

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,*}

¹Molecular Modelling Group, Centro de Biología Molecular Severo Ochoa, CBMSO (CSIC-UAM), E-28049 Madrid, Spain and

²Biosciences Research Institute, School of Experimental Sciences, Universidad Francisco de Vitoria, UFV, E-28223 Pozuelo de Alarcón, Spain.

*To whom correspondence should be addressed.

Associate Editor: XXXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Summary: Understanding biophysical phenomena from the approach of molecular simulation is becoming the state-of-art in many research and technology development fields. Energy surfaces with more than 3 dimensions (2 coordinates and energy) are now computationally accessible, yet interpreting the information they offer is not straightforward and the tasks involved very time-consuming. Here we present MEPSAnd, an open source GUI-based program that natively calculates minimum energy paths across energy surfaces of an arbitrary number of dimensions. In addition to the multidimensional analysis of path through lowest barriers, MEPSAnd can also automatically calculate a finite series of suboptimal paths. To allow the efficient interpretation of results, MEPSAnd offers three distinct plotting solutions: i) energy profiles, ii) coordinate projections and iii) network projections. GUI-independent pipelines are also supported via direct python scripting. Therefore, MEPSAnd is a powerful user friendly tool that streamlines path-finding tasks on n-dimensional energy surfaces.

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

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

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

1 Introduction

The generation of multi-dimensional energy surfaces is becoming broadly used to understand a wide range of biophysical phenomena. Previously, in our lab, we developed a computational tool able to analyze 3D dimensional surfaces (2 coordinates and energy) from a transition theory perspective (Marcos-Alcalde *et al.*, 2015). This program has been widely used since its release in a variety of computational simulation areas, as Quantum Mechanics/Molecular Mechanics (Mendieta-Moreno *et al.*, 2016; Marcos-Alcalde *et al.*, 2017; Fritz *et al.*, 2018; Li *et al.*, 2019; Geronimo *et al.*, 2018), ligand/drug binding to protein structures (Yuan *et al.*, 2018; Duan *et al.*, 2018; Banerjee *et al.*, 2018), protein folding and protein conformational variations (Shao and Zhu., 2019a,b; Shao *et al.*, 2019; Mondal and Reddy, 2019) or molecular dynamics in materials science (Kachmar *et al.*, 2017; Kachmar and Goddard., 2018). The program was

also included into a catalog of open source tools for molecular modeling (Pirhadi *et al.*, 2016).

Since, over the last years, obtaining n-dimensional surfaces is becoming computationally accessible, an n-dimensional minimum energy path-finding tool would facilitate the direct analysis of this kind of data without the need for dimensional reduction techniques.

To tackle this issue we are introducing MEPSAnd, a GUI-based tool for the analysis of n-dimensional energy surfaces from the transition state theory perspective. MEPSAnd can calculate i) the path that passes through the lowest energy barrier/s connecting two arbitrarily chosen points, ii) the region of the surface that has to be sampled to reach those barriers and iii) a series of alternative sub-optimal paths. To facilitate the interpretation of this n-dimensional data, MEPSAnd also offers a range of plotting and projection options.

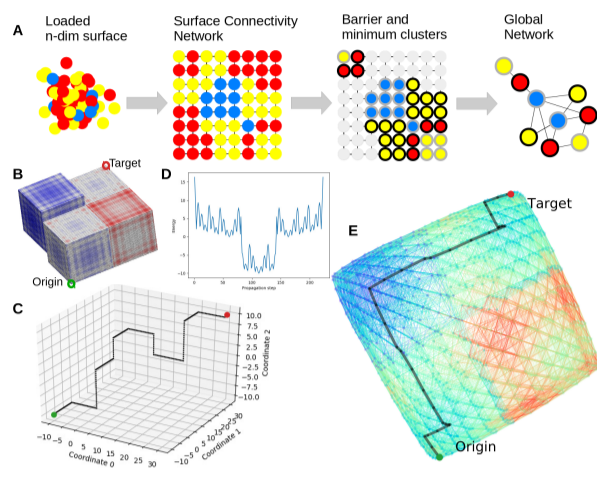


Fig. 1. MEPSAnd overview. **A:** Algorithm steps to build the Surface Connectivity Network and the Global Network. **B:** A 275684 points example 4D surface. Panels **C**, **D** and **E** are native MEPSAnd representations of the resulting simplified minimum energy path from origin (green) to target (red) points of this surface. **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 Global Network representing barrier clusters as edges.

2 Features

Network-based connectivity (grid independent): MEPSAnd functionality relies on two networks representing two different levels of abstraction: a Surface Connectivity Network and a Global Network (Fig. 1A). Surface Connectivity Network describes point-to-point connectivity using a list of neighbors per point that is not directly dependent on the number of dimensions. By default, MEPSAnd builds the Surface Connectivity Network 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.

Global Network is computed over the Surface Connectivity Network, abstracting the whole surface as a network of minima connected by barriers. The Global Network computation consists on:

- **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. This process is called minima propagation. The overlapping regions between different minima propagations are used to define the barrier clusters.
- **Global Network building:** using as vertices the previously defined minimum and barrier clusters, the Global Network is built, in such way that minimum clusters are only connected to barrier clusters and *vice versa*.

Global Network-based path-finding: The new MEPSAnd algorithm searches for the minimum energy path in two sequential steps. First, a path is calculated over the Global Network rendering the visited pairs of minimum and barrier clusters. Second, the paths connecting each of these pairs are computed over the Surface Connectivity Network, yielding a set of path fragments. Joining these path fragments, the minimum energy path is finally reconstructed.

Automatic detection of alternative paths: The use of sequential levels of abstraction during path-finding allows to efficiently propose alternative paths via iterative truncation of the Global Network. After a path is obtained, the edges connecting to the highest and latest barrier/s are removed. Thus, in the next path-finding run, MEPSAnd is forced to sample a different saddle-point. This way it can iteratively propose increasingly sub-optimal paths until all connections between the origin and target regions are finally removed.

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:

- **Energy profile plots:** representation of the energy evolution along a given path (Fig. 1D).
- **Coordinate projections:** 2D and 3D projections of a path over sets of coordinates chosen by the user (Fig. 1C).
- **Network projections:** customizable Global Network graphs over which results can be plotted. GraphML exportation is supported.

Other features: MEPSAnd offers a range of other features such as: i) a computational performance that allows handling surfaces with more than 1 million points in any average computer, ii) varying degrees of path reduction including diagonal connectivity, iii) native support for periodic coordinates (e.g. angles), iv) exhaustive sampling of minimum and barrier plateaus, v) a session-based load/save system or vi) python scripting environment to support GUI-independent pipelines. See Supplementary Material for detailed information.

3 Implementation

MEPSAnd was implemented in Python 3 under an open-source GPLv3 license. The user interface has been built with tkinter standard python GUI library. MEPSAnd depends on the following packages: numpy (Van der Walt *et al.*, 2011), pandas (McKinney, 2010), matplotlib (Hunter, 2007), python-igraph (Csardi and Nepusz, 2006), pycairo (<https://www.cairographics.org>) and the Force Atlas 2 (Jacomy *et al.*, 2014) python port (<https://github.com/bhargavchippada/forceatlas2>), developed by Bhargav Chippada, of the Gephi (Bastian *et al.*, 2009) Force Atlas 2 implementation.

Acknowledgements

The computational support of the "Centro de Computación Científica CCC-UAM" is gratefully recognized.

Funding

This work was supported by grants from the Spanish State Research Agency, partially financed by ERDF Funds (RTI2018-094434-B-I00 and RTC-2017-6494-1 in collaboration with Repessa-Sistemas SA), and by the European Commission project "CONNECT - JPIAMR Virtual Research Institute".

Conflict of Interest: none declared.

References

Bastian, M., Heymann, S. and Jacomy, M. (2009) Gephi: An Open Source Software for Exploring and Manipulating Networks, *International AAAI Conference on Weblogs and Social Media*.

- Csardi, C. and Nepusz T. (2006) The igraph software package for complex network research, *InterJournal*, 1695.
- Banerjee, P., Mondal, S., and Bagchi, B. (2018) Insulin dimer dissociation in aqueous solution: A computational study of free energy landscape and evolving microscopic structure along the reaction pathway, *The Journal of Chemical Physics*, **149**, 114902.
- Duan, J., Hu, C., Guo, J., Guo, L., Sun, J. and Zhao, Z. (2018) A molecular dynamics study of the complete binding process of meropenem to New Delhi metallo- β -lactamase 1, *Physical Chemistry Chemical Physics*, **20(9)**, 6409-6420.
- Fritz, R.A., Alzate-Morales, J.H., Spencer, J., Mulholland, A.J. and van der Kamp, M.W. (2018) Multiscale Simulations of Clavulanate Inhibition Identify the Reactive Complex in Class A β -Lactamases and Predict the Efficiency of Inhibition, *Biochemistry*, **57(26)**, 3560-3563.
- Geronimo, I., Payne, C.M., and Sandgren, M. (2018) Hydrolysis and Transglycosylation Transition States of Glycoside Hydrolase Family 3 β -Glucosidases Differ in Charge and Puckering Conformation, *Journal of Physical Chemistry B*, **122(41)**, 9452-9459.
- Hunter, J.D. (2007) Matplotlib: A 2D graphics environment, *Computing in Science & Engineering*, **9(2)**, 22-30.
- Jacomy M., Venturini T., Heymann S. and Bastian M. (2014) ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software, *PLoS ONE*, **9(6)**, e98679.
- Kachmar, A., Carignano, M., Laino, T., Iannuzzi, M. and Hutter, J. (2017) Mapping the Free Energy of Lithium Solvation in the Protic Ionic Liquid Ethylammonium Nitrate: A Metadynamics Study, *ChemSusChem*, **10**, 3083.
- Kachmar, A. and Goddard, W.A. III (2018) Free Energy Landscape of Sodium Solvation into Graphite, *The Journal of Physical Chemistry C*, **122(35)**, 20064-20072.
- Li, P., Liu, F., Shao, Y. and Mei, Y. (2019) Computational Insights into Endo/Exo Selectivity of the Diels-Alder Reaction in Explicit Solvent at Ab Initio Quantum Mechanical/Molecular Mechanical Level, *The Journal of Physical Chemistry B*, **123(24)**, 5131-5138.
- Marcos-Alcalde, I., Mendieta-Moreno, J.I., Puisac, B., Gil-Rodríguez, M.C., Hernández-Marcos, M., Soler-Polo, D., Ramos, F.J., Ortega, J., Pié, J., Mendieta, J. and Gómez-Puertas, P. (2017) Two-step ATP-driven opening of cohesin head. *Scientific Reports* **7**, 3266.
- Marcos-Alcalde, I., Setoain, J., Mendieta-Moreno, J.I., Mendieta, J., Gómez-Puertas, P. (2015) MEPSA: minimum energy pathway analysis for energy landscapes, *Bioinformatics*, **31(23)**, 3853-3855.
- McKinney, W. (2010) Data Structures for Statistical Computing in Python, *Proceedings of the 9th Python in Science Conference*, **9(2)**, 51-56.
- Mendieta-Moreno, J.I., Trabada, D.G., Mendieta, J. Lewis, J.P., Gómez-Puertas, P. and Ortega, J. (2016) Quantum Mechanics/Molecular Mechanics Free Energy Maps and Nonadiabatic Simulations for a Photochemical Reaction in DNA: Cyclobutane Thymine Dimer, *The Journal of Physical Chemistry Letters*, **7(21)**, 4391-4397.
- Pirhadi S., Sunseri, J. and Koes, D.R. (2016) Open source molecular modeling (2016) *Journal of Molecular Graphics and Modelling*, **69**, 127-143.
- Shao, Q., and Zhu, W. (2019) Ligand binding effects on the activation of the EGFR extracellular domain, *Physical Chemistry Chemical Physics* **21(15)**, 8141-8151.
- Shao, Q., and Zhu, W. (2019) Nonnative contact effects in protein folding, *Physical Chemistry Chemical Physics* **21(22)**, 11924-11936.
- Shao, Q., Yang, L., and Zhu, W. (2019) Selective enhanced sampling in dihedral energy facilitates overcoming the dihedral energy increase in protein folding and accelerates the searching for protein native structure, *Physical Chemistry Chemical Physics* **21(20)**, 10423-10435.
- 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. *The Journal of Physical Chemistry B*, **123(9)**, 1950-1960.
- Van der Walt, S., Colbert, S.C. and Varoquaux, G. (2011) The NumPy Array: A Structure for Efficient Numerical Computation, *Computing in Science & Engineering*, **13(2)**, 22-30.
- Yuan, X., Raniolo, S., Limongelli, V., and Xu, Y. (2018) The Molecular Mechanism Underlying Ligand Binding to the Membrane-Embedded Site of a G-Protein-Coupled Receptor, *Journal of Chemical Theory and Computation*, **14(5)**, 2761-2770.