BIOINFORMATICS APPLICATIONS NOTE Vol. 29 no. 7 2013, pages 950–952
doi:10.1093/bioinformatics/btt051

Structural bioinformatics

Advance Access publication January 30, 2013

MSMExplorer: visualizing Markov state models for biomolecule folding simulations

Bryce Cronkite-Ratcliff[1,*] and Vijay Pande[1,2,3]

[1]Department of Computer Science, [2]Department of Chemistry and [3]Department of Structural Biology, Stanford University, Stanford, CA 94305, USA

Associate Editor: Anna Tramontano

ABSTRACT

Summary: Markov state models (MSMs) for the study of biomolecule folding simulations have emerged as a powerful tool for computational study of folding dynamics. MSMExplorer is a visualization application purpose-built to visualize these MSMs with an aim to increase the efficacy and reach of MSM science.

Availability: MSMExplorer is available for download from https://simtk.org/home/msmexplorer. The source code is made available under the GNU Lesser General Public License at https://github.com/SimTk/msmexplorer.

Contact: pande@stanford.edu

Received on October 22, 2012; revised on December 28, 2012; accepted on January 27, 2013

1 INTRODUCTION

In kinetics-based Markov state models (MSMs) for biomolecule folding simulations, conformation space is partitioned according to natural free-energy barriers, avoiding problematic clustering approximations (Noé and Fischer, 2008). Since their introduction in 2009, programs for generating these models have seen substantial interest and use (Bowman *et al.*, 2009; Senne *et al.*, 2012).

Current visualization tools for MSMs are essentially manual, inhibiting visual analysis of MSMs and slowing down figure production for publications. Indeed, it has become clear that visualization is now a key bottleneck in the use of MSMs to understand kinetic phenomena of interest.

MSMExplorer is an application to address this visualization bottleneck. Specifically, MSMExplorer has three primary aims:

(1) To provide a tool for efficient visual analysis of MSMs.

(2) To streamline the production of publication-ready figures for common MSM network visualization types.

(3) To synthesize disparate MSM research tools into a single intuitive graphical interface.

2 IMPLEMENTATION AND CAPABILITIES

Overview: MSMExplorer is a network visualization application custom-built to visualize MSMs for folding research. The application is built in Java 6 using Swing GUI widgets and makes extensive use of the prefuse visualization toolkit (Heer *et al.*, 2005). The application is organized around a central window, termed GraphView, containing a visualization of the currently loaded MSM, with associated panels used for displaying settings and additional information.

File format support: MSMExplorer has been designed to operate in concert with MSMBuilder (https://simtk.org/home/msmbuilder). As such, MSMExplorer can open raw MSMBuilder output. Additionally, because MSMBuilder projects have a relatively predictable structure, MSMExplorer is able to make educated guesses about the location of additional data. To provide interoperability with other network visualization applications and bundled graph output, MSMExplorer also supports opening and saving files in the GraphML format.

Automatic graph drawing: MSMExplorer makes use of a force-directed layout algorithm for automatic graph drawing, wherein edges act as springs and nodes as masses (Eades, 1984; Fruchterman and Reingold, 1991). In this scheme, the system will generally settle into a local energy minimum, which corresponds to an approximation of a maximally spaced graph within constraints of edge equilibrium length. Simulation parameters can be set by the user. The layout may also be disabled to allow nodes to be placed manually. Additionally, the visibility of nodes may be filtered based on backing data.

Flexible data-based visual encoding variables: In MSMExplorer, many visual encoding variables—node size, color and shape; edge color and weight—are adjustable. In first order, the user may set each of these properties individually for aesthetic control. Moreover, each of these variables may be set to vary with data provided by the user (see example in Fig. 1a). On opening an MSM, MSMExplorer stores node numbers, equilibrium probabilities and transition probabilities, but additional data can be supplied in newline-delimited (for nodes) or matrix (for edges) files; thus, any data imported may be used to vary visual encoding variables.

Scatter plots: MSMExplorer provides facilities to position nodes along 2D axes to generate a scatter plot. Any node data (including user-supplied, as described earlier in the text) may be used for the axes. Axis labels and gridline spacing are fully adjustable. To allow for standard scatter plot production, graph edges may be temporarily hidden. Figure 1b provides an example of a scatter plot layout.

Hierarchical MSMs: MSMBuilder allows for the construction of hierarchical MSMs, which contain multiple models of the same system, each with a different number of macrostates. These hierarchical MSMs may be opened in MSMExplorer.

*To whom correspondence should be addressed.

© The Author 2013. Published by Oxford University Press. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016

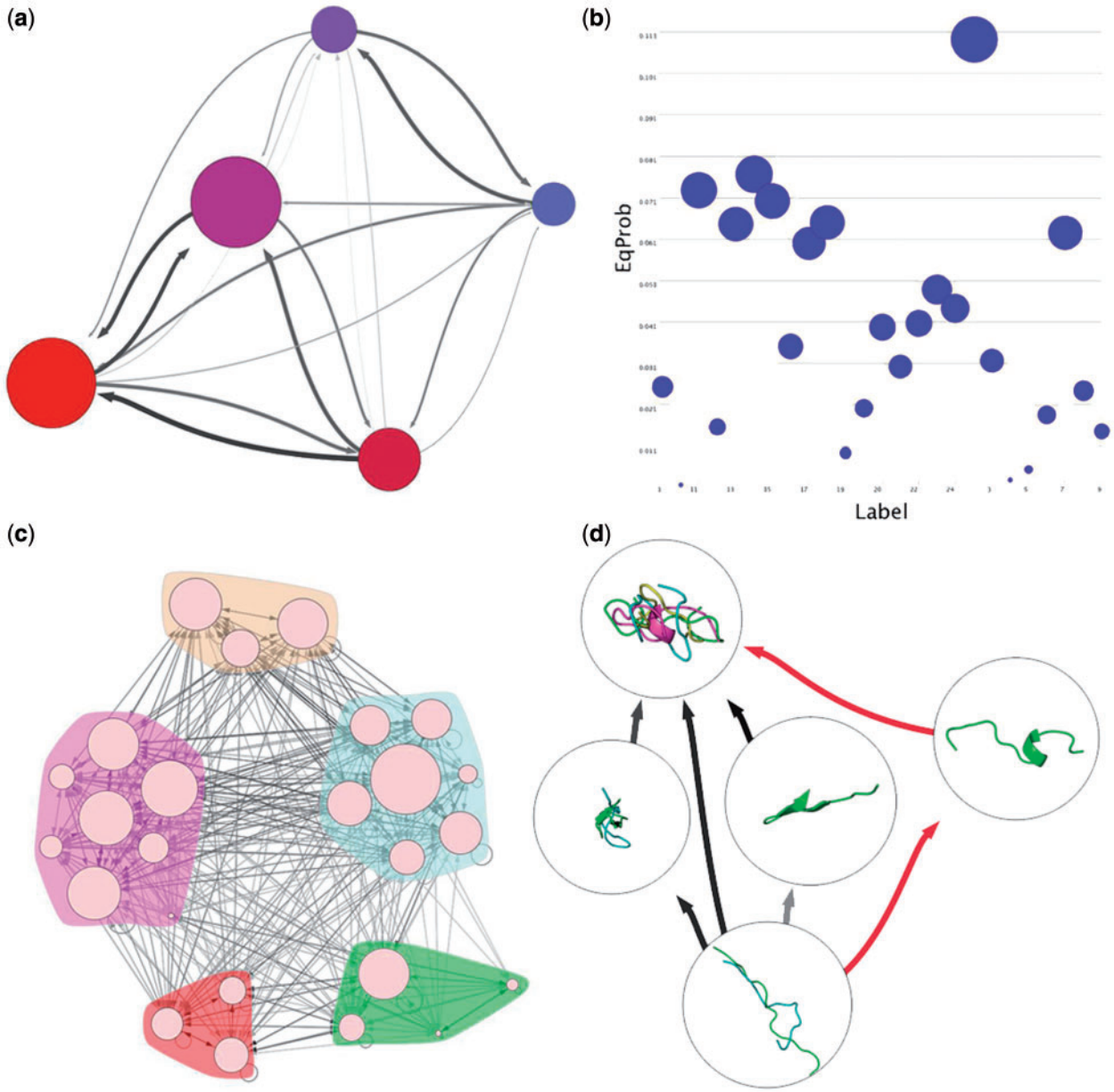

Fig. 1. Visualizations produced with MSMExplorer: (**a**) visual encoding variables can vary with backing data. (**b**) 2D scatter plot layout. (**c**) MSM hierarchy visualization, where the membership of the finer-grained model (tan nodes) in a coarser-grained model is shown. (**d**) A TPT diagram showing the four highest reactive flux paths between two macrostates; images depicting conformations in each macrostate are overlaid on nodes

Users can easily switch between levels of the hierarchy or overlay models in the same hierarchy atop one another to indicate the membership of the states of a more detailed model in the states of a more coarse-grained model. A custom graph-drawing algorithm is applied to group nodes in the same macrostate together. See Figure 1c for an example hierarchical visualization.

Transition path theory: Transition path theory (TPT) is a theoretical framework deployed in biomolecule research to explore folding pathways, in particular those pathways of highest reactive flux between conformations (E and Vanden-Eijnden, 2010; Metzner *et al.*, 2006). In MSMExplorer, the user selects start and end nodes, and a graph depicting the n highest reactive flux paths—where n is set by the user—is automatically generated. Reactive flux through nodes and edges is indicated by node size and edge weight, respectively. Reactive fluxes are saved into the graph backing data and can thus be saved out or used in GraphView to adjust any visual encoding variable. MSMExplorer's implementation of TPT algorithms uses Apache Commons Math matrix facilities and includes an implementation of Dijkstra's algorithm to select highest flux paths after generating a matrix containing reactive fluxes (Dijkstra, 1959).

Image export: As a means to save and share visualizations, MSMExplorer allows export of vector files in SVG format, as

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016

well as export of several raster image file formats. The image can be resized on export.

Support for provided images: MSMExplorer allows users to specify the location of a folder containing images corresponding to nodes. These images may be displayed on top of nodes (see example in Fig. 1d), where they will intelligently rescale with node size, or they are displayed in another window.

PyMol integration: PyMol is a tool for 3D visualization of molecular structures (Schrodinger LLC). On computers with PyMol installed, users may specify a folder containing PDB files. MSMExplorer can then launch PyMol with the PDB file corresponding to the selected node.

2.1 Comparison

Network visualization tools, such as Cytoscape and Gephi, have the most in common with MSMExplorer (Bastian *et al*., 2009; Smoot *et al*., 2011). Both have much more general aims than those of MSMExplorer; thus, they do not provide MSMExplorer's tools specific to biomolecule folding research, including built-in TPT support, PyMol integration, MSM hierarchy support and MSMBuilder integration. Additionally, MSMExplorer is substantially lighter-weight than Gephi or Cytoscape and does away with many features in those programs that would likely be 'cruft' to folding researchers, resulting in a gentler learning curve. Still, MSMExplorer supports full interoperability with these programs via the GraphML format if their interfaces or particular abilities are desired.

3 CONCLUSION

A visualization tool for MSMs for biomolecule folding simulations, called MSMExplorer, has been developed for the purpose of increasing the efficacy of doing folding science with MSMs. The program is currently being used for visual analysis and visualization production for publications in the Pande Lab. Active development will continue to refine MSMExplorer to further achieve its primary aim of advancing folding simulation research.

ACKNOWLEDGEMENTS

The authors thank Jeff Heer and the prefuse team for the prefuse visualization toolkit, MSM scientists in the Pande Lab for their guidance and testing efforts and Lorne Vanatta and Trevor Gokey for contributing code to the project.

Funding: Stanford VPUE through the Stanford Chemistry Department.

Conflict of Interest: none declared.

REFERENCES

Bastian,M. *et al.* (2009) Gephi: an open source software for exploring and manipulating networks. In: *International AAAI Conference on Weblogs and Social Media*.

Bowman,G.R. *et al.* (2009) Using generalized ensemble simulations and Markov state models to identify conformational states. *Methods*, **49**, 197–201.

Dijkstra,E. (1959) A note on two problems in connexion with graphs. *Numerische Mathematik*, **1**, 269–271.

E,W. and Vanden-Eijnden,E. (2010) Transition-path theory and path-finding algorithms for the study of rare events. *Annu. Rev. Phys. Chem.*, **61**, 391–420.

Eades,P. (1984) A heuristic for graph drawing. *Congressus Numerantium*, **42**, 149–160.

Fruchterman,T.M.J. and Reingold,E.M. (1991) Graph drawing by force-directed placement. *Softw. Pract. Exp.*, **21**, 1129–1164.

Heer,J. *et al.* (2005) Prefuse: a toolkit for interactive information visualization. In: *Proceedings of the ACM SIGCHI*.

Metzner,P. *et al.* (2006) Illustration of transition path theory on a collection of simple examples. *J. Chem. Phys.*, **125**, 084110.

Noé,F. and Fischer,S. (2008) Transition networks for modeling the kinetics of conformational change in macromolecules. *Curr. Opin. Struct. Biol.*, **18**, 154–162.

Schrodinger LLC. The PyMol Molecular Graphics System.

Senne,M. *et al.* (2012) EMMA—a software package for Markov model building and analysis. *J. Chem. Theo. Comput.*, **8**, 2223–2238.

Smoot,M.E. *et al.* (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, **27**, 431–432.

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016