

Parametrization of the Elastic Network Model Using High-Throughput Parallel Molecular Dynamics Simulations

Philipp S. Orekhov^{1,2,3}, *Ilya V. Kirillov*¹, *Vladimir A. Fedorov*³,
Ilya B. Kovalenko^{3,4,5,6}, *Nikita B. Gudimchuk*^{3,7}, *Artem A. Zhmurov*^{1,2}

© The Authors 2019. This paper is published with open access at SuperFri.org

Even when modern computational platforms and parallel techniques are used, conventional all-atom simulations are limited both in terms of reachable timescale and number of atoms in the biomolecular system of interest. On the other hand, coarse-grained models, which allow to overcome this limitation, rely on proper and rigorous parametrization of the underlying force field. Here, we present a novel iterative approach for parametrization of coarse-grained models based on direct comparison of equilibrium simulations at all-atom and coarse-grained resolutions. In order to assess the accuracy of our method, we have built and parametrized an elastic network model (ENM) of the tubulin protofilament consisting of four monomers. For this system, our method shows good convergence and the parametrized ENM reproduces protein dynamics in a finer way when compared to ENMs parametrized using the conventional approach. The presented method can be extended to other coarse-grained models with a slight adjustment of the equations describing the iterative scheme.

Keywords: molecular dynamics, coarse-grained models, tubulin, elastic network model, high-throughput simulations, parallel simulations.

Introduction

Despite huge advances in the computational techniques the atomistic simulations of molecular dynamics are still limited in the time and space domains to μs and millions of atoms, respectively, what makes various coarse-grained approaches a valuable alternative. Among the latter, the elastic network models have gained a considerable success: though relatively simple they can yield decent insights into molecular mechanisms of protein function. Particularly, the simple Gō-models have proved themselves advantageous in the studies of protein folding, allostery, etc. [2]. They consider protein as a network of C_{α} -atoms connected with uniform springs when closer than a given cutoff distance. However, homogeneity of such networks, which pursues models with very few fitted parameters, imposes certain limitations on the accuracy of the obtained ENMs [4]. The latter can be overcome using a series of all-atom (AA) simulations produced by modern high-throughput MD approaches in order to bring the CG models to the highest accordance with their AA counterpart.

We suggest a methodology for accurate parametrization of ENMs based on the information derived from a batch of all-atom simulations, which accounts for inconsistency of inter-residue pair contacts, as well as the heterogeneity of their strength. We utilize the developed approach in order to build a CG model of the tubulin protofilament — a structural element of microtubule, which is a cytoskeletal structure essential for intracellular transport, cell division, cell motility,

¹Moscow Institute of Physics and Technology, Dolgoprudny, Russia

²Sechenov University, Moscow, Russia

³M.V. Lomonosov Moscow State University, Moscow, Russia

⁴Federal Research and Clinical Center of Specialized Medical Care and Medical Technologies, Federal Medical and Biological Agency of Russia, Moscow, Russia

⁵Astrakhan State University, Astrakhan, Russia

⁶Scientific and Technological Center of Unique Instrumentation, RAS, Moscow, Russia

⁷Center for Theoretical Problems of Physicochemical Pharmacology, RAS, Moscow, Russia

etc. [5]. We demonstrate that our method leads to a more reliable description of protein dynamics comparing to the conventional ENM approaches. Since our approach is based on the direct comparison of the coarse-grained and all-atom simulations, it can be applied to more complex coarse-grained models. For instance, if the electrostatic energy is explicitly included to the coarse-grained potential energy function, its contribution has to be excluded from potential that is parameterized [3]. In our approach, this will be done automatically throughout iterative process.

1. Methods

All-atom MD simulations of tubulin tetramers were carried out in the GROMACS 5 suite [1] using the CHARMM27 force field. The PDB entry 3J6F, corresponding to a fragment of microtubule bound to GDP, was used in order to build an atomic model for simulations. The following standard protocol was used for simulations: the NPT ensemble controlled by means of Nose–Hoover thermostat ($T_{ref} = 303$ K, $\tau_T = 2$ ps) and Parrinello–Rahman barostat (isotropic pressure, $p_{ref} = 1$ bar, $\tau_p = 2$ ps, compressibility = $4.5 \cdot 10^5$ bar⁻¹); time step — 4 fs, allowed by the mass rescaling; the Verlet cut-off scheme and PME for the long-range electrostatics. In total, two 1 μ s-long trajectories were obtained, last 300 ns of each run was used for further analysis.

To parametrize the ENM, we first extracted mean distances and standard deviation for each interparticle distance from the AA simulations to use as reference values. At each iteration, 4×10^6 steps of CG dynamics is performed, and the same values are extracted from the last 3×10^6 steps. The parameters are then adjusted and procedure is repeated until convergence is reached. The update rules for the parameters are:

$$\frac{1}{K_{i,j}^{n+1}} = \frac{1}{K_{i,j}^n} - \alpha \frac{(\sigma_{i,j}^0)^2 - (\sigma_{i,j}^n)^2}{k_B T}, \quad b_{i,j}^{n+1} = b_{i,j}^n + \beta \frac{b_{i,j}^0 - b_{i,j}^n}{\sigma_{i,j}^0}. \quad (1)$$

Here, $K_{i,j}^n$ is a spring constant for the contact between i -th and j -th particles at n -th iteration, $\sigma_{i,j}^n$ is a dispersion, computed from simulations at n -th iteration, k_B is Boltzmann constant, and T is the temperature; $b_{i,j}^n$ is an average distance between particles i and j throughout the simulation run at iteration n ; zero value for the iteration index indicates that the respective value is computed from the reference AA simulation run; $\alpha = 0.1$ and $\beta = 0.1$ are dimensionless iteration parameters.

Results and Discussion

Defining contacts in CG models solely on the distance cut-off can lead to the certain amount of false-contacts. These not only include the pairs of particles that are not interacting directly, but also those that cannot be properly described by the chosen potential. For instance, double-well potentials cannot be described by harmonic ENMs and thus have to be excluded. To assess the quality of contacts, we compared two independent AA simulation runs. We derived the distributions of interparticle (contact) distances from two runs and computed an overlap between them (Fig. 1a, b). We then excluded the contacts, for which the overlap was less then 50 % (Fig. 1b, c). The distribution of mean distances of such “good” contacts (Fig. 1c, top) allowed us to choose the optimal value for the distance cut-off (8 Å).

The iterative process for optimization of the model parameters converged in ~ 50 iterations, at which point the average distribution overlap reaches the plateau at value of 0.954 ± 0.002

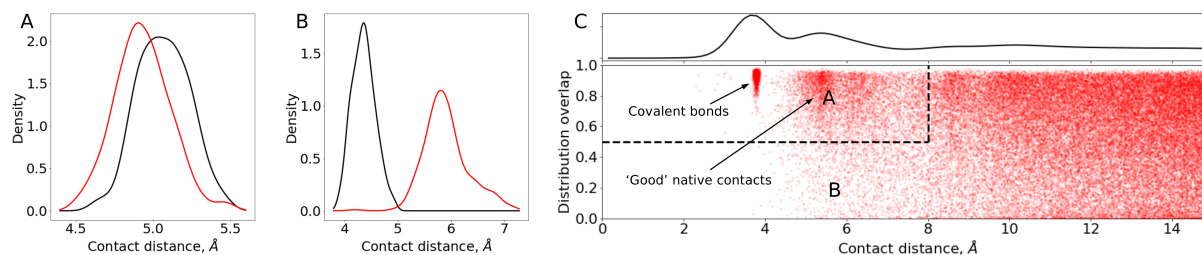


Figure 1. Distributions of interparticle distances for “good” (A) and “bad” (B) contacts obtained from two independent AA simulation runs (red and black curves); normalized distribution of interparticle distances over all contacts in the system with sufficient distribution overlap (> 0.5) (C, top) and a scatter plot of the distances and distribution overlaps for all the contacts (C, bottom) with the “good” contacts embraced in the dashed rectangle

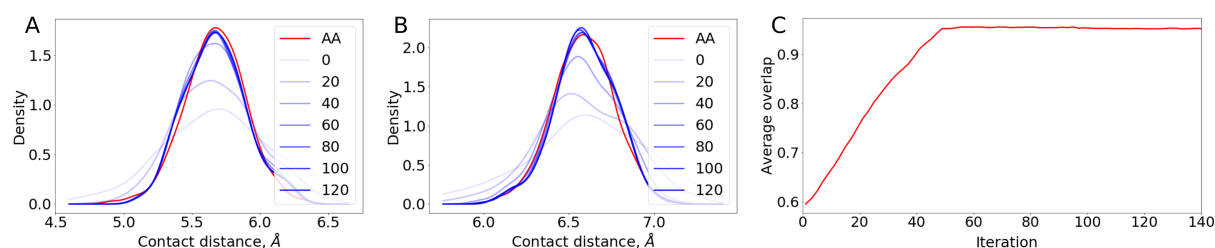


Figure 2. Distributions of the interparticle distances for two arbitrary chosen contacts (A, B): blue curves represent the distributions from CG MD runs at different iterations of the parametrization process, red curve shows the correspondent distribution from the AA simulation; the average overlap of the distributions from the CG simulations with the AA distributions as a function of the iteration number (C)

(Fig. 2c). Visual analysis of interparticle distributions showed a good correspondence between the CG and AA simulations both in distributions of height and dispersion. As expected, the

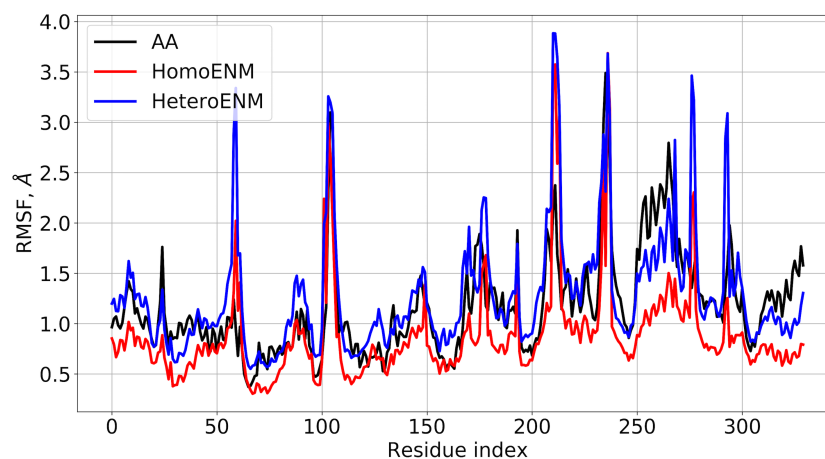


Figure 3. Root mean-square fluctuations computed per residue for one of α -tubulin monomers from AA MD simulations and CG MD simulations using the homogeneous model with the uniform spring constant (homoENM) and the heterogeneous model based on AA simulations (heteroENM)

distributions obtained from CG simulations lacked certain features, observed in AA simulations, e.g. small shoulders in some cases (Fig. 2a, b).

The overall quality of the parametrized model was assessed by comparing the RMSF curves obtained from the AA MD simulations and CG simulations using either the developed ENM or the homogeneous ENM with the uniform spring constant of 10000 kJ/mol·nm². As one can see in Fig. 3, the per-residue fluctuations calculated based on our model match those computed from the AA simulations better comparing to the conventional homogeneous ENM model. The mean-square error equals 0.24 and 0.19 Å² for the homogeneous ENM and our heterogeneous ENM, respectively.

To conclude, we developed a novel approach to parametrize the CG models of biomolecular systems. The method outperforms the conventional approach for parametrization of ENM both in terms of the convergence and the model quality while remaining robust and simple. The proposed procedure can be extended to other CG models and the resulting models can be easily incorporated into the existing MD software, which makes it interesting to anyone using this tool in their research.

Acknowledgments

This study was designed and conducted with support from the Russian Foundation for Basic Research, projects 17-00-00479 and 16-34-60113. Assessment of the quality of interparticle contacts is a part of Russian Science Foundation research project 17-71-10202. The research was carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University supported by the project RFMEFI62117X0011.

This paper is distributed under the terms of the Creative Commons Attribution-Non Commercial 3.0 License which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is properly cited.

References

1. Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., Lindahl, E.: GRO-MACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1, 19–25 (2015), DOI: 10.1016/j.softx.2015.06.001
2. Hills, R.D., Brooks, C.L.: Insights from coarse-grained Gō models for protein folding and dynamics. *International journal of molecular sciences* 10(3), 889–905 (2009), DOI: 10.3390/ijms10030889
3. Hyeon, C., Onuchic, J.N.: Mechanical control of the directional stepping dynamics of the kinesin motor. *Proc. Natl. Acad. Sci. USA* 104(44), 17382–17387 (2007), DOI: 10.1073/pnas.0708828104
4. Lyman, E., Pfaendtner, J., Voth, G.A.: Systematic multiscale parameterization of heterogeneous elastic network models of proteins. *Biophysical journal* 95(9), 4183–4192 (2008), DOI: 10.1529/biophysj.108.139733
5. Nogales, E.: Structural insights into microtubule function. *Annual review of biochemistry* 69(1), 277–302 (2000), DOI: 10.1146/annurev.biochem.69.1.277