Structural bioinformatics

ConfID: an analytical method for conformational characterization of small molecules using molecular dynamics trajectories

Polêto, M. D. [1,2,*,**], **Grisci, B. I.** [3,*,**], **Dorn, M.** [3] **and Verli, H.** [1]

[1]Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 91509-900, Brazil and [2]Departamento de Biologia Geral, Universidade Federal de Viçosa, Viçosa, 36570-000, Brazil and [3]Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 91509-900, Brazil

*These authors contributed equally to this work.

**To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: The conformational space of small molecules can be vast and difficult to assess. Molecular dynamics simulations of free ligands in solution have been applied to predict conformational populations, but their characterization is often based on clustering algorithms or manual efforts.

Results: Here, we introduce *ConfID*, an analytical tool for conformational characterization of small molecules using molecular dynamics trajectories. The evolution of conformational sampling and population frequencies throughout trajectories is calculated to check for sampling convergence while allowing to map relevant conformational transitions. The tool is designed to track conformational transition events and calculate time-dependent properties for each conformational population detected.

Availability: Toolkit and documentation are freely available at `http://sbcb.inf.ufrgs.br/confid`

Contact: marcelo.poleto@ufv.br and bigrisci@inf.ufrgs.br

Supplementary information: Supplementary data are available at *Bioinformatics*

1 Introduction

The conformational space of molecules can be vast, and experimental techniques usually employed, such as X-ray crystallography and NMR, are often limited by its solid-state nature or its sensitivity, respectively (Reynolds, 2014; Skinner *et al*., 2008). Computationally, force field parameters, search algorithms, and geometry sampling are used to predict conformational populations and their derived properties (Riniker *et al*., 2015; Supady *et al*., 2015; Salt *et al*., 2005), generating robust information for ligand-based drug design (LBDD) methods. In this sense, molecular dynamics simulations have been employed to assess the conformational landscape of small molecules and to derive time-dependent properties while accounting for solvation effects (Arantes *et al*., 2019; Dolenc *et al*., 2019). Still, identification and quantification of conformational populations sampled in molecular dynamics (MD) simulations are usually carried by clusterization algorithms, which can be insensitive to small conformational changes or tackled by manual efforts, which can be a source of errors. Here, we present *ConfID*, a free, open-source Python-based tool for characterization of conformational populations of drug-like molecules based on molecular dynamics trajectories.

2 General approach

ConfID is based on dihedral values of torsions of drug-like compounds sampled throughout MD simulations. Details regarding the *ConfID* approach, inputs, and analysis are available in Supplementary Data.

3 *ConfID* applications

3.1 Ligand conformations in solution

Accurate assessment of the conformational space of small molecules can provide substantial information for medicinal chemists when designing new drugs. In this sense, *ConfID* can provide a list of the most common conformations of a lead-compound in water and the impact of possible chemical modifications on their relative abundance in solution. Similar data were reported by (Arantes *et al*., 2019) for chalcones and flavonoids scaffolds using a combination of MD simulations and *ConfID*. Moreover, *ConfID* can be advantageously used to further investigate NMR

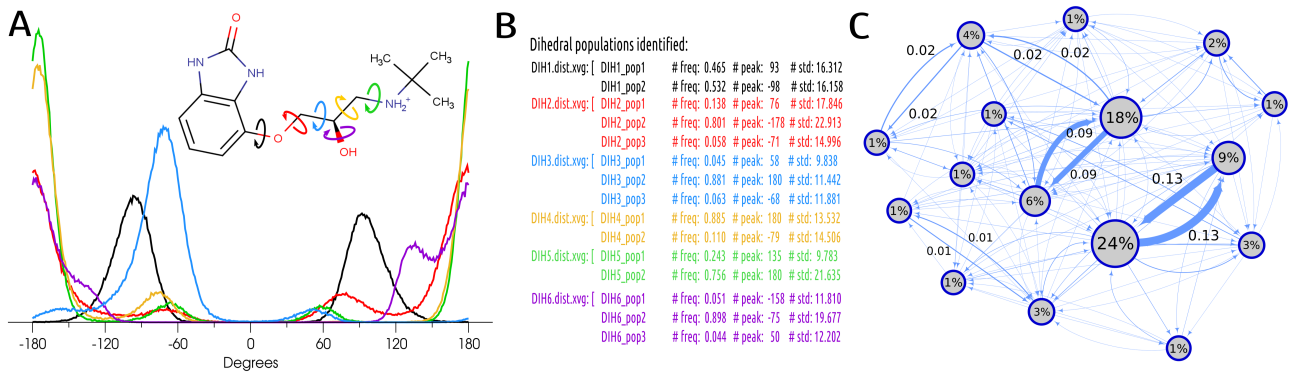

Fig. 1. Schematics of the ConfID approach. Input files contains dihedral values sampled throughout MD simulations (A). Dihedral populations are identified (B), and conformational populations (unique tuple of dihedral populations) are quantified (percentage within nodes), along with transition events (percentages below 1% not shown), which are listed and plotted as graphs (C).

NOESY data by identifying conformational populations of ligands in MD simulations that correlates with experimental results while allowing in-depth comprehension of their chemical features.

3.2 Solvent effects on ligands conformation in solution

Solvent dynamics play a major role in the conformational preferences of small molecules, but their characterization is still challenging in MD studies. By tracking conformers' relative abundance and pinpointing frames of MD trajectories in which each conformer occurs, *ConfID* provides substantial information to investigate the chemical basis of conformational preferences. Arantes *et al*., 2019 reported drastic differences in using aqueous or organic solvent on the conformational ensembles of glycosylated chalcones in MD simulations. The authors observed that some conformations are more frequent in water due to solvent molecules bridging interactions between distant sites (Figures S3 and S4).

3.3 Bound and unbound conformations relationship

When available, experimental structures of ligands bound to their receptor can provide insightful information when compared with their identified conformational populations sampled in MD simulations, as provided by *ConfID* (Arantes *et al*., 2019). The existence of the bound conformation among the conformers in solution is useful for medicinal chemists, whom can suggest chemical modifications to 1) improve molecular interactions between ligand and receptor; or 2) enhance the relative abundance of the bound conformation in solution, favoring its binding entropy (Figure S5).

3.4 Ligand dynamics bound to receptors

Although it is not the primary goal of *ConfID*, our approach can also be applied to the characterization of conformational ensembles of a ligand bound to its receptor. By identifying conformational populations of ligands within the binding site, *ConfID* enables comparisons between bound and unbound dynamics, which might shed light upon molecular recognition mechanisms. However, since achieving ergodicity of MD simulations of large ligand-receptor complexes is difficult, a complete sampling of all ligand conformations is rarely obtained and, therefore, measures of conformers relative abundance or transition events frequencies on such cases may be misleading.

4 Impact of MD simulation parameters

It is important to mention that, as an analysis tool of MD trajectories, *ConfID* fully relies on the correct use of MD simulation parameters and accurate ligand topology parameters and, therefore, users must bear in mind that *ConfID* results are as reliable as the simulations themselves.

5 Conclusion

ConfID can be easily applied for the conformational characterization of drug-like compounds, while taking full advantage of MD simulations, such as accounting for solvation effects or evaluating time-dependent properties. By combining information of transition paths, time-dependent information and population frequencies, *ConfID* can be used to distinguish most relevant conformations from transition intermediates, which can be a challenge for both experimental and computational , while providing crucial information for further LBDD approaches.

Funding

This research received funding by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Finance Code 001, CAPES/Drug Discovery [*23038.007777/2014-87*], CAPES/PROBRAL [*88881.198766/2018-01*], FAPERGS [*16/2551-0000520-6*], Alexander von Humboldt-Stiftung (AvH) [*BRA 1190826 HFST*].

References

Arantes, P.R., Polêto, M.D., John, E.B.O., Pedebos, C., Grisci, B.I., Dorn, M., Verli, H. (2019) Development of GROMOS-Compatible Parameter Set for Simulations of Chalcones and Flavonoids, *Journal of Physical Chemistry B*, **123:5**, 994-1008.

Dolenc, J., van Gunsteren, W.F., Prota, A.E., Steinmetz, M.O., Missimer, J.H. (2019) Conformational Properties of the Chemotherapeutic Drug Analogue Epothilone A: How to Model a Flexible Protein Ligand Using Scarcely Available Experimental Data, *Journal of Chemical Information and Modeling*, 59 (5), 2218-2230.

Reynolds,C.H., (2014) Protein-Ligand Cocrystal Structures: We Can Do Better, *ACS Medicinal Chemistry Letters*, **5:7**, 727-729.

Riniker, S., Landrum, G.A. (2015) Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation, *Journal of Chemical Information and Modeling*, **55:12**, 2562-2574.

Skinner,A.L., Laurence,J.S. (2008) High-field Solution NMR Spectroscopy as a Tool for Assessing Protein Interactions with Small Molecule Ligands, *Journal of Pharmaceutical Sciences*, **97:11**, 4670âĹ“4695.

Salt,D.W., Hudson,B.D., Banting,L., Ellis,M.J., Ford,M.G. (2005) DASH: A Novel Analysis Method for Molecular Dynamics Simulation Data. Analysis of Ligands of PPAR-γ, *Journal of Medicinal Chemistry*, **9:48**, 3214-3220.

Supady,A., Blum,V., Baldauf, C. (2015) First-Principles Molecular Structure Search with a Genetic Algorithm, *Journal of Chemical Information and Modeling*, **55:11**, 2338-2348.

Tesch,R., Becker,C., Müller,M., *et al* (2018) An Unusual Intramolecular Halogen Bond Guides Conformational Selection, *Angewandte Chemie Internacional Edition*, **57:31**, 9970-9975.