

do_x3dna: A tool to analyze structural fluctuations of dsDNA or dsRNA from molecular dynamics simulations

Rajendra Kumar^{1,*} and Helmut Grubmüller¹

¹Department of Theoretical and Computational Biophysics, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, Göttingen, 37077, Germany.

Associate Editor: Prof. Anna Tramontano

ABSTRACT

Summary: The do_x3dna package has been developed to analyze the structural fluctuations of DNA or RNA during molecular dynamics simulations. It extends the capability of the 3DNA package to GROMACS MD trajectories and includes new methods to calculate the global-helical axis of DNA and bending fluctuations during simulations. The package also includes a Python module dnaMD to perform and visualize statistical analyses of complex data obtained from the trajectories.

Availability and Implementation: The source code of the do_x3dna is available at https://github.com/rjdkmr/do_x3dna under GNU GPLv3 license. A detailed documentation, including tutorials and required input data, are freely available at http://rjdkmr.github.io/do_x3dna/.

Contact: rjdkmr@gmail.com or rajendra.kumar@mpibpc.mpg.de

1 INTRODUCTION

Interactions of double helical nucleic acids, namely dsDNA and dsRNA, with proteins are essential for several biological processes including DNA replication (Masai, et al., 2010), transcription (Shandilya and Roberts, 2012), translocation (Croizat and Grainge, 2010) or packaging (Andrews and Luger, 2011; Casjens, 2011). To understand the functional mechanism of these processes, structural properties of nucleic acids have been probed using both experimental and computational studies. In particular, molecular dynamics (MD) simulations have been widely applied to study the roles of nucleic acids in biological process at atomic resolution (Cheatham and Case, 2013; Mackerell and Nilsson, 2008; Rohs, et al., 2009). In this context, several structural descriptors (Lu and Olson, 2003) frequently need to be calculated from the MD simulations. GROMACS (Pronk, et al., 2013), a widely used open-source MD simulation software package, does not offer tools to calculate such structural descriptors of the dsDNA or dsRNA. Presently, two software tools, 3DNA (Lu and Olson, 2008) and Curves+ (Blanchet, et al., 2011), are available to calculate these descriptors of nucleic acids using a PDB input file. These two packages have been extended for use with MD trajectory files in either PDB or AMBER format. However, these packages cannot

be used directly with binary format trajectories (*trr/xtc*) and topology-parameter (*tpr*) files generated by GROMACS.

Therefore, we have developed the do_x3dna package, which directly takes GROMACS input files, calculates structural descriptors frame wise using the 3DNA package, and generates output files in text format. These files can be further used for statistical analyses. However, these output files potentially contain large amounts of data, particularly as microseconds simulations are becoming routine. Consequently, analyses of raw data obtained from do_x3dna can be challenging. To address this problem, we additionally developed the Python module dnaMD, which contains several methods to extract and analyze raw data that were previously calculated using the do_x3dna.

2 IMPLEMENTATION

The do_x3dna package consists of two parts: 1) the do_x3dna tool and 2) the dnaMD Python module (Fig. 1). The do_x3dna tool is written in C and links against GROMACS libraries *libgmx*, *libmd* and *libgmxana*. It uses external 3DNA tools “find_pair” and “analyze” to calculate structural descriptors for each frame, which are on the fly dumped as a PDB file from the trajectory. The do_x3dna tool requires a topology-parameter (*tpr*), a trajectory (*trr/xtc*) file and optionally an atomic index (*ndx*) file as input. The user interface of this program is similar to other GROMACS tools. Additionally, the option for structure alignment is provided to fit each snapshot on the starting structure.

Currently, do_x3dna can be used to calculate structural descriptors such as base-pair parameters (shear, stretch, stagger, buckle, propeller, and opening), base-step parameters (shift, slide, rise, tilt, roll, and twist), or helical parameters (X-displacement, Y-displacement, helical rise, inclination, tip, and helical twist). The option to calculate their average values over time is provided. In addition, major and minor grooves of the double helix, backbone (α , β , γ , δ , ϵ and ξ) and sugar dihedral angles of both strands, helical radii of both strands, and positions of local helical axes can be calculated. Moreover, it generates hydrogen bond map for each base-pair as a function of time.

The dnaMD module uses the external Python modules NumPy (van der Walt, et al., 2011) and SciPy (<http://www.scipy.org/>). A detailed documentation of dnaMD is provided at this website http://rjdkmr.github.io/do_x3dna/apidoc.html. dnaMD implements

*To whom correspondence should be addressed.

a DNA class containing several variables and methods in an object oriented approach. An instance of this class can be initialized with the number of base-pairs present in the dsDNA. Subsequently, the data can be stored in this object variable by reading input files using implemented methods. Subsequently, the data can be analyzed using methods such as `time_vs_parameter`, `parameter_distribution`, `get_mean_error`, `dev_bps_vs_parameter`, `generate_smooth_axis`, `calculate_curvature_tangent` etc. (see details in the above website). Additionally, this module calculates global helical axis by fitting a smooth curve on coordinates of local helical axis using cubic spline interpolation (Fig. 1B).

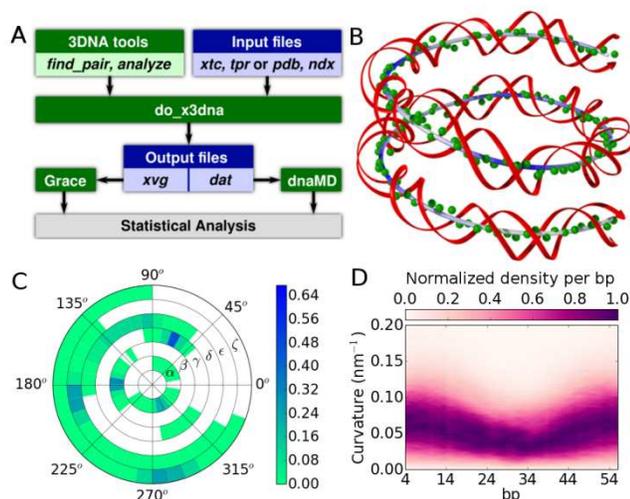


Figure 1: (A) Flow-chart depicting work-flow of the `do_x3dna` package. Grace is an external free plotting tool. (B) Local (green spheres) and global (cylinder) helical axis with its curvature (white-blue: 0.1–0.4 nm⁻¹) of a DNA (red) from a nucleosome (PDB ID: 3UT9). (C) Example distribution of backbone torsion angles as a wheel plot during 330 ns MD simulations of a 60 bp DNA. (D) Example curvature fluctuations of a 60 bp DNA along its global helical axis during MD simulations.

Several tutorials including necessary input files are provided online at http://rjdkmr.github.io/do_x3dna/tutorial.html, which describe how to use the `dnaMD` module for analyses.

3 DISCUSSION

The current implementation was tested with simulation trajectories of a dsDNA, which was either unbound or bound to the protein channel of Phi29 bacteriophage in MD simulations. Subsets of these tests are available online as tutorials at above described website. The tutorials exemplify the use of `dnaMD` to characterize the impact of protein binding on the dsDNA structure.

The `do_x3dna` package extends the efficient and straightforward application of the existing 3DNA package to MD simulations. Note that the Curves+ package contains a program “CANAL”, which was specifically developed for analyzing data obtained from MD simulations. However, this package uses AMBER trajectory and topology files, and generates limited results in text format output files. `do_x3dna` can be used directly for GROMACS, and also for NAMD (Phillips, et al., 2005) or AMBER (Case, et al., 2005) trajectories after converting them into `trr/xtc` or PDB format. A PDB file can be used instead of a GROMACS `tpr` file.

A major advantage of using the `do_x3dna` is the availability of the Python module `dnaMD`, which provides ample flexibility in direct analyses of large and complex raw data. The `dnaMD` module can be used in a Python script in combination with other mathematical modules such as NumPy or SciPy, to perform rapid mathematical and statistical calculations. The obtained results can be visualized by using the Matplotlib module as shown in Fig. 1C and D (Hunter, 2007) as described in the tutorials. Moreover, the `dnaMD` module offers an additional feature over other available tools, namely analysis of helical axis bending by calculating either its curvature (Fig. 1B) or tangent angles using Frenet-Serret formula (Kreyszig, 1991). Overall, the `do_x3dna` package offers an efficient and flexible implementation to quantify and analyze several structural descriptors of dsDNA or dsRNA during MD simulations.

4 CONCLUSION

The `do_x3dna` package allows one to study biological processes that are mediated by nucleic acids through explicit solvent atomistic MD simulations. The application range of the `dnaMD` module could be further extended by features, such as to calculate the elastic property of the dsDNA.

Funding: This work was supported by the European Commission [grant number 211800 SBMPS, FP7/2007-2013].

REFERENCES

- Andrews, A.J. and Luger, K. (2011) Nucleosome structure(s) and stability: variations on a theme. *Annu Rev Biophys*, 40, 99-117.
- Blanchet, C., et al. (2011) CURVES+ web server for analyzing and visualizing the helical, backbone and groove parameters of nucleic acid structures. *Nucleic Acids Res*, 39, W68-73.
- Case, D.A., et al. (2005) The Amber biomolecular simulation programs. *J Comput Chem*, 26, 1668-1688.
- Casjens, S.R. (2011) The DNA-packaging nanomotor of tailed bacteriophages. *Nat Rev Microbiol*, 9, 647-657.
- Cheatham, T.E., 3rd and Case, D.A. (2013) Twenty-five years of nucleic acid simulations. *Biopolymers*, 99, 969-977.
- Crozat, E. and Grainge, I. (2010) FtsK DNA translocase: the fast motor that knows where it's going. *ChemBiochem*, 11, 2232-2243.
- Hunter, J.D. (2007) Matplotlib: A 2D graphics environment. *Comput Sci Eng*, 9, 90-95.
- Kreyszig, E. Chapter II: Theory of Curves. In, *Differential Geometry*. New York: Dover Publications Inc.; 1991. p. 40-42.
- Lu, X.J. and Olson, W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res*, 31, 5108-5121.
- Lu, X.J. and Olson, W.K. (2008) 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat Protoc*, 3, 1213-1227.
- Mackerell, A.D., Jr. and Nilsson, L. (2008) Molecular dynamics simulations of nucleic acid-protein complexes. *Curr Opin Struct Biol*, 18, 194-199.
- Masai, H., et al. (2010) Eukaryotic chromosome DNA replication: where, when, and how? *Annu Rev Biochem*, 79, 89-130.
- Phillips, J.C., et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem*, 26, 1781-1802.
- Pronk, S., et al. (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29, 845-854.
- Rohs, R., et al. (2009) Nuance in the double-helix and its role in protein-DNA recognition. *Curr Opin Struct Biol*, 19, 171-177.
- Shandilya, J. and Roberts, S.G. (2012) The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling. *Biochim Biophys Acta*, 1819, 391-400.
- van der Walt, S., Colbert, S.C. and Varoquaux, G. (2011) The NumPy Array: A Structure for Efficient Numerical Computation. *Comput Sci Eng*, 13, 22-30.