Overview

Essential dynamics: foundation and applications

Isabella Daidone[1] and Andrea Amadei[2]*

Collective coordinates, as obtained by a principal component analysis of atomic fluctuations, are commonly used to predict a low-dimensional subspace in which essential protein motion is expected to take place. The definition of such an essential subspace allows to characterize protein functional, and folding, motion, to provide insight into the (free) energy landscape, and to enhance conformational sampling in molecular dynamics simulations. Here, we provide an overview on the topic, giving particular attention to some methodological aspects, such as the problem of convergence, and mentioning possible new developments.

How to cite this article:
WIREs Comput Mol Sci 2012. doi: 10.1002/wcms.1099

INTRODUCTION

Conformational transitions in proteins are essential for their function, such as substrate binding and product release, allosteric regulation, and many others. Nevertheless, accessing the underlying atomic motions in solution is very challenging. Molecular dynamics (MD) simulations have been used with increasing success to study at the atomic detail conformational dynamics in proteins, for example, secondary structure fluctuations or hydrogen-bonding network dynamical behavior. However, the extraction of functionally relevant motions from simulation results is not straightforward. For example, it is difficult to capture the early stages of the ion-gating process in membrane channels, to reveal conformational changes in the catalytic site of enzymes, or to investigate the folding–unfolding process as occurring in peptides and proteins. A solution to overcome these difficulties is the use of collective coordinates to identify a low-dimensional subspace in which the significant, functional protein motion is expected to take place.

The two most widely used computational methods to identify collective motions are normal mode analysis (NMA)[1,2] and principal component analysis (PCA),[3–7] the latter being the subject of the present overview. NMA is based on a harmonic approximation of the conformational energy surface, that is, assuming a single (parabolic) energy minimum, and independent normal modes are derived by diagonalization of the mass-weighted Hessian matrix of a single structure, corresponding to the energy minimum configuration. It, therefore, ignores the multiple minima nature of the conformational energy surface typically governing the functional motions in solvated proteins. The limitations of the use of normal modes to describe protein collective motions have been the subject of a number of studies.[8,9]

To overcome (partially) the limitations of NMA, a PCA can be carried out on a large number of configurations generated by an MD trajectory [or alternatively, by a Monte Carlo (MC) sampling]. PCA is a multivariate statistical analysis that involves diagonalization of a correlation matrix for a set of observables to yield collective variables. It has been applied to a large variety of very different observables/processes, as, for example, to analyze the genetic history of a group of populations.[10]

One of the first applications of PCA to protein dynamics, the quasi-harmonic analysis, used mass-weighted coordinates of protein atoms to construct the correlation matrix of atomic fluctuations (i.e., the mass-weighted covariance matrix), thus utilizing PCA to reconstruct approximated normal modes.[3–5] This method, based on the assumption of

*Correspondence to: andrea.amadei@uniroma2.it

[1]Department of Chemistry, Chemical Engineering and Materials, University of L'Aquila, L'Aquila, Italy

[2]Department of Chemical Sciences and Technology, University of Rome 'Tor Vergata', Rome, Italy

DOI: 10.1002/wcms.1099

quasi-harmonic internal motions, when applied to the study of processes involving large conformational transitions characterized by complex energy surfaces suffers, in principle, the same limitations of NMA. By using a non-mass-weighted covariance matrix, PCA may properly account for anharmonic internal motions, thus providing the concerted motions associated to the largest collective atomic fluctuations. Typically, more than 90% of the total atomic fluctuation is described by ≈20% of the principal axes (i.e., the covariance matrix eigenvectors). This analysis is often termed optimal dynamic coordinate analysis[6] or essential dynamics (ED) analysis.[7] Excellent reviews on the topic appeared in the years 1999–2000.[11,12] Hence, in this overview we focus on the more recent developments and applications.

THEORETICAL FOUNDATION

Let us consider N dynamical observables $x_1, x_2, \ldots, x_N$ (represented by the N-dimensional column vector $\boldsymbol{x}$) defining the states space of the system of interest. For any distribution of such observables, providing the statistical behavior of $\boldsymbol{x}$ in either the full states space or within a subpart of it, we can define the $N \times N$ correlation matrix $\widetilde{C}$ of the distribution via

$$\widetilde{C} = \langle \Delta\boldsymbol{x}\Delta\boldsymbol{x}^T \rangle, \tag{1}$$

$$\Delta\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}_{\text{ref}}, \tag{2}$$

where $\boldsymbol{x}_{\text{ref}}$ is an arbitrary reference value of $\boldsymbol{x}$ to be chosen according to the type of observables and information considered, $\Delta\boldsymbol{x}^T$ is the transpose of $\Delta\boldsymbol{x}$ (i.e., the N-dimensional row vector), and the angle brackets represent averaging over the distribution. From Eqs (1) and (2), we have

$$[\widetilde{C}]_{l,l'} = \langle \Delta x_l \Delta x_{l'} \rangle \quad l, l' = 1, 2, \ldots, N, \tag{3}$$

clearly showing that when $\boldsymbol{x}_{\text{ref}} = \langle \boldsymbol{x} \rangle$ (i.e., the reference is the mean as in most of the applications on atomic coordinates) the correlation matrix coincides with the covariance matrix, that is, each element of the matrix provides the covariance of two observables. Equation (3) also shows that $\widetilde{C}$ is a symmetric matrix with hence real eigenvalues and orthonormal eigenvectors. Therefore, the transformation matrix $\widetilde{T}$ for the diagonalization of the correlation matrix, with each column given by an eigenvector of $\widetilde{C}$, provides

$$\widetilde{T}^T\widetilde{C}\widetilde{T} = \langle \widetilde{T}^T\Delta\boldsymbol{x}\Delta\boldsymbol{x}^T\widetilde{T} \rangle = \langle \Delta\boldsymbol{q}\Delta\boldsymbol{q}^T \rangle = \Lambda, \tag{4}$$

where Λ is the diagonalized correlation matrix with eigenvalues λ and $\Delta\boldsymbol{q} = \widetilde{T}^T\Delta\boldsymbol{x}$ is the observables vector as expressed in the eigenvectors ($\boldsymbol{\eta}$) basis set. From Eq. (4) it follows that

$$\begin{aligned} \langle \boldsymbol{\eta}_l^T \Delta\boldsymbol{x}\Delta\boldsymbol{x}^T\boldsymbol{\eta}_{l'} \rangle &= \langle \Delta\boldsymbol{x}^T\boldsymbol{\eta}_l\Delta\boldsymbol{x}^T\boldsymbol{\eta}_{l'} \rangle = \langle \Delta q_l \Delta q_{l'} \rangle \\ &= \lambda_l \delta_{l,l'}, \end{aligned} \tag{5}$$

showing that $\widetilde{C}$ is also a positive definite matrix with the eigenvalues given by the mean square projections of $\Delta\boldsymbol{x}$ onto the eigenvectors $\boldsymbol{\eta}$ of the correlation matrix. Such eigenvectors may then serve to define a new orthonormal basis set to describe the states space of the system providing a new set of N dynamical observables ($\Delta q_1, \Delta q_2, \ldots, \Delta q_N$) given by linear combinations of the original ones.

To understand the meaning of such a frame rotation (a schematic two-dimensional example is given in Figure 1), it is convenient to order the eigenvectors and the new observables according to the size of the corresponding eigenvalue, that is, numbering the eigenvectors according to the eigenvalues decreasing order with hence the first eigenvector corresponding to the largest eigenvalue (see Figure 2). If we now consider an arbitrary unit vector $\boldsymbol{v}$ expressed in the eigenvectors basis set

$$\boldsymbol{v} = \sum_{l=1}^{N} a_l \boldsymbol{\eta}_l, \tag{6}$$

$$\sum_{l=1}^{N} a_l^2 = 1, \tag{7}$$

we can easily obtain the mean square projection of $\Delta\boldsymbol{x}$ onto $\boldsymbol{v}$ as

$$\begin{aligned} &\langle (\Delta\boldsymbol{x}^T\boldsymbol{v})^2 \rangle \\ &= \left\langle \left(\Delta\boldsymbol{x}^T \sum_{l=1}^{N} a_l \boldsymbol{\eta}_l \right)^2 \right\rangle = \left\langle \left(\sum_{l=1}^{N} a_l \Delta\boldsymbol{x}^T \boldsymbol{\eta}_l \right)^2 \right\rangle \\ &= \left\langle \left(\sum_{l=1}^{N} a_l \Delta q_l \right)^2 \right\rangle = \left\langle \sum_{l=1}^{N}\sum_{l'=1}^{N} a_l a_{l'} \Delta q_l \Delta q_{l'} \right\rangle \\ &= \sum_{l=1}^{N}\sum_{l'=1}^{N} a_l a_{l'} \langle \Delta q_l \Delta q_{l'} \rangle = \sum_{l=1}^{N} a_l^2 \lambda_l, \end{aligned} \tag{8}$$

which, using Eq. (7), provides

$$\sum_{l=1}^{N} a_l^2 \lambda_l = \lambda_1 + \sum_{l=2}^{N} a_l^2 (\lambda_l - \lambda_1). \tag{9}$$

By noting that from ordering the eigenvalues in decreasing order, that is, $\lambda_1 \geq \lambda_l$, it follows $\sum_{l=2}^{N} a_l^2(\lambda_l - \lambda_1) \leq 0$, we readily obtain

$$\lambda_1 \geq \langle (\Delta\boldsymbol{x}^T\boldsymbol{v})^2 \rangle \tag{10}$$

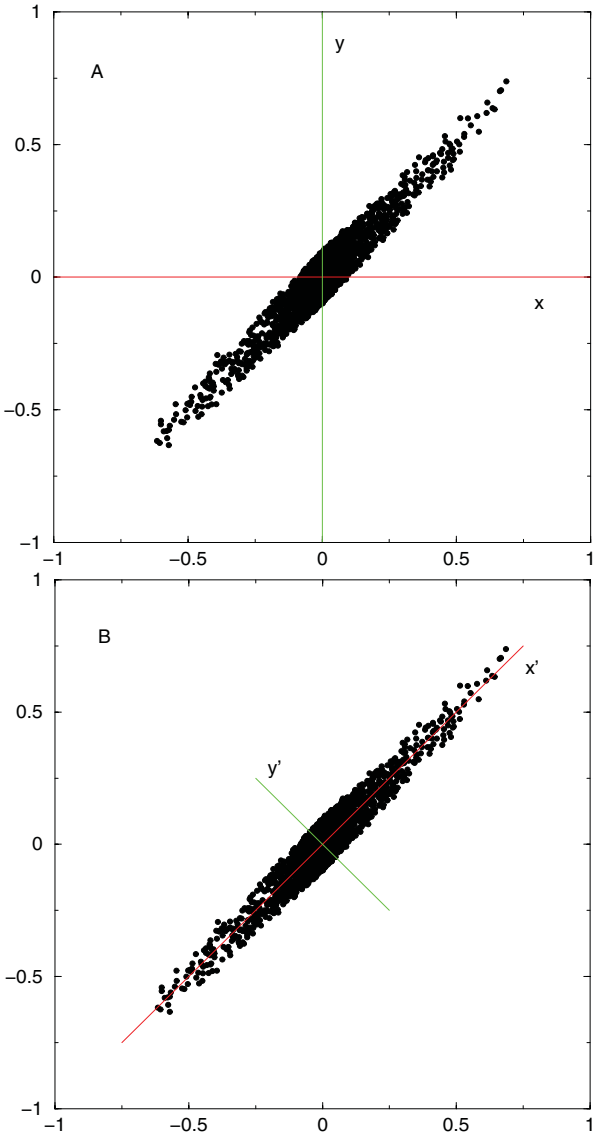

FIGURE 1 | Example of essential dynamics in two dimensions. With a distribution of points as depicted here, two coordinates (x, y) are required to identify a point in the cluster in (A), whereas one coordinate (x′) approximately identifies a point in (B).

and hence

$$\langle(\Delta \boldsymbol{x}^T \boldsymbol{v})^2\rangle \leq \langle(\Delta \boldsymbol{x}^T \boldsymbol{\eta}_1)^2\rangle. \qquad (11)$$

Equation (11) clearly shows that the correlation matrix eigenvector corresponding to the largest eigenvalue provides the direction in the states space maximizing the mean square projection of the observables vector $\Delta \boldsymbol{x}$. With the same derivations for the subspace orthogonal to the first eigenvector, we can obtain the same result for the second eigenvector and, so forth, for any other eigenvector/subspace.

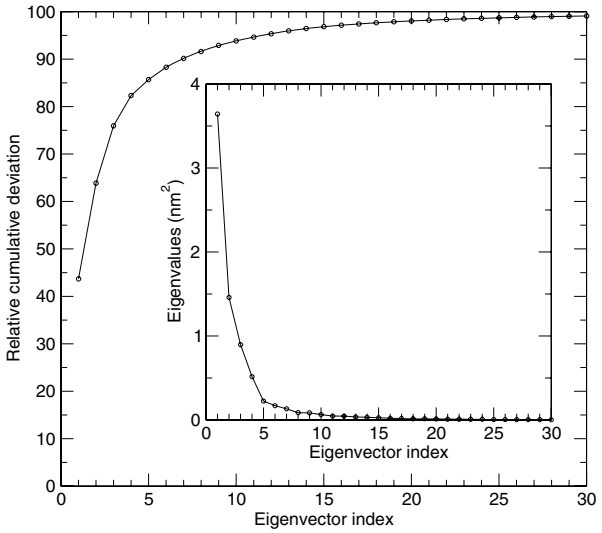

FIGURE 2 | Relative cumulative deviation (i.e., percentage of the cumumlative square fluctuation) up to the first 30 eigenvectors provided by the essential dynamics analysis performed on the C_α atoms of a 25-residue peptide simulated in water. The corresponding eigenvalues are given in the inset. It can be seen that the first two eigenvectors contribute for ≈65% of the total C_α motion.

It is then evident that the correlation matrix eigenvectors furnish a new basis set, equivalent to a rotation of the original axes, fitting the observables distribution at best (i.e., maximizing the $\Delta \boldsymbol{x}$ mean square projections). Hence, for anisotropic distributions such a procedure allows to define a possibly low-dimensional subspace (essential subspace) describing most of the behavior of the system, that is, the properties of the original observables can be largely reconstructed by using a limited set of new dynamical observables as defined by the essential subspace eigenvectors (essential eigenvectors).

CONVERGENCE OF COLLECTIVE COORDINATES

The mathematical derivation described in the section *Theoretical Foundation* is based on the use of an available statistical distribution for the observables considered. In practice, in most cases, we only have at hand a finite sampling distribution as obtained by a MD (or MC) simulation. Therefore, the accuracy of the ED analysis depends on the statistical relevance of the configurational subspace sampled within the simulation.

To evaluate the convergence of the essential eigenvectors, a possible strategy is to divide the simulation trajectory into two or more parts and compare the corresponding essential subspaces. The

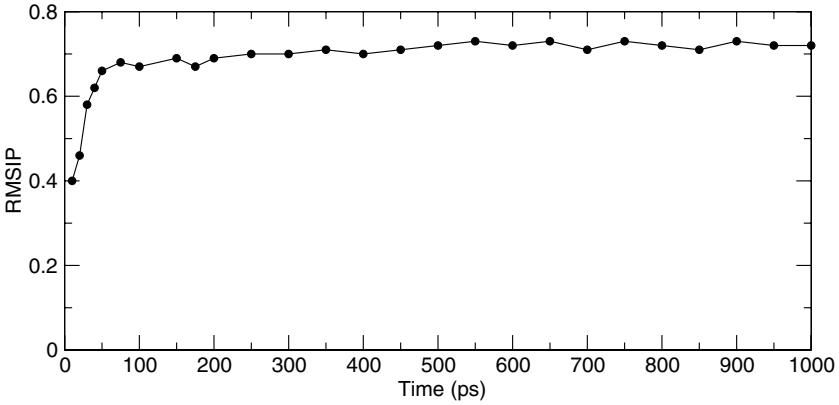

FIGURE 3 | Typical root mean square inner product of the essential subspaces (10 eigenvectors) obtained from two independent subparts of increasing time length as provided by the molecular dynamics simulation of a solvated protein.

degree of overlap between the essential subspaces can be obtained from the root mean square inner product (RMSIP) of the essential eigenvectors of one trajectory subpart with the essential eigenvectors of another trajectory subpart[13–15]:

$$\mathrm{RMSIP} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\sum_{j=1}^{M}(\boldsymbol{\eta}_i \cdot \boldsymbol{\nu}_j)^2}, \qquad (12)$$

where $\boldsymbol{\eta}_i$ and $\boldsymbol{\nu}_j$ are the ith and jth eigenvectors of two different subparts, respectively, and M is the dimension of the subspaces (see Figure 3). Another measure of the overlap related to the RMSIP was proposed[16] in which the dependence on the size of the fluctuations along the considered eigenvectors is introduced. An alternative quantity proposed is the overlap of the covariance matrices as obtained in the different trajectory subparts.[17]

To obtain a quantitative assessment of the statistical significance of the similarity of essential subspaces, it is possible to explicitly evaluate the overlap as provided by a 'random' distribution. The comparison of the distribution of the inner products of the eigenvectors of one part of the trajectory onto the eigenvectors of another part, with the random inner-products distribution allows to quantitatively assess, within a statistical confidence, whether the observed similarity is physically meaningful or not.[14] Another measure of the statistical significance of the overlap of subspaces has been introduced,[9] again based on comparing the results with a random-like distribution.

Other strategies to assess whether protein dynamics is insufficiently sampled were proposed based on comparing a given measured observable (e.g., the cosine content of the principal components) with the corresponding value as obtained by a random walk.[16–18] For example, it was shown that the projections to the essential eigenvectors obtained from short (picoseconds time scale) protein MD trajectories are similar to those of a random walk. In particular, the projections to the first essential eigenvectors show sine- and cosine-shaped curves of large amplitude.[17,18] Although, in principle, the average cosine content of the principal components might be an indicator for bad sampling; in practice, it cannot be used for a quantitative assessment of the sampling because its measure is affected by large errors.[17]

The question of the convergence of the essential eigenvectors was addressed in several papers and led to a controversial discussion.[13,14,16,19–21] In a number of studies that were restricted to short (100 ps to 1 ns) MD simulations, it was concluded that PCA eigenvectors are intrinsically unreliable.[19,20] Instead, in many other studies it was shown that typically for single-domain proteins the eigenvectors converge with the simulation time toward a 'stable' set in the nanoseconds time range,[13,14,16,21] and that such a convergence is statistically significant.[14] Whether this set is really stable beyond the nanoseconds time range, and coincides with the expectation set, is still an open question.[16] To date, protein MD simulations may achieve time lengths which at most reach one microsecond, thus not allowing any explicit evaluation on slower, possible conformational transitions occurring on the microseconds, or higher, timescales. For processes clearly involving conformational transitions on such higher timescales, for example, protein folding–unfolding transitions, the presence of well-distinct conformational basins separated by relatively high free energy barriers implies that the essential eigenvectors evaluated on one basin will not properly describe the fluctuations within the other basins.

BEYOND LINEAR CORRELATIONS

The ED analysis detects only linear correlations, hence it is unable to identify correlations of, for example, concerted rotations of chemical groups (e.g., two methyls simultaneously rotating in a correlated

way). In general, nonlinear correlations between the ED modes do persist.[7] Such higher-order correlations are difficult to detect and require the use of complex procedures, heavily based on iterative numerical methods.

Methods to capture nonlinear correlations have been proposed, as, for example, the nonlinear principal component analysis (NLPCA)[22,23] based on neural networks, the full correlation analysis (FCA)[24,25] based on mutual information, and the Scalable ISOMAP (ScIMAP) algorithm.[26] Nonlinear correlation analysis may reduce the 'essential' dimensionality of the considered systems more efficiently than the ED analysis, that is, a lower number of essential degrees of freedom is needed to achieve a similar level of accuracy in reproducing the motions in configurational space. Nevertheless, nonlinear correlation methods have drawbacks connected to the use of the complex iterative numerical procedures involved, the difficulty to clearly understand the relevance of each obtained degree of freedom, and hence the ordering of the modes, and the often required *a priori* definition of key parameters to be used in the method (e.g., the number of principle components in NLPCA). These drawbacks limit the use of those methods in practice.

Finally, the use of linear correlation analysis (PCA) for non-Cartesian atomic coordinates (e.g., dihedral angles), possibly leading to estimate nonlinear correlations in Cartesian space, has been proposed.[27] However, such methods not always provide useful information on conformational transitions as the metric change associated to the nonlinear coordinates transformation typically implies that the 'essential' modes in the new space do not correspond to any large structural fluctuation and/or transition in the Cartesian space.[28,29]

APPLICATIONS

The drastically reduced dimension of the essential space has been exploited with a great success in (1) functional and folding studies, (2) the construction of peptide folding free energy landscapes, and (3) enhanced sampling techniques.

Functional and Folding Studies

The typical applications of the ED analysis to the study of functional motions, that is, by performing a PCA on unbiased MD simulations, are so many that a comprehensive list is impossible. Therefore, in what follows we will provide some recent examples. The ED analysis is commonly applied to study the large conformational transitions occurring in enzymes and regulatory proteins. For example, ED analysis was used to capture the early stages of the gating process in a potassium channel,[30] or to reveal a gating-like conformational change in the catalytic loop of a HIV-1 integrase,[31] or to suggest that conformational selection, rather than induced-fit, is the dominant mechanism in the molecular recognition dynamics of ubiquitin.[32] Another interesting application was utilized to study the relation between flavin's redox states and protein dynamics. In particular, a dramatic change in the principal components of atomic fluctuations upon reduction of a flavin was revealed.[33] Another application we mention is for the understanding of the solvent-driven dynamical transition in myoglobin that is correlated with the onset of protein function.[34]

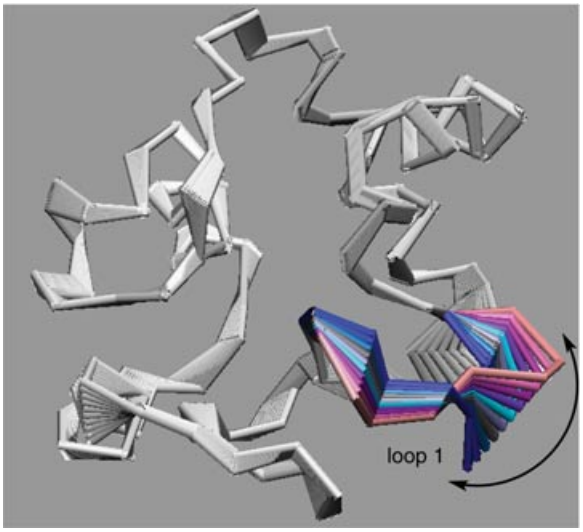

FIGURE 4 | Superimposition of 10 filtered configurations obtained by projecting the Cα motion onto an essential eigenvector of fluctuation involved in the unfolding of cytochrome *c*.

ED analysis can also be used to study protein unfolding. It allowed, for example, to determine the directions of motion that are activated in the early stages of the thermal unfolding of cytochrome *c*[15] (see Figure 4) or to distinguish the different mechanisms taking place in thermal and low-pH-induced unfolding in a prion protein.[35]

Peptide Folding Free Energy Landscapes

The investigation of conformational free energy landscapes is central to the understanding of how peptides, and potentially proteins, fold and function. However, finding a relatively small and appropriate set of coordinates by which to represent the free energy landscape remains challenging for biological macromolecules containing many thousands of degrees of freedom. To this end, PCA is particularly useful in

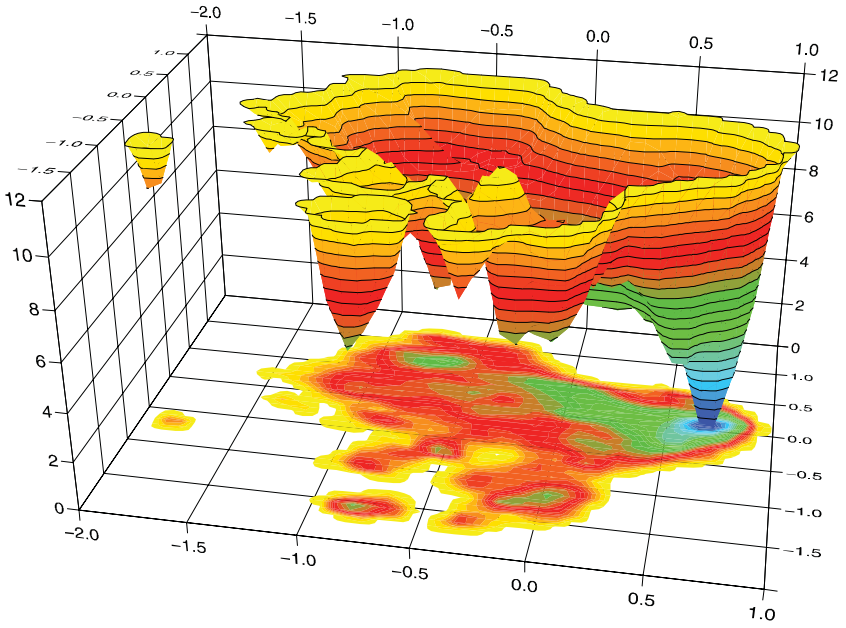

FIGURE 5 | Example of the folding free energy landscape of a peptide in solution as a function of the position along two first essential eigenvectors (q_1, q_2). The corresponding free energy change, $\Delta A(q_1, q_2)$, is given in kJ/mol and q_1, q_2 are given in nm.

providing collective reaction coordinates for the construction of folding free energy surfaces of peptides (and small proteins).

In the case of a sufficiently converged sampling, showing reversible folding/unfolding, the free energy profile as a function of the essential eigenvectors of motion (typically the first, or the first two) can be evaluated as $\Delta A(q) = -RT \ln \rho(q)/\rho(q_{\text{ref}})$, with $\rho(q)$ the equilibrium probability density as a function of the position, q, along the eigenvector (or eigenvectors) and $\rho(q_{\text{ref}})$ the probability density corresponding to the overall free energy minimum position. Surfaces of the internal energy and entropy can be calculated as well[36] and by evaluating within the considered essential subspace the diffusion coefficient a complete (diffusive) kinetic model of the folding–unfolding process can be obtained.[37,38]

Typically, the folding free energy landscape of peptides shows the characteristic features of a funneled landscape, either with a downhill surface toward the folded basin or with a more rugged surface with local minima populating the unfolded basin [36,39–41] (a typical example is given in Figure 5). The degree of roughness was shown to strongly depend on the temperature.[39–41] Comparison of the folding landscapes at different temperatures revealed that a temperature-dependent transition from a funneled free energy landscape (at higher T) to a rugged one (at lower T) occurs for the studied peptides.

The construction of free energy landscapes has been particularly useful for the understanding of the role of solvation in the folding process.[42,43] The effect of solvation on the thermodynamics (and kinetics) can be explored comparing the free energy landscapes in vacuum and in solution. For example, for a polyalanine peptide a dramatic effect of the aqueous solvent on the free energy landscape was observed, resulting in an inverted stability of the α-helical and β-hairpin states.[42] Another important result was obtained comparing explicit- and implicit-solvent-derived landscapes showing that implicit solvent models might fail in reproducing the correct folding thermodynamics.[43] The construction of free energy surfaces was also successfully utilized for the study of the effect of a point mutation on the folding thermodynamics and kinetics of an amyloidogenic peptide.[44]

Enhanced Sampling Techniques

The accessible simulation times of at most hundreds of nanoseconds are much shorter than the micro- to millisecond times scales at which many of the biomolecular processes occur as, for example, ligand binding, molecular recognition, or chemomechanical energy conversion. To overcome this limit, a huge variety of enhanced sampling techniques has been developed (for a review, see Ref 45).

Here, we will focus on the approaches that make use of a sufficiently converged essential subspace, obtained from a relatively short MD simulation, to bias the sampling. These methods can be divided in two categories. In a group of approaches (e.g., local elevation,[46] adaptive umbrella sampling,[47] conformational flooding,[48] or metadynamics[49]) adaptive umbrella potentials are employed to destabilize

the conformations already sampled, and thus gradually elevate the system from the bottom of a conformation minimum until transition into another minimum takes place. In essence, these are search methods with 'memory', recording a history to penalize visiting the same conformation. To improve sampling efficiency, the energy landscape can be lifted selectively in an essential subspace of the biomolecule, for example, a subspace defined by a few essential eigenvectors, as was done in conformational flooding[50] and metadynamics.[51] An example of a successful application of conformational flooding revealed large-scale functional motions in carbonmonoxy myoglobin.[50] In metadynamics, as a byproduct of the biasing process, a quantitative determination of the free energy surface (defined in the essential subspace) can be obtained for relatively simple systems, as, for example, a small peptide.[51] Other biasing techniques exist to construct free energy landscapes (e.g., weighted histogram techniques[52]) but, to the best of our knowledge, they do not make use of principal components as reaction coordinates.

A second group of methods makes use of constraints or restraints along collective degrees of freedom.[53–57] The method of ED sampling (EDS)[53–55] performs constraint dynamics simulations in the reduced space defined by a number of essential eigenvectors. It can be used to increase (expansion procedure) or decrease (contraction procedure) the distance from a reference structure, the expansion or contraction being performed in a selected essential subspace. Recently, the EDS was applied to study the mechanism of protein folding,[55,58] and the accessibility of the closed and open domain conformations in an enzyme.[59] A slightly modified version of EDS, namely, the directed essential dynamics (DED),[57] was successfully applied to determine peptide folding pathways. In another method,[56] restraints in collective coordinate space are applied on an ensemble of parallel MD trajectories. Weak restraints on the ensemble variance are shown to be sufficient for an increase in sampling efficiency along the essential eigenvectors by two orders of magnitudes.

CONCLUSIONS AND OUTLOOK

The great success of PCA for the study of atomic–molecular systems and, in particular, for characterizing the conformational mechanics and dynamics of biomacromolecules, relies on its great simplicity and applicability, also due to the fact that it is not based on any phenomenological parameter or *a priori* information. These characteristics allow this method to provide physically consistent low-dimensional essential subspaces for a wide range of atomic–molecular systems/observables.

As outlined in this overview, during the last two decades ED analysis, that is, PCA applied to the atomic positional fluctuations as provided by molecular simulations, made it possible to collect a huge amount of new data, opening the way to the explicit evaluation of the thermodynamics and kinetics of conformational transitions via atomistic simulations. Despite the enormous number of applications described in literature and its almost a century of existence, PCA is presently not only a widely used tool in most of the fields of atomic–molecular research, but it is still undergoing new developments. Of particular interest for chemical–physical studies are the new applications of PCA on hydrogen-bonding fluctuations to describe folding–unfolding transitions in peptides and, in principle, secondary structure elements of proteins[60] or on electronic state fluctuations to characterize quantum state transitions as provided by perturbation effects.[61]

ACKNOWLEDGMENTS

We acknowledge CASPUR (Consorzio Interuniversitario per le Applicazioni di Supercalcolo Per Universitá e Ricerca) for its financial support with the project 'Theoretical study of electron transfer reactions in complex atomic-molecular systems'.

REFERENCES

1. Brooks B, Karplus M. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin-inhibitor. *Proc Natl Acad Sci USA* 1983, 80:6571–6575.
2. Levitt M, Sander C, Stern PS. The normal modes of a protein: native bovine pancreatic trypsin-inhibitor. *Int J Quantum Chem* 1983, 24:181–199.

3. Karplus M, Kushick JN. Method for estimating the configurational entropy of macromolecules. *Macromolecules* 1981, 14:325–332.
4. Levy RM, Srinivasan AR, Olson WK, McCammon JA. Quasi-harmonic method for studying very low frequency modes in proteins. *Biopolymers* 1984, 23:1099–1112.
5. Kitao A, Hirata F, Go N. The effects of solvent on the conformation and the collective motions of protein: normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum. *J Chem Phys* 1991, 158:447–472.
6. García A. Large-amplitude nonlinear motions in proteins. *Phys Rev Lett* 1992, 66:2696–2699.
7. Amadei A, Linssen ABM, Berendsen HJC. Essential dynamics of proteins. *Proteins* 1993, 17:412–425.
8. 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, 27:425–437.
9. Rueda M, Chacon P, Orozco M. Thorough validation of protein normal mode analysis: a comparative study with essential dynamics. *Structure* 2007, 15:565–575.
10. Cavalli-Sforza LL. Genes, peoples, and languages. *Proc Natl Acad Sci USA* 1997, 94:7719–7724.
11. Kitao A, Go N. Investigating protein dynamics in collective coordinate space. *Curr Opin Struct Biol* 1999, 9:164–169.
12. Berendsen HJ, Hayward S. Collective protein dynamics in relation to function. *Curr Opin Struct Biol* 2000, 10:165–169.
13. de Groot B, van Aalten D, Amadei A, Berendsen H. The consistency of large concerted motions in proteins in molecular dynamics simulations. *Biophys J* 1996, 71:1707–1713.
14. 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, 36:419–424.
15. Roccatano D, Daidone I, Ceruso MA, Bossa C, Di Nola A. Selective excitation of native fluctuations during thermal unfolding simulations: horse heart cytochrome *c* as a case study. *Biophys J* 2003, 84:1876–1883.
16. Lange OF, Grübmuller H. Can principal components yield a dimension reduced description of protein dynamics on long time scales? *J Phys Chem B* 2006, 110:22842–22852.
17. Hess B, Convergence of sampling in protein simulations. *Phys Rev E* 2002, 65:031910–10.
18. Hess B, Similarities between principal components of protein dynamics and random diffusion. *Phys Rev E* 2000, 62:8438–8448.
19. Clarage JB, Romo T, Andrews BK, Pettitt BM, Phillips GN Jr. A sampling problem in molecular dynamics simulations of macromolecules. *Proc Natl Acad Sci USA* 1995, 92:3288–3292.
20. Balsera MA, Wriggers W, Oono Y, Schulten K. Principal component analysis and long time protein dynamics. *J Phys Chem* 1996, 100:2567–2572.
21. Faraldo-Gómez JD, Forrest LR, Baaden M, Bond PJ, Domene C, Patargias G, Cuthbertson J, Sansom MSP. Conformational sampling and dynamics of membrane proteins from 10-nanosecond computer simulations. *Proteins* 2004, 57:783–791.
22. Saegusa R, Sakano H, Hashimoto S. Nonlinear principal component analysis to preserve the order of principal components. *Neurocomputing* 2004, 61:57–70.
23. Nguyen PH. Complexity of free energy landscapes of peptides revealed by nonlinear principal component analysis. *Proteins* 2006, 65:898–913.
24. Lange OF, Grübmuller H. Full correlation analysis of conformational protein dynamics. *Proteins* 2008, 70:1294–1312.
25. Hub JS, de Groot BL. Detection of functional modes in protein dynamics. *PLoS Comput Biol* 2009, 5:e1000480.
26. Das P, Moll M, Stamati H, Kavraki LE, Clementi C. Low-dimensional, free-energy landscapes of protein-folding reactions by nonlinear dimensionality reduction. *Proc Natl Acad Sci USA* 2006, 103:9885–9890.
27. Mu Y, Nguyen PH, Stock G. Energy landscape of a small peptide revealed by dihedral angle principal component analysis. *Proteins* 2005, 58:45–52.
28. van Aalten DM, Conn DA, de Groot BL, Berendsen HJC, Findlay JB, Amadei A. Protein dynamics derived from clusters of crystal structures. *Biophys J* 1997, 73:2891–2896.
29. Maisuradze GG, Leitner DM. Free energy landscape of a biomolecule in dihedral principal component space: sampling convergence and correspondence between structures and minima. *Proteins* 2007, 67:569–578.
30. Grottesi A, Domene C, Hall B, Sansom MSP. Conformational dynamics of m2 helices in kirbac channels: helix flexibility in relation to gating via molecular dynamics simulations. *Biochemistry* 2005, 44:14586–14594.
31. Lee MC, Deng J, Briggs JM, Duan Y. Large-scale conformational dynamics of the HIV-1 integrase core domain and its catalytic loop mutants. *Biophys J* 2005, 88:3133–3146.
32. Lange OF, Lakomek N, Fares C, Schroeder GF, Walter KFA, Becker S, Meiler J, Grubmueller H, Griesinger C, de Groot BL. Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* 2008, 45:4998–5009.
33. Mueller RM, North MA, Yang C, Hati S, Bhattacharyya S. Interplay of flavin's redox states and

protein dynamics: an insight from QM/MM simulations of dihydronicotinamide riboside quinone oxidoreductase 2. *J Phys Chem B* 2011, 115:3632–3641.

34. Tournier A, Smith JC. Principal components of the protein dynamical transition. *Phys Rev Lett* 2003, 91:208106.
35. Gu W, Wang T, Zhu J, Shi Y, Liu H. Molecular dynamics simulation of the unfolding of the human prion protein domain under low pH and high temperature conditions. *Biophys Chem* 2003, 104:79–94.
36. Daidone I, Amadei A, Di Nola A. Thermodynamic and kinetic characterization of a β-hairpin peptide in solution: an extended phase space sampling by molecular dynamics simulations in explicit water. *Proteins* 2005, 59:510–518.
37. Hummer G, García A, Garde S. Conformational diffusion and helix formation kinetics. *Phys Rev Lett* 2000, 85:2637–2640.
38. Daidone I, D'Abramo M, Di Nola A, Amadei A. Theoretical characterization of α-helix and β-hairpin folding kinetics. *J Am Chem Soc* 2005, 127:14825–14832.
39. García A, Sanbonmatsu K. Exploring the energy landscape of β hairpin in explicit solvent. *Proteins* 2001, 42:345–354.
40. Kamiya N, Mitomo D, Shea JE, Higo J. Folding of the 25 residue aβ (12-36) peptide in TFE/water: temperature-dependent transition from a funneled free-energy landscape to a rugged one. *J Phys Chem B* 2007, 111:5351–5356.
41. Maisuradze GG, Liwo A, Scheraga HA. Principal component analysis for protein folding dynamics. *J Mol Biol* 2009, 385:312–329.
42. Levy Y, Hanan E, Solomon B, Becker OM. Helix-coil transition of PrP106-126: molecular dynamic study. *Proteins* 2001, 45:382–396.
43. Zhou R, Berne BJ. Can a continuum solvent model reproduce the free energy landscape of a β-hairpin folding in water? *Proc Natl Acad Sci USA* 2002, 99:12777–12782.
44. Daidone I, Di Nola A, Smith JC. Molecular origin of Gerstmann-Sträussler-Scheinker syndrome: insight from computer simulation of an amyloidogenic prion peptide. *Biophys J* 2011, 100:3000–3007.
45. Tai K. Conformational sampling for the impatient. *Biophys Chem* 2004, 107:213–220.
46. Huber T, Torda AE, van Gunsteren WF. Local elevation: a method for improving the searching properties of molecular dynamics simulation. *J Comput Aided Mol* 1994, 8:695–670.
47. Bartels C, Karplus M. Multidimensional adaptive umbrella sampling: applications to main chain and side chain peptide conformations. *J Comput Chem* 1997, 18:1450–1462.
48. Grübmuller H. Predicting slow structural transitions in macromolecular systems: conformational flooding. *Phys Rev E* 1995, 52:2893–2906.
49. Laio A, Parrinello M. Escaping free-energy minima. *Proc Natl Acad Sci USA* 2002, 99:12562–12566.
50. Schulze BG, Grübmuller H, Evanseck JD. Functional significance of hierarchical tiers in carbonmonoxy myoglobin: conformational substates and transitions studied by conformational flooding simulations. *J Am Chem Soc* 2000, 122:8700–8711.
51. Spiwok V, Lipovová P, Králová B. Metadynamics in essential coordinates: free energy simulation of conformational changes. *J Phys Chem B* 2007, 111:3073–3076.
52. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH. Multidimensional free-energy calculations using the weighted histogram analysis method. *J Comput Chem* 1995, 16:1339–1350.
53. Amadei A, Linssen ABM, de Groot BL, van Aalten DM, Berendsen HJC. An efficient method for sampling the essential subspace of proteins. *J Biomol Struct Dyn* 1996, 13:615–625.
54. de Groot BL, Amadei A, Scheek RM, van Nuland NA, Berendsen HJC. An extended sampling of the configurational space of HPr from *E. coli*. *Proteins* 1996, 26:314–322.
55. Daidone I, Amadei A, Roccatano D, Di Nola A. Molecular dynamics simulation of protein folding by essential dynamics sampling: folding landscape of horse heart cytochrome *c*. *Biophys J* 2003, 85:2865–2871.
56. Abseher R, Nilges M. Efficient sampling in collective coordinate space. *Proteins* 2000, 39:82–88.
57. Chen C, Xiao Y, Zhang L. A directed essential dynamics simulation of peptide folding. *Biophys J* 2005, 88:3276–3285.
58. Narzi D, Daidone I, Amadei A, Di Nola A. Protein folding pathways revealed by essential dynamics sampling. *J Chem Theory Comput* 2008, 4:1940–1948.
59. Daidone I, Roccatano D, Hayward S. Investigating the accessibility of the closed domain conformation of citrate synthase using essential dynamics sampling. *J Mol Biol* 2004, 339:515–525.
60. Zanetti-Polzi L, Daidone I, Anselmi M, Carchini G, Di Nola A, Amadei A. Analysis of infrared spectra of β-hairpin peptides as derived from molecular dynamics simulations. *J Phys Chem B* 2011, 115:11872–11878.
61. Amadei A, D'Alessandro M, D'Abramo M, Aschi M. Theoretical characterization of electronic states in interacting chemical systems. *J Chem Phys* 2009, 130:084109.

