

## Computational Tools

# RedMDStream: Parameterization and Simulation Toolbox for Coarse-Grained Molecular Dynamics Models

Filip Leonarski<sup>1,2,\*</sup> and Joanna Trylska<sup>1,\*</sup>

<sup>1</sup>Centre of New Technologies and <sup>2</sup>Department of Chemistry, University of Warsaw, Warsaw, Poland

**ABSTRACT** Coarse-grained (CG) models in molecular dynamics (MD) are powerful tools to simulate the dynamics of large biomolecular systems on micro- to millisecond timescales. However, the CG model, potential energy terms, and parameters are typically not transferable between different molecules and problems. So parameterizing CG force fields, which is both tedious and time-consuming, is often necessary. We present RedMDStream, a software for developing, testing, and simulating biomolecules with CG MD models. Development includes an automatic procedure for the optimization of potential energy parameters based on metaheuristic methods. As an example we describe the parameterization of a simple CG MD model of an RNA hairpin.

Molecular dynamics (MD) (1) is a technique to study biomolecular motions with a molecular mechanics model and numerical solution of Newton's equations of motions. The first few picosecond-long full-atomic MD of a trypsin inhibitor in a vacuum was carried out in 1977 (2). Recently the dynamics of the same protein was simulated on a millisecond timescale (3). Despite huge progress in increasing MD efficiency and computer power, there is still a large gap between the MD capability and timescales of biological processes. For example, protein translation and folding require up to seconds and involve much larger systems than a trypsin inhibitor.

A possible methodological remedy is to develop reduced MD models, also termed coarse-grained (CG) (4–6) models. In CG models, a biomolecule is represented as a set of interaction centers (beads) that group together multiple atoms. This approach leads to the extension of spatial and temporal scales; the number of degrees of freedom is reduced; and because the simplified interactions usually correspond to the highest-frequency ones, the MD time step can be enlarged. With CG MD models, millisecond timescale simulations for systems much larger than a trypsin inhibitor can be performed on desktop computers. CG MD models have been successfully applied to processes such as nucleic acid thermal denaturation (7), protein folding (8–10), and conformational changes (11), as well as equilibrium dynamics of ribosome (12) or nucleosome (13). However, the common use of CG MD simulations is hindered by the huge effort associated with developing such models.

For full-atomistic MD, there are well-established protocols on how to derive a potential energy function and parameters from quantum chemistry (14). Although there exist

methods for systematic parameterization of CG MD models, such as iterative Boltzmann inversion (15) or renormalization group (16), they are not standardized for systems of biophysical interest. Moreover, the emphasis is on the results of the author-selected problem, and not on the thorough description of the model preparation process (reviewed in (6)). This process, typically only briefly described in publications, is hard to repeat. Although most CG MD models work well, they are limited to a particular problem or molecule class, even though in most cases they could be easily adapted to similar problems or related biomolecules. CG MD models' range of applications can be only extended if protocols for their development are improved and systematized. The obstacle has been the fragmentation of CG MD tools and benchmarks.

A CG MD model itself is not enough to run a simulation; tools to prepare starting files, a CG MD simulation engine, and tools to analyze results are necessary. For full-atomistic MD models, well-tested software packages, such as NAMD/VMD (17,18) or Amber/AmberTools (19), are available. However, in CG modeling, the user often has to combine tools from different sources. Some CG MD models are available only as in-house toolboxes; others are published only as a set of equations without the simulation tools. Another issue has been the benchmarking of these models. For example, comparison of radial distribution functions (or their subsets corresponding to a particular set of bead pairs) is widely used to assess models, although there is no well established method to describe this difference in numerical terms. So providing tools to make the CG MD models more universal has been essential.

We present RedMDStream—an application to develop CG MD force field parameters using a user-specified protocol and perform CG MD simulations. RedMDStream is suitable for the development and simulations of low-resolution (one to three beads per residue) and off-lattice CG MD models.

Submitted December 15, 2014, and accepted for publication March 13, 2015.

\*Correspondence: [f.leonarski@cent.uw.edu.pl](mailto:f.leonarski@cent.uw.edu.pl) or [joanna@cent.uw.edu.pl](mailto:joanna@cent.uw.edu.pl)

Editor: Nathan Baker.

© 2015 by the Biophysical Society  
0006-3495/15/04/1843/5 \$2.00



It is a standalone application written in C and C++ languages, working on Linux machines, parallelized with MPI and OpenMP (see the [Supporting Material](#) for the scalability benchmarks), and freely available under GNU public license from <http://bionano.cent.uw.edu.pl/Software/RedMD>.

### RedMDStream architecture

To develop a CG MD model with RedMDStream, one has to perform the following steps (see [Fig. 1](#)).

- Design CG topology. State the rules to transform an atomistic structure into a CG representation. These rules describe the bead type, placement, mass, and charge, as well as the connectivity network and potential energy terms.
  - Beads can be placed in the positions of particular atoms or residues' centers of masses. See [Fig. 2 b](#) for an example.
  - The connectivity network, including pseudo-bonds, angles, and dihedrals, is set according to user-defined criteria: the sequence (neighbor interactions), distance (distance cut-off in a reference structure), or secondary/tertiary structure (complementary pairing in nucleic acids, salt, or disulphide bridges in proteins). See [Fig. 2, c–d](#).
  - Interactions are assigned for the connectivity network by providing potential energy terms. Users choose between the predefined terms, such as harmonic and Morse potential, or provide their own equation. This equation along with its derivative is tabularized by RedMDStream for better performance. Parameters of the predefined and custom potential terms depend on the properties of a particular interacting pair (triple, quadruple), e.g., beads closer in the reference structure will have stronger bonding, as in elastic network models (12,13). See [Table S1](#) for an example.

All these rules are written in an XML (20) file allowing us to share the developed model with others (see [Fig. S2](#)).
- Provide a simulation protocol. The user describes the implementation of the force field: the files to load, MD simulation method and length, and steps to analyze the trajectory.
  - Molecules can be provided in PDB (21), PDBML (22), and PQR (23) formats. In addition to a coordinate file, one may include structural information from RNAView (24), in a dot-bracket notation for nucleic acids or plain text format. Protein secondary structure annotations are read from PDB and PDBML files. PDBML also includes base pairing annotations (25) for nucleic acids.
  - For MD simulation NVE and NVT (with the Berendsen thermostat), ensembles are available, as well as Langevin and Brownian MD regimes (1). MD is performed using the fully integrated RedMD engine (26).
  - MD trajectory is saved in the DCD binary format and analyzed for pseudo-atoms' root mean square deviation (RMSD), root mean square fluctuations (RMSF), and distance distributions. To efficiently test many CG model modifications, RedMDStream can refrain from saving the trajectory on the disk and use the analyzed measures as a sole outcome of the MD simulation.
- Perform parameter optimization. The optimal topology parameters of the CG model can be determined using a built-in optimization routine. The criteria of the quality of the CG model are selected by the user and the score comes from CG MD simulation(s) outcome. The criteria can be based on comparing the RMSD, RMSF, and distance distributions with a user-selected reference. RedMDStream incorporates metaheuristic optimization methods: evolutionary algorithm, particle swarm optimization, and simplex algorithm. To efficiently find the optimum of the function, the methods follow strategies inspired by nature instead of strict mathematical rules (27). The parameters selected for the optimization can be the same as defined in the potential energy function or some parameters can be either excluded or fixed. Also, one can provide a mathematical expression between a

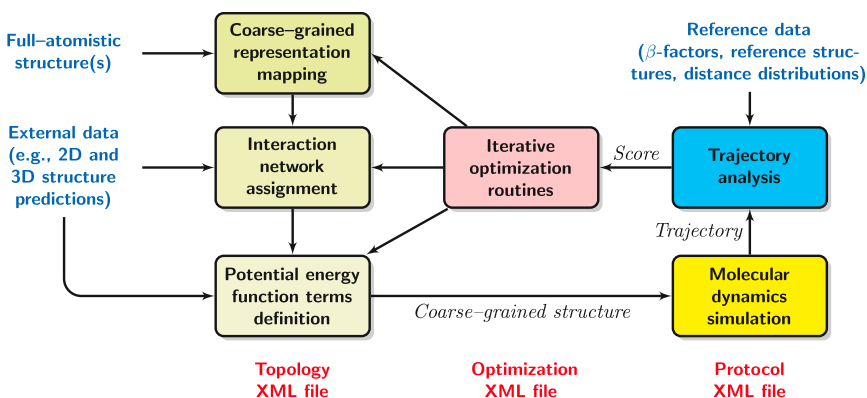


FIGURE 1 RedMDStream workflow. Frames list the tasks performed by RedMDStream. Information read from the input files is in blue font. The control files are in red font. *Italic font* refers to data produced by RedMDStream that can be stored in the output file or memory. To see this figure in color, go online.

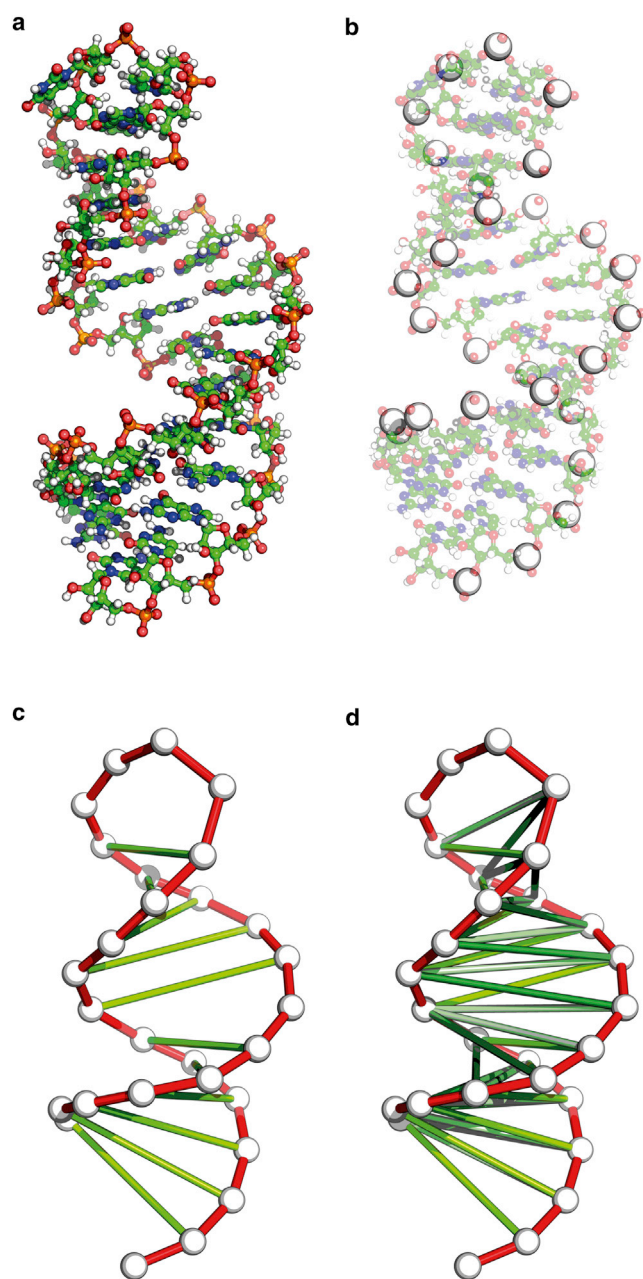


FIGURE 2 (a) A full-atomistic structure of the microROSE element (21) (PDB ID: 2GIO (21), NMR resolved structure). (b) Bead placement (large white spheres) on the phosphorus atom in the CG MD model. (c) Pseudo-bonds in variant 1 of the force field model. (d) Pseudo-bonds in variants 2 and 3. To see this figure in color, go online.

parameter used in the optimization and one or many parameters defining the energy function of the CG MD model.

### RNA one-bead force field

As an example, we optimized a one-bead-per-nucleotide CG MD model (28) for an RNA hairpin, a fragment of the

repression of heat shock gene expression element, called microROSE (29) (Fig. 2 a). As a reference, a 100 ns MD simulation with a full-atomistic force field was performed, see Supporting Material for details.

Three variants of the CG model were tested (Fig. 2, c–d, Table S1, and Fig. S3), differing in the interactions of complementary pairs: variant 1 had a single pseudo-bond per one complementary pair; variant 2 had three bonds per one complementary pair; and variant 3 had the same as in variant 2 but one of the bonds was sequence specific. We wanted to verify if one bond per complementary pair is sufficient, and if adding the base pair type specificity improves the model. To find the best parameters for each variant, we optimized the model (with particle swarm optimization and evolutionary algorithm) based on 5 ns CG MD simulations of microROSE (for the protocol, see Supporting Material). Optimization was scored by comparing RMSD, RMSF, and distance distributions with the full-atomic trajectory. For the scoring criteria, see Table S2. Input files necessary to repeat this optimization are in the file *microrose.RedMDStream.zip* in the Supporting Material.

The best optimization scores are shown in Table 1. The results depend on the Langevin dumping constant used in CG MD simulations whose increase drops the best score by 10% to 20%. This constant affects mostly RMSF (Fig. S4), but the low dumping constant gives high unstable RMSD in the simulation of variant 1 (Fig. S5).

Further, we verified the performance of the CG models optimized based on 5 ns CG MDs in longer 100 ns simulations. The simple model (variant 1) became unstable in the first few ns with RMSD up to 15 Å (see Fig. S6). However, the additional bonds per complementary pair in variants 2 and 3 provided stable 100 ns simulations. The distance distributions (Figs. S7–S10) suggest that instability and low score on short timescales are connected. Additional complementary bonds provide extra-stabilization, but for the cost of too narrow distributions (see Figs. S9 and S10). No extra bonds ensure better flexibility, which better resembles the distributions from an all-atom MD but leads to the unnatural unfolding of the structure. On the other hand, the effect of adding the base pair specificity to the potential is minor; it neither improves the score nor distributions. As shown in Fig. S8 (variant 3,  $\gamma = 5.0 \text{ ps}^{-1}$ ) this addition has even an adverse effect on the distributions, resulting in a split of the single P–P  $i:j$  peak for different base pairs (A–U, C–G, and wobble). This effect does not appear

TABLE 1 Best score for each of the force field variants, depending on the Langevin dumping constant

	$\gamma = 5 \text{ ps}^{-1}$	$\gamma = 20 \text{ ps}^{-1}$
Variant 1	0.27	0.21
Variant 2	0.27	0.23
Variant 3	0.24	0.22

Scores are normalized to 0.0 to 1.0 scale. Lower values correspond to better models. Scoring criteria are described in Table S2.

in a model optimized with a higher Langevin dumping constant. This is consistent with the distributions from the full-atomistic model that share the same maximum for all the base types. Overall, variant 2 seems optimal, as it provides the best score without introducing too many parameters.

Besides analyzing the best model, we looked for a general relationship between the parameter value and score (Fig. S11). Overall, the three CG model variants give similar preferable bonding parameters for neighboring beads (both the force constant and equilibrium distance). However, variant 1 requires stronger angle and dihedral force constants to compensate for the lack of extra pseudo-bonds. In variants 2 and 3, the loop region could be even unrestricted in terms of angle and dihedral (force constant approaching to zero). Finally, variant 3 correctly predicted that CG pairs are stronger (higher force constant) and closer than AU pairs. For details of the methodology, such as the choice of evolutionary algorithm parameters and examples of CG models, please refer to (28).

## CONCLUSIONS

In a single software package RedMDStream allows defining a CG model (mapping and force field), its further optimization and simulation, as well as its analysis. Combining all the tools provides a consistent environment for a CG MD model developer. Using the software may replace hours of manual trial-and-error tests, as well as provide an overview of CG MD model properties.

Although RedMDStream aids in the CG force field design and optimization process, it does not operate as a blackbox. Especially, the optimization results should be visually inspected to verify if the scoring function was well defined and does not accept unphysical models. This effect can be minimized by using physically reasonable parameter ranges, e.g., based on the Boltzmann inversion. Tests on longer timescales should be also performed. Note, that the outcome depends on the used structure set. If a CG force field aims to describe a whole class of molecules, it would be better to include many representative molecules in the optimization protocol (28), see [Supporting Material](#) for the discussion on the optimization procedure.

RedMDStream uses the XML format to define the CG topology. This topology definition could work as interoperable format to store CG MD models, so authors could share their XML files with others. Future releases will extend the trajectory quality measures with force matching (30) and include LAMMPS (31) support.

## SUPPORTING MATERIAL

Supporting Materials and Methods, 12 figures, two tables, and RedMDStream files are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(15\)00284-2](http://www.biophysj.org/biophysj/supplemental/S0006-3495(15)00284-2).

## AUTHOR CONTRIBUTIONS

F.L. and J.T. designed the research; F.L. wrote the software, performed the simulations, and analyzed the data; J.T. supervised the research; F.L. and J.T. wrote the article.

## ACKNOWLEDGMENTS

The authors thank Valentina Tozzini, Giuseppe Carducci, and Paolo Mereghetti for useful discussions. F.L. and J.T. were supported by the University of Warsaw (ICM/KDM/G31-4 and CeNT/BST), National Science Centre (DEC-2011/03/N/NZ2/02482 and DEC-2012/05/B/NZ1/00035), Ministry of Science and Higher Education (Canaletto Programme), and HPC Infrastructure for Grand Challenges of Science and Engineering POWIEW project, co-financed by European Regional Development Fund operated within the Innovative Economy Operational Programme.

## SUPPORTING CITATIONS

References (32–34) appear in the [Supporting Material](#).

## REFERENCES

- Schlick, T. 2010. *Molecular Modeling and Simulation: An Interdisciplinary Guide*, 2nd ed. Springer-Verlag, New York.
- McCammon, J. A., B. R. Gelin, and M. Karplus. 1977. Dynamics of folded proteins. *Nature*. 267:585–590.
- Shaw, D. E., P. Maragakis, ..., W. Wriggers. 2010. Atomic-level characterization of the structural dynamics of proteins. *Science*. 330: 341–346.
- Levitt, M. 1976. A simplified representation of protein conformations for rapid simulation of protein folding. *J. Mol. Biol.* 104:59–107.
- Tozzini, V. 2005. Coarse-grained models for proteins. *Curr. Opin. Struct. Biol.* 15:144–150.
- Leonarski, F., and J. Trylska. 2014. Modeling nucleic acids at the residue-level resolution. *In* *Computational Methods to Study the Structure and Dynamics of Biomolecules and Biomolecular Processes, Vol. 1* A. Liwo, editor. Springer, Berlin, pp. 109–149.
- Sambriski, E. J., D. C. Schwartz, and J. J. de Pablo. 2009. A mesoscale model of DNA and its renaturation. *Biophys. J.* 96:1675–1690.
- Erman, B. 2001. Analysis of multiple folding routes of proteins by a coarse-grained dynamics model. *Biophys. J.* 81:3534–3544.
- Pokarowski, P., A. Kolinski, and J. Skolnick. 2003. A minimal physically realistic protein-like lattice model: designing an energy landscape that ensures all-or-none folding to a unique native state. *Biophys. J.* 84:1518–1526.
- Ding, F., S. V. Buldyrev, and N. V. Dokholyan. 2005. Folding Trp-cage to NMR resolution native structure using a coarse-grained protein model. *Biophys. J.* 88:147–155.
- Chu, J.-W., and G. A. Voth. 2007. Coarse-grained free energy functions for studying protein conformational changes: a double-well network model. *Biophys. J.* 93:3860–3871.
- Trylska, J., V. Tozzini, and J. A. McCammon. 2005. Exploring global motions and correlations in the ribosome. *Biophys. J.* 89:1455–1463.
- Voltz, K., J. Trylska, ..., J. Langowski. 2012. Unwrapping of nucleosomal DNA ends: a multiscale molecular dynamics study. *Biophys. J.* 102:849–858.
- Wang, J., R. M. Wolf, ..., D. A. Case. 2004. Development and testing of a general amber force field. *J. Comput. Chem.* 25:1157–1174.
- Reith, D., M. Pütz, and F. Müller-Plathe. 2003. Deriving effective mesoscale potentials from atomistic simulations. *J. Comput. Chem.* 24:1624–1636.

16. Savelyev, A., and G. A. Papoian. 2009. Molecular renormalization group coarse-graining of polymer chains: application to double-stranded DNA. *Biophys. J.* 96:4044–4052.
17. Phillips, J. C., R. Braun, ..., K. Schulten. 2005. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26:1781–1802.
18. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33–38, 27–28.
19. Case, D. A., T. E. Cheatham, 3rd, ..., R. J. Woods. 2005. The Amber biomolecular simulation programs. *J. Comput. Chem.* 26:1668–1688.
20. Bray, T., J. Paoli, ..., F. Yergeau, editors. 2008. Extensible Markup Language (XML) 1.0. W3C, 5th ed. <http://www.w3.org/TR/2008/REC-xml-20081126/>.
21. Berman, H. M., T. Battistuz, ..., C. Zardecki. 2002. The Protein Data Bank. *Acta Crystallogr. D Biol. Crystallogr.* 58:899–907.
22. Westbrook, J., N. Ito, ..., H. M. Berman. 2005. PDBML: the representation of archival macromolecular structure data in XML. *Bioinformatics.* 21:988–992.
23. Baker, N. A., D. Sept, ..., J. A. McCammon. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA.* 98:10037–10041.
24. Yang, H., F. Jossinet, ..., E. Westhof. 2003. Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res.* 31:3450–3460.
25. Leontis, N. B., and E. Westhof. 2001. Geometric nomenclature and classification of RNA base pairs. *RNA.* 7:499–512.
26. Górecki, A., M. Szypowski, ..., J. Trylska. 2009. RedMD—reduced molecular dynamics package. *J. Comput. Chem.* 30:2364–2373.
27. Haupt, R. L., and S. E. Haupt. 2004. Practical Genetic Algorithms. 2nd ed. John Wiley & Sons, Hoboken, New Jersey.
28. Leonarski, F., F. Trovato, ..., J. Trylska. 2013. Evolutionary algorithm in the optimization of a coarse-grained force field. *J. Chem. Theory Comput.* 9:4874–4889.
29. Chowdhury, S., C. Maris, ..., F. Narberhaus. 2006. Molecular basis for temperature sensing by an RNA thermometer. *EMBO J.* 25:2487–2497.
30. Ercolessi, F., and J. B. Adams. 1994. Interatomic potentials from first-principles calculations: the force-matching method. *Europhys. Lett.* 26:583.
31. Plimpton, S. 1995. Fast parallel algorithms for short-range molecular dynamics. *J. Comp. Phys.* 117:1–19.
32. Cornell, W. D., P. Cieplak, ..., P. A. Kollman. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117:5179–5197.
33. Pérez, A., I. Marchán, ..., M. Orozco. 2007. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.* 92:3817–3829.
34. Banáš, P., D. Hollas, ..., M. Otyepka. 2010. Performance of molecular mechanics force fields for RNA simulations: stability of UUCG and GNRA hairpins. *J. Chem. Theory Comput.* 6:3836–3849.