MDM-TASK-web: A web platform for protein dynamic residue networks and modal analysis

Olivier Sheik Amamuddy, Michael Glenister, and Özlem Tastan Bishop*

Research Unit in Bioinformatics (RUBi), Department of Biochemistry and Microbiology, Rhodes University, Makhanda 6140, South Africa.

*Corresponding author: o.tastanbishop@ru.ac.za

Abstract

MDM-TASK-web is the web server for the MD-TASK and MODE-TASK software suites. The web server simplifies the set-up for computing and visualizing results from dynamic residue networks, perturbation-response scanning, dynamic cross-correlation, essential dynamics, normal mode analysis and coordination-propensity. Upgrades and new functionalities to these software suites have also been introduced via the web server. An embedded work-flow, integrated documentation and 3D visualization functionality allow for a more intuitive and easy-to-use web platform.

The web server is available at https://mdmtaskweb.rubi.ru.ac.za/. It is Django-based using a MySQL database and is compatible with all major web browsers.

1. Introduction

Molecular dynamics (MD) simulations and normal mode analysis (NMA) are two key computational methods to study protein dynamics, and their structural and functional behavior (Liang *et al.,* 2020; Sheik Amamuddy *et al*., 2020). Traditional MD trajectory analysis tools provide limited information to identify behavior such as allosteric or mutation effects. We, previously, proposed pipelines to analyze mutation effects (Brown and Tastan Bishop 2017; Brown *et al*., 2017a), and allosteric behavior (Penkler *et al*., 2017) using MD trajectories; and established MD-TASK (Brown *et al*., 2017b). Our second software suit, MODE-TASK (Ross *et al*., 2018) comprises NMA and Principal Component Analysis (PCA). Both suites have been highly utilized.

The required technical knowledge and software dependencies may act as a hurdle against more widespread usage of the tools and techniques. MDM-TASK-web bridges this gap by providing access to both tools while introducing new functionalities. It has been designed with a simple and intuitive interface that is supported by any recent web browser. The need for additional software, complex dependencies and command line expertise are greatly reduced. MDM-TASK-web includes new features such as additional DRN metrics, a communication propensity (CP) tool (Chennubhotla and Bahar, 2007; Penkler and Tastan Bishop, 2019), an aggregator of residue contact maps, comparative essential dynamics (ED), an anisotropic network model (ANM) workflow, and integrated 2D & 3D visualization. All tools have been ported to Python 3.

2. Methods

MDM-TASK-web is a single-page web application enabled by the Django web framework (Django, 2013). It uses the Bootstrap (Bootstrap, 2020) framework and the Knockout.js (Knockout.js, 2020) library to implement a dynamic and responsive front-end interface. Jobs are handled by the Job Management System (JMS) (Brown *et al.*, 2015) tool, and NGL (version 2) (Rose *et al.*, 2018) is used for 3D visualization. Extracts from MDM-TASK-web are presented in Figure 1.

2.1 Trajectory management

Transfer and storage of MD data can be a challenge for web servers as working with these large files can be offset by bandwidth and storage limitations. MDM-TASK-web can re-use trajectories and suggests preliminary solvent removal from the trajectory and topology files. A coarse-graining tool (https://github.com/oliserand/MD-TASK-prep), compatible with MDTraj,

PYTRAJ, MDAnalysis, GROMACS, VMD and CPPTRAJ, reduces trajectory size by retaining only C_α and C_β atoms. Trajectory and topology files can be provided as URLs, so that remote simulation data can be analyzed without the need for specialized hardware. User data is privately stored on the server, while trajectories are automatically removed after 30 days.

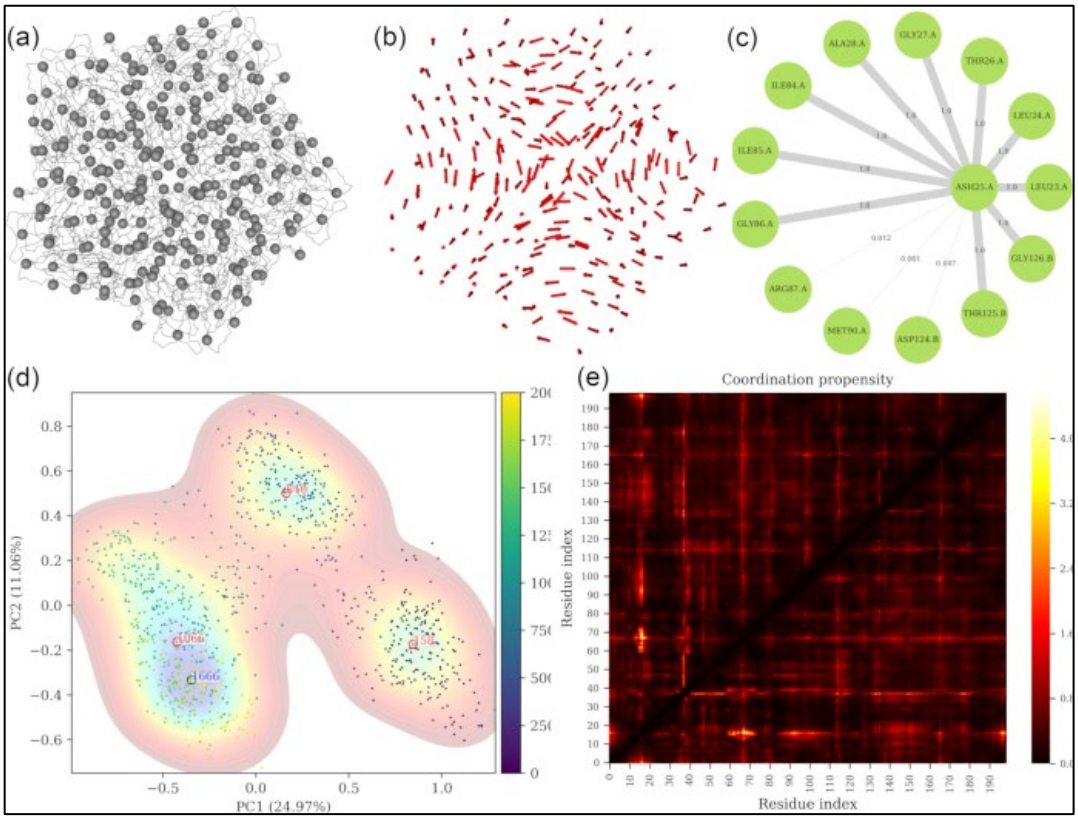

Figure 1: Extracts from the MDM-TASK-web interface, showing (a) a coarse-grained viral capsid pentamer, (b) the normal mode quiver plot for the coarse-grained pentamer, (c) a weighted residue contact map, (d) an essential dynamics plot and (e) a coordination propensity heat map.

2.2 Molecular visualization

The NGL Viewer is used for molecular visualization due to its memory-efficient rendering capability. Various metrics (correlations from PRS, centrality values, coarse-grained atom positions, and modes from ANM) are mapped onto a user-provided protein structure to facilitate interpretation. A viewer is also embedded to enable trajectory visualization.

2.3 MD-TASK

MD-TASK provides tools for performing DRN, DCC and PRS calculations. The previous implementation of DRN has been upgraded to eight metrics (See Figure 2). Each of the network centrality metrics (Hagberg *et al.*, 2008) is computed for each MD frame, and their centrality values are aggregated residue-wise as medians or time-averages. Mapped 3D structures can be directly visualized/ compared and are saved in PDBx/mmCIF format. CSV files of the DRN metrics are generated. A residue contact heat map tool now facilitates the comparison of a relatively large number of weighted contact maps, e.g. a common site or a mutation locus across several mutants. The algorithm for PRS is unchanged; but the interface now requires only a trajectory, its topology (the initial conformation) and a target conformation to generate an interactive 3D map of residue correlations. A new pairwise CP tool, based on the notion that signal transduction events are directly related to distance fluctuations, describes the efficiency of communication between all residue pairs. It is computed as the mean-square fluctuation of inter-residue distance (C_α atoms).

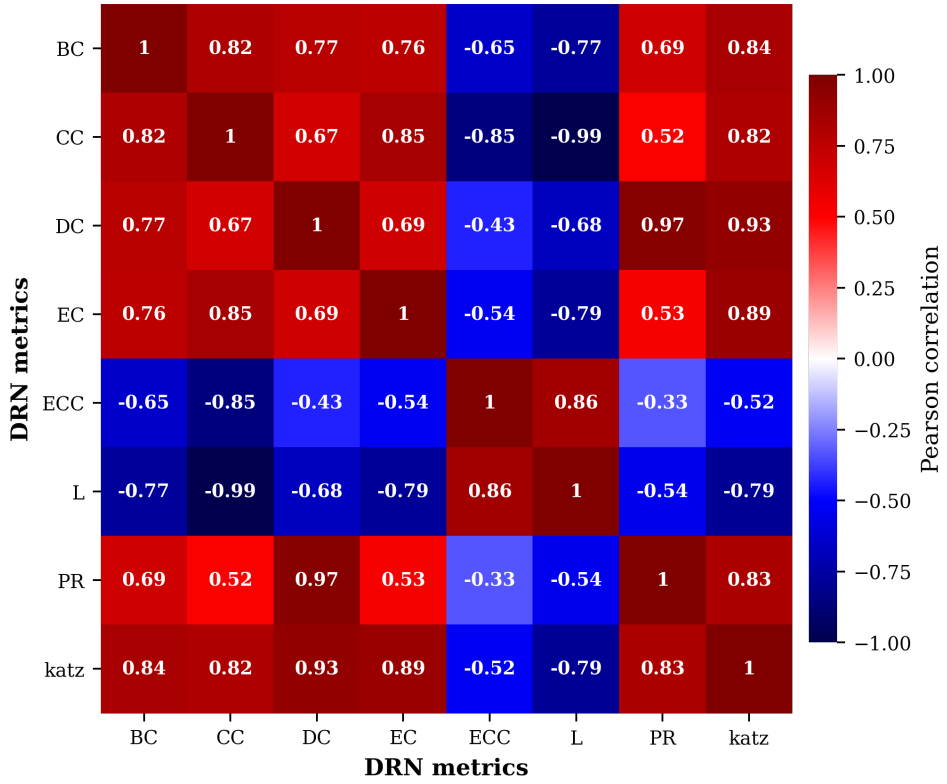

Figure 2: Relationships between DRN metrics inferred from HIV protease using pairwise Pearson correlations.. The previous implementation of DRN in MD-TASK has been upgraded to a total of 8 centrality metrics [betweenness centrality (BC), average shortest path lengths (*L*), closeness centrality (CC), eccentricity (ECC), degree centrality

(DC), eigencentrality (EC), PageRank (PR) and Katz (KC) centrality]. Each of the network centrality metrics is computed for each MD frame, and their centrality values are aggregated residue-wise as medians or time-averages. In this case, the average was used. **Briefly, these are the definitions of the metrics:** (1) BC is the fraction of shortest paths that go through an intervening node. It is normalized by the number of geodesics for all node pairs. (2) CC is highest when the average path length connecting a node all other nodes in a graph is shortest. (3) Conversely, *L* (the inverse of CC) is maximized when the average path length is longest. (4) ECC is the longest path from a node to any other other node in a graph. (5) DC computes the number of residue neighbors (degree) and normalizes by the total number of nodes in the graph. (6) EC is obtained from the eigenvector corresponding to the largest eigenvalue obtained by decomposing the adjacency matrix. This is an extension of DC by virtue of attributing node importance on the basis of the degree of their adjacent nodes. (7) KC is a generalization of EC, which via an adjacency damping coefficient alpha and a basal adjacency beta, assigns centrality on the basis of the node's immediate connectivity, including those beyond one degree of separation. (8) PR is an adjusted version of KC, which assigns centrality values by factoring in the degree of their neighbors and the importance of their connections.

2.4 MODE-TASK

MODE-TASK enables the calculation of protein ED, and the estimation of normal modes from coarse-grained static proteins under the assumptions of the elastic network model. The user is guided from the initial (optional) coarse-graining step, to solving and visualizing the normal modes. Mean square fluctuation can be computed. PCA tools are also integrated. Comparative ED aligns one or more trajectories to a reference trajectory before performing a single decomposition to lay out all conformations on a common set of principal axes. Comparative ED features automated conformation extraction from lowest energy basins and applies k-means to sample centroid conformations from the first 2 principal components in standard PCA. N and C terminal residues may be deselected before the structural alignment step, to reduce unwanted noise a improve performance. Residue selection is enabled, and is applied post global fitting of the C_α atoms.

3. Performance

HIV protease was used to test the server-side run time of each tool (Table 1), except for ANM where a capsid pentamer was used. Test data can be found at https://github.com/oliserand/MD-TASK-prep/example_data.

Table 1. Tool performance evaluations

Tool	Average run time (secs)
PRS (198 residues, 20 frames, 100 perturbations)	246
DCC (198 residues, 1001 frames)	608
CP (198 residues, 1001 frames)	1020

DRN (average or median; 198 residues; 20 frames)	
BC	32
L	71
Degree	28
Closeness	42
Eccentricity	40
Eigenvector	28
PageRank	42
Katz	40
Residue contact map (1001 frames)	22
Contact heat map (6 variants)	10
Essential dynamics (198 residues, 1001 frames)	
Standard PCA	25
Internal PCA	112
MDS	234
t-SNE	263
Comparative ED (2 proteins)	38
ANM from the enterovirus 71 capsid pentamer (PDB ID: 3VBS)	
Coarse-graining	48
ANM construction (270 residues)	64
3D visualization + MSF	17

Acknowledgements

We thank the Centre for High Performance Computing (CHPC); Dr D. Penkler for the CP script, and the RUBi members for their valuable suggestions.

Funding

This work is supported by the National Institutes of Health Common Fund under grant number U41HG006941 to H3ABioNet. The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the funders.

References

Bootstrap (2020) [Internet]. http://getbootstrap.com/.

Brown, D.K. *et al.* (2015) JMS: An Open Source Workflow Management System and Web-Based Cluster Front-End for High Performance Computing. PLoS One, 10, e0134273.

Brown, D.K. *et al.* (2017a) Structure-based analysis of single nucleotide variants in the renin-angiotensinogen complex. Glob. Heart, pii: S2211–8160(17)30006-6.

Brown, D.K. and Tastan Bishop, Ö. (2017) Role of Structural Bioinformatics in Drug Discovery by Computational SNP Analysis. Glob. Heart, pii: S2211-8160(17)30009-1.

Brown, D.K. *et al.* (2017b) MD-TASK: a software suite for analyzing molecular dynamics trajectories. Bioinformatics, 33, 2768–2771.

Chennubhotla, C. and Bahar, I. (2007) Signal Propagation in Proteins and Relation to Equilibrium Fluctuations. PLoS Comput. Biol., 3, e172.

Django (2013) [Computer Software]. Django Softw. Found.

Hagberg, A.A. *et al.* (2008) Exploring network structure, dynamics, and function using NetworkX. In, Varoquaux,G. et al. (eds), Proceedings of the 7th Python in Science Conference (SciPy2008). Pasadena, CA USA, pp. 11–15.

Knockout.js (2020) [Internet]. http://knockoutjs.com/.

Liang, Z. *et al.* (2020) Integration of network models and evolutionary analysis into high-throughput modeling of protein dynamics and allosteric regulation: theory, tools and applications. Brief Bioinform. 2020 May 21;21(3):815-835.

Penkler, D. *et al.* (2017) Perturbation-Response Scanning Reveals Key Residues for Allosteric Control in Hsp70. J Chem Inf Model. Jun 26;57(6):1359-1374.

Penkler, D.L. and Tastan Bishop, Ö. (2019) Modulation of Human Hsp90α Conformational Dynamics by Allosteric Ligand Interaction at the C-Terminal Domain. Sci. Rep., 9, 1600.

Rose, A.S. *et al.* (2018) NGL viewer: web-based molecular graphics for large complexes. Bioinformatics, 34, 3755–3758.

Ross, C. *et al.* (2018) MODE-TASK: large-scale protein motion tools. Bioinformatics, 34, 3759–3763.

Sheik Amamuddy O, *et al.* (2020) Integrated Computational Approaches and Tools forAllosteric Drug Discovery. Int J Mol Sci. 2020 Jan 28;21(3):847.