Path-metadynamics: A computational study of conformational transitions in proteins

Díaz Leines, G.

Publication date
2013

Summary

Proteins are considered the machinery of life. They are involved in basically all fundamental functions of the cell, such as sensing and signalling, catalysis of metabolic reactions, structural support, transport of molecules from one location to another and replication of DNA. To meet their functional roles, proteins are required to undergo conformational changes and therefore, in the last decades it has been proposed that protein functions are governed ultimately by their dynamical character, rather than their static structures. In this view, proteins can sample a large ensemble of conformations around the average structure as a result of thermal energy and they are rooted on a multidimensional free energy landscape that defines the relative probabilities of the conformational states (thermodynamics) and the energy barriers between them (kinetics).

One of the most fundamental examples of conformational changes that illustrates the dynamical nature of proteins, is the folding of proteins into their compact three dimensional structures. After proteins have been synthesized by the ribosome as linear chains, they fold into specific three-dimensional structures in order to become functional. A failure to fold correctly, or remain correctly folded, usually produces inactive or dysfunctional proteins that can give rise to different neurodegenerative diseases. Other examples of conformational changes, where proteins meet different functional roles, include allosteric transitions in enzymes, force generation by motor proteins, the opening and closing of ion channels, and then conformational changes induced by ligand binding to enzymes and receptors. This means that proteins are highly dynamical objects and a detailed molecular description of structural changes would allow us to understand the underlying mechanisms of their functions, and therefore, develop predictive models of proteins that help in understanding, for instance, the origin of the different neurodegenerative disorders, or allow us to design ligands that can modulate the equilibria of conformational transitions and their rates.

Molecular Dynamics (MD) simulation has become a powerful tool for the theoretical studies of complex transitions, providing insight in the stability and the dynamics of many molecular systems. In particular, MD simulation can complement experiments by modelling the dynamical time evolution of biomolecular systems in atom-

istic detail. However, even though straightforward all-atom MD can now access the millisecond regime, many conformational changes take place on timescales that are very long compared to the molecular time accessible for straightforward MD. Therefore, simulating the kinetics of folding and large conformational changes in proteins has been limited to relatively small model proteins and short time scales. These so-called rare event processes are often associated to high free energy barriers between metastable states which hamper the efficient sampling of pathways with conventional MD. The study of rare events requires the development of special rare methods that allow the crossing of large free energy barriers for the study of complex transitions.

The aim of this thesis is to address the development and the application of a new rare event methodology, path-metadynamics, to explore free energy landscapes and determine transition pathways between metastable states. Path-metadynamics is an alternative approach to understand the mechanism of dynamical processes occurring on time scales that are very long compared to molecular time scales. Although there are a lot of examples of these so-called rare event processes, such as conformational changes in biomolecules, nucleation events in phase transitions and chemical reactions, in this thesis, we are particularly interested in the study of folding and unfolding events of proteins and we have employed path-metadynamics to explore the free energy landscape of different protein systems.

Path-metadynamics extends the existing Metadynamics method with new algorithms to (1) bias the system along a parametrized curve that is a function of other relevant order parameters or collective variables, and (2) evolve on-the-fly this curve to the most likely reaction pathway. This is done by adding an extra variable that is a function of all the other collective variables and represents a guess reaction pathway on the FES. By steering the dynamics along this variable with the growing metadynamics bias potential, the path efficiently evolves towards the locally most-likely reaction pathway, while simultaneously the biasing potential tends to an estimator of the free energy profile along the converged pathway. The computational expense of the combined method no longer depends on the number of collective variables. Especially the latter allows us to study rather complex transitions that require more collective variables than can be dealt with in a standard metadynamics simulation. Standard metadynamics has an exponential scaling of the computational cost with the dimensionality, which allows for at most 3 or 4 collective variables. In the first part of this thesis (chapter 3), we present the path-metadynamics method and show how it is used to simultaneously find the reaction mechanism and the free energy profile between the stable states in a single simulation. I illustrate the method on a classical molecular dynamics simulation on a conformational transition in alanine-dipeptide.

In chapter 4 we assess the theory and performance of path-metadynamics in more detail by extending the method with an additional progress variable that spans the complete CV space into the family of hyperplanes perpendicular to the path. This variable represents the transition tube between two stable states. By introducing the effect of a 'tube' potential applied on this variable, we show that the path-

metadynamics method can recover two limits: the average transition path (average placing scheme or finite temperature limit) and the minimum free energy path (gradient descent scheme or the zero temperature limit). Additionally, we present a recipe to use the path-metadynamics method. This recipe is summarized from different tests to fine-tune the parameters for the convergence of the path and the free energy profile.

The second part of this work (chapters 5 and 6) shows the application of path-metadynamics in studying conformational transitions in proteins. In chapter 4 we show that the path-Metadynamics method can successfully be applied to study the millisecond unfolding process of a relevant photoreceptor, the Photoactive Yellow Protein (PYP), where it is able to find not just the free energy profiles along the unfolding transition, but also provide unique molecular insight along the reaction paths obtained. The results provide a complementary view to the conclusion of Vreede et. al. about the existence of intermediate metastable states present during the unfolding process of PYP and the relevant CVs important for the description of the transitions. Additionally, we find that the highest barrier of the process appears during the solvent exposure of Glu^{46} and that the relevant CVs describe an important role of residue Asn^{43} during the transition to the signalling state of the protein. We have implemented the path metadynamics approach in the development version of PLUMED package. PLUMED is a plugin for free energy calculations, such as metadynamics, to be used in combination with molecular dynamics simulation packages, such as Gromacs, which is the code used for our simulations and can allow us to compute the mechanisms and rates as a function of all collective variables that are relevant to describe the unfolding process of PYP. Additionally, we propose an strategy to locate relevant collective variables and estimate accurate folding barriers using path-metadynamics and subsequent analysis of the free energy profiles.

In chapter 6, we complete the thesis by applying the path-metadynamics method for the prediction of the formation-dissociation mechanism of one of the most studied coiled coils in globular proteins, the leucine zipper protein domain of the yeast transcription factor GCN4. This coiled coil has motivated many studies working into the fundamental relation between the amino acid sequence and protein folding. We have identified some relevant degrees of freedom participating in the process of formation of the complex. Our results indicate that the transition does not occur along a single robust pathway but exhibits transition state heterogeneity. Moreover, the free energy profiles obtained along the average transition pathways indicate that the most likely mechanism occurs through an intermediate characterized by the dissociation of the N-terminal and the partial loss of helical structure of the dimer. Experimental studies have confirmed the strong stability of the C-terminal and have suggested a probable pathway through this intermediate state.

In conclusion, the path-metadynamics method opens up a way for the investigation of rare event transitions on high-dimensional free energy landscapes. The computational time spent in the simulations performed in this thesis, is orders of magnitude less than using straightforward MD, while the dimensionality of the set of CVs se-

lected to describe the transitions is an order of magnitude larger than those included when using the standard metadynamics approach. This provides a step forward in the study of complex rare event transitions. We stress that although the method allows the selection of a high-dimensional space of CVs, a recipe for an a priori selection of all relevant reaction coordinates that describe the transition, is still the main challenge in the study of complex rare event transitions. Therefore, this remains a main limitation for the path-metadynamics method. In our knowledge, there is not yet a precise recipe for the selection of relevant reaction coordinates, although approaches like the likelihood maximization analysis in combination with methods like transition path sampling, can throw some light into the initial selection of a set of CVs. It is important to note that although path-metadynamics cannot solve the problem of an a priori selection of CVs, it is possible to employ a strategy to test the sufficiency of the set of CVs based on the one-dimensional free energy picture obtained out of an intrinsically multidimensional event. Moreover, despite this limitation, the path-metadynamics algorithm has shown to enable the study of locally probable pathways on high-dimensional spaces with a relatively robust performance. It is also important to mention that although the scope of this thesis is the study of conformational transitions in proteins, path-metadynamics can be applied in a wide range of disciplines that go from material science to biophysics, where reactive transitions can occur on very long timescales. Furthermore, although the current version of the path-metadynamics algorithm is local, in the sense that it is limited to the exploration of a single transition tube between stable states, future extensions of the method could include a bias exchange of pathways between different transitions tubes, allowing the exploration of multiple state transitions where there is not a unique solution for the average transition path. This extension would enable the exploration of even more complex free energy landscapes, providing a possibility to study, for instance, bigger protein systems.