Bioinformatics, 35(18), 2019, 3505–3507
doi: 10.1093/bioinformatics/btz104
Advance Access Publication Date: 19 February 2019
Applications Note

OXFORD

Structural bioinformatics

eBDIMS server: protein transition pathways with ensemble analysis in 2D-motion spaces

Laura Orellana[1,*], Johan Gustavsson[2], Cathrine Bergh[3], Ozge Yoluk[3,†] and Erik Lindahl[1,3,*]

[1]Department of Biochemistry and Biophysics, Science for Life Laboratory, Stockholm University, Stockholm 106 91, Sweden, [2]Department of Computational Science and Technology, KTH Royal Institute of Technology, Stockholm 100 44, Sweden and [3]Department of Applied Physics, Swedish e-Science Research Center, KTH Royal Institute of Technology, Solna 171 21, Sweden

*To whom correspondence should be addressed.
[†]Present address: Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, MD 21201, USA

Associate Editor: Alfonso Valencia

Received on August 18, 2018; revised on December 30, 2018; editorial decision on February 7, 2019; accepted on February 15, 2019

Abstract

Summary: Understanding how proteins transition between different conformers, and how conformers relate to each other in terms of structure and function, is not trivial. Here, we present an online tool for transition pathway generation between two protein conformations using Elastic Network Driven Brownian Dynamics Importance Sampling, a coarse-grained simulation algorithm, which spontaneously predicts transition intermediates trapped experimentally. In addition to path-generation, the server provides an interactive 2D-motion landscape graphical representation of the transitions or any additional conformers to explore their structural relationships.

Availability and implementation: eBDIMS is available online: http://ebdims.biophysics.se/ or as standalone software: https://github.com/laura-orellana/eBDIMS, https://github.com/cabergh/eBDIMS.

Contact: laura.orellana@scilifelab.se or erik.lindahl@gmail.com

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Protein conformational changes are essential to understand biological function, but are difficult to tackle both experimentally and computationally. Cryo-EM and X-ray techniques are however providing a wealth of new structures trapped in different conformations. Yet, understanding how different conformers relate to each other is often non-trivial and requires information on the actual pathways connecting experimental end-states. Public web servers that calculate transitions are therefore common (Zheng and Wen, 2017) but at most evaluate the generated pathways in terms of stereochemical quality. Typically, validation of the biological feasibility of *in silico* pathways is complex (Weiss and Levitt, 2009), and relies on *ad hoc* reaction coordinates (e.g. distances, angles, etc.). We have demonstrated that minimal but 'structurally rich' X-ray ensembles provide Principal Components (PCs) which are powerful intrinsic coordinates for path-validation (Orellana *et al.*, 2016). Using projections onto the PC-space (Fig. 1), we showed that Elastic Network Driven Brownian Dynamics Importance Sampling (eBDIMS) coarse-grained simulations generate transitions along X-ray encoded motions, spontaneously predicting multiple on-pathway experimental intermediates. Unlike other methods, eBDIMS tracks both forward and reverse paths, which together define a low-energy area sampled by costly molecular dynamics (MD) simulations. Here we present an online tool for eBDIMS-pathways generation between two user-defined end-states, and its integrated 2D-motion projection, either in PC-space if an ensemble is provided, or by default in normal mode (NM) space, to facilitate path and conformer analysis.

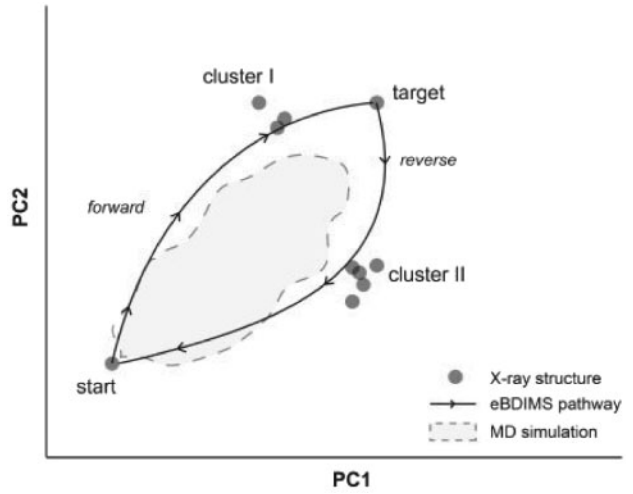

Fig. 1. Scheme of eBDIMS path-projections onto the PC-motion space. eBDIMS generates two transitions between the start and target structures (forward and reverse). Projection onto the PCs of an experimental X-ray ensemble (dots) shows path asymmetry and how they visit different on-pathway intermediate clusters. Note that the area sampled by MD typically is contained within the low-energy area defined by both paths

2 Materials and methods

2.1 The eBDIMS algorithm

The eBDIMS algorithm is based on the Essential Dynamics-refined Elastic Network Model (ED-ENM) force field (Orellana *et al.*, 2010), which reduces proteins to a Cα network connected by harmonic springs. The ED-ENM potential energy (U) for a given atom (i) is:

$$U_i = \frac{1}{2}\sum_{j=1}^{N} K_{ij}(r_{ij} - r_{ij}^0)^2 \quad (1)$$

where N is the number of residues, r_{ij} is the distance between Cα-atoms i and j and K_{ij} is an MD-refined force constant. The network is subject to Langevin simulation (Orozco *et al.*, 2011), where random forces (ξ) act as thermal fluctuations, and a friction coefficient (γ) as the solvent:

$$m\ddot{r}_i = F_i - \gamma\dot{r}_i + \xi(t) \quad (2)$$

Equation (2) is integrated with the *Verlet* algorithm, which is biased by accepting every k steps ($k = 10$, default) moves that decrease the progress variable:

$$\Gamma_i = \sum_{i,j=1}^{N} (r_{ij} - r_{ij}^{\text{target}})^2 \quad (3)$$

Until convergence (99.5%) or maximal time (4 h) is reached. Further ED-ENM/eBDIMS details are described elsewhere (Orellana *et al.*, 2016).

2.2 Pathway projections onto 2D-motion spaces

Transition pathways are projected onto a 2D-motion space defined by the PCs of a structural ensemble (if provided) and the NMs of the start structure (default). The NMs overlap with ensemble PCs from X-ray or MD and efficiently track conformational changes (López-Blanco and Chacón, 2016). Ideally, uploaded ensembles should contain at least three distinct conformers (Orellana *et al.*, 2016). Tools like Bio3D (Grant *et al.*, 2006), ProDy (Bakan *et al.*, 2011) or our tutorial with the pdbParser package (https://github.com/ozyo/pdbParser) are useful to prepare Protein Data Bank (PDB) ensembles. Start coordinates (r_0) are used as reference for alignment, PCs and NMs calculation (Orozco *et al.*, 2011). Then, each structure is represented by its projections onto two representative NM/PCs:

$$M_k = |\Delta r_{i-0}|.\cos(M_k.r_{i-0}) \quad k = 1, 2 \quad (4)$$

where $\Delta r_{i-0} = (r - r_0)/||r - r_0||$ is the $3N$-difference transition vector between the coordinates of structure r and the reference, r_0 and M_k is either a PC or a NM. To be meaningful, PCs and NMs need to, respectively, capture >70% of the ensemble structural variance or >80% of the transition (measured by the dot product between the transition vector and each eigenvector M_k). However, the main significance criterion is their ability to separately cluster different functional states, which needs to be evaluated by the user (more details in Supplementary Fig. S1 and Supplementary Methods).

3 Implementation

The eBDIMS code is written in C and parallelized with OpenMP for shared memory multi-threading. Jobs are scheduled linearly and executed on a server equipped with Quad-Core AMD Opteron 64-bit Processors 2374 HE at 2.20 GHz (8 cores) and 16 GB RAM. Interactive 2D- and 3D-plots are generated in the browser with d3.js and NGL viewer (Rose and Hildebrand, 2015), respectively. See scheme of the workflow in Supplementary Figure S2. Documentation, tutorials and examples including those in Orellana *et al.* (2016) with sample files (Supplementary Table S1) are provided online.

Input data

End-state structures (required): two conformations (referred to as start and target, respectively) are uploaded as coordinate files in strict PDB-format, or retrieved by their IDs (e.g. 4NPQ: A, chain A of PDB 4NPQ).

Conformational ensemble (optional): one file with coordinate sets in PDB-format and delimited by MODEL and ENDMDL labels that strictly match the reference end-states in residue number.

Output results

Successful job submission redirects to a unique result page displaying job status. Active jobs can be monitored in real-time charts reporting the transition progress (%) and RMSD to the target (Å). A typical run takes from minutes to hours depending on system size and transition complexity. Finished jobs present interactive 3D time-resolved molecular representations linked to PC or NM 2D-spaces that show eBDIMS transitions (*forward*: start → target; *reverse*: target → start) and all input structures.

Funding

This work was supported by the Sven and Lilly Lawskis Foundation (to L.O.); the Knut and Alice Wallenberg foundation (KAW), Vetenskapsrådet and the Swedish e-Science Research Center (to E.L.), and computational resources from the Swedish National Infrastructure for Computing (SNIC).

Conflict of Interest: none declared.

References

Bakan,A. *et al.* (2011) ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics*, **27**, 1575–1577.

Grant,B.J. *et al*. (2006) Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics*, **22**, 2695–2696.

López-Blanco,J.R. and Chacón,P. (2016) New generation of elastic network models. *Curr. Opin. Struct. Biol.*, **37**, 46–53.

Orellana,L. *et al*. (2010) Approaching elastic network models to molecular dynamics flexibility. *J. Chem. Theory Comput.*, **6**, 2910–2923.

Orellana,L. *et al*. (2016) Prediction and validation of protein intermediate states from structurally rich ensembles and coarse-grained simulations. *Nat. Commun.*, **7**, 12575.

Orozco,M. *et al*. (2011) Coarse-grained representation of protein flexibility. Foundations, successes, and shortcomings. *Adv. Protein Chem. Struct. Biol.*, **85**, 183–215.

Rose,A.S. and Hildebrand,P.W. (2015) NGL Viewer: a web application for molecular visualization. *Nucleic Acids Res.*, **43**, W576–W579.

Weiss,D.R. and Levitt,M. (2009) Can morphing methods predict intermediate structures? *J. Mol. Biol.*, **385**, 665–674.

Zheng,W. and Wen,H. (2017) A survey of coarse-grained methods for modeling protein conformational transitions. *Curr. Opin. Struct. Biol.*, **42**, 24–30.