Bioinformatics, 38(1), 2022, 273–274
doi: 10.1093/bioinformatics/btab538
Advance Access Publication Date: 21 July 2021
Applications Note

OXFORD

Structural bioinformatics

MDContactCom: a tool to identify differences of protein molecular dynamics from two MD simulation trajectories in terms of interresidue contacts

Chie Motono[1,2,*], Shunsuke Yanagida[3], Miwa Sato[3] and Takatsugu Hirokawa[1,4,5,*]

[1]Cellular and Molecular Biotechnology Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tokyo 135-0064, Japan, [2]Computational Bio Big-Data Open Innovation Laboratory (CBBD-OIL), AIST, Waseda University, Tokyo 169-0072, Japan, [3]Bioscience & Healthcare Engineering Division, DX Infrastructure Engineering Unit 1, Mitsui Knowledge Industry Co., Ltd, Tokyo 164-0003, Japan, [4]Division of Biomedical Science, Faculty of Medicine, University of Tsukuba, Ibaraki 305-8575, Japan and [5]Transborder Medical Research Center, Faculty of Medicine, University of Tsukuba, Ibaraki 305-8575, Japan

*To whom correspondence should be addressed.
Associate Editor: Jinbo Xu

Received on December 23, 2020; revised on June 28, 2021; editorial decision on July 16, 2021; accepted on July 20, 2021

Abstract

Summary: Comparing results from multiple MD simulations performed under different conditions is essential during the initial stages of analysis. We propose a tool called MD Contact Comparison (MDContactCom) that compares residue-residue contact fluctuations of two MD trajectories, quantifies the differences, identifies sites that exhibit large differences and visualizes those sites on the protein structure. Using this method, it is possible to identify sites affected by varying simulation conditions and reveal the path of propagation of the effect even when differences between the 3D structure of the molecule and the fluctuation RMSF of each residue is unclear. MDContactCom can monitor differences in complex protein dynamics between two MD trajectories and identify candidate sites to be analyzed in more detail. As such, MDContactCom is a versatile software package for analyzing most MD simulations.
Availability and implementation: MDContactCom is freely available for download on GitLab. The software is implemented in Python3. https://gitlab.com/chiemotono/mdcontactcom.
Contact: c-motono@aist.go.jp or t-hirokawa@md.tsukuba.ac.jp
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

The dynamics of protein molecules are critical for their biochemical function and molecular recognition. MD calculations are useful for sampling the 3D structures of a protein molecule and assessing its dynamics. Contact maps can comprehensively encode the 3D structural information of the molecule in a 2D matrix. This approach has been used for the exhaustive description of intramolecular interactions, 3D structure reconstruction of a protein molecule (Vassura *et al.*, 2008), and the prediction of protein structure including protein complex formation (Pulim *et al.*, 2008). Recently, dynamics information obtained from MD simulations displayed in a contact map has become a useful descriptive method of MD trajectory (Mercadante *et al.*, 2018). An unsupervised neural network-based method has also been developed to detect allosteries by comparison of time fluctuations of protein structures in the form of distance matrices (Tsuchiya *et al.*, 2019).

When analyzing the function of a protein molecule, two or more simulations are usually performed in parallel with different system setups. For example, different temperature or pressure in the ligand-binding state (apo/holo), in the ligand molecule or mutations of protein residues. Comparison of these trajectories is critical. At the initial stage of analysis, the productivity or convergence of simulations are checked, then differences between the trajectories are verified with root mean square deviations (RMSD), root mean square fluctuations (RMSF) and secondary structural changes. At the advanced stage, major dynamics are extracted and compared by PCA (Kitao *et al.*, 1991) or more sophisticated methodologies like PLS-DA (Peters and de Groot, 2012) or LDA-ITER (Sakuraba and Kono, 2016).

Here, we propose MD Contact Comparison (MDContactCom), a tool that compares the residue-residue contact fluctuation of two MD trajectories, quantifies the difference as similarity indices, and visualizes the sites where the difference in the index is large. The tool is highly automated and executed with a single command to extract affected sites and visualize them during the initial evaluation. With recent advances in structural analysis, particularly cryo-electron microscopy, the number of protein structures is increasing

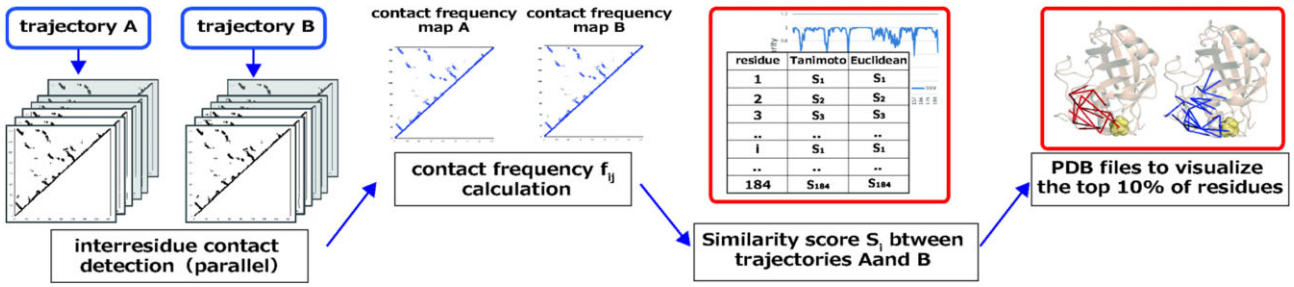

Fig. 1. Workflow of MDContactCom in default mode. Input and output files are shown in blue and red squares, respectively. MDContactCom detects contacts in each structure of the input trajectories, then calculates a contact frequency *fij* between two residues through a trajectory. Contact frequency *fij* of each residue i is compared between two trajectories to give a similarity index *Si* (Tanimoto coefficient and Euclidean distance). In addition to text data and a plot of *Si* against *i*, MDContactCom provides PDB files for visualization of residues with significant similarity indices

exponentially. MDContactCom can be used by structural biologists to rapidly detect differences in protein dynamics.

2 Features

2.1 Depiction of algorithms

The MDContactCom workflow in default mode is presented in Figure 1. Details of inputs, outputs and formulas are provided in Supplementary Data (Supplementary Appendices S1 and S2).

When two MD trajectories A and B [in pdb format or output of Amber (Case *et al.*, 2021), CHARMM (Brooks *et al.*, 2009), Desmond (Bowers *et al.*, 2006), GROMACS (Abraham *et al.*, 2015) or NAMD (Phillips *et al.*, 2020)] are inputted to MDContactCom, they are processed as follows:

i. Contact frequency calculation. Interresidue contacts are detected for each structure frame in an MD trajectory. A contact frequency f_{ij} between residue *i* and residue *j* is then calculated.
ii. Comparison of contact frequencies between two trajectories. To compare two trajectories A and B, similarity coefficients (Tanimoto coefficient and Euclidean distance) S_{iAB} of residue i are calculated and the output is presented as a table and graph.
iii. Visualization of residues with large differences. PDB files are created to highlight residues with significant S_{iAB} and their contacts on the 3D structure. This information is useful for identifying regions of the protein to focus on after MD simulations are performed.

2.2 An example of the application of MDContactCom

We applied MDContactCom to analyze the MD trajectories of Cyclophilin A and its variant V29L. The mutation is reported to have an allosteric effect upon the distal binding site without accompanying conformational changes (Doshi *et al.*, 2016; Holliday *et al.*, 2017). Details of the analysis are described in Supplementary Data (Supplementary Appendix S3). MDContactCom detected residues in the pathways where the mutation effects propagate (Supplementary Fig. S4).

3 Conclusion

MDContactCom compares two MD trajectories on an interresidue contact basis, quantifies the differences in contact frequency for each residue, and visualizes the sites with large differences and their contacts on a 3D structure. This method is a versatile tool for the analysis of MD calculations with a wide range of applications for trajectory comparison under different simulation conditions. Applications include equilibrium versus non-equilibrium state, analysis of unfolding, mutations and association-dissociation of ligands or biomolecules. Moreover, structural biologists will find MDContactCom is easily accessible and simple to use.

Funding

This work was supported as a Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS) by Japan Agency for Medical Research and Development (AMED) [JP18am0101114 to C.M. and T.H.].

Conflict of Interest: none declared.

References

Abraham,M.J. *et al.* (2015) Gromacs: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, **1-2**, 19–25.
Bowers,K.J. *et al.* (2006) Scalable algorithms for molecular dynamics simulations on commodity clusters, In *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06), Tampa, Florida, 2006, November 11-17*.
Brooks,B.R. *et al.* (2009) CHARMM: the biomolecular simulation program. *J. Comput. Chem.*, **30**, 1545–1614.
Case,D.A. *et al.* (2021) *Amber 2021*, University of California, San Francisco.
Doshi,U. *et al.* (2016) Dynamical network of residue–residue contacts reveals coupled allosteric effects in recognition, catalysis, and mutation. *Proc. Natl. Acad. Sci. USA*, **113**, 4735–4740.
Holliday,M.J. *et al.* (2017) Networks of dynamic allostery regulate enzyme function. *Structure*, **25**, 276–286.
Kitao,A. *et al.* (1991) The effects of solvent on the conformation and the collective motions of protein: normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum. *Chem. Phys.*, **158**, 447–472.
Mercadante,D. *et al.* (2018) CONAN: A tool to decode dynamical information from molecular interaction maps. *Biophys. J.*, **114**, 1267–1273.
Peters,J.H. and de Groot,B.L. (2012) Ubiquitin dynamics in complexes reveal molecular recognition mechanisms beyond induced fit and conformational selection. *PLoS Comput. Biol.*, **8**, e1002704.
Phillips,J.C. *et al.* (2020) Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.*, **153**, 044130.
Pulim,V. *et al.* (2008) Optimal contact map alignment of protein-protein interfaces. *Bioinformatics*, **24**, 2324–2328.
Sakuraba,S. and Kono,H. (2016) Spotting the difference in molecular dynamics simulations of biomolecules. *J. Chem. Phys.*, **145**, 074116.
Tsuchiya,Y. *et al.* (2019) Autoencoder-based detection of dynamic allostery triggered by ligand binding based on molecular dynamics. *J. Chem. Inf. Model.*, **59**, 4043–4051.
Vassura,M. *et al.* (2008) FT-COMAR: fault tolerant three-dimensional structure reconstruction from protein contact maps. *Bioinformatics*, **24**, 1313–1315.