Bioinformatics, 36(3), 2020, 956–958
doi: 10.1093/bioinformatics/btz649
Advance Access Publication Date: 16 August 2019
Applications Note

Structural bioinformatics

MEPSAnd: minimum energy path surface analysis over *n*-dimensional surfaces

Iñigo Marcos-Alcalde[1,2], Eduardo López-Viñas[2,*] and Paulino Gómez-Puertas[1,*]

[1]Molecular Modelling Group, Centro de Biología Molecular Severo Ochoa, CBMSO (Consejo Superior de Investigaciones Científicas–Universidad Autónoma de Madrid, CSIC-UAM), Cabrera, CL Nicolás 1, E-28049 Madrid, Spain and [2]Biosciences Research Institute, School of Experimental Sciences, Universidad Francisco de Vitoria, UFV, Building E, Ctra. M-515 Pozuelo-Majadahonda Km 1,800, Pozuelo de Alarcón, Madrid 28223, Spain

*To whom correspondence should be addressed.

Associate Editor: Arne Elofsson

Received on June 27, 2019; revised on July 30, 2019; editorial decision on August 13, 2019; accepted on August 14, 2019

Abstract

Summary: *n*-dimensional energy surfaces are becoming computationally accessible, yet interpreting their information is not straightforward. We present minimum energy path surface analysis over n-dimensional surfaces (MEPSAnd), an open source GUI-based program that natively calculates minimum energy paths across energy surfaces of any number of dimensions. Among other features, MEPSAnd can compute the path through lowest barriers and automatically provide a set of alternative paths. MEPSAnd offers distinct plotting solutions as well as direct python scripting.

Availability and implementation: MEPSAnd is freely available (under GPLv3 license) at: http://bioweb.cbm.uam.es/software/MEPSAnd/.

Contact: eduardo.lopez@ufv.es or pagomez@cbm.csic.es

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

1 Introduction

Multi-dimensional energy surfaces are broadly present in a range of biophysical research. In 2015, we published a computational tool (Marcos-Alcalde *et al.*, 2015) able to analyze 3D surfaces (2 coordinates and energy), which has been widely used to study: enzyme reaction mechanisms (Fritz *et al.*, 2018; Geronimo *et al.*, 2018; Li *et al.*, 2019; Marcos-Alcalde *et al.*, 2017; Mendieta-Moreno *et al.*, 2016), ligand binding (Banerjee *et al.*, 2018; Duan *et al.*, 2018; Yuan *et al.*, 2018), protein folding and conformational sampling (Mondal and Reddy, 2019; Shao *et al.*, 2019; Shao and Zhu, 2019a,b) as well as solvation in materials science (Kachmar *et al.*, 2017; Kachmar and Goddard, 2018). It was also included into a catalog of molecular modeling tools (Pirhadi *et al.*, 2016).

Now that *n*-dimensional surfaces are computationally accessible, a path-finding tool capable of working with this type of data without the need for dimensional reductions could dramatically facilitate their use. We are introducing minimum energy path surface analysis over *n*-dimensional surfaces (MEPSAnd), a GUI-based tool that natively handles *n*-dimensional energy surfaces and is capable of calculating (i) the path connecting two points through the lowest energy barrier/s, (ii) the region of the surface sampled to reach those barriers and (iii) a series of alternative sub-optimal paths. Also, a range of plotting and projection options is offered to facilitate the interpretation of the *n*-dimensional results. The main advantages of MEPSAnd over the previous program (Marcos-Alcalde *et al.*, 2015) are the ability to work with surfaces with any number of dimensions, the handling of data without the need for predefined grids, the consideration of all equivalent trajectories at the same time, the automatic detection of multiple sub-optimal trajectories, the exhaustive sampling of minimum and barrier plateaus, providing a better description of topologically complex surfaces, the use of flexible graphic solutions for the interpretation of *n*-dimensional results, better calculation performance and scripting support.

2 Features

Network-based connectivity (grid independent): MEPSAnd abstracts surfaces using two networks (Fig. 1A): a Surface Connectivity Network (SCN) and a Global Network (GN). The SCN describes point-to-point connectivity using a list of neighbors per point that is not directly dependent on the number of dimensions. MEPSAnd can build the SCN from a given *n*-dimensional surface via cutoff-based connectivity inference. Alternatively, the user may provide any previously defined connectivity obtained by third party solutions. This approach allows MEPSAnd to natively handle surfaces with an arbitrary number of dimensions. GN is computed over the SCN, abstracting the surface as a network of minima connected by barriers. The GN computation consists on:

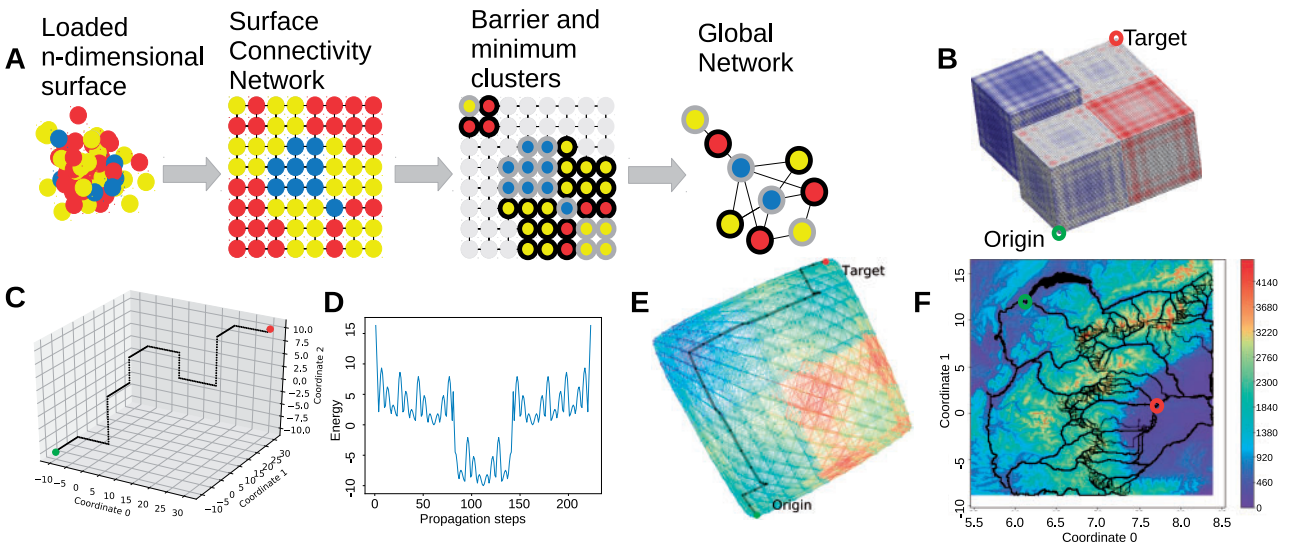

Fig. 1. MEPSAnd overview. (**A**) Algorithm steps to build the SCN and the GN. (**B**) An example 4D surface with 275 684 points. Panels (**C**, **D** and **E**) are MEPSAnd representations of the resulting simplified minimum energy path from origin (green) to target (red). C: Path projection over coordinates 0, 1 and 2 (i.e. in this case X, Y and Z). D: Path energy profile. E: Network projection of the path over the GN representing barrier clusters as edges. (**F**) Example contour plot of a 3D surface showing the 392 paths found from origin (green) to target (red) sampling alternative saddle points (Color version of this figure is available at *Bioinformatics* online.)

Minima detection: Minima in MEPSAnd can be single points or groups of points with the same energy value. All of them are called minimum clusters. A cluster is considered as a minimum cluster if none of its constituent points has a neighbor with lower energy.

Barrier detection: Every minimum cluster is extended to any higher energy point to which it is connected, until no more higher points can be reached (minima propagation). The overlapping regions between different minima propagations are used to define the barrier clusters.

GN building: Using the previously defined minimum and barrier clusters as vertices the GN is built in such way that minimum clusters are only connected to barrier clusters and vice versa.

GN-based path-finding: The new MEPSAnd algorithm searches for the minimum energy path in two sequential steps. First, once the GN path has been calculated, MEPSAnd computes paths over this network. This path returns a list of visited pairs of barrier and minimum clusters. Second, the paths connecting each of these pairs are computed over the SCN. Each of these real point paths is computed in the following manner: propagation to the lowest energy neighbors is performed from minimum to barrier annotating the loop step in which each point is occupied to later perform a steepest descent trace back from barrier to minimum minimizing the steps taken. Each of these paths is called a 'path fragment'. MEPSAnd connects the ends of these fragments to build a 'fragmentwise connectivity' and path fragments are combined to reconstruct what is referred to as 'fragmentwise path', i.e. the complete minimum energy path. The algorithm in more extensively described in the Supplementary Material (user manual).

Automatic detection of alternative paths: MEPSAnd proposes alternative paths via iterative truncation of the GN (Fig. 1F). After a path is obtained, the edges connecting to the highest and latest barrier/s are removed, forcing to sample a different saddle-point. This process runs iteratively until the target cannot be reached anymore.

Data projections to facilitate the visualization of *n*-dimensional results: In order to provide interpretable representations of *n*-dimensional results, MEPSAnd offers three different plotting systems: (i) energy profile plots: representation of the energy evolution along a given path (Fig. 1D), (ii) coordinate projections: 2D and 3D projections of a path over sets of coordinates chosen by the user (Fig. 1C), (iii) network projections: customizable GN graphs over which results can be plotted (Fig. 1E). GraphML exportation is supported.

Other features: (i) Capacity of handling more than one million points in any average computer, (ii) varying degrees of path reduction including diagonal connectivity, (iii) periodic coordinates support (e.g. angles), (iv) session saving system and (vi) python scripting environment.

3 Implementation

MEPSAnd is written in Python 3 and depends on: numpy (Van der Walt *et al.*, 2011), pandas (McKinney, 2010), matplotlib (Hunter, 2007), pycairo (https://www.cairographics.org), python-igraph (Csardi and Nepusz, 2006) and the python port (https://github.com/bhargavchippada/forceatlas2) of the Force Altas 2 (Jacomy *et al.*, 2014) Gephi (Bastian *et al.*, 2009) implementation, developed by Bhargav Chippada.

Acknowledgements

Grants from the Spanish State Research Agency—ERDF Funds (RTI2018-094434-B-I00 and RTC-2017-6494-1 with Repessa-Sistemas SA), and from the EC project 'CONNECT—JPIAMR VRI'. Computational support of the 'CCC-UAM' is acknowledged.

Conflict of Interest: none declared.

References

Banerjee,P. *et al.* (2018) Insulin dimer dissociation in aqueous solution: a computational study of free energy landscape and evolving microscopic structure along the reaction pathway. *J. Chem. Phys.*, **149**, 114902.

Bastian,M. *et al.* (2009) Gephi: an open source software for exploring and manipulating networks. In: *Proceedings of the Third International Conference on Weblogs and Social Media, ICWSM 2009, San Jose, California, USA, May 17-20, 2009*. The AAAI Press 2009.

Csardi,C. and Nepusz,T. (2006) The igraph software package for complex network research. *Inter. J. Complex Syst.*, **1695**, 1–9.

Duan,J. *et al.* (2018) A molecular dynamics study of the complete binding process of meropenem to New Delhi metallo–lactamase 1. *Phys. Chem. Chem. Phys.*, **20**, 6409–6420.

Fritz,R.A. *et al.* (2018) Multiscale simulations of clavulanate inhibition identify the reactive complex in class A-lactamases and predict the efficiency of inhibition. *Biochemistry*, **57**, 3560–3563.

Geronimo,I. *et al.* (2018) Hydrolysis and transglycosylation transition states of glycoside hydrolase family 3 glucosidases differ in charge and puckering conformation. *J. Phys. Chem. B*, **122**, 9452–9459.

Hunter,J.D. (2007) Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.*, **9**, 22–30.

Jacomy,M. *et al.* (2014) ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS One*, **9**, e98679.

Kachmar,A. *et al*. (2017) Mapping the free energy of lithium solvation in the Protic ionic liquid ethylammonuim nitrate: a metadynamics study. *ChemSusChem*, **10**, 3083.

Kachmar,A. and Goddard,W.A.,III. (2018) Free energy landscape of sodium solvation into graphite. *J. Phys. Chem. C*, **122**, 20064–20072.

Li,P. *et al*. (2019) Computational insights into endo/exo selectivity of the diels-alder reaction in explicit solvent at ab initio quantum mechanical/molecular mechanical level. *J. Phys. Chem. B*, **123**, 5131–5138.

Marcos-Alcalde,I. *et al*. (2015) MEPSA: minimum energy pathway analysis for energy landscapes. *Bioinformatics*, **31**, 3853–3855.

Marcos-Alcalde,I. *et al*. (2017) Two-step ATP-driven opening of cohesin head. *Sci. Rep*., **7**, 3266.

McKinney,W. (2010) Data structures for statistical computing in Python. *Proc. 9th Python Sci. Conf*., **9**, 51–56.

Mendieta-Moreno,J.I. *et al*. (2016) Quantum mechanics/molecular mechanics free energy maps and nonadiabatic simulations for a photochemical reaction in DNA: cyclobutane thymine dimer. *J. Phys. Chem. Lett*., **7**, 4391–4397.

Mondal,B. and Reddy,G. (2019) Cosolvent effects on the growth of protein aggregates formed by a single domain globular protein and an intrinsically disordered protein. *J. Phys. Chem. B*, **123**, 1950–1960.

Pirhadi,S. *et al*. (2016) Open source molecular modeling. *J. Mol. Graph. Model*., **69**, 127–143.

Shao,Q. *et al*. (2019) Selective enhanced sampling in dihedral energy facilitates overcoming the dihedral energy increase in protein folding and accelerates the searching for protein native structure. *Phys. Chem. Chem. Phys*., **21**, 10423–10435.

Shao,Q. and Zhu,W. (2019a) Ligand binding effects on the activation of the EGFR extracellular domain. *Phys. Chem. Chem. Phys*., **21**, 8141–8151.

Shao,Q. and Zhu,W. (2019b) Nonnative contact effects in protein folding. *Phys. Chem. Chem. Phys*., **21**, 11924–11936.

Van der Walt,S. *et al*. (2011) The NumPy array: a structure for efficient numerical computation. *Comput. Sci. Eng*., **13**, 22–30.

Yuan,X. *et al*. (2018) The molecular mechanism underlying ligand binding to the membrane-embedded site of a GProtein- coupled receptor. *J. Chem. Theory Comput*., **14**, 2761–2770.