Available online at www.sciencedirect.com

SciVerse ScienceDirect

Physics of Life Reviews 10 (2013) 1–26

PHYSICS of LIFE reviews

www.elsevier.com/locate/plrev

Review

Comparing proteins by their internal dynamics: Exploring structure–function relationships beyond static structural alignments

Cristian Micheletti

Scuola Internazionale Superiore di Studi Avanzati, via Bonomea 265, Trieste, Italy

Available online 26 October 2012

Communicated by E. Shakhnovich and E. Di Mauro

Abstract

The growing interest for comparing protein internal dynamics owes much to the realisation that protein function can be accompanied or assisted by structural fluctuations and conformational changes. Analogously to the case of functional structural elements, those aspects of protein flexibility and dynamics that are functionally oriented should be subject to evolutionary conservation. Accordingly, dynamics-based protein comparisons or alignments could be used to detect protein relationships that are more elusive to sequence and structural alignments. Here we provide an account of the progress that has been made in recent years towards developing and applying general methods for comparing proteins in terms of their internal dynamics and advance the understanding of the structure–function relationship.

Text

1. Introduction

Over the past decades enormous efforts have been made to clarify the sequence → structure → function relationships for proteins and enzymes. In particular the sequence → structure connection has been extensively probed by dissecting the detailed physico-chemical mechanisms that assist and guide the folding process of several proteins [29,42,43,84]. The more general aspects of this relationship are, however, better captured by analysing the degenerate mapping between the ensembles of naturally-occurring protein sequences and their corresponding folds [26–28,44,67,69,84,99]. For instance, the current ∼85,000 entries can be clustered in about 20,000 non-redundant sequence sets but cover only 1500 distinct structural folds [113,122].

The introduction of general quantitative schemes for comparing, or aligning, protein sequences and protein structures has played a crucial role for framing the observed many-to-one sequence–structure relationship in the context of molecular evolution [117,139]. In particular, by following the impact that evolutionary sequence divergence has on native structural changes [28] it has been possible to identify general properties of peptide chains, amino acid hydrogen-bonding patterns, thermodynamic stability etc. that govern the sequence–structure relationship by constraining the repertoire of viable structural changes that are evolutionary accessible [27,34,95,96,101,147,156,170,176].

E-mail address: michelet@sissa.it.

1571-0645/$ – see front matter
http://dx.doi.org/10.1016/j.plrev.2012.10.009

As a result, remote evolutionary relationships are more confidently obtained from structure-based comparative methods than sequence based ones.

Besides the above general constraints, additional and stronger ones are imposed by functional requirements. In fact, it has long been known that enzymes that have evolutionarily diverged and that catalyse different reactions, tend to conserve very precisely functional structural elements and the location of the active site where different amino acids can be recruited for different function [10,28,116,123,169]. More recently it has also emerged that specific features of protein internal dynamics that impact biological activity and functionality can also be subject to evolutionary conservation [21,87,137,181,182].

By analogy with the sequence–structure case, one may therefore envisage that quantitative methods apt for comparing function-oriented properties in different proteins could advance the capability of detecting protein evolutionary relationships that may be elusive to sequence- or structure-based investigations.

Here we shall review recent studies which focused on the comparison of protein internal dynamics, which is arguably one of the many aspects that often, though not always, assist or influence protein function over a wide range of time scales [14,37,104]. For example, concerted structural movements in enzymes, either "innate" or triggered by ligand binding, have been argued to be important for enzymes to achieve a catalytically-competent state, promote catalytic efficiency, for allosteric signal propagation and protein–protein interactions [1,11,12,24,32,37,38,52,55,59,60,71,87,102,106,107,110,114,124,126,132,137,155,160,168,179,181,183].

We shall accordingly report on the progress that has been made in recent years towards developing and exploiting quantitative numerical strategies for comparing the internal dynamics of proteins and explore its connection with structural and functional similarities.

The material presented in the review is organised as follows. Because these approaches are virtually all based on numerical characterisations of protein internal dynamics we shall first provide a self-contained methodological summary of the theoretical/computational techniques used to characterise and compare protein internal dynamics. Next we shall overview the contexts where dynamics-based comparisons, with different resolution and scope, have been applied. We shall further provide an in depth discussion of a number of selected instances where dynamics-based similarities have been detected within structurally-heterogeneous members of specific protein families, and even across protein families.

2. Comparing protein internal dynamics: Methodological aspects

In this section we provide a self-contained overview of the quantitative numerical approaches employed to characterise and compare the internal dynamics of proteins. In particular, we first review the essential dynamics analysis techniques which are commonly applied to atomistic molecular dynamics simulations or phenomenological coarse-grained models (elastic networks) to single out the collective degrees of freedom that best account for protein's internal motion in thermal equilibrium. Next we shall discuss how the essential dynamical spaces and other dynamics-related quantities can be used for comparative purposes.

2.1. Protein internal dynamics: Essential dynamics analysis of MD trajectories

The wealth of information produced by extensive atomistic molecular dynamics (MD) simulations of globular proteins is typically described and rationalised by identifying the few collective degrees of freedom that best capture the internal protein dynamics. Arguably, the most commonly used technique is represented by the principal component analysis [48] of amino acid pairwise displacements.

This technique relies on the spectral decomposition of the matrix of pairwise correlations of the displacements of amino acids, represented by their C_α atoms, from their reference positions.

In the following we shall indicate with $\mathbf{r}_i(t)$ the three-dimensional position at simulation time t of the ith C_α atom and with $\delta\mathbf{r}(t) \equiv \mathbf{r}_i(t) - \langle \mathbf{r}_i \rangle$ the associated vector displacement from the average reference position. A generic entry of the matrix of pairwise displacement correlations, C, is accordingly defined as

$$C_{ij,\mu\nu} = \left\langle \delta r_{i,\mu}(t) \delta r_{j,\nu}(t) \right\rangle \tag{1}$$

where $\delta r_{i,\mu}(t)$ is the μth Cartesian component of the vector displacement of the ith amino acid and $\langle\,\rangle$ denotes the average over simulation time. For proteins consisting of N amino acids, the symmetric covariance matrix C has linear size equal to $3N$.

It is important to notice that the matrix element of Eq. (1) can be equivalently rewritten as:

$$C_{ij,\mu\nu} = \sum_{l=1}^{3N} \lambda_l v^l_{i,\mu} v^l_{j,\nu} \tag{2}$$

where $\lambda_1, \lambda_2, \ldots$ are the eigenvalues of C ranked by decreasing magnitude and $\mathbf{v}^1, \mathbf{v}^2, \ldots$ are the corresponding orthonormal eigenvectors.

Because the protein overall mean square fluctuation is given by

$$\sum_{i,\mu} \langle \delta r_{i,\mu}(t)^2 \rangle = \sum_{i,\mu} C_{ii,\mu\mu} = \sum_l \lambda_l \tag{3}$$

one has that top ranking eigenvectors of C embody the independent degrees of freedom that most contribute to the internal dynamics of the protein. Indeed, for most globular proteins of 100–200 amino acids, the top 10 eigenvectors suffice to capture most of the protein mean square fluctuation [48]. For this reason, considerations are typically restricted to the linear space spanned by the top eigenvectors of C, which is commonly termed the essential dynamical space [5].

The structural deformations entailed by the essential eigenvectors, or essential modes, are typically found to embody concerted, collective displacements of protein subportions consisting of several amino acids [48,162]. As a matter of fact, the large-scale collective conformational changes that many proteins and enzymes need to sustain in order to carry out their biological functionality have been shown to lie in the essential dynamical space [3,33,40,105,128,141,150,158,181].

These observations provide an *a posteriori* justification for considering the essential dynamical spaces as providing key information into functionally-oriented aspects of proteins.

We conclude by noting that one relevant technical point of the essential dynamics analysis regards the definition of the reference amino acids positions from which the instantaneous displacements $\delta\mathbf{r}$ are calculated. For proteins that have an overall rigid-like character, these positions can be obtained by averaging the conformers sampled by the MD simulation after optimally superposing them. The structural superposition is necessary to remove the overall rotations and translations of the molecules. It is important to stress that this step is not trivially accomplished when proteins have an appreciable internal flexibility character (e.g. due to the presence of mobile subdomains) [184]. In this case, to avoid artefactual results, it is crucial to identify the correct frame of reference for describing and computing the internal structural fluctuations of the protein, see e.g. the discussion of Refs. [60,61] and related supporting material.

However, it must be noted that the relative displacements of domains in multidomain proteins can be so large that protein movements cannot be reliably described by a linear superposition of a limited number of essential dynamical spaces, even if obtained with the above-mentioned procedure. A prototypic example is offered by the relative rotation of protein domains by a finite angle. In this case the directions of instantaneous rotations of the two extreme positions can project very poorly on the difference vector of the latter (see Fig. 3 in Ref. [151]). In such cases the salient degrees of freedom of protein internal dynamics can be identified by decomposing the protein of interest into quasi-rigid domains [2,16,54,56,66,79,131,175] and next considering their relative roto-translations [109], see also Section 3.9.

2.2. *Essential dynamical spaces from elastic network models*

The collective character of the top eigenvectors of the covariance matrix C obtained from atomistic MD simulations suggests that the essential dynamical spaces could be reliably identified by coarse-grained protein models.

This observation, which was stimulated by the seminal work of M. Tirion [162] has in fact lead to the introduction of the well-known elastic network models which, despite adopting a simplified description of a protein's structure and its native amino acid interactions, can reliably identify the essential dynamical spaces of globular proteins with a negligible computational expenditure [7,8,33,63,100,102,157].

In these approaches, each amino acid is described by one or few centroids (e.g. the C_α atom for the main chain [7] and an additional centroid for the side chain [102]) the model potential energy is constructed by introducing quadratic

penalties for the deviations from the native values of the distance of all pairs of centroids that are in contact in the native state. Accordingly, for a protein consisting of N amino acids, the resulting potential energy has the form:

$$U = \frac{1}{2} \sum_{ij,\mu\nu} \delta r_{i,\mu} M_{ij,\mu\nu} \delta r_{j,\nu} \tag{4}$$

where M is a symmetric matrix of linear size $3N$. In the following we shall indicate with $\tau_0, \tau_1, \ldots, \tau_{3N}$ the eigenvalues of M ranked for *increasing* magnitude, and with $\mathbf{w}^0, \mathbf{w}^1, \ldots, \mathbf{w}^{3N}$ the associated orthonormal eigenvectors. The eigenvalues $\{\tau_l\}$ are all positive except for the six attributed to the global rotations and translations of the molecule. It is evident that Eq. (4) bears strong analogies with the normal mode analysis of proteins [85,162].

Because of the quadratic character of the model potential energy of Eq. (4) canonical equilibrium properties of the elastic network can be calculated exactly. In particular, a generic entry of the model covariance matrix C is given by

$$C_{ij,\mu\nu} = \kappa_B T \tilde{M}^{-1}_{ij,\mu\nu} \tag{5}$$

where $\kappa_B T$ is the thermal energy at the temperature of interest, T, and the tilde superscript denotes the pseudoinversion operation, i.e. the removal of the zero-eigenvalue space prior to the inversion of M. Equivalently, C can be written as

$$C_{ij,\mu\nu} = {\sum_{l}}' \frac{\kappa_B T}{\tau_l} w^l_{i,\mu} w^l_{j,\nu} \tag{6}$$

where the prime indicates the omission of the eigenspaces associated to the zero eigenvalues. The above expression clarifies that the degrees of freedom that most account for the proteins' fluctuations in thermal equilibrium correspond to the modes of protein deformation associated to the smallest eigenvalues, i.e. those that cost least energy to excite.

If the proteins dynamics were described by an overdamped Langevin scheme, these low-energy modes would also be those having the slowest relaxation time. Although the harmonic character of the near-native free energy well and the white noise Langevin description apply only limitedly to proteins [15,64,72,97,104,127], the observation is qualitatively consistent with the fact that collective low-energy modes in proteins occur over long time scales (and hence are occasionally referred to as "low-frequency" modes). These observations motivate the practice, adopted in this review too, of regarding the principal components of equilibrium structural fluctuations as embodying the salient internal dynamical properties.

We conclude by mentioning that in recent years alternative formulations of elastic network models have been proposed including versions based on the matching of observables obtained from atomistic MD simulations [81] and on the use of internal coordinates, which are commonly used in normal mode analysis too [53,90,91,98,118].

2.3. Anharmonicity of proteins free energy landscape

The viability of elastic network models to capture the salient traits of protein conformational fluctuations is justified *a posteriori* by the good accord between the essential covariance matrices of elastic network models and of extensive atomistic MD simulations. For example in Ref. [102] it was compared the covariance matrices of HIV-1 protease with a bound ligand obtained from a 14-ns MD simulation with an atomistic force-field and explicit solvent and the beta-Gaussian elastic network model, which employs two centroids per amino acids (for main- and side-chain, respectively). The linear correlation coefficient of the ∼20,000 corresponding distinct entries of the two matrices was significant (equal to 0.8) like the consistency of the two sets of essential dynamical spaces. A more recent example of the good accord of protein structural fluctuations computed with elastic network models and MD atomistic simulations is provided by the work of Romo and Grossfield on GPCRs membrane proteins [142]. This study showed that a suitably-parametrised model can match the essential dynamical spaces and their relative weight observed in microsecond-long simulations.

This agreement is noteworthy in consideration of the highly complex free energy landscape explored by folded proteins can explore in thermal equilibrium. In fact, this landscape presents several tiers of local minima [45,46,171] with low barriers (compared to the thermal energy $\kappa_B T$) separating conformational states with local structural differences such as the rotameric state of a sidechain while large ones separate conformational ensembles with major subdomains rearrangement, such as for open and closed conformation of certain enzymes. In turn, the hierarchical organisation of these minima reflects in a broad range of time-scales, from the ps to the ms and beyond, over which the mentioned structural changes can occur as observed in NMR and single-molecule experiments [14,59,60,104,173].

From these general considerations and from the detailed analysis of the protein conformational substates visited over MD trajectories of hundreds of ns [128,138] it emerges that the harmonic approximation on which elastic network models rely may be a highly simplified parametrisation of even the near-native free energy landscape.

While this limitation, that may be more or less severe depending on the molecule rigidity, must be clearly be borne in mind, it should be noted that the free-energy landscape of a few proteins has been shown to be endowed with particular properties that make the harmonic, or quasi-harmonic [57,65,73,86] free-energy approximation still informative even when dealing with major and slow conformational changes. Specifically, computational studies of lysozyme [76], protein G [127] and adenylate kinase [128] has clarified that the principal directions of the free energy minima associated to the substates populated by each of these proteins are very consistent with each other and also very similar to the difference vectors connecting the substates themselves. This indicates that, despite their structural differences, different substates of the same protein tend to have very similar modes of conformational fluctuations and that the latter, in turn, predispose the observed conformational changes between substates. Indeed, by analysing and comparing the covariance matrices of longer and longer MD trajectories of protein G [127], it was seen that while the trace of the matrix tended to increase (due to the breadth of visited conformational space), the consistency of the essential spaces remained highly significant.

From these results it emerges that the essential dynamical spaces calculated from a relatively short MD simulation or from an elastic network model, would still bear information on the conformational fluctuations sustained by the proteins over time-scales where the harmonic approximation is invalid. The fact that these considerations might hold more in general and not only for the proteins investigated in Refs. [76,127,128] is reinforced by the fact that the difference vector bridging pairs of different protein conformers (such as open and closed forms of several enzymes) has been shown to overlap significantly with the essential dynamical spaces calculated from elastic network models for either conformer [33,78,105,124,158]. Analogous conclusions were drawn more recently by Liu et al. who compared the consistency of essential dynamical spaces of cyanovirin-N obtained from atomistic simulations of varying duration [88].

2.4. Essential dynamical spaces of protein sub-portions

For the purpose of comparing the essential dynamical spaces of proteins with different length and/or architecture it is necessary to identify the essential dynamical spaces of specific protein subparts.

This is straightforward to do in the context of atomistic molecular dynamics simulations. In fact, one simply needs to restrict considerations to the amino acids of interest when calculating the average reference structure and the covariance matrix. The top eigenvectors of this "reduced" covariance matrix (whose entries are clearly not equal to the corresponding ones in the matrix computed for the full protein) accordingly provide the generalised degrees of freedom that best capture the internal motion of the amino acid of interest.

A different approach is however needed for elastic network models. In this case, the reduced covariance matrix of the amino acids of interest must be obtained by the thermodynamic integration of the degrees of freedom of the remainder amino acids. For completeness of notation we assume that the N protein amino acids have been grouped in two sets, a and b. Set a gathers all the n amino acids of interest. The interaction matrix M, after the row/columns reordering following the amino acid groupings, can be partitioned in blocks as follows:

$$M = \left(\begin{array}{c|c} M_a & V \\ \hline V^T & M_b \end{array}\right) \tag{7}$$

where the submatrices M_a and M_b capture the elastic network interactions involving pairs of amino acids in set a and b, respectively, matrix V contains the elastic network couplings of amino acids in the two sets and T denotes the transpose. Matrices M_a and M_b are square and symmetric (of linear size $3n$ and $3(N-n)$, respectively) while matrix V is, in general, rectangular.

Because of the quadratic character of the energy function U it is possible to calculate exactly the reduced matrix effective interactions for amino acids in set a which is equal to:

$$M_a^{\text{eff}} = M_a - V M_b^{-1} V^T \tag{8}$$

and finally, the covariance matrix of set a is obtained by taking the pseudoinverse of M_a^{eff} [19,65,105]. It is important to point out that the second term in the right-hand side of the above equation allows for taking into account the

influence of the remaining amino acids from those of interest. This term is also crucial to ensure that the dynamics of amino acids in set a is described in the proper reference system where the roto-translations of set a alone (and not the whole protein) are extracted.

We conclude by mentioning that, in the same spirit of Eq. (8), one can obtain effective interaction (and covariance) matrices for few generalised degrees of freedom that depend linearly on amino acid Cartesian coordinates. One such example is offered by the study of Ref. [17] where the structural fluctuations of a large set of EF-hand proteins was studied in terms of the relative motion of the axes of their four helices. A further relevant avenue where the degrees-of-freedom integration can be profitably applied is represented by proteins embedded in a constraining matrix. A notable instance is represented by membrane proteins whose conformational plasticity can have important functional implications [39,145]. For such proteins, Romo and Grossfield [142] have recently shown that Eq. (8) can be generalised and used to define effective inter-amino acid interactions which taken into account the influence of embedding bilayer.

2.5. Measures of similarities of two sets of essential dynamical spaces

The information about protein internal dynamics that can be gleaned by applying the methods described in the previous section, can be used in quantitative approaches for the dynamics-based comparison, or alignment of proteins.

We start by discussing the case where the two proteins of interest, A and B, are so similar that sequence or structural alignments suffice to establish extensive one-to-one correspondences between all of their amino acids or a subset of them.

The consistency of the dynamics of the two sets of amino acids marked for alignment can be assessed by the standard root mean square inner product (RMSIP) of their essential dynamical spaces. Customarily, the comparison is restricted to the top 10 essential modes, which are usually sufficient to cover most of the global mean square fluctuation of a protein observed in MD simulations [48]. Accordingly, the RMSIP is defined as:

$$RMSIP = \sqrt{\frac{1}{10}\sum_{l,m=1}^{10}\left[\sum_{i=1}^{n}\sum_{\mu} v_{i,\mu}^{l} w_{i,\mu}^{m}\right]^2} \tag{9}$$

$$= \sqrt{\frac{1}{10}\sum_{l,m=1}^{10}\left|\mathbf{v}^{l}\cdot\mathbf{w}^{m}\right|^2} \tag{10}$$

where $\mathbf{v}^l$ and $\mathbf{w}^l$ denote the lth essential mode of the marked amino acids in protein A and B, respectively, and we have further assumed that matching amino acids carry the same index, $i = 1, \ldots, n$, in the two proteins. Because of the orthonormality of each of the two basis sets $\{\mathbf{v}\}$'s and $\{\mathbf{w}\}$'s, the RMSIP takes on values in the 0–1 range.

The RMSIP measure was introduced for the purpose of assessing the convergence of an MD simulation by comparing the essential dynamical spaces of e.g. the first and second half of the trajectory [4]. Although a simple quantitative criterion for its statistical significance is lacking, it is generally held that RMSIP values larger than 0.7 imply meaningful dynamical similarities [62]. For completeness we mention that other measures of dynamical similarity and MD simulation convergence are available, see e.g. Refs. [18,47,134,143,144].

We finally point out that, for the purpose of profiling the contribution of individual amino acids to the overall mean square inner product one can consider the quantity, which is invariant for changes of the basis of the essential dynamical spaces [18]:

$$Q_i = \frac{1}{10}\sum_{l,m=1}^{10}\left[\sum_{\mu} v_{i,\mu}^{l} w_{i,\mu}^{m}\right]\left[\mathbf{v}^{l}\cdot\mathbf{w}^{m}\right] \tag{11}$$

where i is the index of the amino acid of interest, or its square root $q_i = \sqrt{Q_i}$.

2.6. Best-matching essential dynamical spaces

The RMSIP of Eq. (10) measures the overall consistency of the essential dynamical spaces and therefore is invariant upon change of the basis vectors for the two linear spaces, $\{\mathbf{v}\}$'s and $\{\mathbf{w}\}$.

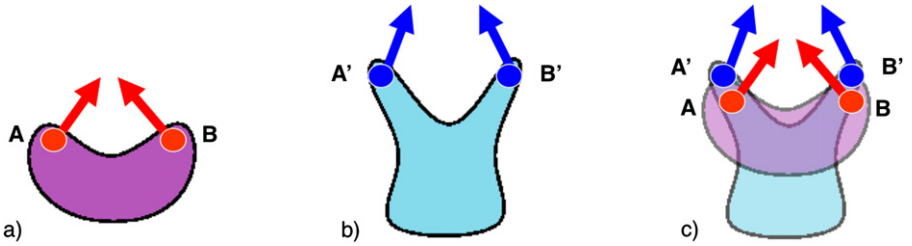

Fig. 1. Example of dynamics-based alignment. The two cartoon structures in panels (a) and (b) have dissimilar shapes. Yet, their internal movements, schematically indicated by the arrows, are consistent and can provide valuable clues for superposing the two structures, as shown in panel (c).

This property can be exploited to replace the $\{\mathbf{v}\}$'s and $\{\mathbf{w}\}$'s with two new sets of orthonormal vectors $\tilde{\mathbf{v}}^1, \tilde{\mathbf{v}}^2, \ldots, \tilde{\mathbf{v}}^{10}$ and $\tilde{\mathbf{w}}^1, \tilde{\mathbf{w}}^2, \ldots, \tilde{\mathbf{w}}^{10}$ which are ranked for decreasing mutual consistency (magnitude of the scalar product) [128].

To do so, one constructs the 10×10 asymmetric matrix D whose entries are $D_{ij} = \mathbf{w}_i \cdot \mathbf{v}_j$. Next one solves the eigenvalue problems [128]:

$$D^T D \mathbf{a}^i = \mu_i \mathbf{a}^i \tag{12}$$

$$D D^T \mathbf{b}^i = \mu_i \mathbf{b}^i \tag{13}$$

Assuming that the eigenvalues have been ranked by decreasing order $\mu_1 > \mu_2 > \cdots > \mu_{10}$, and that $\mathbf{a}^i$ and $\mathbf{b}^i$ have unit norm, one has that the new basis vectors are given by

$$\tilde{\mathbf{v}}^i = \sum_{j=1}^{10} a^i_j \mathbf{v}^j \tag{14}$$

$$\tilde{\mathbf{w}}^i = \sum_{j=1}^{10} b^i_j \mathbf{w}^j \tag{15}$$

The newly defined orthonormal basis, $\{\tilde{\mathbf{v}}\}$ and $\{\tilde{\mathbf{w}}\}$ have the following remarkable properties:

- the ith vector in one set is orthogonal to all vectors in the *other* set with index different from i, i.e. $\tilde{\mathbf{v}}_i \cdot \tilde{\mathbf{w}}_j = 0$ if $i \neq j$;
- the scalar products $\tilde{\mathbf{v}}_i \cdot \tilde{\mathbf{w}}_i$ have magnitude that decreases with i;

therefore the new basis vectors are optimally ranked for decreasing mutual consistency and are ideally suited to represent the most consistent (or inconsistent) subspaces spanned by the $\{\mathbf{v}\}$'s and $\{\mathbf{w}\}$ [128].

Once more we stress that, as the $\{\tilde{\mathbf{v}}\}$ and $\{\tilde{\mathbf{w}}\}$ provide alternative basis for the same spaces spanned by the $\{\mathbf{v}\}$'s and $\{\mathbf{w}\}$'s, the RMSIP of $\{\tilde{\mathbf{v}}\}$ and $\{\tilde{\mathbf{w}}\}$ is the same as for the $\{\mathbf{v}\}$'s and $\{\mathbf{w}\}$.

2.7. Beyond structural alignment: Dynamics-based protein alignment

2.7.1. Aligning proteins by matching their essential dynamical spaces

The previous approach needs to be suitably generalised in contexts where one wishes to detect dynamics-based correspondences in different proteins without relying on their prior sequence or structure alignment.

A prototypical situation is illustrated in Fig. 1 where two cartoon structures with different shape are sketched in panels (a) and (b). Despite the overall shape difference, the structural deformation modes described by the arrows, are well-consistent and can provide the basis for aligning the two structures, see panel (c).

As first noted by Zen et al. [177], the example in Fig. 1 clarifies that meaningful dynamics-based alignments cannot be simply obtained by purely rewarding the similarity of directionality and magnitude of the essential dynamical spaces of any of two sets of amino acids in the proteins of interest. In fact, the alignment shown in panel (c) is intuitively perceived as viable because the origins of the paired arrows, A–A′ and B–B′, are nearby in space. If

the origins had been arbitrarily dislocated in space, then the paired arrows would not have implied any consistent structural modulations of the two shapes (but motions of very large amplitude can significantly change the geometrical relationships of dynamically-corresponding regions, see Section 3.9).

Prompted by the above considerations, Zen et al. [177] introduced and applied a dynamics-based alignment scheme which simultaneously rewarded the consistency of the essential dynamical spaces of matching amino acids as well as their spatial proximity. Specifically, in this alignment technique the score to be maximised over the possible sets of corresponding amino acids pairs was based on distance-weighted generalisation of the root mean square inner product,

$$\sqrt{\frac{1}{10}\sum_{l,m=1}^{10}\left[\sum_{i=1}^{n}\sum_{\mu} v_{i,\mu}^{l} w_{i,\mu}^{m}\right]\left[\sum_{i=1}^{n}\sum_{\mu} v_{i,\mu}^{l} w_{i,\mu}^{m} f(d_i)\right]} \tag{16}$$

where $i = 1, \ldots, n$ runs over the n aligned amino acids, d_i is the distance between the ith (matching) amino acids in proteins A and B after an optimal superposition over the putative matching region, and $f(d) = [1 - \mathrm{tgh}((d - d_c)/\Delta)]/2$ is a sigmoidal distance weighting factor where $d_c = 4$ Å and $\Delta = 2$ Å.

Notice that, as for the RMSIP, the measure (16) is independent of the choice of the bases spanning the linear space of the top 10 essential dynamical modes.

The sought dynamics-based alignment is accordingly obtained by maximising the measure of Eq. (16) (after a suitable n-dependent regularisation, see Ref. [177]) over the space of possible amino acid pairings in the two proteins, and finally by assessing its statistical significance by comparing it against a null reference case.

Clearly, the combinatorial space of matching amino acids is very large and, because each attempted alignment involves the re-calculation of the essential dynamical spaces, the computational effort entailed by this comparison is significant and can take several minutes on present-day computers for two proteins of ∼100–200 amino acids.

By heuristically restricting the search of matching amino acids and by using approximate but faster calculations of the alignment score, the original algorithm of Zen et al. [177] was sped up sufficiently for interactive use via the Aladyn web-server [130]. The results of this publicly-available server will be frequently referred to in the remainder of this article.

2.7.2. Aligning proteins by matching pairwise distance fluctuations

An alternative method to align proteins based on their internal dynamics properties was recently proposed by Biggin and coworkers [112]. In this method one exclusively considers the pairwise distance fluctuations of amino acids, with no explicit reference to the spatial coordinates of the latter, nor to the detailed information contained in the top essential dynamical spaces. This scheme is based on the idea that, if a set of amino acids $\{\alpha\}$ in protein A has similar movements to a corresponding set of amino acids $\{\beta\}$ in protein B then the matrices of pairwise distance fluctuations of the two sets, F_α and F_β, should be similar too.

In the approach of Münz et al. [112] a generic entry of the F matrix is defined as

$$F_\alpha(i, j) = \mathrm{std.dev}(d_{\alpha_i, \alpha_j}) \tag{17}$$

where the right-hand side is the standard deviation of the distance of amino acids i and j in set α calculated over a converged molecular dynamics trajectory.

Next, one calculates the relative difference of each corresponding matrix entry,

$$d(i, j) = \frac{|F_\alpha(i, j) - F_\beta(i, j)|}{(F_\alpha(i, j) + F_\beta(i, j))/2} \tag{18}$$

and an overall dynamical score $S^{AB}(\alpha, \beta)$ is constructed by weighting the contribution of all $d(i, j)$'s.

As in the previous approach, the best dynamics-based alignment of the two proteins is found by maximising $S^{AB}(\alpha, \beta)$ (again after a suitable length-regularisation procedure) over all possible choices of $\{\alpha\}$ and $\{\beta\}$. In the study of Ref. [112], the exploration of the vast combinatorial space of a.a. pairings was carried out within a Monte Carlo optimisation scheme.

2.7.3. Aligning proteins by matching the mean square fluctuation profiles

The possibility to align proteins by detecting correspondences in the amplitudes of amino acids motions in different proteins was first explored by Keskin et al. [75]. In this study, which is covered in Section 3.1, the one-to-one

correspondences of amino acids in a set of structurally-related proteins was based on a supervised matching of the amplitude of amino acid fluctuations computed from an isotropic elastic network model [8].

An automatic implementation of this alignment strategy was recently introduced by Tobi [163]. In this study, the one-dimensional character of the quantity to be matched (mean square fluctuation) was exploited, as in sequence alignments, within the dynamical-programming alignment of Needleman and Wunsch [115].

3. Comparative studies of protein internal dynamics

Early systematic dynamics-based comparisons were all targeted to groups of proteins known to be significantly related from the sequence, structural or functional point of view. In such contexts, in fact, the assessment and interpretation of the comparisons is more straightforward. Accordingly, we shall first discuss these comparative investigations of proteins whose relatedness is known *a priori*. We shall next report on studies which considered proteins with limited structural relatedness as well as investigations targeted at understanding more general (and possibly evolutionary) dynamics-based aspects of the structure/function relationship. When appropriate, the results of these earlier studies will be revisited using the dynamics-based alignment of Ref. [177] as implemented in the publicly available Aladyn web-server [75].

3.1. Common fluctuation patterns in proteins with a Rossmann-like fold

We first discuss the case of proteins adopting a Rossmann-like fold which were addresses in the studies of Keskin et al. [75] and Pang et al. [120].

In the study of Ref. [75], which is arguably the first dynamics-based comparative investigation, Keskin et al. considered six proteins each consisting of two linked globular domains with a Rossmann-like fold. The proteins covered two homologous groups: the first one (CATH [122] code 3.40.190.10) included cofactor binding fragment of CysB, the lysine/arginine/ornithine-binding protein (LAO), the enzyme porphobilinogen deaminase (PBGD), the N-terminal lobe of ovotransferrin (OVOT) while the second one (CATH code 3.40.50.2300) comprised the ribose-binding protein (RBP) and the leucine/isoleucine/valine-binding protein (LIVBP).

The internal dynamics of these proteins was characterised by using a simplified (isotropic) Gaussian network model [8] to compute their mean-square fluctuation profiles and the lowest energy modes. The authors observed that the latter mostly entailed a hinge-bending motion of the two domains around the linker and the predicted motion amplitude varied significantly between the unliganded and liganded state of the molecules. In connection to this latter result it is worth noting that for several other proteins it has been shown that the internal dynamics sensitively depends on substrates and cofactors. A prototypical example is offered by dihydrofolate reductase where dynamical properties, arguably linked to catalysis, has been shown numerically to strongly depend on the type of bound ligand [136].

The similarities of the modes amplitude profiles across the six proteins, further prompted Keskin et al. to attempt a manually-curated alignment of the proteins by matching the modes shape in a gapless portion of one of the two domains. The amino acid correspondences were next extended to the remainder of the proteins by inspecting both their FSSP structural alignments [68] and, again, the modes shape. These supervised alignments returned very good superpositions of the modes amplitude profiles across the considered proteins and, because of the limited use of structural correspondences, the RMSD after an optimal superposition of the corresponding amino acids was about 7 Å.

From the consistency of the modes' profiles the authors concluded that members of the same fold can share common dynamical features on a global, collective scale and further envisaged that fully-automated dynamics-based alignments of proteins might have been feasible.

The implications of structural relatedness for the similarity of protein internal dynamics were next explored by Pang et al. [120] by using atomistic molecular dynamics simulations on a set of four periplasmic binding proteins in various forms: apo, holo and crystallised in different conditions.

The monomeric units of these entries, which included the LAO protein considered by Keskin et al. [75], comprised about 230 amino acids and consisted, again, of two Rossmann-like domains connected by a linker. Based on DALI [67] alignments Pang et al. identified a core of 100 amino acids (i.e. spanning about 40–45% of the proteins) common to the four proteins.

Table 1
Representatives of the seven common protease folds, A–G. The list includes proteases with different catalytic chemistry (aspartic-, serine-, cysteine- and metallo-proteases). For convenience of comparative purposes, because the active site is comprised within the monomeric units of 3C-like proteinase, assemblin and dipeptidyl peptidase I, we did not consider the multimeric biological form of these entries. Conversely, because the catalytic aspartic dyad of HIV-1 protease straddles the dimeric interface, we retained its full dimer. The corresponding structures are represented in Fig. 2.

Tag	Protein	PDBid	CATH code
A	endothiapepsin (ASP)	1er8E	2.40.70.10
B	HIV-1 protease (ASP)	1nh0AB	2.40.70.10
C	3C-like proteinase (SER)	1uk4A	2.40.10.10 1.10.1840.10
D_1	adenain (CYS)	1avp	3.40.395.10
D_2	sedolisin (SER)	1ga6	3.40.50.200
D_3	pyroglutamyl peptidase I (CYS)	1ioi	3.40.630.20
E	assemblin (SER)	1jq7A	3.20.16.10
F_1	dipeptidyl peptidase I (CYS)	1k3bA	2.40.128.80
F_2	cruzipain	1me4	3.90.70.10
G_1	atrolysin E (Zn)	1kuf	3.40.390.10
G_2	carboxy peptidase A1 (Zn)	8cpa	3.40.630.10

The comparison of the internal dynamics was carried out on the common core amino acids and regarded various quantities calculated from 10- or 20-ns long molecular dynamics simulations. In particular, the comparison included: the amino acids' mean square fluctuations, the overlap of the covariance matrices and the overlap (RMSIP) of the two essential dynamical spaces.

By comparing the properties of the same protein but in liganded and unliganded forms, Pang et al. observed clear differences in the molecules' internal dynamics, consistently with the findings of Keskin et al. reported above.

Regarding the comparison of different proteins, the authors reported a significant overlap of all dynamical properties computed over the common core. In particular, throughout the set of periplasmatic binding proteins, the first and second essential dynamical modes systematically corresponded to, respectively, the hinge-bending and twisting motions of the linked domains.

However, by examining how the overlap of the covariance matrix and essential dynamical spaces increased with simulation time, the authors observed that each protein tended to occupy specific regions of the essential dynamical space. It was concluded that these differences reflected protein-specific features, arguably encoded in their sequence. While, it cannot be ruled out *a priori* that the observed differences could be ascribable to the several non-aligned amino acids, the observation of Pang et al. is very interesting and relevant in the present context, because it points to specific dynamics-based features which can be beyond reach of sequence-independent approaches, such as elastic network models.

3.2. Dynamics-based alignment of proteases

Proteases, enzymes that cleave peptide chains, account for about 2% of the genome of various organisms [135,140,153]. In view of this representative weight and biological importance, they have been systematically investigated and compared.

The comprehensive survey carried out by Tyndall et al. [165], identified 7 common structural folds for this family of enzymes. Various representatives for the seven common folds were identified by Carnevale et al. [19] and are listed in Table 1 and shown in Fig. 2.

As reported in Table 1, the various representatives cover 4 different architectures and 9 different topologies of the CATH classification scheme [122]. Notice that the two aspartic proteases, the endothiapepsin and HIV-1 PR share the full CATH code, implying that they have detectable sequence homology despite their marginal sequence identity, different length and different oligomeric state (monomeric for endothiapepsin and dimeric for HIV-1 PR) [13,22,159].

Besides this ASP-protease pair, other pairs of entries listed in Table 1 have significant overall structural similarities. In particular the six possible distinct pairings between pyroglutamyl peptidase I, atrolysin E, sedolisin and carboxy peptidase A1 are all significant according to the DALI statistical criteria [67]. Interestingly, the simultaneous multiple

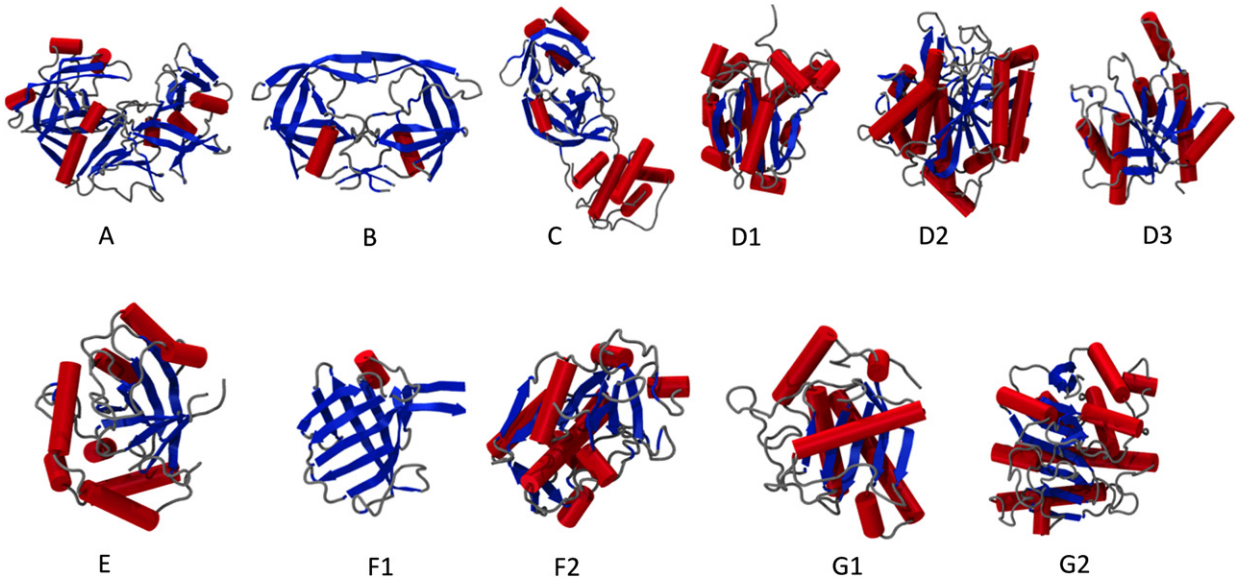

Fig. 2. Representative structures of the common protease folds listed in Table 1. This illustration and subsequent ones were prepared with the VMD graphical package [70].

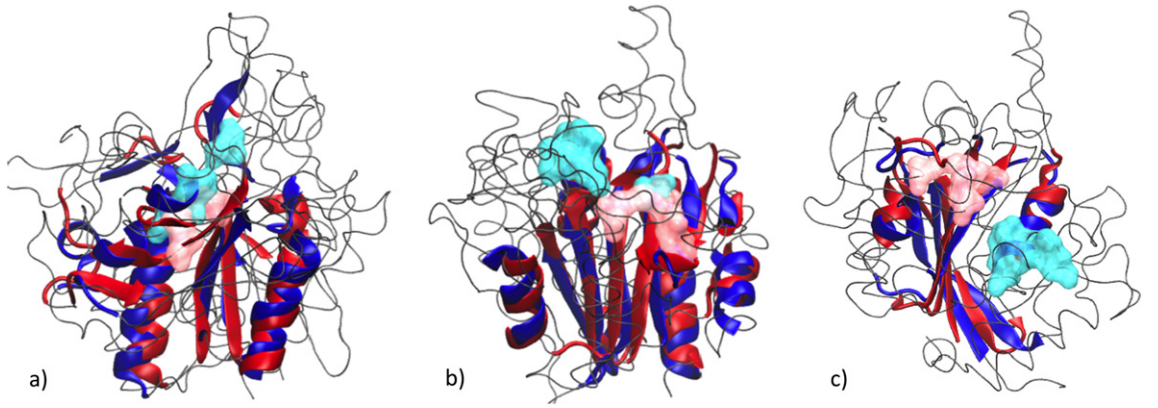

Fig. 3. Structural alignments of pyroglutamyl peptidase I with (a) sedolisin, (b) carboxy peptidase A1 and (c) atrolysin E. In all panels the pyroglutamyl peptidase I is shown in red, while the partner proteins are shown in blue. Aligned regions are shown with thick ribbons and known active sites [129] are highlighted with Van-der-Waals surfaces. The trace of non-aligned regions is shown as a thin grey curve.

alignment of these four entries is poor and involves several short fragments for a total of about 30 amino acids (consistently for both Mistral and Multiprot [103,148]).

The top structural alignments within this group involved the entry pyroglutamyl peptidase I and are shown in Fig. 3. As it was reported in Ref. [19] (see Fig. 3 therein) the alignments, involve several disconnected matching fragments comprising the active site and the surrounding region within 7–10 Å of it.

The good structural superposition of the active sites in panels (a) and (b) of Fig. 3 provides evidence for the existence of functionally-related traits that are shared by proteases that are non-homologous and rely on different catalytic chemistry (serine-, cysteine- and metallo-proteases).

The fact that functional activity of various proteases is known to be impacted by their large-scale internal dynamics [13,124–126], which can involve mechanical couplings between the active site and distal regions at the protein surface [102,124–126], poses the question of whether dynamics-based alignments can be used to identify further relationships between proteases that are elusive to the pure structural comparison. The possibility to do so is illustrated in Fig. 4, which illustrates the dynamics based alignment of HIV-1 PR and endothiapepsin.

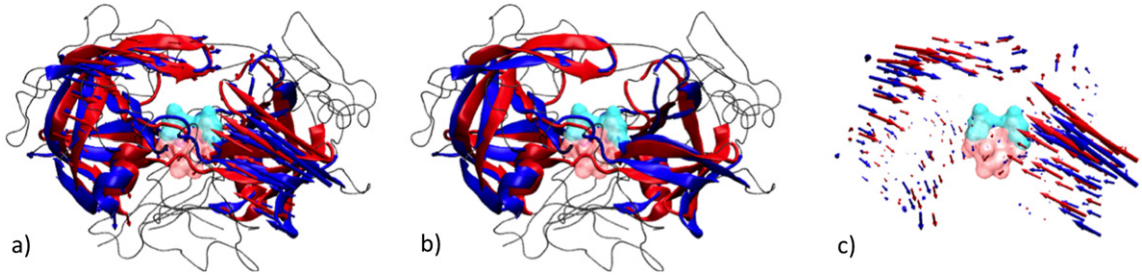

Fig. 4. (a) Dynamics-based alignment of HIV-1 protease (red) and endothiapepsin (blue) obtained with the Aladyn web-server [130]. A thick ribbon is used to highlight aligned regions and known active sites are highlighted with Van-der-Waals surfaces. The trace of non-aligned regions is shown as a thin grey curve while the arrows represent the three best matching essential modes. The ribbons and the modes are shown separately in panels (b) and (c), respectively.

Following the spirit of Ref. [19], we have used the Aladyn algorithm to align all pairs of entries in Table 1. In addition to the previously mentioned significant structural pairings, the Aladyn algorithm identifies 8 additional significant alignments (p-value < 0.02, corresponding to the incidence of less than one false positive in the set of all pairwise alignments of the entries in Table 1). These pairs are shown in Fig. 5. Notice that calpain, adenain, atrolysin E, and HIV-1 PR (corresponding respectively to tags F_2, D_1, G_1 and B in Table 1) constitute a notable dynamically-alignable "clique" because all pairings of these proteins (with the sole exception of cruzipain–HIV-1 PR which involves only 30 amino acids) are significant.

The structural and dynamical consistency of the 8 aligned pairs is shown in Fig. 5. It is striking to see that the active sites of the compared proteins are very well superposed or in contact, with the exception of two alignments, assemblin–HIV-1 PR (E–B) and assemblin–atrolysin E (E–G_1) where the active sites are at a distance of 10 Å. The overall RMSD of the matching amino acids is $\sim$3.0 Å.

It is also noticed that the corresponding modes, tend to outline a shearing deformation of region surrounding the active site. This result is in accord with the general functional features common to proteases, which consists of the shearing of the bound peptide into a beta extended conformation prior to cleavage [165]. More generally, the finding is consistent with the observed property that active sites in enzymes tend to be located at the interface of quasi-rigid domains, as this can ensure a fairly rigid geometry of the catalytic region located at the interface combined with an appreciable modulation of the surrounding region which ought to aid the substrate recognition and processing [131,146].

For the specific case of proteases, the dependence of the enzymatic activity and catalytic rate on the global conformational fluctuations of the proteins has been advocated for HIV-1 PR [126] (but this does not occur for furin, a serine protease [20]). The proposed mechanism for HIV-1 PR has been corroborated by recently experimental findings [30]. Further examples of the coupling between the modulation of the geometry of the region near the active site and the global protein motions are provided by triose phosphate isomerase [80] and dihydrofolate reductase [1,138].

We emphasise that all the pairings identified with the dynamics based alignment shown in Fig. 5 are not deemed significant in DALI alignments. The findings therefore suggest that, for certain proteins and enzymes, some functionally-oriented features can be more confidently identified using dynamics-based alignments than with sequence- or structure-based alignment approaches.

3.3. Dynamics-based alignment of PDZ domains

We next discuss the dynamical similarities of members of the PDZ domain family. PDZ domains are structural moduli commonly associated to ion channels and receptors or otherwise involved in signal transduction pathways [41,111,112,152].

They are typically 80–100 amino acids long and adopt an overall globular fold comprising two α helices and 6β strands, see Fig. 6b. The interaction with a partner protein usually occurs through the accommodation of its C-terminal segment in the β_2–α_2 cleft. In fact, the observed mobility of helix α_2 relative to the PDZ-domain core has been argued to be important for ligand binding and recognition [31,77,112]. Although PDZ-domains sustain modest structural changes after ligand binding, see panels (a) and (b) in Fig. 6, experimental and numerical evidence suggest that there exist allosteric pathways running internally to the molecule that signal the binding event to regions that are opposite on

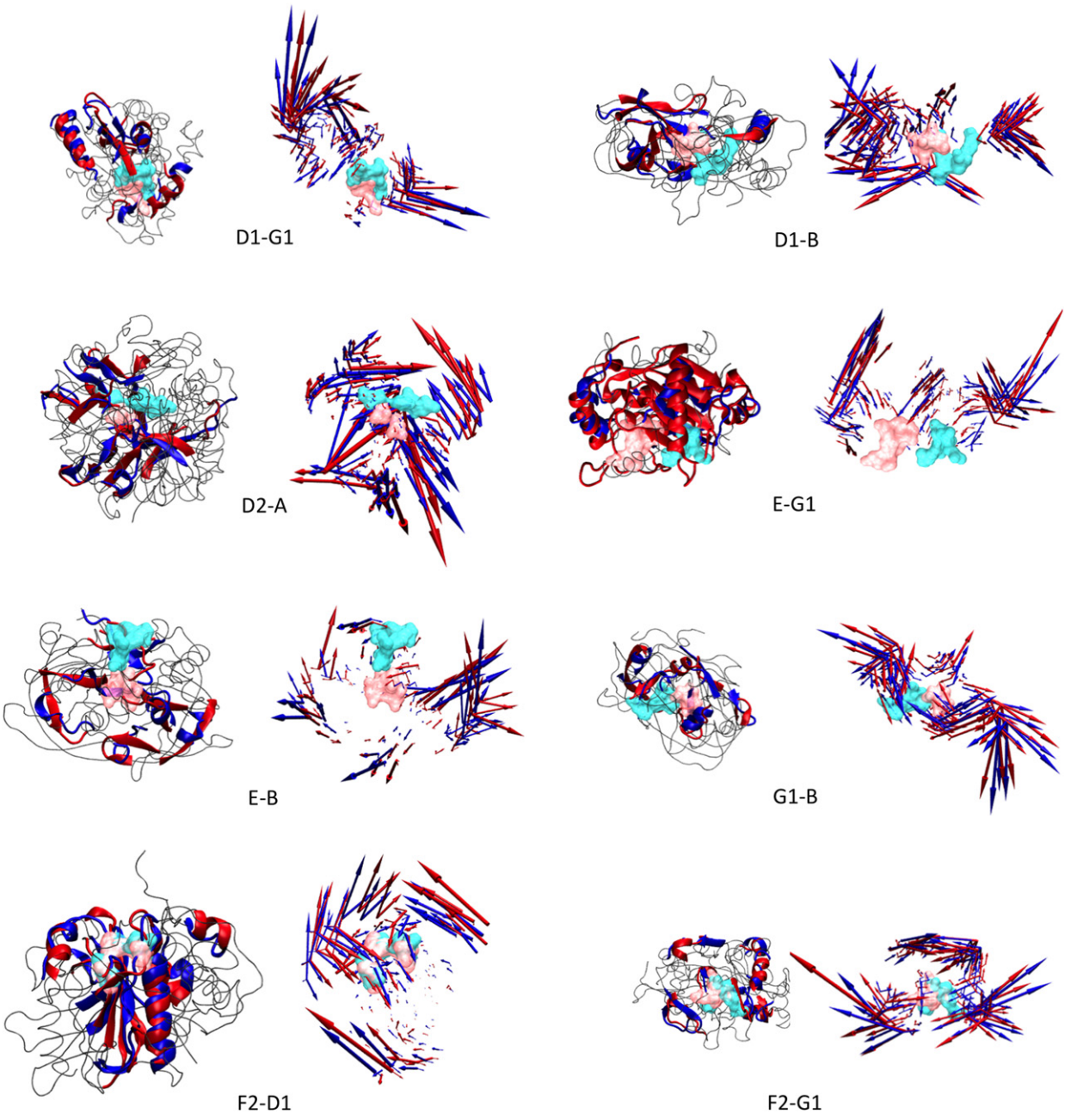

Fig. 5. Significant dynamics-based alignments of various pairs of proteases. The pairs are tagged as in Table 1. For each pair we report separately the structural superposition of the aligned regions (ribbons) and of the top three best-matching modes (arrows). Aligned elements are shown in blue for the first entry of the pair and in red for the second. The active sites are shown in cyan and pink for the first and second entry of the pair, respectively.

the protein surface respect to the binding cleft [31,77,82,89]. While key aspects of the signal propagation mechanism are still controversial [25] various evolutionary aspects of the allosteric mechanism and the binding mode have been actively investigated using a variety of techniques including bioinformatics [89], NMR [82], elastic network linear response theory [49] and molecular dynamics simulations [112].

In particular, Biggin and coworkers [112] have recently introduced and systematically applied the dynamics-based alignment outlined in Section 2.7.2, to compare the mainchain dynamics of 10 PDZ domains from both unicellular

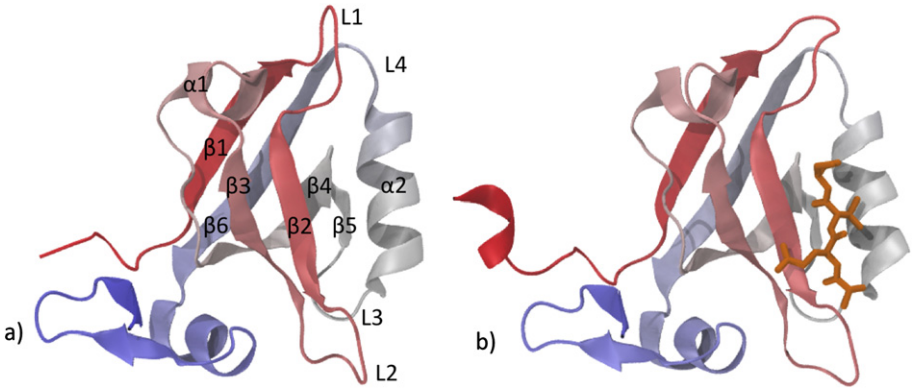

Fig. 6. (a) Apo and (b) holo forms of a PDZ domain. The PDBid of the shown entries is 1bfe for the apo form and 1be9 for the holo one. The ligand bound to the α_2–β_2 cleft of the holo form is highlighted in orange. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
List of PDZ domains considered in Ref. [112]. The blank line separates PDZ domain from multicellular organisms (above line) from unicellular ones (below line).

Compound	CATH domain	CATH code
Postsynaptic density protein 95 (PSD-95)	1be9A00	2.30.42.10
Nitric-oxide synthase (nNOS)	1qauA00	2.30.42.10
Alpha-1 syntrophin	1qavA00	1.14.13.39
Inactivation-no-after-potential D protein (Inad)	1ihjA00	2.30.42.10
Segment polarity protein dishevelled homolog DVL-2 (DVL2)	2f0aA00	
Glutamate receptor interacting protein 2 (GRIP2)	1x5rA00	
Tricorn protease	1k32A04	
Type II secretion system protein C	2i6vA00	3.4.21
Hypothetical serine protease rv0983	1y8tA03	
Photosystem II D1 protease	1fc6A02	2.30.42.10

and multicellular organisms, see Table 2. The dynamics-based comparison, was based on the analysis of pairwise distance fluctuations of amino acids calculated from 20-ns long atomistic molecular dynamics simulations.

Within this set of sequence- and structurally-related PDZ domains Münz et al. observed the largest dynamical consistency among the domains from multicellular organisms. In fact, significant dynamics-based similarities were found almost exclusively among entries from multicellular organisms (particularly pairs nNOS–PSD95, nNOS–Alpha-1 syntrophin, nNOS–DVL2, Inad–Alpha-1 syntrophin, Inad–DVL2, DVL2–Alpha-1 syntrophin).

One such pair, PSD95 and nNOS, was analysed in-depth to highlight the differences of sequence, structure and dynamics-based alignments. Through this comparative investigation, the authors noticed that dynamical correspondences were particularly poor in the α_2 region, which is otherwise structurally well-alignable. Because the mobility of this helix arguably impacts the binding of ligands it was concluded that the dynamical differences could reflect subtle differences in the functionality of PSD95 and nNOS [112].

The findings of Münz et al. are illustrated and revisited here through the dynamics-based alignment method of Zen et al. as implemented in the Aladyn web-server. The Aladyn alignment of PSD95 and nNOS is shown in Fig. 7 and illustrates the good consistency of the essential dynamical spaces of the aligned regions. Interestingly, the contribution of the various corresponding amino acids to the good RMSIP value, which is equal to 0.74, is rather uneven.

This is illustrated in Fig. 7(b) which portrays the residue-wise contribution to the mean square inner product, Q_i (see Eq. (11)) along with the mean-square residue fluctuations. It is seen that the Q profile is peaked in correspondence of the loops L_1, L_2, L_3 and L_4 which are also associated to peaks of the crystallographic B-factor profiles. Although the comparison of computed mean-square fluctuations with B-factors is not perfectly transparent (the latter are affected by crystal packing and disorder [45]), the accord of the two sets of peaks is consistent with the intuition that, given the overall accord of the essential modes, the highest values of Q should be observed in correspondence with regions

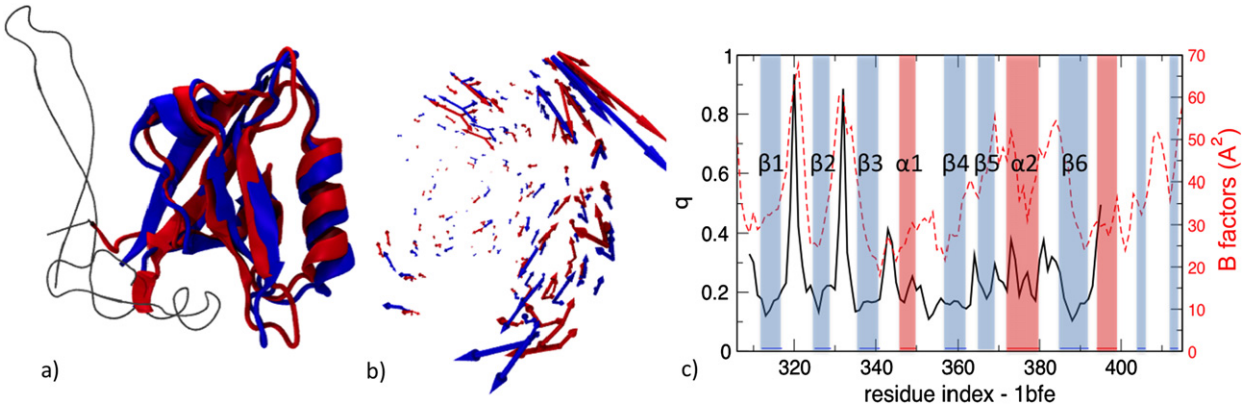

Fig. 7. Dynamics-based alignment [130] of the two PDZ domains discussed in Ref. [112]. The alignment was obtained with the Aladyn web-server [130] and consists of an uninterrupted stretch 87 amino acids (ARG309–GLU395 for 1bfe and ASN14–GLU101 for 1qau) at an RMSD of 2.2 Å and with an RMSIP of 0.74. The structural superposition is shown in panel (a) and the top three matching modes are shown in panel (b). Corresponding elements for entry 1bfe are shown in red while those for entry 1qau are shown in blue. The crystallographic B-factors and the local essential dynamics space overlap, $q = \sqrt{Q}$ (see Eq. (11)), of 1bfe are shown respectively with a dashed and a solid line in panel (c). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of high mobility (where the norm of the essential modes concentrates). By the same token, one would have expected to observe a peak of the Q profile in correspondence of the mobile helix α_2 and the nearby portions of the flanking strands β_5 and β_6. By contrast, however, the relative contribution of these regions to the RMSIP is small. This is therefore indicative of a poor consistency of the generalised direction of motion of this region in the two proteins of interest, thus confirming the findings of Münz et al. from a different dynamics-based perspective.

3.4. Conservation of general dynamical patterns in protein families and superfamilies

Besides the previous investigations that aimed at elucidating specific functionally-related aspects in different proteins by using dynamics-based alignment strategies, there have been a number of studies where more general dynamical properties were compared across various protein families and superfamilies.

In recent years Echave and coworkers have carried out several such studies with the purpose of assessing the extent to which features such as mean-square fluctuation profiles and overall shape (amplitude modulation) of the essential modes have been evolutionarily conserved [92–94].

The first of such analyses was carried out for a set of 18 members of the globin family [93]. The considered globins typically consisted of 130–150 amino acid and shared a structural core of 68 amino acids [103].

The comparison of Maguid et al. [93] was focused on the set of about 100 corresponding amino acids that were identified by the multiple (CLUSTAL [161]) sequence alignment of the 18 globins.

The dynamics of the globins was next characterised by the mean-square fluctuation profiles and molecules' lowest energy modes which were computed using the isotropic Gaussian network model [8]. In this model, the presence of the heme group was not taken into account.

The comparison of the dynamics across the different globins was carried out by measuring the linear correlation coefficient between the fluctuation amplitudes of corresponding amino acids or between their displacements in the top modes. For comparative purposes, the latter were reranked so as to have maximally compatible sets of first modes, second modes etc. across the globins. The main differences of this comparative strategy from the one described in Section 2.4 is that the dynamics of the corresponding amino acids is obtained by neglecting the effect of non-aligned amino acids (equivalent to omitting the second term in Eq. (8)) and for the use of reranked top modes in place of identifying the most consistent directions in the linear space spanned by the top modes.

After carrying out these comparative steps, Maguid et al. [93] concluded that both the mean-square fluctuations and the shape (amplitude modulation) of the top reranked modes were highly consistent across the various members of the globin family.

Building on these findings, Maguid et al. [92,94] extended the analysis to a comprehensive set of ∼1000 protein entries from several hundred families superfamilies of the HOMSTRAD database [108,133,154]. The studies followed

the same comparative pathway outlined above for the globins, with the significant modifications that corresponding amino acids were identified in pairwise MAMMOTH [119] structural alignments and the anisotropic beta-Gaussian elastic network model was used in place of the isotropic one. Furthermore, the degree of collectivity of the modes was also assessed and compared.

The studies of Refs. [92,94] reported that the dynamical similarity (mean-square-fluctuation profiles, mode shape and mode collectivity) within members of the same family and superfamily was significantly larger compared to pairs of unrelated protein entries. In addition, the similarity within the same family was stronger than within the same superfamily.

From this series of studies, Echave and coworkers concluded that general dynamical properties of proteins tend to be preserved in the course of evolution and are quantitatively detectable.

3.5. Conservation of specific functionally-oriented dynamics in enzymes

In the recent study of Ref. [137] Ramanathan et al. addressed, by means of atomistic molecular dynamics simulations, the extent to which enzymes with the same function but different degree of homology rely on the same functionally-oriented dynamics.

The study considered a few members for each of three different types of enzymes: the CypA peptidyl–prolyl isomerase, the DHFR oxidoreductase and ribonuclease A (RNaseA).

For each member, extensive molecular dynamics simulations were carried out. The authors next compared the dynamics-based features that directly impacted the known rate-limiting step of the enzyme catalytic activity. This important technical step allowed Ramanathan et al. to address in a direct and precise way the functionally-oriented dynamical aspects of the proteins without relying on their dynamics-based alignment or considering general aspects of the internal dynamics that are inconsequential for biological functionality [137].

By these means Ramanathan et al. ascertained that the reaction-coupled motions of the members of each of the three types of enzymes were highly similar. Because the members were picked from different species it was further concluded that the detailed functionally-oriented dynamical aspects have been evolutionarily conserved.

The analysis established two further notable features. First, the dynamical similarities found for the homologous CypA entries were found to extend to the non-homologous PIN1 peptidyl–prolyl isomerase. In consideration of the structural differences of the modelled structure of Pin1 and CypA it was concluded that the reaction-coupled motions of the enzymes were conserved despite the structural differences. Secondly, it was observed that the dynamical aspects influencing the functional activity involved regions that are not necessarily near the active site, thus pointing out at an overall interplay of local and global aspects in the functional "mechanics" of the enzymes. The fact that these features might hold for several other enzymes is reinforced by the consistency with the findings reported earlier for members of the proteases family as well as by instances such as R67 dihydrofolate reductase where enzyme flexibility has been argued to impact the catalysed reaction [74].

3.6. Comparison of general dynamical patterns in members of the SCOP database

Besides the above-mentioned studies, a comparative investigation of mean-square fluctuation profiles and mode shapes was recently undertaken by Tobi [163] for an extensive set of entries from the SCOP/Astral database [6,23]. A distinctive point of the analysis of Ref. [163] is the fact that the set of amino acids over which the dynamical properties are automatically compared is not identified by sequence or structural alignments, but by matching the fluctuation (or mode) amplitude profile itself, as first envisaged by Keskin et al. [75].

A key ingredient of this comparative approach is the use of the isotropic Gaussian network model [8]. Because this phenomenological model does not possess the full rotational–translational invariance of the three-dimensional elastic networks, its essential dynamical spaces have a one-dimensional character. By restricting considerations to the one-dimensional profile of a single mode (or of the mean-square fluctuation) Tobi used a dynamics-based programming strategy to identify corresponding amino acids for various pairs of proteins.

Significant matches were reported for pairs of proteins with different overall structural organisation. Consistently with the isotropic character of the elastic network model, the lowest energy mode of these matching proteins typically exhibited a single node located approximately in the middle of the matching subchain, thus entailing a hinge-bending motion. This motion was prototypically illustrated in Ref. [163] for two pairs of entries: OPRTase (PDBid, 1s7o

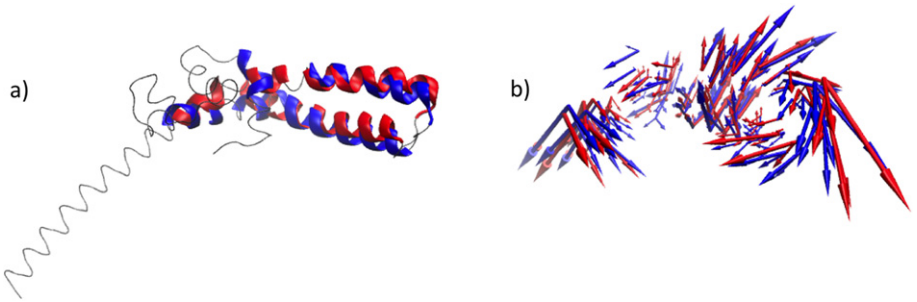

Fig. 8. Dynamics-based alignment of two OPRTase (PDBid: 1s7o chain A) and Mediator complex subunit 21 (PDBid: 1ykh chain A) discussed in Ref. [163]. The structural superposition of the aligned regions (ribbons) and three best-matching modes (arrows) are shown in panels (a) and (b), respectively. Aligned elements of OPRTase are shown in blue, while those of Mediator complex subunit 21 are shown in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

chain A) with Mediator complex subunit 21 (PDBid 1ykh chain A fragment 111–205) and Baseplate wedge protein 9 (PDBid, 1s2e chain A) with transcarboxylase (PDBid 1rqh chain A fragment 307–474).

Notably, the former of these two pairs has also a significant dynamics-based alignment according to the scheme of Zen et al. which employs a three-dimensional elastic network model as well as the integration of the dynamics of non-corresponding amino acids). The corresponding Aladyn alignment is shown in Fig. 8.

3.7. Comparison of the structural variability in a protein superfamily with the internal dynamics of its members

An interesting problem regards the extent to which evolutionary conformational drifts observed in proteins superfamilies occurs along the essential dynamical spaces of the family members.

This question was first posed by Leo-Macias et al. [83] who considered 35 representative protein families. For each family, the members were first structurally aligned to identify the common core and then a principal component analysis was carried our to obtain the main deformation modes. The latter were finally compared with the essential dynamical spaces obtained from elastic network models. The comparison of the two sets of spaces, which nowadays can be largely automated with the aid of bioinformatic tools such as ProDy [9], indicated a good mutual consistency.

The investigation of Leo-Macias et al. was recently extended by Velazquez-Muriel et al. [167] who considered a larger set of 55 families and used atomistic MD simulations. This study reported that the conformational space explored in MD simulations at constant-temperature has a smaller breadth than that spanned by known members of the same superfamily. However, the complexity of the explored space is significantly larger for MD simulations than for the internal variability of protein superfamilies. In this study the complexity was defined and measured as the minimal number of essential modes required to account for the same fraction of the global mean-square fluctuation of the superfamily or MD trajectory.

Based on these findings, Velazquez-Muriel et al. [167] concluded that the structural evolution of superfamilies has occurred in diverse and much richer ways than those kinetically accessible in thermal equilibrium to any of the superfamily members. Yet, such enhanced conformational variability was constrained in fewer generalised directions, compared to those that are *a priori* kinetically accessible.

These conclusions, in turn, prompted the speculation that the restrictions to the viable superfamily "conformational complexity" reflect the evolutionary pressure to preserve certain patterns of structural fluctuations/motion that cannot be arbitrarily modified without compromising dynamics-based aspects relevant to function. The effect was most evident for enzymes, where the largest restrictions of the conformational variability was observed [167].

The possibility that physics-based constraints may also promote the consistency of the evolutionary deformation modes and essential dynamical spaces was explored by Echave and coworkers in Refs. [35,36]

3.8. Dynamics-based alignment of proteins with different structure and function

We now report on the studies of Zen et al. [177] who carried out comparisons of the internal dynamics of a comprehensive set of 76 enzymes covering the six main functional groups (oxydoreductases, transferases, hydrolases, lyases, ligases).

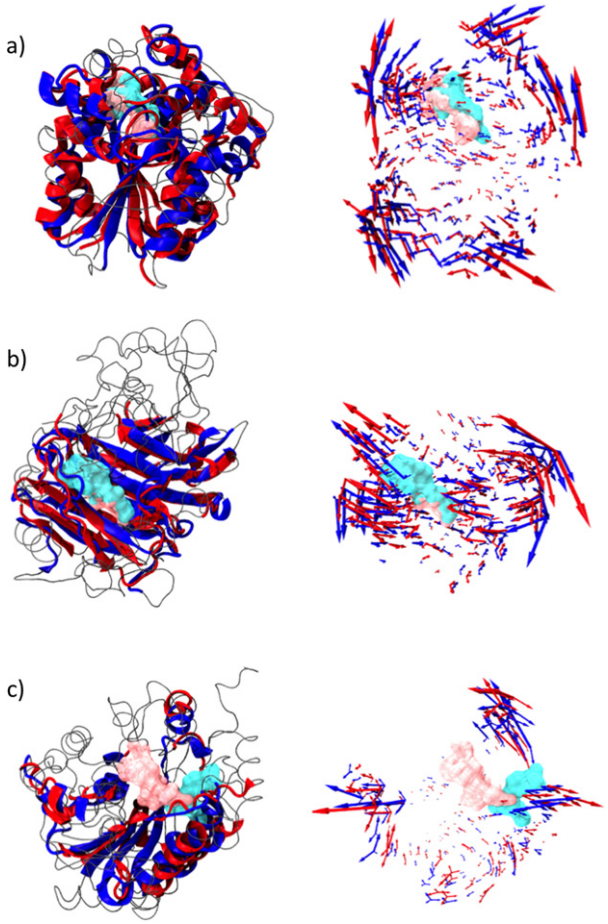

Fig. 9. Examples of significant dynamics-based alignments of proteins with different degree of structural and functional similarities (captured by the CATH code and primary EC number, respectively). The examples are taken from Ref. [177] and the alignments were produced with the Aladyn web-server. The aligned proteins in panel (a) have the same fold (they share the full CATH code) but have different function. The pair in panel (b) have the same function but different CATH architecture. The pair in panel (c) differ by CATH architecture and function. The pair in panel (a) involves a haloalkane dehalogenase (PDBid 2had, CATH: 3.40.50.1820, EC: 4) and an (s)-acetone-cyanohydrin lyase (PDBid: 1yb7, CATH: 3.40.50.1820, EC: 3). The pair in panel (b) involves a Cellobiohydrolase i (PDBid: 1dy4, CATH: 2.70.100.10, EC: 3) and a glucanase (PDBid: 2ayh, CATH: 2.60.120.200, EC: 3). The pair in panel (c) involves an exonuclease (PDBid: 1ako, CATH: 3.60.10.10, EC: 3) and an Enoyl-reductase (PDBid: 1d7o, CATH: 3.40.50.720, EC: 1). For each pair we report separately the structural superposition of the aligned regions (ribbons) and of the top three best-matching modes (arrows). Aligned elements are shown in blue for the first entry of the pair and in red for the second. The active sites are shown in cyan and pink for the first and second entry of the pair, respectively.

The analysis of Zen et al. was aimed at ascertaining whether similar functionally-oriented dynamical properties (arising from either evolutionary conservation or convergence) could be found in enzymes with major sequence and structure differences.

The study entailed the dynamics-based alignment (in the spirit of Section 2.7.1) of all the possible pairings of such enzymes. About 30 of such pairings were singled out as being outstanding for statistical significance. Two thirds of such pairings involved enzymes with detectable sequence homology or structural similarity as resulting by global or partial structural superposition using the DALI alignment program. One such example is offered by the pair 1yb7–2had which share the full CATH code, despite the different function. The dynamics-based alignment of this pair is shown in Fig. 9(a) where one can observe the remarkable structural superposition of the molecules' active sites.

Interestingly, the remaining third of the significant pairings involved entries whose structural relatedness was not significant by standard alignment criteria and occasionally involved enzymes with different function, i.e. different primary Enzyme Commission (EC) number.

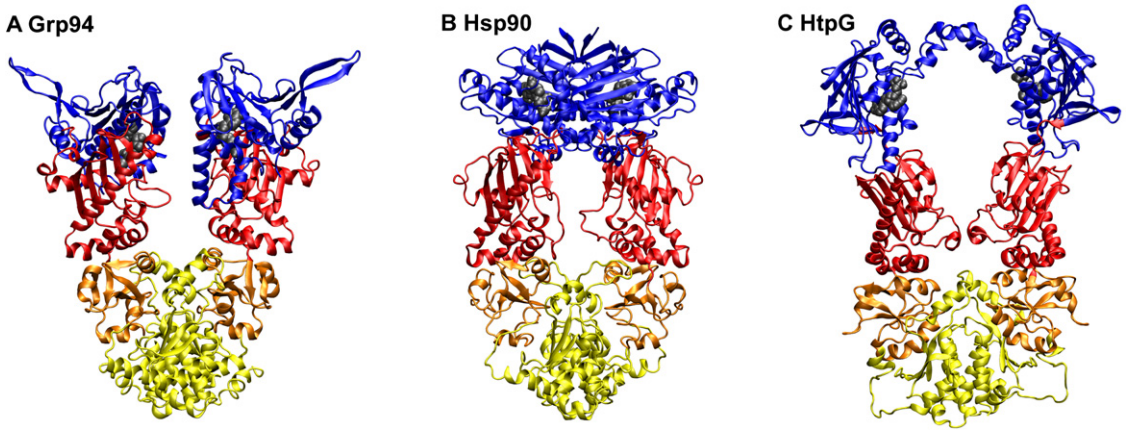

Fig. 10. Crystallographic structures of three Hsp90 conformers used in the comparative dynamics study of Ref. [109]. The structures correspond to: (A) canine ATP-bound Grp94 structure, PDBid: 2o1u; (B) yeast ATP-bound Hsp90 structure, PDBid: 2cg9; (C) HtpG structure, PDBid 2iop.pdb. Different colours are used to highlight the various structural subdomains: blue, N-terminal domains; red, M-large domains; orange, M-small domains; yellow, C-terminal domains. Reproduced from Fig. 1 of Ref. [109]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Two such pairs are respectively, 1dy4–2ayh and 1ako–1d7o, which are respectively shown in panels (b) and (c) of Fig. 9. It is seen that while the overall structural correspondence is limited (and in fact aligned regions can have different secondary structure content), the alignment reflects a very good consistency of the matching modes as well as the superposition of the known active regions.

As for the previously discussed case of proteases, the match of the latter and the fact that the matching modes entail the modulation of the region surrounding the active site, support the notion that common functionally-oriented dynamics-based properties can be detected in proteins that possibly differ by structure and even detailed catalytic chemistry [58,137].

3.9. Comparing large-scale movements of multidomain proteins

As anticipated at the end of Section 2.1, a particularly challenging case for characterising protein internal dynamics, as well as comparing it, is represented by proteins comprising mobile domains.

For such molecules, in fact, the relative displacements of the mobile domains can be so large that the motion is only poorly described by linearly superimposing a few essential modes onto a reference structure, see Fig. 3 in Ref. [151]. A familiar example is offered by the opening of a door: the larger the opening angle, the poorer the directional consistency of the initial displacement of the door's edge and the difference vector of the initial and final edge positions. As a consequence, the essential dynamical spaces calculated for a short trajectory, or by applying elastic network models on a specific protein conformer, can only limitedly capture and describe large-amplitude motions in such complexes. Furthermore, the very same calculation of essential dynamical spaces from extensive MD simulations can be problematic because they rely on the use of rigid-structural alignments which cannot well superimpose the visited conformers over all their amino acids.

At least for some proteins with mobile subdomains, using internal angular coordinates instead of Cartesian displacements can provide a viable alternative for describing the large-amplitude protein motion [90,98,118].

The fact that suitably-defined angular coordinates can be used for comparing the dynamics of proteins articulated in several domains was recently illustrated by Morra et al. in Ref. [109]. This study considered three homodimeric multidomain Hsp90 chaperones, namely mammalian Grp94, yeast Hsp90 and *E. coli* HtpG. The three chaperones, which are represented in Fig. 10, have a mutual sequence identity of $\sim$45% and most of their amino acids can be put into one-to-one correspondence by using flexible structural alignment [174].

The internal dynamics of the chaperones was characterised by extensive molecular dynamics simulations started from different initial conformers which differed by the presence and type of bound ligand. Next, to extract the large-

scale dynamical features that are shared by the chaperones, considerations were restricted to the extensive set of corresponding amino acids.

The motion of such set was found to be well approximated by the relative rigid-like movements of three quasi-rigid domains (similar, but not equal, to the structural ones). As a matter of fact, for all three chaperones it was possible to identify two consensus hinges and axes of motion controlling the rotation of the side-domains relative to the core of each protomer. Notably, one of the hinges (the one at the boundary of the N-terminal and Middle domain) occurs in correspondence of a site that had been previously shown to be important to chaperone functionality. In fact, it was validated as a as a potential target for Hsp90 inhibition [164,166]. Based on the detailed analysis of the same simulations carried out in Ref. [109] it was further concluded that an analogous role could be played by the site accommodating the second hinge.

The study of Ref. [109] therefore suggests that comparative dynamical analysis based on quasi-rigid protein domain movements could represent a promising avenue for identifying functional relationships in multidomain proteins and possibly protein complexes too.

3.10. A dynamics-based metric for protein space

We conclude the overview by reporting on the recent work of Hensen et al. [58] who considered a set of ∼100 proteins covering the main known folds and compared their structural features and especially a comprehensive series of dynamical observables calculated from 100-ns long atomistic MD simulations. In particular, to each protein entry, Hensen et al. associated a dynamical "fingerprint" consisting of a multidimensional array whose components were dynamics-based scalars. These scalar quantities included the spread of the essential dynamics eigenvalues, the roughness of the free energy landscape, the root-mean-square deviation from the crystallographic structure, the root-mean-square fluctuations from the average structure etc.

At variance with the studies mentioned earlier, which aimed at detecting detailed dynamical correspondences among proteins, the investigation of Ref. [58] was mostly targeted to establishing the overall features of the space spanned by the dynamical fingerprints. In particular, Hensen et al. meant to introduce a dynamics-based metric to explore the occupation of the fingerprint space (termed the "dynasome space") and understand e.g. whether structurally or functionally similar proteins can be clustered.

From this survey, the authors concluded that in the considered dynamical space, proteins are not partitioned in distinct clusters but are distributed rather continuously. This interesting aspect therefore parallels the findings of recent studies which support the view that structural properties cover a continuum rather than a discrete succession of conformers [121,149,172,180].

The analysis has further revealed the strong connection between dynamical and structural similarities, consistently with the studies, mentioned earlier in this review, where the structural relatedness has been frequently associated to strong dynamical implications.

It is interesting to observe that, as in the study of Zen et al. [177] described in the previous section, the analysis of Hensen et al. [58] has highlighted the possible existence of appreciable dynamical similarities in proteins with limited structural relatedness. The example offered by the authors pertained to the pairing of two hydrolases, serralysin and rhizopuspepsin (PDB codes 1sat and 2apr). Their structural alignment is non-significant according to DALI statistical criteria while in the dynamic metric space considered by the authors they have a strong dynamical proximity. Consistently with this finding the Aladyn alignment of this pair, which involves 79 amino acids, is statistically significant too as the observed RMSIP = 0.66 and the associated p-value is 0.025.

Finally, by examining the dynamic fingerprint of functionally-related proteins Hensen et al. [58] concluded that it ought to be possible to reliably establish and assign proteins function based on their neighbours in the metric dynamic space. Indeed, the possibility to carry out functional assignments on the basis of dynamics-based data represents a very interesting avenue with several practical ramifications.

As a related issue we report that pairwise dynamics-based alignments have been previously carried out with the purpose of predicting the active site of proteins for which standard homology-based approaches are not applicable. In particular, this approach was undertaken to predict the nucleic-acids binding sites of proteins adopting non-canonical OB-folds, as discussed in Ref. [178].

4. Conclusions

Over the past decades, several bioinformatics tools and computational methods have been introduced and systematically applied to clarify aspects of the relationship between structure and dynamics for protein and enzymes.

Many such studies contributed to clarifying how the interplay of structure and internal dynamics of various proteins impacts their biological functionality. The latter, in fact, is often – though not always – associated with the innate capability of these biomolecules to sustain concerted, large-scale conformational changes so to bind ligands, change oligomeric state etc.

In recent years, besides the well-established approach of dissecting such properties for specific, individual proteins and enzymes, there has been a growing interest for comparative studies of proteins' internal dynamics.

In such studies, covered by this review, the key dynamics-based properties of proteins are singled out by identifying those features (such as essential dynamical spaces, mean square fluctuation profiles, relaxation times etc.) that are shared by proteins with different degrees of sequence, structure and functional similarities.

Such comparative investigations have been carried out with two main purposes: characterising functionally-oriented mechanisms for specific groups of proteins and understanding the more general organisation of the "protein universe" by complementing the sequence and structural perspectives with a dynamics-based one.

For the first objective, detailed comparative tools have been developed, including the so-called dynamics-based alignments which use dynamics-based properties to establish one-to-one correspondences of amino acids in different proteins. These strategies have been used to identify common hinge-bending motions in multidomain proteins, to complement sequence and structural alignment in singling out functionally-relevant regions in proteins with different degrees of homologies, and to highlight common large-scale movements in proteins that differ significantly by fold and/or function.

The latter aspect, is tightly connected to the second objective, namely the development and use of dynamics-based criteria to trace elusive evolutionary relationships and group/classify proteins by their internal dynamics [17,50,51]. This perspective has been pursued so far to highlight the degree of conservation of the amplitude of amino acid fluctuations in protein families and superfamilies, to clarify the extent to which the structural variations accumulated within protein superfamilies have occurred along the "innate" directions of structural fluctuations of its members, and even to introduce a metric to quantify how evenly are proteins distributed in a generalised dynamics-space. The latter perspective can have important implications for functional assignment.

In conclusion, the valuable findings provided by the recent introduction of methods for comparing detailed or general dynamical properties of proteins suggest that they could be profitably used in conjunction with classic comparative methods to characterise proteins at the various steps of the sequence → structure → function ladder.

Arguably, the progress towards this goal would be greatly aided by the development of unsupervised methods to single out those dynamical features that are more likely attributed to the biological functionality of a given protein and by the more systematic investigation of evolutionary relationships from a detailed dynamics-based perspective.

Acknowledgements

I am grateful to P. Agarwal, G. Bussi, V. Carnevale, J. Echave, A. Finkelstein, H. Gruebmueller, R. Jernigan, A. Lesk, D. Tobi and A. Zen for very valuable discussions. We acknowledge support from the Italian Ministry of Education, grant PRIN 2010HXAW77.

References

[1] Agarwal PK, Billeter SR, Rajagopalan PT, Benkovic SJ, Hammes-Schiffer S. Network of coupled promoting motions in enzyme catalysis. Proc Natl Acad Sci USA 2002;99:2794–9.

[2] Aleksiev T, Potestio R, Pontiggia F, Cozzini S, Micheletti C. PiSQRD: a web server for decomposing proteins into quasi-rigid dynamical domains. Bioinformatics 2009;25:2743–4.

[3] Alexandrov V, Lehnert U, Echols N, Milburn D, Engelman D, Gerstein M. Normal modes for predicting protein motions: a comprehensive database assessment and associated Web tool. Protein Sci 2005;14:633–43.

[4] Amadei A, Ceruso MA, Di Nola A. On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations. Proteins 1999;36:419–24.

[5] Amadei A, Linssen ABM, Berendsen HJC. Essential dynamics of proteins. Proteins 1993;17:412–25.

[6] Andreeva A, Howorth D, Chandonia JM, Brenner SE, Hubbard TJ, Chothia C, et al. Data growth and its impact on the SCOP database: new developments. Nucleic Acids Res 2008;36:419–25.
[7] Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J 2001;80:505–15.
[8] Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a single parameter harmonic potential. Fold Des 1997;2:173–81.
[9] Bakan A, Meireles LM, Bahar I. Prody: protein dynamics inferred from theory and experiments. Bioinformatics 2011;27:1575–7.
[10] Bartlett GJ, Borkakoti N, Thornton JM. Catalysing new reactions during evolution: economy of residues and mechanism. J Mol Biol 2003;331:829–60.
[11] Bavro VN, De Zorzi R, Schmidt MR, Muniz JR, Zubcevic L, Sansom MS, et al. Structure of a KirBac potassium channel with an open bundle crossing indicates a mechanism of channel gating. Nat Struct Mol Biol 2012;19:158–63.
[12] Bhabha G, Lee J, Ekiert DC, Gam J, Wilson IA, Dyson HJ, et al. A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. Science 2011;332:234–8.
[13] Blundell T, Srinivasan N. Symmetry, stability, and dynamics of multidomain and multicomponent protein systems. Proc Natl Acad Sci USA 1996;93:14243–8.
[14] Boehr DD, Dyson HJ, Wright PE. Conformational relaxation following hydride transfer plays a limiting role in dihydrofolate reductase catalysis. Biochemistry 2008;47:9227–33.
[15] Brooks BR, Janezic D, Karplus M. Harmonic analysis of large systems I. Methodology. J Comput Chem 1995;16(12):1522–42.
[16] Camps J, Carrillo O, Emperador A, Orellana L, Hospital A, Rueda M, et al. FlexServ: an integrated tool for the analysis of protein flexibility. Bioinformatics 2009;25:1709–10.
[17] Capozzi F, Luchinat C, Micheletti C, Pontiggia F. Essential dynamics of helices provide a functional classification of EF-hand proteins. J Proteome Res 2007;6:4245–55.
[18] Carnevale V, Pontiggia F, Micheletti C. Structural and dynamical alignment of enzymes with partial structural similarity. J Phys Condens Matter 2007;19:285206.
[19] Carnevale V, Raugei S, Micheletti C, Carloni P. Convergent dynamics in the protease enzymatic superfamily. J Am Chem Soc 2006;128:9766–72.
[20] Carnevale V, Raugei S, Micheletti C, Carloni P. Large-scale motions and electrostatic properties of furin and HIV-1 protease. J Phys Chem A 2007;111:12327–32.
[21] Cascella M, Micheletti C, Rothlisberger U, Carloni P. Evolutionarily conserved functional mechanics across pepsin-like and retroviral aspartic proteases. J Am Chem Soc 2005;127:3734–42.
[22] Cascella M, Micheletti C, Rothlisberger U, Carloni P. Evolutionarily conserved functional mechanics across pepsin-like and retroviral aspartic proteases. J Am Chem Soc 2005;127:3734–42.
[23] Chandonia JM, Hon G, Walker NS, Lo Conte L, Koehl P, Levitt M, et al. The astral compendium in 2004. Nucleic Acids Res 2004;32:189–92.
[24] Chennubhotla C, Bahar I. Signal propagation in proteins and relation to equilibrium fluctuations. PLoS Comput Biol 2007;3:1716–26.
[25] Chi CN, Elfström L, Shi Y, Snäll T, Engström A, Jemth P. Reassessing a sparse energetic network within a single protein domain. Proc Natl Acad Sci USA 2008;105:4679–84.
[26] Chothia C. One thousand families for the molecular biologist. Nature 1992;357:543–4.
[27] Chothia C, Finkelstein AV. The classification and origins of protein folding patterns. Annu Rev Biochem 1990;59:1007–39.
[28] Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. EMBO J 1986;5:823–6.
[29] Creighton T. Proteins, structure and molecular properties. 2nd edition. New York: W.H. Freeman and Company; 1993.
[30] Das A, Mahale S, Prashar V, Bihani S, Ferrer JL, Hosur MV. X-ray snapshot of HIV-1 protease in action: observation of tetrahedral intermediate and short ionic hydrogen bond SIHB with catalytic aspartate. J Am Chem Soc 2010;132:6366–73.
[31] De los Rios P, Cecconi F, Pretre A, Dietler G, Michielin O, Piazza F, et al. Functional dynamics of PDZ binding domains: a normal-mode analysis. Biophys J 2005;89:14–21.
[32] del Sol A, Tsai CJ, Ma B, Nussinov R. The origin of allosteric functional modulation: multiple pre-existing pathways. Structure 2009;17:1042–50.
[33] Delarue M, Sanejouand YH. Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: the elastic network model. J Mol Biol 2002;320:1011–24.
[34] DePristo MA, Weinreich DM, Hartl DL. Missense meanderings in sequence space: a biophysical view of protein evolution. Nat Rev Genet 2005;6:678–87.
[35] Echave J. Why are the low-energy protein normal modes evolutionarily conserved? Pure Appl Chem 2012;84:1931–7.
[36] Echave J, Fernández FM. A perturbative view of protein structural variation. Proteins 2010;78:173–80.
[37] Eisenmesser EZ, Bosco DA, Akke M, Kern D. Enzyme dynamics during catalysis. Science 2002;295:1520–3.
[38] Eisenmesser EZ, Millet O, Labeikovsky W, Korzhnev DM, Wolf-Watz M, Bosco DA, et al. Intrinsic dynamics of an enzyme underlies catalysis. Nature 2005;438:117–21.
[39] Engel A, Gaub HE. Structure and mechanics of membrane proteins. Annu Rev Biochem 2008;77:127148.
[40] Falke JJ. Enzymology: a moving story. Science 2002;295:1480–1.
[41] Fanning AS, Anderson JM. PDZ domains: fundamental building blocks in the organization of protein complexes at the plasma membrane. J Clin Invest 1999;103:767–72.
[42] Fersht AR. Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding. New York: W.H. Freeman; 1999.
[43] Finkelstein A. Protein physics. Singapore: World Scientific–Academic Press; 2002.
[44] Fleming K, Kelley LA, Islam SA, MacCallum RM, Muller A, Pazos F, et al. The proteome: structure, function and evolution. Philos Trans R Soc Lond B Biol Sci 2006;361:441–51.

[45] Frauenfelder H, Parak F, Young RD. Conformational substates in proteins. Annu Rev Biophys Biophys Chem 1988;17:451–79.
[46] Frauenfelder H, Sligar SG, Wolynes PG. The energy landscapes and motions of proteins. Science 1991;254:1598–603.
[47] Fuglebakk E, Echave J, Reuter N. Measuring and comparing structural fluctuation patterns in large protein datasets. Bioinformatics 2012. http://dx.doi.org/10.1093/bioinformatics/bts445.
[48] Garcia A. Large-amplitude nonlinear motions in proteins. Phys Rev Lett 1992;68:2696–9.
[49] Gerek ZN, Ozkan SB. Change in allosteric network affects binding affinities of PDZ domains: analysis through perturbation response scanning. PLoS Comput Biol 2011;7:e1002154.
[50] Gerstein M, Krebs W. A database of macromolecular motions. Nucleic Acids Res 1998;26:4280–90.
[51] Gerstein M, Lesk AM, Chothia C. Structural mechanisms for domain movements in proteins. Biochemistry 1994;33:6739–49.
[52] Glembo TJ, Farrell DW, Gerek ZN, Thorpe MF, Ozkan SB. Collective dynamics differentiates functional divergence in protein evolution. PLoS Comput Biol 2012;8:e1002428.
[53] Go N, Noguti T, Nishikawa T. Dynamics of a small globular protein in terms of low-frequency vibrational modes. Proc Natl Acad Sci USA 1983;80:3696–700.
[54] Golhlke H, Thorpe MF. A natural coarse graining for simulating large biomolecular motion. Biophys J 2006;91:2115–20.
[55] Hammes-Schiffer S, Benkovic SJ. Relating protein motion to catalysis. Annu Rev Biochem 2006;75:519–41.
[56] Hayward S, Kitao A, Berendsen HJC. Model-free methods of analyzing domain motions in proteins from simulation: a comparison of normal mode analysis and molecular dynamics simulation of lysozyme. Proteins 1997;27:425–37.
[57] Hayward S, Kitao A, Go N. Harmonic and anharmonic aspects in the dynamics of BPTI: a normal mode analysis and principal component analysis. Protein Sci 1994;3:936–43.
[58] Hensen U, Meyer T, Haas J, Rex R, Vriend G, Grubmüller H. Exploring protein dynamics space: the dynasome as the missing link between protein structure and function. PLoS ONE 2012;7:e33931.
[59] Henzler-Wildman K, Kern D. Dynamic personalities of proteins. Nature 2007;450:964–72.
[60] Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M, Kern D. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. Nature 2007;450:913–6.
[61] Henzler-Wildman KA, Thai V, Lei M, Ott M, Wolf-Watz M, Fenn T, et al. Intrinsic motions along an enzymatic reaction trajectory. Nature 2007;450:838–44.
[62] Hess B. Convergence of sampling in protein simulations. Phys Rev E 2002;65:031910.
[63] Hinsen K. Analysis of domain motions by approximate normal mode calculations. Proteins 1998;33:417–29.
[64] Hinsen K, Kneller GR. Solvent effects in the slow dynamics of proteins. Proteins 2008;70:1235–42.
[65] Hinsen K, Petrescu AJ, Dellerue S, Bellisent-Funel MC, Kneller G. Harmonicity in slow protein dynamics. Chem Phys 2000;261:25–37.
[66] Hinsen K, Thomas A, Field MJ. Analysis of domain motion in large proteins. Proteins 1999;34:369–82.
[67] Holm L, Park J. Dalilite workbench for protein structure comparison. Bioinformatics 2000;16:566–7.
[68] Holm L, Sander C. The FSSP database of structurally aligned protein fold families. Nucleic Acids Res 1994;22:3600–9.
[69] Holm L, Sander C. Mapping the protein universe. Science 1996;273:595–603.
[70] Humphrey W, Dalke A, Schulten K. VMD – visual molecular dynamics. J Mol Graph 1996;14:33–8.
[71] Jackson CJ, Foo JL, Tokuriki N, Afriat L, Carr PD, Kim HK, et al. Conformational sampling, catalysis, and evolution of the bacterial phosphotriesterase. Proc Natl Acad Sci USA 2009;106:21631–6.
[72] Janezic D, Brooks BR. Harmonic analysis of large systems ii. comparison of different protein models. J Comput Chem 1995;16(12):1543–53.
[73] Janezic D, Venable R, Brooks BR. Harmonic analysis of large systems III. Comparison with molecular dynamics. J Comput Chem 1995;16(12):1544–56.
[74] Kamath G, Howell EE, Agarwal PK. The tail wagging the dog: insights into catalysis in R67 dihydrofolate reductase. Biochemistry 2010;49:9078–88.
[75] Keskin O, Jernigan RL, Bahar I. Proteins with similar architecture exhibit similar large-scale dynamic behavior. Biophys J 2000;78:2093–106.
[76] Kitao A, Hayward S, Go N. Energy landscape of a native protein: jumping-among-minima model. Proteins 1998;33:496–517.
[77] Kong Y, Karplus M. Signaling pathways of PDZ2 domain: a molecular dynamics interaction correlation analysis. Proteins 2009;74:145–54.
[78] Krebs W, Alexandrov V, Wilson C, Echols N, Yu H, Gerstein M. Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistics. Proteins 2002;48:682–95.
[79] Kundu S, Sorensen DC, Phillips GN. Automatic domain decomposition of proteins by a Gaussian network model. Proteins 2004;57:725–33.
[80] Kurkcuoglu O, Jernigan RL, Doruker P. Loop motions of triosephosphate isomerase observed with elastic networks. Biochemistry 2006;45:1173–82.
[81] Orellana L, Rueda M, Ferrer-Costa C, Lopez-Blanco JR, Chacon P, Orozco M. Approaching elastic network models to atomistic molecular dynamics. J Chem Theor Comput 2010;6:2910–23.
[82] Law AB, Fuentes EJ, Lee AL. Conservation of side-chain dynamics within a protein family. J Am Chem Soc 2009;131:6322–3.
[83] Leo-Macias A, Lopez-Romero P, Lupyan D, Zerbino D, Ortiz AR. An analysis of core deformations in protein superfamilies. Biophys J 2005;88:1291–9.
[84] Lesk AM. Introduction to protein science: architecture, function and genomics. UK: Oxford University Press; 2004.
[85] Levitt M, Sander C, Stern PS. Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. J Mol Biol 1985;181:423–47.
[86] Levy RM, Srinivasan AR, Olson WK, McCammon JA. Quasi-harmonic method for studying very low frequency modes in proteins. Biopolymers 1984;23:1099–112.
[87] Liu Y, Bahar I. Sequence evolution correlates with structural dynamics. Mol Biol Evol 2012;29:2253–63.

[88] Liu L, Gronenborn AM, Bahar I. Longer simulations sample larger subspaces of conformations while maintaining robust mechanisms of motion. Proteins 2011;80:616625.
[89] Lockless SW, Ranganathan R. Evolutionarily conserved pathways of energetic connectivity in protein families. Science 1999;286:295–9.
[90] Lopéz-Blanco JR, Garzón JI, Chacón P. iMod: multipurpose normal mode analysis in internal coordinates. Bioinformatics 2011;27:2843–50.
[91] Lu M, Poon B, Ma J. A new method for coarse-grained elastic normal-mode analysis. J Chem Theory Comput 2006;2(3):464–71.
[92] Maguid S, Fernandez-Alberti S, Echave J. Evolutionary conservation of protein vibrational dynamics. Gene 2008;422:7–13.
[93] Maguid S, Fernandez-Alberti S, Ferrelli L, Echave J. Exploring the common dynamics of homologous proteins. Application to the globin family. Biophys J 2005;89:3–13.
[94] Maguid S, Fernández-Alberti S, Parisi G, Echave J. Evolutionary conservation of protein backbone flexibility. J Mol Evol 2006;63:448–57.
[95] Maritan A, Micheletti C, Banavar JR. Role of secondary motifs in fast folding polymers: a dynamical variational principle. Phys Rev Lett 2000;84:3009–12.
[96] Maritan A, Micheletti C, Trovato A, Banavar JR. Optimal shapes of compact strings. Nature 2000;406(6793):287–90.
[97] McCammon JA, Gelin BR, Karplus M, Wolynes PG. The hinge-bending mode in lysozyme. Nature 1976;262:325–6.
[98] Mendez R, Bastolla U. Torsional network model: normal modes in torsion angle space better correlate with conformation changes in proteins. Phys Rev Lett 2010;104:228103.
[99] Meyerguz L, Kleinberg J, Elber R. The network of sequence flow between protein structures. Proc Natl Acad Sci USA 2007;104:11627–32.
[100] Micheletti C, Banavar JR, Maritan A. Conformations of proteins in equilibrium. Phys Rev Lett 2001;87:088102.
[101] Micheletti C, Banavar JR, Maritan A, Seno F. Protein structures and optimal folding from a geometrical variational principle. Phys Rev Lett 1999;82:3372–5.
[102] Micheletti C, Carloni P, Maritan A. Accurate and efficient description of protein vibrational dynamics: comparing molecular dynamics and Gaussian models. Proteins 2004;55:635–45.
[103] Micheletti C, Orland H. Mistral: a tool for energy-based multiple structural alignment of proteins. Bioinformatics 2009;25:2663–9.
[104] Min W, Luo G, Cherayil BJ, Kou SC, Xie XS. Observation of a power-law memory kernel for fluctuations within a single protein molecule. Phys Rev Lett 2005;94:198302.
[105] Ming D, Wall ME. Allostery in a coarse-grained model of protein dynamics. Phys Rev Lett 2005;95:198103.
[106] Miyashita O, Onuchic JN, Wolynes PG. Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. Proc Natl Acad Sci USA 2003;100:12570–5.
[107] Miyashita O, Wolynes PG, Onuchic JN. Simple energy landscape model for the kinetics of functional transitions in proteins. J Phys Chem B 2005;109:1959–69.
[108] Mizuguchi K, Deane CM, Blundell TL, Overington JP. HOMSTRAD: a database of protein structure alignments for homologous families. Protein Sci 1998;7:2469–71.
[109] Morra G, Potestio R, Micheletti C, Colombo G. Corresponding functional dynamics across the Hsp90 chaperone family: insights from a multiscale analysis of MD simulations. PLoS Comput Biol 2012;8.
[110] Morra G, Verkhivker G, Colombo G. Modeling signal propagation mechanisms and ligand-based conformational dynamics of the Hsp90 molecular chaperone full-length dimer. PLoS Comput Biol 2009;5:e1000323.
[111] Münz M, Hein J, Biggin PC. The role of flexibility and conformational selection in the binding promiscuity of PDZ domains. PLoS Comput Biol 2012;8:e1002749.
[112] Münz M, Lyngso R, Hein J, Biggin PC. Dynamics based alignment of proteins: an alternative approach to quantify dynamic similarity. BMC Bioinform 2010;11:188.
[113] Murzin A, Brenner S, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995;247:536–40.
[114] Nashine VC, Hammes-Schiffer S, Benkovic SJ. Coupled motions in enzyme catalysis. Curr Opin Chem Biol 2010;14:644–51.
[115] Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 1970;48:443–53.
[116] Ojha S, Meng EC, Babbitt PC. Evolution of function in the "two dinucleotide binding domains" flavoproteins. PLoS Comput Biol 2007;3(7):e121.
[117] Orengo CA, Thornton JM. Protein families and their evolution—a structural perspective. Annu Rev Biochem 2005;74:867–900.
[118] Orozco M, Orellana L, Hospital A, Naganathan AN, Emperador A, Carrillo O, et al. Coarse-grained representation of protein flexibility. Foundations, successes, and shortcomings. Adv Protein Chem Struct Biol 2011;85:183–215.
[119] Ortiz AR, Strauss CE, Olmea O. Mammoth (matching molecular models obtained from theory): an automated method for model comparison. Protein Sci 2002;11:2606–21.
[120] Pang A, Arinaminpathy Y, Sansom MSP, Biggin PC. Comparative molecular dynamics—similar folds and similar motions? Proteins 2005;61:809–22.
[121] Pascual-García A, Abia D, Ortiz AR, Bastolla U. Cross-over between discrete and continuous protein structure space: insights into automatic classification and networks of protein structures. PLoS Comput Biol 2009;5:e1000331.
[122] Pearl F, Todd A, Sillitoe I, Dibley M, Redfern O, Lewis T, et al. The CATH domain structure database and related resources Gene3D and DHS provide comprehensive domain family information for genome analysis. Nucl Acids Res 2005;33:D247–51.
[123] Pegg SC, Brown SD, Ojha S, Seffernick J, Meng EC, Morris JH, et al. Leveraging enzyme structure–function relationships for functional inference and experimental design: the structure–function linkage database. Biochemistry 2006;45:2545–55.
[124] Perryman AL, Lin J-H, McCammon JA. HIV-1 protease molecular dynamics of a wild-type and of the V82F/I84V mutant: possible contributions to drug resistance and a potential new target site for drugs. Protein Sci 2004;13:1108–23.
[125] Piana S, Carloni P, Parrinello M. Role of conformational fluctuations in the enzymatic reaction of HIV-1 protease. J Mol Biol 2002;319:567–83.

[126] Piana S, Carloni P, Rothlisberger U. Drug resistance in HIV-1 protease: flexibility-assisted mechanism of compensatory mutations. Protein Sci 2002;11:2393–402.
[127] Pontiggia F, Colombo G, Micheletti C, Orland H. Anharmonicity and self-similarity of the free energy landscape of protein G. Phys Rev Lett 2007;98:048102.
[128] Pontiggia F, Zen A, Micheletti C. Small- and large-scale conformational changes of adenylate kinase: a molecular dynamics study of the subdomain motion and mechanics. Biophys J 2008;95:5901–12.
[129] Porter CT, Bartlett GJ, Thornton JM. The catalytic site atlas: a resource of catalytic sites and residues identified in enzymes using structural data. Nucl Acids Res 2004;32:D129–33.
[130] Potestio R, Aleksiev T, Pontiggia F, Cozzini S, Micheletti C. Aladyn: a web server for aligning proteins by matching their large-scale motion. Nucleic Acids Res 2010;38:41–5.
[131] Potestio R, Pontiggia F, Micheletti C. Coarse-grained description of protein internal dynamics: an optimal strategy for decomposing proteins in rigid subunits. Biophys J 2009;96:4993–5002.
[132] Provasi D, Artacho MC, Negri A, Mobarec JC, Filizola M. Ligand-induced modulation of the free-energy landscape of g protein-coupled receptors explored by adaptive biasing techniques. PLoS Comput Biol 2011;7:e1002193.
[133] Pugalenthi G, Bhaduri A, Sowdhamini R. GenDiS: genomic distribution of protein structural domain superfamilies. Nucleic Acids Res 2005;33:252–5.
[134] Pérez A, Blas JR, Rueda M, López-Bes JM, de la Cruz X, Orozco M. Exploring the essential dynamics of b-B-DNA. J Chem Theory Comput 2005;1:790–800.
[135] Quesada V, Ordóñez GR, Sánchez LM, Puente XS, López-Otín C. The Degradome database: mammalian proteases and diseases of proteolysis. Nucleic Acids Res 2009;37:239–43.
[136] Radkiewicz JL, Zipse H, Clarke S, Houk KN. Neighboring side chain effects on asparaginyl and aspartyl degradation: an ab initio study of the relationship between peptide conformation and backbone NH acidity. J Am Chem Soc 2001;123:3499–506.
[137] Ramanathan A, Agarwal PK. Evolutionarily conserved linkage between enzyme fold, flexibility, and catalysis. PLoS Biol 2011;9:e1001193.
[138] Ramanathan A, Savol AJ, Langmead CJ, Agarwal PK, Chennubhotla CS. Discovering conformational sub-states relevant to protein function. PLoS ONE 2011;6:e15827.
[139] Ranea JA, Sillero A, Thornton JM, Orengo CA. Protein superfamily evolution and the last universal common ancestor (LUCA). J Mol Evol 2006;63:513–25.
[140] Rawlings ND, Barrett AJ, Bateman A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res 2012;40:343–50.
[141] Rod TH, Radkiewicz JL, Brooks CL. Correlated motion and the effect of distal mutations in dihydrofolate reductase. Proc Natl Acad Sci USA 2003;100:6980–5.
[142] Romo TD, Grossfield A. Validating and improving elastic network models with molecular dynamics simulations. Proteins 2011;79:2334.
[143] Rueda M, Chacón P, Orozco M. Thorough validation of protein normal mode analysis: a comparative study with essential dynamics. Structure 2007;15:565–75.
[144] Rueda M, Ferrer-Costa C, Meyer T, Pérez A, Camps J, Hospital A, et al. A consensus view of protein dynamics. Proc Natl Acad Sci USA 2007;104:796–801.
[145] Sachs JN, Engelman DM. Introduction to the membrane protein reviews: the interplay of structure, dynamics, and environment in membrane protein function. Annu Rev Biochem 2006;75:707712.
[146] Sacquin-Mora S, Laforet E, Lavery R. Locating the active sites of enzymes using mechanical properties. Proteins 2007;67:350–9.
[147] Shakhnovich BE, Deeds E, Delisi C, Shakhnovich E. Protein structure and evolutionary history determine sequence space topology. Genome Res 2005;15:385–92.
[148] Shatsky M, Nussinov R, Wolfson H. A method for simultaneous alignment of multiple protein structures. Proteins 2004;56:143–56.
[149] Skolnick J, Arakaki AK, Lee SY, Brylinski M. The continuity of protein structure space is an intrinsic property of proteins. Proc Natl Acad Sci USA 2009;106:15690–5.
[150] Smith GR, Sternberg MJE, Bates PA. The relationship between the flexibility of proteins and their conformational states on forming protein–protein complexes with an application to protein–protein docking. J Mol Biol 2005;347:1077–101.
[151] Song G, Jernigan RL. An enhanced elastic network model to represent the motions of domain-swapped proteins. Proteins 2006;63:197–209.
[152] Songyang Z, Fanning AS, Fu C, Xu J, Marfatia SM, Chishti AH, et al. Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. Science 1997;275:73–7.
[153] Southan C. Assessing the protease and protease inhibitor content of the human genome. J Pept Sci 2000;6:453–8.
[154] Stebbings LA, Mizuguchi K. HOMSTRAD: recent developments of the homologous protein structure alignment database. Nucleic Acids Res 2004;32:203–7.
[155] Stein A, Rueda M, Panjkovich A, Orozco M, Aloy P. A systematic study of the energetics involved in structural changes upon association and connectivity in protein interaction networks. Structure 2011;19:881–9.
[156] Suel GM, Lockless SW, Wall MA, Ranganathan R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. Nat Struct Biol 2003;10:59–69.
[157] Sulkowska J, Kloczkowski A, Sen T, Cieplak M, Jernigan R. Predicting the order in which contacts are broken during single molecule protein stretching experiments. Proteins 2007;71:45–60.
[158] Tama F, Sanejouand YH. Conformational change of proteins arising from normal mode calculations. Protein Eng 2001;14:1–6.
[159] Tang J, James MN, Hsu IN, Jenkins JA, Blundell TL. Structural evidence for gene duplication in the evolution of the acid proteases. Nature 1978;271:618–21.
[160] Teilum K, Olsen JG, Kragelund BB. Functional aspects of protein flexibility. Cell Mol Life Sci 2009;66:2231–47.

[161] Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL-X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 1997;25:4876–82.
[162] Tirion MM. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. Phys Rev Lett 1996;77:1905–8.
[163] Tobi D. Dynamics alignment: comparison of protein dynamics in the SCOP database. Proteins 2012;80:1167–76.
[164] Tsutsumi S, Mollapour M, Graf C, Lee CT, Scroggins BT, Xu W, et al. Hsp90 charged-linker truncation reverses the functional consequences of weakened hydrophobic contacts in the n domain. Nat Struct Mol Biol 2009;16:1141–7.
[165] Tyndall JD, Nall T, Fairlie DP. Proteases universally recognize beta strands in their active sites. Chem Rev 2005;105:973–99.
[166] Vasko RC, Rodriguez RA, Cunningham CN, Ardi VC, Agard DA, McAlpine SR. Mechanistic studies of Sansalvamide A-amide: an allosteric modulator of Hsp90. ACS Med Chem Lett 2010;1:4–8.
[167] Velázquez-Muriel JA, Rueda M, Cuesta I, Pascual-Montano A, Orozco M, Carazo JM. Comparison of molecular dynamics and superfamily spaces of protein domain deformation. BMC Struct Biol 2009;9:6.
[168] Vishveshwara S, Ghosh A, Hansia P. Intra and inter-molecular communications through protein structure network. Curr Protein Pept Sci 2009;10:146–60.
[169] Weinreich DM, Delaney NF, Depristo MA, Hartl DL. Darwinian evolution can follow only very few mutational paths to fitter proteins. Science 2006;312:111–4.
[170] Williams SG, Lovell SC. The effect of sequence evolution on protein structural divergence. Mol Biol Evol 2009;26:1055–65.
[171] Wolynes PG, Onuchic JN, Thirumalai D. Navigating the folding routes. Science 1995;267:1619–20.
[172] Xie L, Bourne PE. Detecting evolutionary relationships across existing fold space, using sequence order-independent profile–profile alignments. Proc Natl Acad Sci USA 2008;105:5441–6.
[173] Yang H, Luo G, Karnchanaphanurach P, Louie T, Rech I, Cova S, et al. Protein conformational dynamics probed by single-molecule electron transfer. Science 2003;302:262–6.
[174] Ye Y, Godzik A. FATCAT: a web server for flexible structure comparison and structure similarity searching. Nucleic Acids Res 2004;32:582–5.
[175] Yesylevskyy SO, Kharkyanen VN, Demchenko AP. Hierarchical clustering of correlation patterns: new method of domain identification in proteins. Biophys Chem 2006;119:84–93.
[176] Zeldovich KB, Shakhnovich EI. Understanding protein evolution: from protein physics to Darwinian selection. Annu Rev Phys Chem 2008;59:105–27.
[177] Zen A, Carnevale V, Lesk AM, Micheletti C. Correspondences between low-energy modes in enzymes: dynamics-based alignment of enzymatic functional families. Protein Sci 2008;17:918–29.
[178] Zen A, de Chiara C, Pastore A, Micheletti C. Using dynamics-based comparisons to predict nucleic acid binding sites in proteins: an application to ob-fold domains. Bioinformatics 2009;25:1876–83.
[179] Zen A, Micheletti C, Keskin O, Nussinov R. Comparing interfacial dynamics in protein–protein complexes: an elastic network approach. BMC Struct Biol 2010;10:26.
[180] Zhang Y, Hubner IA, Arakaki AK, Shakhnovich E, Skolnick J. On the origin and highly likely completeness of single-domain protein structures. Proc Natl Acad Sci USA 2006;103:2605–10.
[181] Zheng W, Brooks B, Thirumalai D. Allosteric transitions in the chaperonin GroEL are captured by a dominant normal mode that is most robust to sequence variations. Biophys J 2007;93:2289–99.
[182] Zheng W, Brooks BR, Doniach S, Thirumalai D. Network of dynamically important residues in the open/closed transition in polymerases is strongly conserved. Structure 2005;13:565–77.
[183] Zheng W, Brooks BR, Thirumalai D. Allosteric transitions in biological nanomachines are described by robust normal modes of elastic networks. Curr Protein Pept Sci 2009;10:128–32.
[184] Zhou Y, Cook M, Karplus M. Protein motions at zero-total angular momentum: the importance of long-range correlations. Biophys J 2000;79:2902–8.