Bioinformatics, 32(10), 2016, 1568–1570
doi: 10.1093/bioinformatics/btw020
Advance Access Publication Date: 14 January 2016
Applications Note

OXFORD

Structural bioinformatics

SIRAH tools: mapping, backmapping and visualization of coarse-grained models

Matías R. Machado* and Sergio Pantano*

Biomolecular Simulations, Institut Pasteur De Montevideo, Montevideo 11400, Uruguay

*To whom correspondence should be addressed.

Associate Editor: Anna Tramontano

Received on 9 December 2015; revised on 7 January 2016; accepted on 8 January 2016

Abstract

Summary: Coarse-grained (CG) models reduce the cost of molecular dynamics simulations keeping the essence of molecular interactions. Still, the diversity of CG representations (sizes, connectivity, naming, etc.) hampers the handling and visualization of such models. SIRAH Tools comprises a set of utilities to convert all-atoms coordinates to arbitrary residue-based CG schemes, write GROMACS' topological information at any resolution into PSF format and a VMD plugin to visualize, analyze and retrieve pseudo-atomistic information from CG trajectories performed with the SIRAH force field. These tools facilitate the use of intricate CG force fields outside the small developer's community.

Availability and implementation: Different utilities of SIRAH Tools are written in Perl, Tcl, or R. Documentation and source codes are freely distributed at http://www.sirahff.com.

Contact: mmachado@pasteur.edu.uy or spantano@pasteur.edu.uy

1 Introduction

Molecular dynamics (MD) simulations constitute a reliable approach for the study of bimolecular structures, dynamics and interactions (Karplus and Lavery, 2014). However, its computational cost has prompted the development of coarse-grained (CG) models aimed to reach ever-growing size and timescales. In general, different mapping schemes are used to provide CG representations from all-atoms (AA) molecular systems (Ingólfsson *et al.*, 2014), making difficult the creation and editing of molecular topology files. Moreover, arbitrary CG schemes can be challenging for general-purposes visualization programs to yield correct rendering, molecular connectivity, beads' size, charges, etc. To overcome such difficulties we created a series of utilities named SIRAH Tools, which are distributed within the SIRAH force field package for CG and multiscale simulations (see Darré *et al.*, 2015 and references therein). SIRAH Tools is freely available along with tutorials for running and analyzing CG simulations at http://www.sirahff.com.

2 SIRAH tools features

2.1 Mapping AA to CG models

SIRAH Tools contains the Perl script cgconv.pl that takes as input AA files in PDB or PQR format and transforms residue coordinates to CG following mapping rules described in, so called, map files. Three mapping functions are currently available: CMASS, CGEOM and MAP. For CMASS and CGEOM, the atoms specified are mapped to their center of mass or geometry, respectively. If MAP is used, the positions of single atoms are given to CG beads. Though SIRAH mapping is used by default, user-created files can be read from the command line, the flag –h displays all available options. Mapping can also be restricted to sets of residues to build multiresolution systems (Machado and Pantano, 2015).

2.2 Converting GROMACS' topologies to PSF

Molecular visualization programs guess the connectivity of AA systems from standard interatomic distances and angles. However, this is difficult for arbitrary CG representations and topological information must be supplied in additional files, e.g. AMBER's topology (Salomon-Ferrer *et al.*, 2013) or X-PLOR PSF files (Brooks *et al.*, 2009). However, GROMACS' topologies (Pronk *et al.*, 2013) are not currently supported by visualization software. SIRAH's g_top2psf.pl converts GROMACS' topologies to PSF files interpreting the '#include' statements used to consider chains, solvent molecules, ligands, etc., being equally useful for AA, multiscale or arbitrary CG topologies.

2.3 Visualization and structural analysis of CG systems

The sirah_vmdtk.tcl plugin improves the visualization and analysis of CG trajectories with VMD (Humphrey *et al.*, 1996). This plugin sets correct van der Waals radii, coloring code by atom and residue types and macros for selecting molecular components on SIRAH trajectories (Fig. 1A). The utility sirah_ss assigns secondary structures to CG proteins in SIRAH classifying amino acids in α-helix (H), extended β-sheet (E) or, otherwise, coil (C) conformations, based on the instantaneous values of the backbone's torsional angles and Hydrogen bond-like (HB) interactions (Darré *et al.*, 2015). Comparing the assignments of sirah_ss with STRIDE (Frishman and Argos, 1995), we find ∼90% of agreement between both methods (Fig. 1B). Sirah_ss produces ASCII files of average and by-frame results, which can be visualized as a color matrix using the R script ssmtx2png.R (see, for instance, Fig. 7C in Darré *et al.*, 2015).

2.4 Backmapping CG coordinates to AA

The sirah_vmdtk.tcl plugin also includes the module sirah_backmap to retrieve pseudo-atomistic information from CG models. The atomistic positions are built on a by-residue basis following the geometrical reconstruction proposed by (Parsons *et al.*, 2005). Briefly, atomic positions are generated from the location of CG beads or previously reconstructed atoms using internal coordinates. Bond distances and angles are derived from rough organic chemistry considerations stored in backmapping libraries. Currently, backmapping libraries contain instructions for proteins and DNA of SIRAH, but user-defined procedures can be easily included for other models. Although plane AA reconstructions may result in poor stereochemistry conformations, this is readily fixed performing energy minimization using an AA force field, making the procedure robust and independent of the fine details of the backmapping library. The present implementation runs 100 steps of steepest-descent followed by 50 steps of Conjugated Gradient minimization in vacuum using the sander module of AmberTools (Salomon-Ferrer *et al.*, 2013). Clearly, the accuracy of the reconstruction depends on the CG model. Backmapping of four SIRAH CG trajectories from (Darré *et al.*, 2010, 2015) results in good structural descriptors (Fig. 1C). As a stringent test, we calculated the HB conservation (Hydrogens are often lost in CG models). The conservation of HB calculated with a Donor-Acceptor cutoff of 0.4 nm and 120° between Donor-H-Acceptor varied from 12% (s.d. 1) for 1QYO up to 24% (s.d. 3) for both 1RA4 and 1TQG. These numbers raised to 45% (s.d. 4), 59% (s.d. 5) and 64% (s.d. 4), for 1QYO, 1RA4 and 1TQG, respectively, if we only consider the backbone interaction associated to the secondary structure, while the DNA kept a 74% (s.d. 5) of HB in the Watson–Crick region.

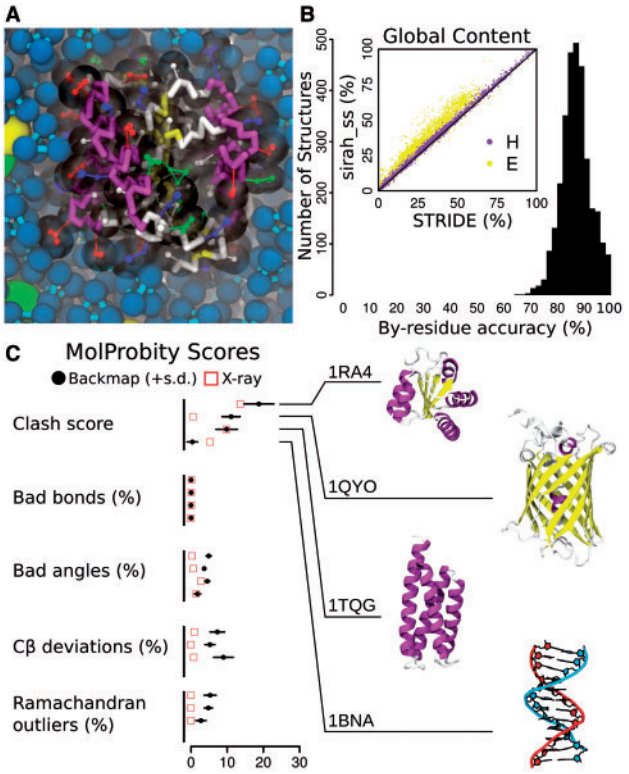

Fig. 1. SIRAH Tools features. (**A**) Molecular visualization of protein 1RA4 at CG level embedded in explicit solvent showing the connectivity of the system. The backbone is colored by secondary structure (H: purple, E: yellow, C: white) as calculated by sirah_ss. Amino acids are shown in CPK and depicted by physicochemical characteristics according to VMD standard coloring code. Excluded volumes are shown as semitransparent balls. (**B**) Percentage of residues with same secondary structure assignment by sirah_ss and STRIDE. The inset compares sirah_ss against STRIDE to estimate global content of H or E conformations. The analysis was performed on 3000 proteins from the PDB with resolution lower than 0.2 nm, identity less than 90% and no modified residues. (C) Structural quality of backmapped trajectories as evaluated by MolProbity (Chen *et al.*, 2010). Reported values correspond to the average calculated on 1 μs of MD simulation (Color version of this figure is available at *Bioinformatics* online.)

3 Conclusion

CG models are considerably expanding. Yet, their visualization and analysis still suffers from limitations that restrict their use to highly skilled users. SIRAH tools aims to bridge this gap by providing easy and flexible mapping utilities, processing of topologies and data analysis.

Funding

This work was partially funded by FOCEM (MERCOSUR Structural Convergence Fund), COF 03/11. M.R.M. and S.P. belong to the SNI program of ANII.

Conflict of Interest: none declared.

References

Brooks,B.R. *et al.* (2009) CHARMM: the biomolecular simulation program. *J. Comput. Chem.*, **30**, 1545–1614.

Chen,V.B. *et al.* (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 12–21.

Darré,L. *et al.* (2010) Another coarse grain model for aqueous solvation: WAT FOUR? *J. Chem. Theory Comput.*, **6**, 3793–3807.

Darré,L. *et al.* (2015) SIRAH: a structurally unbiased coarse-grained force field for proteins with aqueous solvation and long-range electrostatics. *J. Chem. Theory Comput.*, **11**, 723–739.

Frishman,D. and Argos,P. (1995) Knowledge-based protein secondary structure assignment. *Proteins*, **23**, 566–579.

Humphrey,W. *et al.* (1996) VMD: visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38.

Ingólfsson,H.I. *et al.* (2014) The power of coarse graining in biomolecular simulations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, **4**, 225–248.

Karplus,M. and Lavery,R. (2014) Significance of molecular dynamics simulations for life sciences. *Isr. J. Chem*, **54**, 1042–1051.

Machado,M.R. and Pantano,S. (2015) Exploring LacI–DNA dynamics by multiscale simulations using the SIRAH force field. *J. Chem. Theory Comput.*, **11**, 5012–5023.

Parsons,J. *et al.* (2005) Practical conversion from torsion space to Cartesian space for in silico protein synthesis. *J. Comput. Chem.*, **26**, 1063–1068.

Pronk,S. *et al.* (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, **29**, 845–854.

Salomon-Ferrer,R. *et al.* (2013) An overview of the Amber biomolecular simulation package. *Wiley Interdiscip. Rev. Comput. Mol. Sci*, **3**, 198–210.,