**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, \(^3\)Mitsui Knowledge Industry Co., Ltd, Tokyo 135-0064 Japan, \(^4\)Division of Biomedical Science, Faculty of Medicine, University of Tsukuba, Ibaraki 305-8575 Japan, \(^5\)Transborder Medical Research Center, Faculty of Medicine, University of Tsukuba, Ibaraki 305-8575 Japan

*To whom correspondence should be addressed.

Received on XXXXX; revised on XXXXX; accepted on XXXXX

**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:** 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

**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 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 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 to visualize them during the initial evaluation. With recent advances in structural analysis, particularly cryo-electron microscopy, the number of protein structures is increasing 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 (Appendix 1 and 2).

When two MD trajectories A and B (in pdb format or output of Amber (Case et al., 2021), CHARMM (Brooks et al., 2009), Desmond (Schrödinger, 2021), 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 a 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) \( S_{AB} \) 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 \( S_{AB} \) 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 (Appendix 3). MDContactCom detected residues in the pathways where the mutation effects propagate (Figure 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.
Brooks,B.R. et al. (2009) CHARMM: The biomolecular simulation program. J. Comput. Chem., 30, 1545–1614.
Doshi,U. et al. (2016) Dynamical network of residue-residue contacts reveals coupled allosteric effects in recognition, catalysis, and mutation. Proc. Natl. Acad. Sci., 113, 4735–4740.
Case, D.A et al., (2021), Amber 2021, University of California, San Francisco.
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.
Phillips, J.C. et al. (2020) Scalable molecular dynamics on CPU and GPU architectures with NAMD. J. Chem. Phys., 153.
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.
Schrödinger Release 2021-2: Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2021. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2021.
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.