{m,i}^T \mathbf{v}_{m,i} \quad (7)$$

where the angle brackets denote the mean and tr denotes the trace, or diagonal sum of the matrix. Here $\mathbf{v}_{m,i}$ is a vector of size 3 containing the elements of \mathbf{v}_m that describe the motion of atom i . To obtain a scalar quantification of the correlation of two nodes, a correlation matrix is commonly calculated, following Ichiye and Karplus [49]:

$$P_{ij} = \frac{\text{tr}(\mathbf{C}_{ij})}{(\text{tr}(\mathbf{C}_{ii})\text{tr}(\mathbf{C}_{jj}))^{1/2}} \quad (8)$$

Here, the numerator is proportional to the expected inner product of displacements, which depends on both the magnitudes and the angles between node displacements, whenever $i \neq j$.

As mentioned above, the inner product in Eq. (5) quantifies the contribution of a mode to a small displacement from the energetic minimum. As a means to identify a few normal modes that approximate the displacement well, the squared overlap and related measures are commonly calculated [15,50]. The squared overlap O_m , of a normalised displacement vector \mathbf{d} and a normal mode \mathbf{v}_m is the squared inner product:

$$O_m(\mathbf{d}) = (\mathbf{d}^T \mathbf{v}_m)^2 \quad (9)$$

with

$$\sum_{m=1}^{3N} O_m(\mathbf{d}) = 1 \quad (10)$$

since the normal modes are orthonormal. Such approaches are often applied even when the displacements analysed are not strictly infinitesimal. They have been important in validating the ENMs, by comparing with experimentally observed displacements. Large overlaps with low-energy modes indicate that the model favours a displacement, even if the energy cannot be determined accurately.

2.2. Parameterisation: force constants and cut-offs

Apart from the choice of granularity and coordinate system used to represent the protein as an elastic network, the different ENMs proposed over the years mainly differ in how the force constants are determined (the function determining k_{ij} in Eq. (1)). While this function is commonly chosen to be a function of interatomic distance in the equilibrium conformation, model developers have not reached a consensus on which mathematical formalism is more appropriate, or which benchmarking standards should be used. The simplest approach, following Tirion's initial model, uses uniform force constants for atoms or nodes that are within a given cut-off distance from each other. Springs longer than this cut-off are then assigned a force constant of zero, which is equivalent to omitting the spring from the model. Other formalisms connect all nodes and set the force constants proportional to a function decaying with distance. Fig. 3 illustrates the two

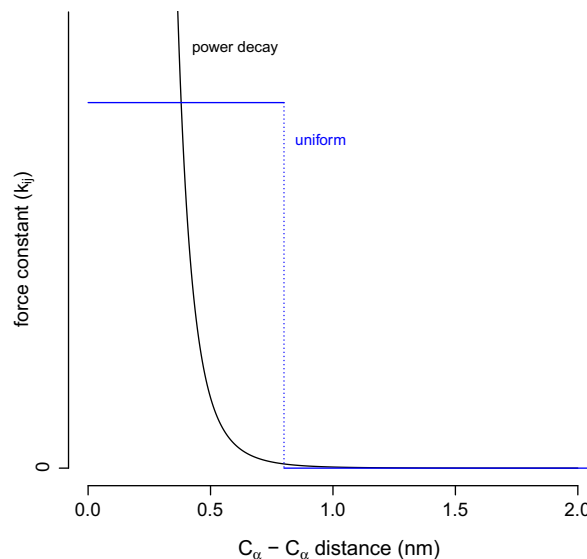


Fig. 3. Distance dependence of force constants. Illustrates two common schemes for determining force constants based on interatomic distance. Black line: force constants proportional to smoothly decaying function of interatomic distance (here an inverse power of 6). Grey line: uniform force constants for atoms closer than a cut-off distance (here 0.8 nm).

approaches. Since energies of individual modes are typically only interpreted after normalisation, the exact values of the force constants are not important, only the contrast between them. While different choices of mathematical formalisms can be brought to close agreement through careful parameterisation [51], it is important to choose appropriate benchmarks to parameterise against. Fig. 2 illustrates the difference between common parameterisations of a uniform force constant model.

The parameterisation of ENMs was initially motivated by comparison to detailed chemical potentials [3,4] analysis of MD-trajectories [52] and radial distribution analysis of the coordination between residues in the protein core [2,5]. Taking advantage of the vast amount of structural data available, it has also become custom to parameterise model predictions against crystallographic B-factors and ensemble variation in NMR models. This practice does not come without assumptions, however, as neither of these are direct observations of thermal motion, and in the case of B-factors, the experimental conditions do not reflect the solvent environment for which its application might be intended. In recent years attempts have been made to carefully quantify how these assumptions affect the parameterisation [53–56]. Notably, a wide range of cut-off values (from 8 to 15 Å) has been used in cut-off based models to represent the interatomic interactions. Interestingly, the directions (eigenvectors) of intrinsic motion are well captured even when each node is connected to equally many neighbours with springs of uniform strength [33,57]. The intrinsic motion of structures represented by discretisation of low-resolution density maps has also proven to yield useful information about structural flexibility [58,59]. These results emphasize that the global shape of the protein largely determines the principal directions of motion. This might explain the apparent robustness of the ENMs to the choice of cut-off, as even very long cut-offs can discriminate protein surface from protein interior. In order to also predict the energetic separation between normal modes, a narrower range of parameters are found to be appropriate [53,56]. The lack of uniformity in ENM parameterisation is partly a consequence of differing opinions on benchmarking procedures. The studies mentioned above have shown that the performance of different ENMs depend on the benchmark chosen, and thus, researchers should carefully consider which benchmark they trust for their application, and choose or define their model accordingly. ENMs can also be modelled to reflect a crystalline environment [60], and parameterisations obtained for such models can potentially help in parameterising single protein ENMs by

comparing with B-factors. Even so, exact interpretation should be made cautiously, as B-factors are heavily influenced by non-thermal contributions [54,55].

3. Validation

Early studies comparing coarse-grained NMA and experimental structural data, or molecular dynamics simulations were used to validate the elastic network approach. Validation against detailed molecular mechanics force fields performed on large protein datasets has shown that ENMs reproduce the slow dynamics obtained from molecular simulations well (see references [11–14,39,40,44,56,61]).

Furthermore, several studies have focused on validation against experimental data; they evaluated the number of low-energy modes necessary to describe the structural difference between two distinct x-ray structures (for example, the open and closed-forms) of the same protein using the overlap between the calculated set of modes and the structure difference vector as a quality measure. Using measures like Eq. (9), these studies show that in many cases a few low-energy normal modes account for most of the structure differences [15–18]. Hinsen et al. [16] compared the characterisation of domains from an ENM with those obtained from internal distance differences in experimentally determined conformations of citrate synthase, HIV-1 reverse transcriptase and aspartate transcarbamylase. Sanejouand and coworkers systematically analysed the agreement between low-energy normal modes and small data sets of experimentally determined structures in different conformational states [17,62]. Krebs et al. showed that more than half of a set of 3800 protein motions could be described by only two of the lowest energy normal modes [18]. Using the large number of structures determined for some proteins, the structural variation can be decomposed into principal components and compared with normal modes, as demonstrated by, for example, Yang et al. [14]. In all of these studies the conformational changes of the proteins were found to be well described by the lower energy normal modes intimately linked to the protein's structure.

In addition ENMs have been used as a tool for characterisation in many case studies of proteins and macromolecular complexes. In many such studies, the normal mode analysis is validated by comparing with conformational change, or by testing the insights obtained by independent means [36,63–65]. Comparison of predictions from ENMs with molecular dynamics simulations has also been used to validate and benchmark models [12,39,40,44,56,61].

4. Comparing intrinsic dynamics: getting quantitative

Comparisons of principal modes of motion have been performed successfully by manual inspection and expert judgement of ENM predictions. In recent years efforts have been made to assess the similarity of motion quantitatively. In addition to the different similarity measures proposed, many general methodological concerns and pitfalls have been identified.

4.1. Similarity measures

ENMs can predict atomic fluctuations through Eq. (7), and these can be compared to fluctuations obtained from other structures or models by an appropriate association measure, such as the squared inner product, SIP:

$$\text{SIP}(\mathbf{a}, \mathbf{b}) = \frac{\left(\sum_{i=1}^N a_i b_i\right)^2}{\left(\sum_{i=1}^N a_i^2\right)\left(\sum_{i=1}^N b_i^2\right)} = \left(\frac{\mathbf{a}^T \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|}\right)^2 \quad (11)$$

where \mathbf{a} and \mathbf{b} are vectors of size N with elements quantifying the atomic fluctuation of each atom in the model. Correlation measures have also been commonly applied to compare such fluctuation vectors. Since

exact energies are not reliably predicted by ENMs, any quantity that factors in the eigenvalues, such as the atomic fluctuations, have to be compared in a normalised fashion.

As mentioned above, motions calculated from ENMs are only valid for small displacements from the equilibrium, and the inference to large deformations involves the assumption that the interatomic couplings are relevant for longer timescales. The atomic fluctuations only reflect these couplings indirectly. Even if the atomic fluctuations have been the most common way to compare intrinsic motion quantitatively, it is therefore preferable to compare the normal modes or the covariance matrices of the ENMs. In recent work, we contrast comparison of fluctuations with those of normal modes and covariance matrices [56,66].

For comparing sets of normal modes, the root mean squared inner product (RMSIP) of the lowest modes has been a common choice:

$$\text{RMSIP}(\mathbf{V}, \mathbf{W}) = \left(\frac{1}{n} \sum_{m=7}^{n+6} \sum_{l=7}^{n+6} (\mathbf{v}_m^T \mathbf{w}_l)^2\right)^{\frac{1}{2}} \quad (12)$$

where \mathbf{V} and \mathbf{W} are sets of normal modes or principal components, and the sum runs over the nontrivial modes with the lowest energy or the highest variance. The constant n defines, somewhat arbitrarily, a subspace of protein motion that is considered accessible by low-energy motion. The RMSIP quantifies how similar the directions of these low-energy subspaces are for two protein models. Since the modes are orthonormal, the RMSIP would be exactly 1 if the summation was extended to the entire set of modes. Typically, this measure has been applied with $n = 10$, following Amadei et al. [67]. As the RMSIP does not represent the energetic separation between modes in the sets, measures that incorporate eigenvalues as well have been proposed. Hess [68] defined an overlap function, OV:

$$\text{OV}(\mathbf{A}, \mathbf{B}) = 1 - \left(\frac{\text{tr}\left(\left(\mathbf{A}^{\frac{1}{2}} - \mathbf{B}^{\frac{1}{2}}\right)^2\right)}{\text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})}\right)^{\frac{1}{2}} \quad (13)$$

where \mathbf{A} and \mathbf{B} are covariance matrices. $\mathbf{A}^{1/2}$ is the matrix that decomposes into the same orthonormal eigenvectors but with eigenvalues that are the square root of those in \mathbf{A} .

Here the normalisation is realised by the sum of matrix traces in the denominator. The trace of a covariance matrix is equal to the total variation in the system. We also used trace normalising covariance matrices [66] and applied our similarity measure to normalised matrices $\tilde{\mathbf{C}}$:

$$\tilde{\mathbf{C}} = \frac{1}{\text{tr}(\mathbf{C})} \mathbf{C} \quad (14)$$

Applying the OV to these normalised matrices reveals the similarity with the RMSIP:

$$\text{OV}(\tilde{\mathbf{A}}, \tilde{\mathbf{B}}) = 1 - \left(1 - \sum_{m=7}^{3N} \sum_{l=7}^{3N} (\kappa_m \mu_l)^{\frac{1}{2}} (\mathbf{v}_m^T \mathbf{w}_l)^2\right)^{\frac{1}{2}} \quad (15)$$

where $\tilde{\mathbf{A}}$ is decomposed into eigenvectors \mathbf{v}_m and eigenvalues κ_m , $\tilde{\mathbf{B}}$ into eigenvectors \mathbf{w}_l and μ_l and the sums run over the nontrivial modes. A similar measure was also proposed by Carnevale et al. [69].

Considering the problem of comparing the intrinsic deformations in proteins as a matter of comparing their Boltzmann distribution, the field of multivariate statistics provides many measures of distance or similarity. Of those, the Bhattacharyya coefficient and the closely related Bhattacharyya distance have been adapted for comparing internal deformations of ENMs [56,66]. The Bhattacharyya coefficient, BC is defined as:

$$\text{BC}(p_a, p_b) = \int p_a(\mathbf{r})^{\frac{1}{2}} p_b(\mathbf{r})^{\frac{1}{2}} d\mathbf{r} \quad (16)$$

where p_a and p_b denote probability density functions (PDFs) for a multivariate random variable. For comparing internal deformations of proteins, the distributions can be taken to be mean-centred, and for ENMs, the PDFs will be Gaussian with the covariance matrix specified in Eq. (6). For mean-centred Gaussian distributions with trace normalised covariance matrices, BC has the closed form:

$$BC(p_a, p_b) = \frac{|\tilde{\mathbf{A}}|^{1/2} |\tilde{\mathbf{B}}|^{1/2}}{|\frac{1}{2}(\tilde{\mathbf{A}} + \tilde{\mathbf{B}})|^{1/2}} \quad (17)$$

where \mathbf{A} is the covariance matrix of p_a and \mathbf{B} is the covariance matrix of p_b , and vertical bars denote the matrix determinant. However, the measure is only defined for positive-definite covariance matrices, and an approximation to $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$ has to be made due to the presence of the trivial modes. This has been solved by projecting the matrices to a lower dimensional subspace chosen from the eigendecomposition of the mean matrix in the denominator [66].

The information used by the different measures is illustrated in Fig. 4. These simple bivariate distributions can be taken to represent deformations of a molecule represented by two coordinates. The SIP ignores any directionality of motion, and simply adds up the total variance of each variable. The RMSIP considers the agreement in direction for all pairs of eigenvectors that rank amongst the n highest principal components (corresponding to lower energy normal modes). For this simplified example only an RMSIP with $n = 1$ can be considered, which amounts to only comparing the principal component with the maximal variation of each distribution. Note that the comparisons

with the other principal components of either distribution are represented with dotted lines in Fig. 4. The OV compares all pairs of eigenvectors, but factors in the variance along each direction. This is illustrated by vectors with lengths proportional to the standard deviation along the principal directions, and can be contrasted with the normalised vectors considered by the RMSIP. The BC quantifies the similarity of the PDFs, which is here illustrated by the overlapping region of the two elliptical distributions.

4.2. Structural alignment

When the intrinsic motions of non-identical structures are compared, it is necessary to first obtain a description of the parts of the different structures that are to be compared with each other. For example, a structural alignment can describe the amino acid residues that are in structural correspondence to each other, between two or more structures. Comparing distant homologues provides a challenge in defining which parts of the proteins to compare. This is commonly solved by structural alignment, which is a challenging problem, particularly for the simultaneous alignment of sets of proteins [70–73].

When considering an alignment for comparing multiple structures, sequence identity and volumetric differences tend to pose a big challenge for finding equivalent atomic coordinates between them. The optimum solution between two structures, let alone many, tends to scale with sequence length and variability. Moreover, there is also the question of the most reasonable way of assessing a resulting solution, even though root mean squared deviation (RMSD) is generally accepted as the standard across different tools within the field. Due to the dynamic

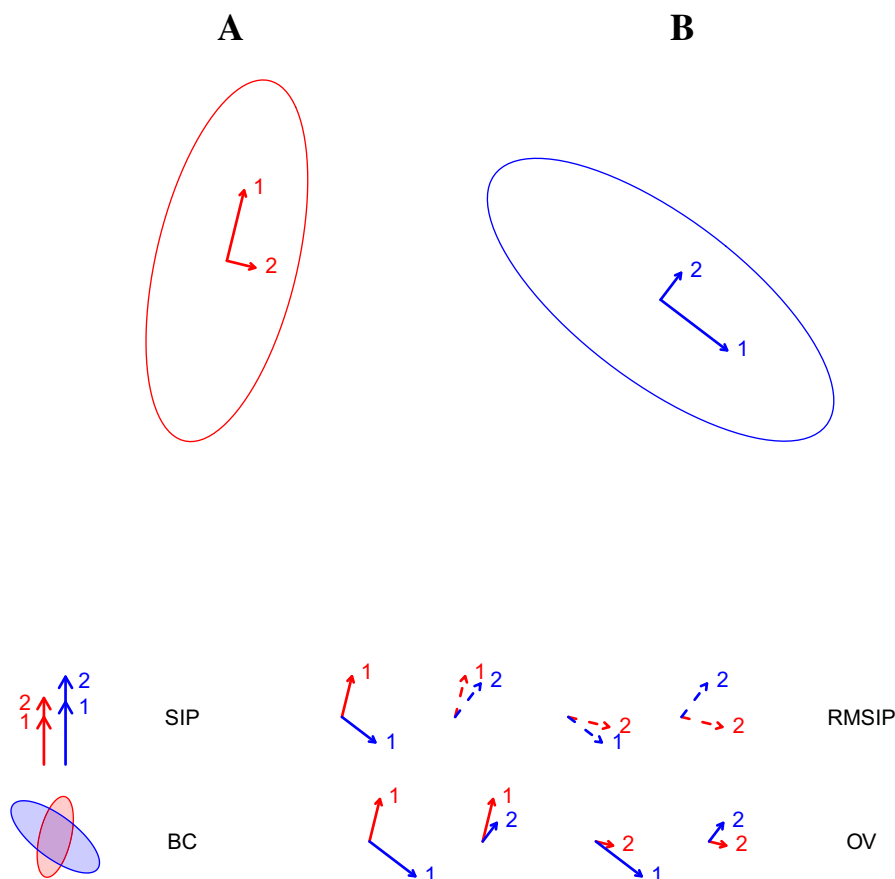


Fig. 4. Graphical interpretation of the different similarity measures. The figure illustrates what information is utilised by different measures of similarity. Top: two bivariate distributions representative of simplified molecular deformation with principal components drawn as red and blue vectors respectively. The principal components are displayed with length proportional to the standard deviation along the principal directions. The illustrations of different measures for comparing this molecular deformation is depicted by the abbreviation for the measures as defined in the text. (SIP): comparison of magnitudes only. (RMSIP): comparison of direction of only the n most principal component(s), all vectors are normalised to equal length. (OV): comparison of direction and magnitude. (BC): comparison of distribution overlap.

nature of structures, many alignment solutions involve a component of flexibility to achieve a better fit between structures. Yet these solutions are mostly available for pair-wise alignments; A thorough discussion about defining comparable regions of a protein and on some strategies for aligning pairs of proteins using models of their intrinsic flexibility is described in a review of Micheletti [30].

Most multiple structural alignment methods involve computing all pair-wise alignments between a set of structures, before producing a consensus between all of them [72]. The differences between multiple structure alignment programs involve the choice of geometric reference points (such as secondary structure or C α atoms), algorithm for aligning them in a pairwise fashion or identifying a consensus core alignment to optimise iteratively, and the scoring method used to assess the solution. Problems unique to multiple structure alignments involve the length of consensus alignment between multiple structures, and pairwise RMSDs within the set [74]. We find that in order to compare structures effectively, it is essential to have a robust alignment that is able to take into account natural and yet unique variations within a set of proteins.

Previously, we have found that the results of comparative analysis are sensitive to the quality of the alignment [66], especially if the set contains structures that are related at the SCOP family and superfamily levels [75].

To illustrate this, we constructed multiple structural alignments for a large set of proteins with triosephosphate isomerase-like domains (SCOP c.1, TIM alpha/beta barrel dataset used in Fuglebakk et al. [66]) using two popular programs, STAMP [73] and MUSTANG [76]. The set consists of structures from two different superfamilies, which subdivide into two families each. The triosephosphate isomerase possesses a fold that is tricky to align, as it is completely symmetrical in its enclosed barrel-like configuration that consists of 8 beta strands and 8 alpha helices. As such, it is a challenge for even the most sophisticated algorithms to align, especially when the sequence identity is low, due to the abundance and diversity of this fold [77]. Visual inspection of the superimpositions provided by STAMP shows that it is heavily biased towards the N-terminus, where the alignment is optimised, losing symmetry in the points of common reference towards the C-terminus

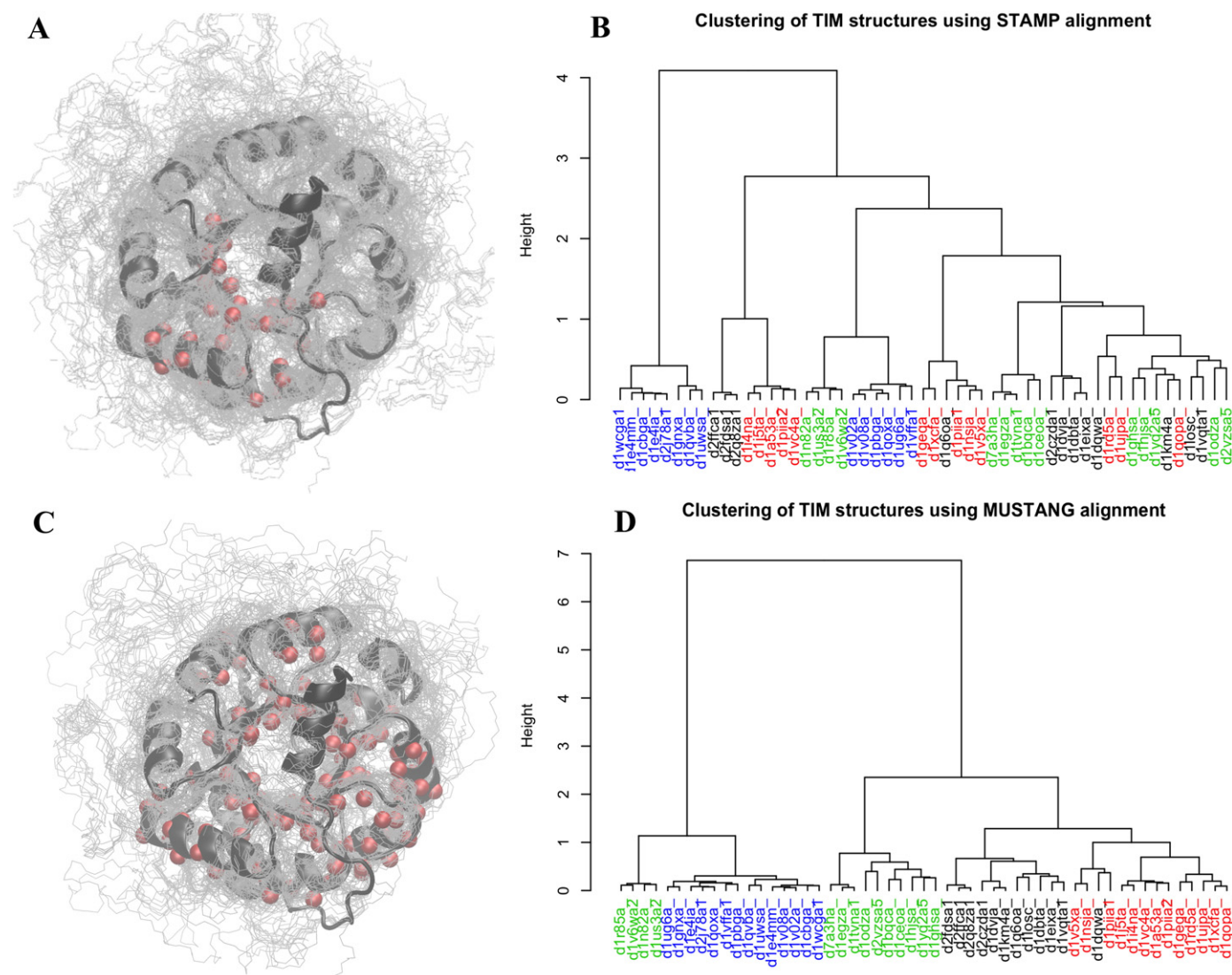


Fig. 5. Influence of the alignment methods of 53 structures with the TIM-barrel fold on similarity measures. (A): STAMP and (C): MUSTANG alignments of 53 structures with the TIM-barrel fold. The light grey lines show the superimposition of the structures, while the dark grey cartoon of the secondary structure shows one of them as a representative. The red spheres highlight the points on the structure that are conserved throughout the alignment with a bias towards the N-terminus. K-means clustering (with $k = 4$) of the Bhattacharya score analysis comparing the covariances using the (B): STAMP and (D): MUSTANG alignments. The SCOP identifiers on the plots are coloured signify structures that are from the same SCOP family, and are grouped such that red and black are from one SCOP superfamily, and blue and green are from the other. We see a heterogenous clustering of the structures across superfamilies and families with STAMP while we see good groupings with respect to the family level, but a less distinct separation at the superfamily level with respect to the green group.

(Fig. 5A and C). On the other hand, the MUSTANG alignment is able to provide a superimposition that is less sparse, with regard to these points in common, distributed all over the structures (Fig. 5B and D).

STAMP [73] relies on an algorithm that assesses pairwise alignments within a set of proteins and extends it in a progressive manner with the aid of hierarchical clustering. In this process, there is a chance that an error made in introducing a gap is propagated to the final output. MUSTANG [76] employs a similar algorithm that differs in the way that it refines the alignment once the pairwise alignments are completed by introducing an intermediate step where the residue position equivalences are assessed globally (in the context of the other proteins in the set).

The effect of the difference in quality between these alignments propagates when we try to cluster these structures to their family and superfamily levels based on their intrinsic dynamics, using the Bhattacharya coefficient score (Eq. (17)) [66]. The difference is striking: we see some separation between the structures at the family and superfamily levels which is not totally identical to the SCOP annotation, but performs better than the mixed clustering that we see resulting from STAMP. Since we find a large overlap between structural similarity and dynamics in most cases, we find that using an appropriate tool to align multiple structures is important in having reliable results when comparing their intrinsic dynamics.

4.3. Comparing only the conserved regions

Once the issue of determining the corresponding parts of each protein within a set is resolved, we can proceed to define a subset or a core between all of these proteins that can be compared dynamically. In such a procedure, each protein can be partitioned into a subset of core atoms A, and a subset of excluded peripheral atoms B. While only A has a corresponding part in all the structures compared, B is still linked to the dynamics of A and should be retained in the calculation of motion to preserve its influence. This needs to be observed when measures are normalised. In the case of comparing normal modes or covariance matrices, lower dimensional matrices describing the motion of only A needs to be obtained. Many deformations of the proteins can be consistent with A acting as a rigid body, while B deforms internally. Since B is defined to not be comparable between structures, it is desirable to express the internal deformations of A in a way that is consistent with how it deforms in the context provided by B. Mathematically, the problem is manifested by the fact that the parts of the eigenvectors corresponding to only A are not generally orthogonal. One common way to deal with this problem is to define a potential for A, which is restrained by the presence of B [52,69]. Assuming that B deforms along the direction of minimal energy, such a restrained potential can be obtained by differentiating Eq. (3) with respect to deformations of B. Substituting these minimal energy deformations of B back into Eq. (3) gives the Hessian of the constrained potential:

$$\mathbf{H}^{\text{con}} = \mathbf{H}_{aa} - \mathbf{H}_{ab} \mathbf{H}_{bb}^{-1} \mathbf{H}_{ab}^T \quad (18)$$

where the Hessian of the full potential is partitioned so that \mathbf{H}_{aa} reflects interactions in A, \mathbf{H}_{ab} reflects interactions between A and B, and \mathbf{H}_{bb} reflects interactions internal to B. This method was originally introduced for ENMs, but now has also been extended to all-atom and hybrid quantum mechanics/molecular mechanics potentials, for its recognised potential as an analysis method [78].

When normal modes or covariance matrices are expressed in rotational variant coordinate systems, like the formalism in Cartesian coordinates described above, it is necessary to ensure that they are expressed in identical or similar rotations. When comparing models of identical proteins, this can be solved exactly. In cases where different proteins are compared, an approximation to a common rotation is typically obtained by rotating to minimise the sum of squared distances between aligned residues. For comparing proteins with very different

equilibrium structures, the validity of such an approximation to a common reference frame might in principle become a concern. Possible solutions include considering internal coordinates [45] or comparing rotationally invariant properties of the normal modes, like the correlation matrix in Eq. (8).

5. Strategies and applications of comparative analysis

Comparing multiple structures is a natural extension to the study of intrinsic dynamics. In the case of ENMs, where it is computationally efficient and inexpensive, it has been seen as a logical choice for analysis of large sets of structures in many ways. The mode vectors produced from ENMs are informative on their own, and provide a good qualitative description of a protein's inherent flexibility. Luo and Bruice did one example of such a qualitative analysis, where they conducted a visual inspection of the normal mode vectors on six structures along the dihydrofolate reductase reaction mechanism [79]. With this analysis, they focussed on regions (such as the M20 loop and subdomain rotations) that were seen to undergo conformational changes during catalysis, and found that they were consistent with the principal motions from the low-energy modes, and were able to further validate these findings against NMR, kinetic and molecular dynamics studies found in literature.

In general, we find that such analyses benefit from quantification of dynamical properties, and here, we outline a selected list of examples where ENMs have been used to compare the dynamics of multiple structures, whether from the same sequence or not, related by homology or fold. Some of the early examples discussed have been used to validate the ENM as a viable model for intrinsic dynamics analysis, yet the strategies employed are applicable to most comparative applications. In most of the examples, atomic fluctuation profiles (Eq. (7)) were compared with crystallographic B-factors, using measures like that in Eq. (11). The use of the overlap analysis (such as the squared overlap, Eq. (9)) was also very common. In addition, we find that comparing covariance/correlation matrices, using similarity measures (such as the RMSIP, Eq. (12)) and perturbation response methods [80] are also useful techniques when comparing dynamics.

5.1. Comparing multiple structures of the same protein sequence: conformational changes

The squared overlap between modes and the structural difference from one conformation to another has been introduced as a way to understand the transformation between two states of an enzyme (Eq. (9)). This analysis allows for the identification of modes that contribute to the conformational change seen [15,17,65]. Traditional dynamics studies that compare two extreme states, e.g. fully open ligand-free conformations vs. fully closed ligand-bound conformations, lead to interpretation that the modes with high overlap with the difference between conformations are the ones important for conformational change. These modes tend to be interpreted as the transition path between active and inactive states. In general, mapping conformation transition paths is a much more complex affair that requires more detailed and rigorous calculations than the overlap analysis to estimate [81–83].

Even so, the method has been useful in understanding the changes in flexibility between states. Extending pair-wise comparisons can come in the form of performing serialised overlaps between multiple pairs of structures of different conformational intermediates, or a large-scale survey of conformational transitions. An example of a large-scale survey of transitions is the work from Stein and colleagues who performed serial overlap analyses of unbound to bound conformational changes (multiple pairs in some cases, where more than one ligand-bound conformation was available for an enzyme) from a total of over 12,000 structures [84]. This was one of the analyses used to assess the cost of conformational change upon ligand binding, and whether they fit the lock-key, induced-fit or conformational selection binding models.

ENMs can be used to produce covariance matrices (Eq. (6)) that can be compared between them to understand the difference in dynamics for different states. The work of Seckler et al. [19] is one such example, where the authors use ENMs in addition to structural comparison; the authors retrieved 52 structures of HIV-1 reverse transcriptase, and compared them to reference structures using a measure of dynamics similarity called the covariance complement (a form of the OV measure, Eq. (13)). The structures differed in state such that some were ligand-free while others had DNA, RNA, adenosine triphosphate (ATP) and various inhibitors bound. Further they found linear variation of RMSD with the covariance complement to be a signature of functional state, and showed that the ratio between the two measures can be used to cluster these 52 structures into three main levels. These levels corresponded to their level of activity based on the ligand-types. This is an example where dynamics is used to distinguish between the effects of ligands.

Allosteric effects of ligands and their ability to cause changes in flexibility have often been explored using ENMs [85–87]. While the allosteric effect is often thought of as a large conformational change in a structure, Rodgers et al. explored the hypothesis that the low-frequency normal modes are able to propagate allosteric signals without causing a large conformational change in a family of transcription factors called CRP/FNR [20]. They constructed ENMs for structures of the Catabolite Activator Protein (CAP) from *Escherichia coli* representing unliganded, single and double-liganded forms and introduced mutations outside the substrate-binding pocket by varying the spring constants of all springs attached to a single residue. This was a strategy for probing changes to the free energy of substrate binding. They found that the regions that experienced the greatest change in cooperativity were not necessarily adjacent to the substrate-binding site. They used their method to predict residues involved in allosteric signalling in CAP and validated their findings on a homologue GlxR, through a combination of ENM, MD and experimental results.

The large increase in the number of X-ray structures has led to the opportunity of analysing preferred conformations with multivariate statistical analyses such as principal component analysis (PCA). PCA has the advantage of reducing the dimensionality of large-scale data into basis vectors (or principal components), ordered based on how much of the structural variance they describe. The technique has been used for characterising principal directions of fluctuation in Molecular Dynamics trajectories, as first introduced by García [88]. However, information about the conformational landscape of proteins can also be obtained by comparing structures obtained under varying experimental condition, or under small variations in sequence. Datasets of related structures can then be decomposed through PCA or similar techniques [31,32]. The principal components in this case describe the direction of the structural variance in the dataset (due to experimental conditions or evolutionary relationships rather than the thermal forces in ENM or MD calculations). One strategy employing PCA involves calculating principal components for a large collection of structures for a given protein, and comparing the resulting principal components to modes obtained from ENM calculations on representative structures. Yang and colleagues have shown that normal modes can directly be compared with principal components extracted from a large set of structures from HIV-1 protease, providing a direct comparison of calculated values to experimental data [14]. Both Bakan and Bahar [89] and Katebi et al. [90] applied this strategy to different enzymes; structures of HIV-1 reverse transcriptase, p38 MAP kinase and cyclin-dependent kinase 2, with and without inhibitors, were analysed by Bakan and Bahar and all available structures of triosephosphate isomerase (TIM) by Katebi et al. They were able to relate the variation in the structural space to intrinsic dynamics and further to function.

5.2. Comparing dynamics between different oligomeric/multimeric states

The variations in the dynamics of monomers often translate to changes in global structure, whether they are perturbed by mutations

or ligands. As a recommendation, when comparing oligomeric structures, one should be careful of the effect of calculating the modes of proteins in different oligomeric states before drawing conclusion on the dynamics of the system or making a one-to-one comparison. Eigenvectors calculated on entire structures with different oligomeric states are not necessarily comparable, due to the pronounced effect of complexation of the subunits on the construction of its ENM. As a result, the lowest modes are different from one oligomeric state to the other, where subunit–subunit motions make up the lowest energy modes in multimeric assemblies. Since monomers usually possess differences in conformation based on their oligomeric state, comparing monomers extracted from different states provides sufficient basis for observing changes in their dynamics [25]. Zen et al. [91] compared the flexibility of interfacial residues in obligate and non-obligate dimers. They found that the mobility of amino acids experiences non-uniform changes depending on the type of interface, size and obligate nature of the complex. They were able to gain this insight with the use of root mean squared fluctuations (similar to Eq. (7)) of the bound and *in silico* free forms of the monomers, where the bound form was modelled with Eq. (18) when comparing the states.

Another notable example of working with multiple subunits is in the study of Alzheimer's A β (1–40) amyloid fibrils, where Xu and colleagues constructed models of these protein assemblies based on two forms of naturally occurring symmetries and varying lengths of the fibrils [92]. They were able to characterise the effects of the fibril size on the overall flexibility of these large structures, and changes to the low-energy motions. Similarly, Polles et al. explored the flexibility in different assemblies of a heterogeneous set of virus capsids based on fluctuations-based analyses and domain decomposition [93]. Other studies dedicated to the comparison of dynamics to changes in oligomeric state have been reported for monomeric, dimeric and tetrameric states of GPCRs [94], dimeric and hexameric (trimer-of-dimer) states of the serine receptor Tsr [95], and monomeric and dimeric states of the p53 protein [96].

5.3. Comparing dynamics between more distantly related proteins

Structure comparison has been long established as a means to understanding the evolution of proteins, as has the conservation between sequence and structure since the work of Lesk and Chothia [97]. In many cases, even from visual inspection, a structure is not just seen to encode dynamic information, but also a historical time-point in the evolution of the family to which it belongs. Some sequence mutations, insertions and deletions can be accommodated by the plastic deformations of a common architectural core and retain the precise geometry of the active site, even if peripheral regions or accessory domains vary [71]. Hence it has been of great interest in linking observable structural and sequence evolution in the conservation of dynamics, especially in a family of proteins.

The observation that low-energy normal modes so frequently appear in functional motion [18,33] has motivated investigations into their evolutionary conservation across protein families. The globins have served as a good example of a well conserved yet diverged group of proteins and has been a subject of quantitative comparisons by Maguid and colleagues [24]. They developed a method to quantify similarities between the collective modes using a singular value decomposition approach to find representative vectors to describe the dynamics of an aligned core of structures. This work laid the foundation for exploring the evolutionary conservation of dynamics in the lowest energy modes, which they further developed and validated across larger standard datasets from the family to superfamily levels [98,99]. While this work clearly confirms that low-energy normal modes are conserved between structurally conserved proteins, it is interesting to note that this conservation can be explained as the structural response to random perturbations, rather than necessarily selective pressure on certain kinds of motion [35,100,101].

In 2006, Carnevale et al. [69,102] introduced the idea of performing pair-wise alignments and comparing dynamics for the regions covered by giving each pair an overall score such as RMSIP (Eq. (11)), which they used on pairs of proteases with very low structural similarity and sequence identity. Partial pair-wise alignments between the proteases allowed them to conclude that often the dynamical conservation far exceeded the structural conservation.

Others have explored the idea that the effect of sequence changes throughout the evolution of a protein structure would be along its principal modes. Leo-Macias et al. [31,32] performed large-scale deformation analysis on the multiple alignments of the cores of 35 protein families and compared them against their evolutionary deformations gained from PCAs of the corresponding structures via the RMSIP score. They were able to relate regions with the greater evolutionary variability with regions that experience larger thermal fluctuations. This work was further validated with comparisons against MD simulations [27] and further reinforced by work done on the Ras GTPase superfamily [26].

In a more recent study combining allostery and evolutionary conservation, Kolan et al. report the role of the lowest energy modes on the mechanical motions and conformational changes of six members of the GPCR family [21]. When comparing between the ligand-bound and ligand-free states, they found that the slowest modes are well suited to describe components of the activation mechanism. They compared the overlap of their slowest modes by calculating the normalised mean squared displacements of the aligned C α atoms as a correlation score and showed that all the GPCR members except rhodopsin agree well, and concluded that rhodopsin was not representative of all GPCRs in its motions. They concluded that the ENM calculations were able to capture the long-range mechanism of GPCR activation, where binding in the extracellular domain can cause a conformational change in the cytoplasmic domain.

Another example of linking the conservation of certain structural/sequence motifs to function is displayed by the work of Lukman and Grant [23]. They surveyed maltose transporters and characterised a network of residues that have an influence on the overall dynamics of the proteins in different conformational states. This work is an example of analysis inspired by the developments in the perturbation-response analysis by Zheng et al. [80], who had earlier studied the conservation of dynamics in distantly-related motor proteins by comparing the conformational changes experienced by myosin, F1-F0 ATPase and kinesin [28]. In this case, the authors concluded that while the large conformational change seen in myosin and F1-F0 ATPase was consistent with motions that can be described as a power-stroke type movement, while the kinesin followed a Brownian ratchet-type mechanism. The perturbation-response method has gained greater traction as seen in the efforts to develop a useful metric [103,104] to describe a single residue's response to an applied force in a given position, as a predictive tool.

Studies like this have also prompted initiatives to categorise protein structures dynamically [105]. Further databases storing results from normal mode analysis using ENMs on large number of structures have been built, such as ProMode Elastic [46] or MolmovDB (Database of Macromolecular Movement) [106]. These show the interest of the community and the potential of ENMs for the characterisation of intrinsic dynamics in a way that can complement existing structural classifications systems. In addition, there have been efforts in using dynamic information as a means of aligning different proteins, and their developments have provided insight into comparing dynamics in general [69,107–109].

5.4. Comparing dynamics of proteins with different folds

In the paradigm where the conservation of dynamics is due to structural similarity and not vice versa, the comparison of dynamics based on shape and fold, independent of sequence similarity or conservation, has also been a topic of great interest [34,110]. Since proper folding of the

protein is a requirement for function in many cases, it is natural to seek to understand how the fold affects function, and their principal modes of motion is an important ingredient in understanding functional properties of the fold.

In a systematic survey of functional dynamics in enzymes with low structural similarity, Zen et al. [109] identified structural correspondences between proteins with loose analogies in secondary and tertiary structure. In several of these structurally diverse enzymes they identify similarities in protein dynamics that can be linked to the protein function. The work was made possible through the development of a scheme for dynamical alignment, which was later applied to the study of a range of studies and compared with results from other such approaches [30,111].

Hollup et al. showed that computer-generated models based on ideal structures, stripped of influences of sequence conservation and evolutionary links, could be used reliably in the analysis of dynamics [29]. They showed that the spatial arrangement of secondary structures in a protein is an important component of the low-energy modes, while the loops connecting these elements play a minor role. Another study characterised the motions of two proteins with cylindrical symmetry, the beta-barrel Dronpa and the toroidal DNA-clamp, as ideal structures and compared them both qualitatively to find similarities in their global motions [112].

6. Computational tools and frameworks

The simplicity of the ENMs makes them relatively easy to implement if routines for the necessary linear algebra are provided. This makes it easy to integrate ENMs with other kinds of structural analysis. Most model developers also make implementations available online or upon request. In addition the interested user can choose from a range of tools and frameworks available for computing and analysing ENMs. The Molecular Modelling Toolkit [113] and ProDy [114] are libraries for the programming language Python that support normal mode decomposition, analysis and visualisation of ENMs. For the statistical computing software R, ENMs are integrated into packages for analysing molecular dynamics data like LOOS [115] and Bio3D [116]. $\Delta\Delta$ P [117] is a collection of scripts for ENM and principal component analysis that allows the application of a range of ENMs without requiring familiarity with programming. A range of web servers are also available such as WEBnma [110], EINemo [118], ANM webserver [119], KOSMOS [120], NMSim [121], and NOMAD-Ref [122]. These provide a variety of analysis on ENM normal modes, typically making the analysis accessible for an audience less experienced with computational analysis. Another initiative that aims to take ENMs to a wider audience is the software Maven, provided as a standalone application for analysis and visualisation of ENMs [123].

Of the web servers, WEBnma is the only one currently supporting comparative analysis with the use of structural alignment information along with the submitted structures. It also provides an easy access to the Bhattacharyya coefficient score. ProDy allows for the comparison of sequence evolution data (with the implementation of a co-evolutionary analysis tool Evol) and intrinsic dynamic information from ENMs [124]. The latest releases of Bio3D provide an implementation that provides the framework for “automated ensemble analysis methods”, which includes multiple sequence alignments and a selection of similarity measures and correlations analysis. Amongst the algorithms proposed for dynamics-based alignment of proteins, the approach of Zen et al. [111] has also been made available as a web-server [125].

7. Conclusion/perspectives

The use of ENMs for comparative analysis of protein dynamics has led to greater understanding of the conservation of dynamics across structures with different conformations and within a protein family. Moreover, there have been more and more evidence that comparing

dynamics is a viable way for gaining greater understanding of the mechanisms employed by proteins for their function. Efforts have been made lately to evaluate the effect of the choice of similarity measures, the ENM parameter and the structural alignments. The results of these studies, summarised in this manuscript will be useful for users getting started on comparing the dynamics of proteins in a wide variety of settings. We believe that it is important that users are aware of the potential impact of choices in the different steps of their computational strategies when performing comparative analysis of protein dynamics, and advocate for detailed reporting of the latter in scientific publications.

References

- [1] P.J. Fleming, F.M. Richards, Protein packing: dependence on protein size, secondary structure and amino acid composition, *J. Mol. Biol.* 299 (2000) 487–498.
- [2] B. Halle, Flexibility and packing in proteins, *Proc. Natl. Acad. Sci. U. S. A.* 99 (2002) 1274–1279.
- [3] M.M. Tirion, Large amplitude elastic motions in proteins from a single-parameter, atomic analysis, *Phys. Rev. Lett.* 77 (1996) 1905–1908.
- [4] K. Hinsen, Analysis of domain motions by approximate normal mode calculations, *Proteins* 33 (1998) 417–429.
- [5] A.R. Atilgan, S.R. Durell, R.L. Jernigan, M.C. Demirel, O. Keskin, I. Bahar, Anisotropy of fluctuation dynamics of proteins with an elastic network model, *Biophys. J.* 80 (2001) 505–515.
- [6] B. Brooks, M. Karplus, Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme, *Proc. Natl. Acad. Sci. U. S. A.* 82 (1985) 4995–4999.
- [7] N. Go, T. Noguti, T. Nishikawa, Dynamics of a small globular protein in terms of low-frequency vibrational modes, *Proc. Natl. Acad. Sci. U. S. A.* 80 (1983) 3696–3700.
- [8] M. Levitt, C. Sander, P.S. Stern, Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme, *J. Mol. Biol.* 181 (1985).
- [9] T. Noguti, N. Go, Collective variable description of small-amplitude conformational fluctuations in a globular protein, *Nature* 296 (1982) 776–778.
- [10] B.R. Brooks, M. Karplus, Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor, *Proc. Natl. Acad. Sci. U. S. A.* 80 (1983) 6571–6575.
- [11] F. Pontiggia, G. Colombo, C. Micheletti, H. Orland, Anharmonicity and self-similarity of the free energy landscape of protein G, *Phys. Rev. Lett.* 98 (2007) 048102.
- [12] M. Rueda, P. Chacón, M. Orozco, Thorough validation of protein normal mode analysis: a comparative study with essential dynamics, *Structure* 15 (2007) 565–575.
- [13] L. Skjaerven, A. Martinez, N. Reuter, Principal component and normal mode analysis of proteins; a quantitative comparison using the GroEL subunit, *Proteins* 79 (2011) 232–243.
- [14] L. Yang, G. Song, A. Carriquiry, R.L. Jernigan, Close correspondence between the motions from principal component analysis of multiple HIV-1 protease structures and elastic network modes, *Structure* 16 (2008) 321–330.
- [15] O. Marques, Y.H. Sanejouand, Hinge-bending motion in citrate synthase arising from normal mode calculations, *Proteins* 23 (1995) 557–560.
- [16] K. Hinsen, A. Thomas, M.J. Field, Analysis of domain motions in large proteins, *Proteins* 34 (1999) 369–382.
- [17] F. Tama, Y.H. Sanejouand, Conformational change of proteins arising from normal mode calculations, *Protein Eng.* 14 (2001) 1–6.
- [18] W.G. Krebs, V. Alexandrov, C.A. Wilson, N. Echols, H. Yu, M. Gerstein, Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic, *Proteins* 48 (2002) 682–695.
- [19] J.M. Seckler, N. Leioatts, H. Miao, A. Grossfield, The interplay of structure and dynamics: insights from a survey of HIV-1 reverse transcriptase crystal structures, *Proteins* 81 (2013) 1792–1801.
- [20] T.L. Rodgers, P.D. Townsend, D. Burnell, M.L. Jones, S.A. Richards, T.C. McLeish, E. Pohl, M.R. Wilson, M.J. Cann, Modulation of global low-frequency motions underlies allosteric regulation: demonstration in CRP/FNR family transcription factors, *PLoS Biol.* 11 (2013) e1001651.
- [21] D. Kolan, G. Fonar, A.O. Samson, Elastic network normal mode dynamics reveal the GPCR activation mechanism, *Proteins* 82 (2014) 579–586.
- [22] M. Laberge, T. Yonetani, Common dynamics of globin family proteins, *IUBMB life* 59 (2007) 528–534.
- [23] S. Lukman, G.H. Grant, A network of dynamically conserved residues deciphers the motions of maltose transporter, *Proteins* 76 (2009) 588–597.
- [24] S. Maguid, S. Fernández-Alberti, J. Echave, Exploring the common dynamics of homologous proteins. Application to the globin family, *Biophys. J.* 89 (2005) 3–13.
- [25] E. Marcos, R. Crehuet, I. Bahar, Changes in dynamics upon oligomerization regulate substrate binding and allostery in amino acid kinase family members, *PLoS Comput. Biol.* 7 (2011) e1002201.
- [26] F. Raimondi, M. Orozco, F. Fanelli, Deciphering the deformation modes associated with function retention and specialization in members of the Ras superfamily, *Structure* 18 (2010) 402–414.
- [27] J.A. Velazquez-Muriel, M. Rueda, I. Cuesta, A. Pascual-Montano, M. Orozco, Comparison of molecular dynamics and superfamily spaces of protein domain deformation, *BMC Struct. Biol.* 9 (2009) 6.
- [28] W. Zheng, S. Doniach, A comparative study of motor-protein motions by using a simple elastic-network model, *Proc. Natl. Acad. Sci. U. S. A.* 100 (2003) 13253–13258.
- [29] S.M. Hollup, E. Fuglebakk, W.R. Taylor, N. Reuter, Exploring the factors determining the dynamics of different protein folds, *Protein Sci.* 20 (2011) 197–209.
- [30] C. Micheletti, Comparing proteins by their internal dynamics: exploring structure–function relationships beyond static structural alignments, *Phys. Life Rev.* (2013) 10.
- [31] A. Leo-Macias, P. Lopez-Romero, D. Lupyán, D.I. Zerbino, A.R. Ortiz, Core deformations in protein families: a physical perspective, *Biophys. Chem.* 115 (2005) 125–128.
- [32] A. Leo-Macias, P. Lopez-Romero, D. Lupyán, D. Zerbino, A.R. Ortiz, An analysis of core deformations in protein superfamilies, *Biophys. J.* 88 (2005) 1291–1299.
- [33] S. Nicolay, Y.-H. Sanejouand, Functional modes of proteins are among the most robust, *Phys. Rev. Lett.* 96 (2006) 4.
- [34] F. Tama, C.L. Brooks, Symmetry, form, and shape: guiding principles for robustness in macromolecular machines, *Annu. Rev. Biophys. Biomol. Struct.* 35 (2006) 115–133.
- [35] J. Echave, F.M. Fernández, A perturbative view of protein structural variation, *Proteins* 78 (2010) 173–180.
- [36] W. Zheng, B.R. Brooks, D. Thirumalai, Allosteric transitions in the chaperonin GroEL are captured by a dominant normal mode that is most robust to sequence variations, *Biophys. J.* 93 (2007) 2289–2299.
- [37] W. Zheng, B.R. Brooks, D. Thirumalai, Low-frequency normal modes that describe allosteric transitions in biological nanomachines are robust to sequence variations, *Proc. Natl. Acad. Sci. U. S. A.* 103 (2006) 7664–7669.
- [38] I. Bahar, A.R. Atilgan, B. Erman, Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential, *Fold. Des.* 2 (1997) 173–181.
- [39] K. Moritsugu, J.C. Smith, Coarse-grained biomolecular simulation with REACH: realistic extension algorithm via covariance Hessian, *Biophys. J.* 93 (2007) 3460–3469.
- [40] K. Moritsugu, V. Kurkal-Siebert, J.C. Smith, REACH coarse-grained normal mode analysis of protein dimer interaction dynamics, *Biophys. J.* 97 (2009) 1158–1167.
- [41] J.I. Jeong, Y. Jang, M.K. Kim, A connection rule for alpha-carbon coarse-grained elastic network models using chemical bond information, *J. Mol. Graph. Model.* 24 (2006) 296–306.
- [42] M.H. Kim, S. Seo, J.I. Jeong, B.J. Kim, W.K. Liu, B.S. Lim, J.B. Choi, M.K. Kim, A mass weighted chemical elastic network model elucidates closed form domain motions in proteins, *Protein Sci.* 22 (2013) 605–613.
- [43] Y. Dehouck, A.S. Mikhailov, Effective harmonic potentials: insights into the internal cooperativity and sequence-specificity of protein dynamics, *PLoS Comput. Biol.* 9 (2013) e1003209.
- [44] C. Micheletti, P. Carloni, A. Maritan, Accurate and efficient description of protein vibrational dynamics: comparing molecular dynamics and Gaussian models, *Proteins* 55 (2004) 635–645.
- [45] R. Mendez, U. Bastolla, Torsional network model: normal modes in torsion angle space better correlate with conformational changes in proteins, *Phys. Rev. Lett.* 104 (2010) 228103–228104.
- [46] H. Wako, S. Endo, Ligand-induced conformational change of a protein reproduced by a linear combination of displacement vectors obtained from normal mode analysis, *Biophys. Chem.* 159 (2011) 257–266.
- [47] F. Tama, F.X. Gadea, O. Marques, Y.H. Sanejouand, Building-block approach for determining low-frequency normal modes of macromolecules, *Proteins* 41 (2000) 1–7.
- [48] K. Hinsen, G.R. Kneller, Solvent effects in the slow dynamics of proteins, *Proteins* 70 (2008) 1235–1242.
- [49] T. Ichiye, M. Karplus, Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations, *Proteins* 11 (1991) 205–217.
- [50] K. Hinsen, G.R. Kneller, Projection methods for the analysis of complex motions in macromolecules, *Mol. Simul.* 23 (2000) 275–292.
- [51] N. Leioatts, T.D. Romo, A. Grossfield, Elastic network models are robust to variations in formalism, *J. Chem. Theory Comput.* 8 (2012) 2424–2434.
- [52] K. Hinsen, A.J. Petrescu, S. Dellerue, M.C. Bellissent-Funel, G.R. Kneller, Harmonicity in slow protein dynamics, *Chem. Phys.* 261 (2000) 25–37.
- [53] D. Riccardi, Q. Cui, G.N. Phillips Jr., Evaluating elastic network models of crystalline biological molecules with temperature factors, correlated motions, and diffuse X-ray scattering, *Biophys. J.* 99 (2010) 2616–2625.
- [54] K. Hinsen, Structural flexibility in proteins: impact of the crystal environment, *Bioinformatics* 24 (2008) 521–528.
- [55] R. Soheilifard, D.E. Makarov, G.J. Rodin, Critical evaluation of simple network models of protein dynamics and their comparison with crystallographic B-factors, *Phys. Biol.* 5 (2008) 026008.
- [56] E. Fuglebakk, N. Reuter, K. Hinsen, Evaluation of protein elastic network models based on an analysis of collective motions, *J. Chem. Theory Comput.* 9 (2013) 5618–5628.
- [57] M.K. Kim, G.S. Chirikjian, R.L. Jernigan, Elastic models of conformational transitions in macromolecules, *J. Mol. Graph. Model.* 21 (2002) 151–160.
- [58] F. Tama, W. Wrighers, C.L. Brooks 3rd, Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory, *J. Mol. Biol.* 321 (2002) 297–305.
- [59] D. Ming, Y. Kong, M.A. Lambert, Z. Huang, J. Ma, How to describe protein motion without amino acid sequence and atomic coordinates, *Proc. Natl. Acad. Sci. U. S. A.* 99 (2002) 8620–8625.
- [60] S. Kundu, J.S. Melton, D.C. Sorensen, G.N. Phillips Jr., Dynamics of proteins in crystals: comparison of experiment with simple models, *Biophys. J.* 83 (2002) 723–732.
- [61] K. Moritsugu, J.C. Smith, REACH coarse-grained biomolecular simulation: transferability between different protein structural classes, *Biophys. J.* 95 (2008) 1639–1648.
- [62] M. Delarue, Y.H. Sanejouand, Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: the elastic network model, *J. Mol. Biol.* 320 (2002) 1011–1024.
- [63] H. Valadie, J.J. Lacapere, Y.H. Sanejouand, C. Etchebest, Dynamical properties of the MscL of *Escherichia coli*: a normal mode analysis, *J. Mol. Biol.* 332 (2003) 657–674.

- [64] F. Tama, M. Valle, J. Frank, C.L. Brooks, Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy, *Proc. Natl. Acad. Sci. U. S. A.* 100 (2003) 9319–9323.
- [65] N. Reuter, K. Hinsen, J.-J. Lacapere, Transconformations of the SERCA1 Ca-ATPase: a normal mode study, *Biophys. J.* 85 (2003) 99020.
- [66] E. Fuglebakk, J. Echave, N. Reuter, Measuring and comparing structural fluctuation patterns in large protein datasets, *Bioinformatics* 28 (2012) 2431–2440.
- [67] A. Amadei, M.A. Ceruso, A. Di Nola, On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations, *Proteins* 36 (1999) 419–424.
- [68] B. Hess, Convergence of sampling in protein simulations, *Phys. Rev. E* 65 (2002) 031910.
- [69] V. Carnevale, F. Pontiggia, C. Micheletti, Structural and dynamical alignment of enzymes with partial structural similarity, *J. Phys. Condens. Matter* 19 (2007) 285206–285214.
- [70] C. Berbalk, C.S. Schwaiger, P. Lackner, Accuracy analysis of multiple structure alignments, *Protein Sci.* 18 (2009) 2027–2035.
- [71] H. Hasegawa, L. Holm, Advances and pitfalls of protein structural alignment, *Curr. Opin. Struct. Biol.* 19 (2009) 341–348.
- [72] M.A. Marti-Renom, I.N. Shindyalov, P.E. Bourne, in: J.B. Gu, P.E. (Eds.), 2nd ed., *Struct. Bioinformatics* John Wiley and Sons, NJ, 2009, pp. 397–417.
- [73] R.B. Russell, G.J. Barton, Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels, *Proteins* 14 (1992) 309–323.
- [74] J. Ma, S. Wang, Algorithms, applications, and challenges of protein structure alignment, *Adv. Protein Chem. Struct. Biol.* 94 (2014) 121–175.
- [75] A.G. Murzin, S.E. Brenner, T. Hubbard, C. Chothia, SCOP: a structural classification of proteins database for the investigation of sequences and structures, *J. Mol. Biol.* 247 (1995) 536–540.
- [76] A.S. Konagurthi, J.C. Whisstock, P.J. Stuckey, A.M. Lesk, MUSTANG: a multiple structural alignment algorithm, *Proteins* 64 (2006) 559–574.
- [77] N. Nagano, C.A. Orengo, J.M. Thornton, One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions, *J. Mol. Biol.* 321 (2002) 741–765.
- [78] H.L. Woodcock, W. Zheng, A. Ghysels, Y. Shao, J. Kong, B.R. Brooks, Vibrational subsystem analysis: a method for probing free energies and correlations in the harmonic limit, *J. Chem. Phys.* 129 (2008) 214109.
- [79] J. Luo, T.C. Bruice, Envisioning the loop movements and rotation of the two subdomains of dihydrofolate reductase by elastic normal mode analysis, *J. Biomol. Struct. Dyn.* 27 (2009) 245–258.
- [80] W. Zheng, B.R. Brooks, S. Doniach, D. Thirumalai, Network of dynamically important residues in the open/closed transition in polymerases is strongly conserved, *Structure* 13 (2005) 565–577.
- [81] P. Maragakis, M. Karplus, Large amplitude conformational change in proteins explored with a plastic network model: adenylate kinase, *J. Mol. Biol.* 352 (2005) 807–822.
- [82] P.C. Whitford, J.N. Onuchic, P.G. Wolynes, Energy landscape along an enzymatic reaction trajectory: hinges or cracks? *HFSP J.* 2 (2008) 61–64.
- [83] Y. Togashi, T. Yanagida, A.S. Mikhailov, Nonlinearity of mechanochemical motions in motor proteins, *PLoS Comput. Biol.* 6 (2010) e1000814.
- [84] A. Stein, M. Rueda, A. Panjkovich, M. Orozco, P. Aloy, A systematic study of the energetics involved in structural changes upon association and connectivity in protein interaction networks, *Structure* 19 (2011) 881–889.
- [85] R. Tehver, J. Chen, D. Thirumalai, Allosteric wiring diagrams in the transitions that drive the GroEL reaction cycle, *J. Mol. Biol.* 387 (2009) 390–406.
- [86] H.N. Motlagh, J.O. Wrabl, J. Li, V.J. Hilser, The ensemble nature of allostery, *Nature* 508 (2014) 331–339.
- [87] S. Mittelmacht, I.N. Berezovsky, Coherent conformational degrees of freedom as a structural basis for allosteric communication, *PLoS Comput. Biol.* 7 (2011) e1002301.
- [88] A.E. Garcia, Large-amplitude nonlinear motions in proteins, *Phys. Rev. Lett.* 68 (1992) 2696–2699.
- [89] A. Bakan, I. Bahar, The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding, *Proc. Natl. Acad. Sci. U. S. A.* 106 (2009) 14349–14354.
- [90] A.R. Katebi, R.L. Jernigan, The critical role of the loops of triosephosphate isomerase for its oligomerization, dynamics, and functionality, *Protein Sci.* 23 (2014) 213–228.
- [91] A. Zen, C. Micheletti, O. Keskin, R. Nussinov, Comparing interfacial dynamics in protein-protein complexes: an elastic network approach, *BMC Struct. Biol.* 10 (2010) 26.
- [92] Z. Xu, R. Paparcone, M.J. Buehler, Alzheimer's amyloid fibrils feature size-dependent mechanical properties, *Biophys. J.* 98 (2010) 2053–2062.
- [93] G. Polles, G. Indelicato, R. Potestio, P. Cermelli, R. Twarock, C. Micheletti, Mechanical and assembly units of viral capsids identified via quasi-rigid domain decomposition, *PLoS Comput. Biol.* 9 (2013) e1003331.
- [94] M.Y. Niv, M. Filizola, Influence of oligomerization on the dynamics of G-protein coupled receptors as assessed by normal mode analysis, *Proteins* 71 (2008) 575–586.
- [95] B.A. Hall, J.P. Armitage, M.S. Sansom, Mechanism of bacterial signal transduction revealed by molecular dynamics of Trs dimers and trimers of dimers in lipid vesicles, *PLoS Comput. Biol.* 8 (2012) e1002685.
- [96] N. Kantarci, P. Doruker, T. Haliloglu, Cooperative fluctuations point to the dimerization interface of p53 core domain, *Biophys. J.* 91 (2006) 421–432.
- [97] C. Chothia, A.M. Lesk, The relation between the divergence of sequence and structure in proteins, *EMBO J.* 5 (1986) 823–826.
- [98] S. Maguid, S. Fernández-Alberti, G. Parisi, J. Echave, Evolutionary conservation of protein backbone flexibility, *J. Mol. Evol.* 63 (2006) 448–457.
- [99] S. Maguid, S. Fernández-Alberti, J. Echave, Evolutionary conservation of protein vibrational dynamics, *Gene* 422 (2008) 7–13.
- [100] J. Echave, Evolutionary divergence of protein structure: the linearly forced elastic network model, *Chem. Phys. Lett.* 457 (2008) 413–416.
- [101] J. Echave, Why are the low-energy protein normal modes evolutionarily conserved? *Protein Sci.* 84 (2012) 1931–1937.
- [102] V. Carnevale, S. Raugei, C. Micheletti, P. Carloni, Convergent dynamics in the protease enzymatic superfamily, *J. Am. Chem. Soc.* 128 (2006) 9766–9772.
- [103] C. Atilgan, A.R. Atilgan, Perturbation-response scanning reveals ligand entry–exit mechanisms of ferric binding protein, *PLoS Comput. Biol.* 5 (2009) e1000544.
- [104] Z. Nevin Gere, S. Kumar, S. Banu Ozkan, Structural dynamics flexibility informs function and evolution at a proteome scale, *Evol. Appl.* 6 (2013) 423–433.
- [105] N. Warren, A. Strom, B. Nicolet, K. Albin, J. Albrecht, B. Bausch, M. Dobbe, M. Dudek, S. Firsigs, C. Fritsche, et al., Comparison of the intrinsic dynamics of aminoacyl-tRNA synthetases, *Protein J.* 33 (2014) 184–198.
- [106] M. Gerstein, W. Krebs, A database of macromolecular motions, *Nucleic Acids Res.* 26 (1998) 4280–4290.
- [107] M. Davis, D. Tobi, Multiple Gaussian network modes alignment reveals dynamically variable regions: the hemoglobin case, *Proteins* 82 (9) (2014) 2097–2105.
- [108] D. Tobi, Dynamics alignment: comparison of protein dynamics in the SCOP database, *Proteins* 80 (2012) 1167–1176.
- [109] A. Zen, V. Carnevale, A.M. Lesk, C. Micheletti, Correspondences between low-energy modes in enzymes: dynamics-based alignment of enzymatic functional families, *Protein Sci.* 17 (2008) 918–929.
- [110] S.M. Hollup, G. Salensminde, N. Reuter, WEBnm@: a web application for normal mode analyses of proteins, *BMC Bioinform.* 6 (2005).
- [111] A. Zen, C. de Chiara, A. Pastore, C. Micheletti, Using dynamics-based comparisons to predict nucleic acid binding sites in proteins: an application to OB-fold domains, *Bioinformatics* 25 (2009) 1876–1883.
- [112] G. Hu, S. Michielsens, S.L. Moors, A. Ceulemans, The harmonic analysis of cylindrically symmetric proteins: a comparison of Dronpa and a DNA sliding clamp, *J. Mol. Graph. Model.* 34 (2012) 28–37.
- [113] K. Hinsen, The Molecular Modeling Toolkit: a new approach to molecular simulations, *J. Comput. Chem.* 21 (2000) 79–85.
- [114] A. Bakan, L.M. Meireles, I. Bahar, ProDy: protein dynamics inferred from theory and experiments, *Bioinformatics* 27 (2011) 1575–1577.
- [115] T.D. Romo, A. Grossfield, 31st Annual International Conference of the IEEE EMBS, 2009, pp. 2332–2335.
- [116] B.J. Grant, A.P.C. Rodrigues, K.M. ElSawy, J.A. McCammon, L.S.D. Caves, Bio3d: an R package for the comparative analysis of protein structures, *Bioinformatics* 22 (2006) 2695–2696.
- [117] T.L. Rodgers, D. Burnell, P.D. Townsend, E. Pohl, M.J. Cann, M.R. Wilson, T.C. McLeish, ΔAPT: a comprehensive toolbox for the analysis of protein motion, *BMC Bioinform.* 14 (2013) 183.
- [118] K. Suhre, Y. Sanejouand, ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement, *Nucleic Acids Res.* 32 (2004) 610–614.
- [119] E. Eyal, L.W. Yang, I. Bahar, Anisotropic network model: systematic evaluation and a new web interface, *Bioinformatics* 22 (2006) 2619–2627.
- [120] S. Seo, M.K. Kim, KOSMOS: a universal morph server for nucleic acids, proteins and their complexes, *Nucleic Acids Res.* 40 (2012) W531–W536.
- [121] D.M. Kruger, A. Ahmed, H. Gohlke, NMSim web server: integrated approach for normal mode-based geometric simulations of biologically relevant conformational transitions in proteins, *Nucleic Acids Res.* 40 (2012) W310–W316.
- [122] E. Lindahl, C. Azuara, P. Koehl, M. Delarue, NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis, *Nucleic Acids Res.* 34 (2006) W52–W56.
- [123] M.T. Zimmermann, A. Kloczkowski, R.L. Jernigan, MAVENS: motion analysis and visualization of elastic networks and structural ensembles, *BMC Bioinform.* 12 (2011).
- [124] A. Bakan, A. Dutta, W. Mao, Y. Liu, C. Chennubhotla, T.R. Lezon, I. Bahar, Evol and ProDy for bridging protein sequence evolution and structural dynamics, *Bioinformatics* 30 (18) (2014) 2681–2683.
- [125] R. Potestio, T. Aleksiev, F. Pontiggia, S. Cozzini, C. Micheletti, ALADYN: a web server for aligning proteins by matching their large-scale motion, *Nucleic Acids Res.* 38 (2010) W41–W45.
- [126] C. Toyoshima, H. Nomura, T. Tsuda, Lumenal gating mechanism revealed in calcium pump crystal structures with phosphate analogues, *Nature* 432 (2004) 361–368.
- [127] P. Gourdon, X.Y. Liu, T. Skjorrjinge, J.P. Morth, L.B. Moller, B.P. Pedersen, P. Nissen, Crystal structure of a copper-transporting PIB-type ATPase, *Nature* 475 (2011) 59–64.
- [128] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, *J. Mol. Graph.* 14 (1996) 33–38.
- [129] K.R. Skowronek, F. Guo, Y. Zheng, N. Nassar, The C-terminal basic tail of RhoG assists the guanine nucleotide exchange factor trio in binding to phospholipids, *J. Biol. Chem.* 279 (2004) 37895–37907.