Biophysical Chemistry 208 (2016) 62–67

Contents lists available at ScienceDirect

Biophysical Chemistry

journal homepage: http://www.elsevier.com/locate/biophyschem

Computing Reaction Pathways of Rare Biomolecular Transitions using Atomistic Force-Fields

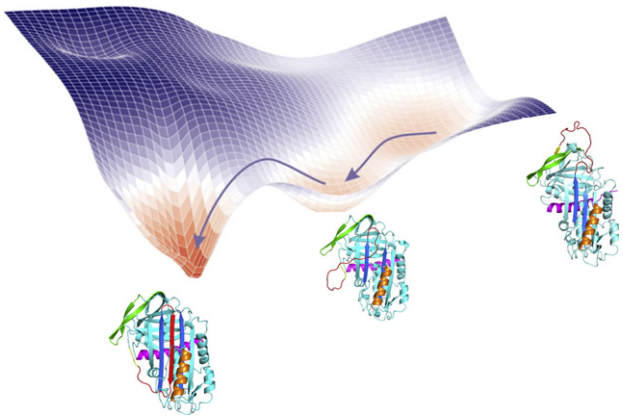
P. Faccioli [a,b,*], S. a Beccara [c,b]

[a] Dipartimento di Fisica, Università degli Studi di Trento, Via Sommarive 14 Povo, Trento, 38123 Italy
[b] Trento Institute for Fundamental Physics and Applications (INFN-TIFPA), Via Sommarive 14 Povo, Trento, 38123 Italy
[c] European Centre for Theoretical Studies in Nuclear Physics and Related Areas (ECT*-FBK), via Sommarive 18, I-38123 Trento, Italy

HIGHLIGHTS

- Numerical simulations of protein folding are extremely difficult by plain MD and, to date, only limited to relatively small chains, performed on a special-purpose computer (ANTON)
- The dominant reaction pathways is a path integral based approximate variational approach which focuses on the reactive part of the dynamics. In this method, folding trajectories for arbitrary proteins can be obtained using state-of-theart all-atom force fields
- The approach was validated against plain MD results performed on ANTON and then applied to larger systems, which cannot be investigated using MD, even on ANTON.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 30 April 2015
Received in revised form 23 June 2015
Accepted 23 June 2015
Available online 14 August 2015

Keywords:
Molecular Simulations
Path Sampling
Rare events
Computational Biophysics
Protein
Folding

ABSTRACT

The Dominant Reaction Pathway (DRP) method is an approximate variational scheme which can be used to compute reaction pathways in conformational transitions undergone by large biomolecules (up to ~10^3 amino-acids) using realistic all-atom force fields. We first review the status of development of this method. Next, we discuss its validation against the results of plain MD protein folding simulations performed by the DE-Shaw group using the *Anton* supercomputer. Finally, we review a few representative applications of the DRP approach to study reactions which are far too complex and rare to be investigated by plain MD, even on the *Anton* machine.

* Corresponding author.
E-mail address: faccioli@science.unitn.it (P. Faccioli).

http://dx.doi.org/10.1016/j.bpc.2015.06.014
0301-4622/

1. Introduction

To date, computer simulations represent the only existing approach through which it is possible to attain a complete atomistic characterisation of conformational reactions of biomolecules. In particular, Molecular Dynamics (MD) provide a theoretically sound framework to predict the time-evolution of important biomolecules such as proteins and nucleic acids and to compute a large class of observables, both in- and out- thermal equilibrium.

In general, the reliability or feasibility of MD simulations depends on a number of factors, including (i) the accuracy of the force field, (ii) the computational cost of sampling the configuration space and identify the relevant reaction pathways and (iii) the procedure adopted to reduce the atomic-resolution raw data and extract the relevant physico-chemical information.

The development of the special-purpose *Anton* supercomputer has significantly extended the time intervals which can be covered by simulations up a few milliseconds for proteins consisting of less than 100 amino-acids [1]. This has allowed Shaw and co-workers to demonstrate that the existing all-atom force fields are sufficiently accurate to attain the protein native structures starting from denatured configurations and to observe many spontaneous unfolding-refolding events [1–3].

The ultra-long atomistic simulations performed on *Anton* have also been used to test and further develop the theoretical schemes to condense the huge amount of information encoded in the simulation raw data. In particular, Anton data have been used to assess the quality of specific reaction coordinates (see e.g. [5,6]) and have been used to build coarse-grained representations of the dynamics based on Markov state models [7].

In spite of these important achievements, to date the computational cost of performing an exhaustive sampling of the space of reaction pathways stands as the major limiting factor, which prevents computer simulations to be applicable to a wider range of biological problems. Indeed, most reactions of biological interest occur at time scales many orders of magnitude larger than the millisecond, thus are inaccessible to any theoretical scheme based on the direct integration of the equations of motion, even using *Anton*. Consequently, it is important to continue the development of more efficient algorithms to sample the reactive pathways space (see e.g. [4] and references therein). In particular, in the following section we report on the state of development of an approach called Dominant Reaction Pathways (DRP) [11,12]. This method is based on a variational principle derived in a number of works [8–10], combined with some advanced path-sampling algorithm [13,14]. As we shall see, this specific combination of algorithms makes it possible to investigate reactions which take place over time scales which are many orders of magnitude larger than those which can be covered by MD using Anton, using realistic all-atom atomistic force fields.

2. The DRP approach

Sampling the reaction pathways using the DRP method is a three-step procedure (see Fig. 1). In the **first step**, one uses plain MD simulations to generate an ensemble of equilibrium configurations in the reactant and product states starting from some representative model configuration. For example, in protein folding DRP simulations, one typically generates a number of initial unfolded structures by performing plain MD unfolding simulations at high temperature starting from the crystal native structure, followed by relaxation simulations at room temperature. On the other hand, configurations in the native states are straightforwardly obtained by short room temperature MD simulations from the crystal native structure. We emphasise that this first step requires some prior knowledge of the structure of the reactant and product stats and, in some cases, may involve model-dependent assumptions.

In the **second step**, one generates a large number of *trial* reaction pathways which start from the *same reactant configuration* and reach the *same product configuration* in a computationally affordable time interval. To this end, we found to be particularly convenient to adopt the so-called ratchet-and-pawl (rMD) algorithm [13,14] which exploits a particular history-dependent biasing force to drive the system towards the target configuration. The feature of the rMD algorithm which distinguishes it from other biased-MD schemes is that the biasing force only sets in whenever the system attempts to backtrack towards the reactant — defined in terms of some position-dependent reaction coordinate—. Conversely, no bias is applied whenever the system spontaneously takes a step towards the product. So, in rMD, the physical spontaneous fluctuations (and not the bias) make the system advance towards the target. Several tests on toy models of different complexity (see e.g. the supplementary material of Ref. [12]) have shown that, in contrast to what happens with pulled or steered MD, rMD produces realistic reaction pathways even when the bias is defined using a sub-optimal reaction coordinate. For molecular transitions, a convenient reaction coordinate is the one introduced in [14] which measures the distance between the instantaneous contact map and the target contact map.

In general, the trial reactions pathways generated in the second step are all subject to uncontrolled systematic errors, introduced by the biasing force. To partially compensate for this problem, in the **third step**, one relies on a variational principle to identify the least-biased path within the ensemble of trial reaction pathways with identical boundary

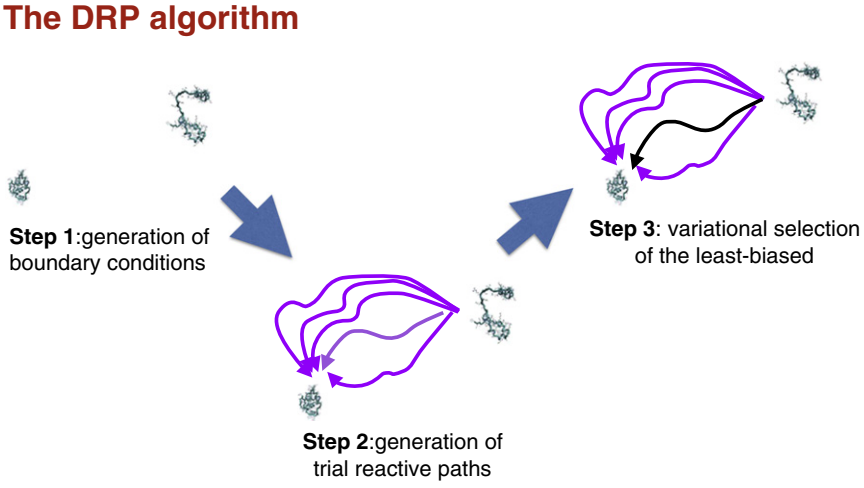


Fig. 1. Cartoon representation of the three-step procedure involved in a DRP calculation.