Implementation and Testing of Multiple Walkers Approach Based Free Energy Calculations in the Grid Environment

Zora Střelcová[1], Petr Kulhánek[1], Jan Kmuníček[2], Jaroslav Koča[1] and Luděk Matyska[2]

[1] National Centre for Biomolecular Research, Masaryk University, Kotlářská 2, CZ-61137 Brno, Czech Republic
stre@chemi.muni.cz
[2] CESNET, z. s. p. o., Zikova 4, CZ-16000 Praha 6, Czech Republic
kmunicek@ics.muni.cz

Abstract. The calculation of free energy is a challenging task because it requires very long molecular simulations to obtain converged and reliable results. To overcome this so-called sampling problem, several methods have been developed. In this work, Adaptive Biasing Force (ABF) method and Multiple Walkers Approach (MWA) are used in the study of molecular shuttle in pseudorotaxane complexes and water molecule transit through cucurbit[n]uril cavity. ABF method improves sampling by biasing system along a predefined reaction coordinate, which describes studied process. The sampling is further improved by MWA, which exchanges reconstructed free energy potential among several independent simulations (walkers) biased by ABF. Exchanges with all walkers need not be necessarily synchronized therefore it is possible to speed up the whole calculation in almost linear fashion. Due to this feature the combination of both methods is suitable for utilization within the production grid environment. Our tests were performed in the two distinct grid environments - VOCE and MetaCentrum – and subsequently validated that MWA/ABF approach is promising tool benefiting from run in any Grid.

Keywords: free energy, adaptive biasing force method, multiple walkers approach, pseudorotaxanes, molecular shuttle, cucurbit[n]urils, VOCE, MetaCentrum

1 Introduction

The free energy is an important physico-chemical quantity related to chemical equilibrium and kinetics of (bio)chemical processes. Its theoretical prediction based on computer simulations can be used for deeper understanding of aforementioned processes, especially, if they cannot be examined by experimental means. Unfortunately, direct calculations of free energy from molecular simulations are not feasible because of very huge disproportion between real and simulation times. Due to this fact, most of studied processes are very “rare” in computer simulations even if

they occur very often in reality. To overcome this so called sampling problem, several methods were introduced in past. All of them bias a system in such a way that studied processes occur more frequently during molecular simulations. Several methods varying in the biasing strategy are available: constrained dynamics [1], umbrella sampling [2], metadynamics [3], and Adaptive Biasing Force (ABF) method [4]. These methods belong into the group of Potential of Mean Force (PMF) methods [5] where the system is biased along a prescribed reaction coordinate. The reaction coordinate is usually some geometrical parameter, which defines a path among states of interest, for example between reactants and products of studied chemical reaction or process.

In some cases, these methods might still provide unsatisfactory results. This usually happens if a motion along the reaction coordinate is coupled with some low motion mode, which is typical behavior of biochemical or very flexible systems. The solution of this problem is either to run very long molecular simulations or to use more advanced methods improving the sampling. Significant improvement has been shown using replica exchange molecular dynamics [6], string method [7], and Multiple Walkers Approach (MWA) [8]. All these approches simulate several replicas of the same system either under different conditions (usually temperature) or in the different parts of configuration space. Since communication overhead among replicas is usually very small, all of these methods are suitable for large scale calculations.

In this work, the applicability of ABF method together with MWA will be demonstrated in the study of molecular shuttle occurring in pseudorotaxane complexes (Fig. 1) and water molecule transit through cucurbit[n]uril cavity utilizing the grid computational resources. The ABF method was chosen due to its simplicity and the possibility to handle complicated reaction coordinates. Moreover, its coupling with MWA was not too complicated.

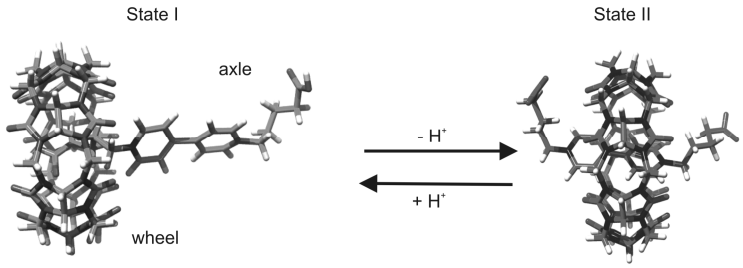

Fig. 1. Schematic representation of studied molecular shuttle

Pseudorotaxanes [9] are interlocked molecules composed of the linear molecule (so called axle) threaded into the macrocyclic molecule (so-called wheel). The axle is hold in the wheel cavity by non-covalent interactions, which prevents the complex from dissociation but, on the other had, allows the motion of both components to each other. In this work, the molecular shuttle (Fig. 1) was studied in the pseudorotaxane complex formed by cucurbit[7]uril (wheel) and 1,1'-bis(4-carboxybutyl)-4,4'-bipyridinium (axle). This complex has been selected as a test case because the molecular shuttle in the complex was thoroughly studied by experimental methods [10, 11]. After methodology validation, the study was extended to an analysis of

water molecule transit through cucurbit[n]uril (CB[n]) cavity (n=5–8), which cannot be studied by any currently available experimental methods.

2 Computational Details

2.1 System Preparation

All CB[n] and the axles were artificially built and further optimized by Gaussian 03 [12] on Hartree-Fock level using 6-31G* basis set. At the same level of theory, electrostatic potential was calculated and used in derivation of atomic charges using R.E.D. II program [13]. Input topologies and coordinates for molecular dynamics simulations were prepared by tLeap module from AMBER 9 package [14]. Eight systems were prepared: four pseudorotaxane complexes differing in the axle position towards the wheel and in different protonation states and four CB[n] with n from 5 to 8. The systems were then immersed into water. Periodic boundary conditions were applied using truncated octahedral box with inscribed largest sphere radius about 25 Å. The system electro-neutrality was achieved by adding chloride ions whenever it was necessary. Pseudorotaxane complexes were also prepared in NaCl solution with concentration about 0.12 M to be consistent with experimental condition [10]. Both wheels and axles were described by GAFF force field [15], sodium and chlorine ions by parm99SB force field [16] and water molecule by TIP3P model [17].

Built systems were slowly heated to 300 K and further equilibrated at this temperature and at constant pressure 1 bar. The temperature was maintained by Langevin stochastic thermostat and the pressure was maintained by weakly coupled barostat. Long-range interactions were treated by particle-mesh Ewald method [18] with direct summation cut-off set to 9.0 Å. The length of all bonds containing hydrogen atoms was constrained by SHAKE algorithm [19], which allows using integration time step of 2 fs.

Equilibrated systems were simulated for additional 10 ns (pseudorotaxanes) and 16 ns (CB[n]). During this period, intermediate atom coordinates and velocities (restart files) were collected every 500 ps. They served as starting points for individual MWA walkers. All simulations were performed using pmemd program from AMBER package.

2.2 Free Energy Calculations

Free energy calculations were done using modified pmemd program with implemented ABF method. Accumulated PMF was applied when 200 samples were collected in a particular bin. About 40 individual walkers were run with starting configurations taken from intermediate restart files of production runs. Every walker exchanges accumulated data with MWA server every 250 time steps. Proper selection of reaction coordinate representation and implementation details of MWA are discussed in depth in the section 3.

2.3 Tested Grid Environments

VOCE (Virtual Organization for Central Europe) provides a complete grid infrastructure under the EGEE project umbrella. It was established to serve the requirements from all researchers belonging to the Central Europe region as defined by the EGEE project. VOCE was set up as the very first catch-all VO with regional scope spanning the whole Central Europe federation. Currently, the portfolio of participating countries involves Austria, Belarus, Croatia, Czech Republic, Hungary, Poland, Slovakia, and Slovenia that indicates well the heterogeneity in partners as the main common regional characteristic. Nowadays, it is composed of more than 8 000 CPU cores available, 100 TBs of disc space, and it serves around 150 end users. According to the activity monitoring the main areas of interests for the researchers within VOCE environment are computational chemistry studies, numerical astrophysics simulations, phylogeny studies and plasma analysis. Moreover, it is also heavily used for running sets of testing jobs (collections, parametric or DAG jobs), development of various portlets, and generic activities related to porting any applications to the EGEE infrastructure.

MetaCentrum is the Czech education and scientific e-infrastructure which covers majority of activities concerning super-, cluster- and grid computing and/or high performance computing in general in the Czech Republic. Currently, MetaCentrum operates and manages distributed computing infrastructure consisting of computing and storage resources owned by Czech Education and Scientific Network operator (CESNET) as well as those of co-operative academic centers within the Czech Republic. MetaCentrum is responsible for building the National Grid and its integration to related international activities, especially in the European Union. In addition, it is actively involved in many international Grid projects such as EGEE, EGI DS, EUAsiaGRID and others. Project MetaCentrum is the key activity of CESNET and officially represents the interests of the national Grid community towards other national and international bodies. Simultaneously MetaCentrum activity deals with necessary research and development to ensure optimal functionality, security and performance of the infrastructure. The activities include virtualization of physical infrastructure - computing resources, storage capacity and computer networks; broadening of computational and disc capacity of distributed PC clusters; development of MetaCentrum infrastructure and provided services; continued integration into international Grid-related projects; providing of large informational services including dynamical ones through MetaCentrum portal; other development of authorization infrastructure, middleware, etc.

3 Results and Discussion

3.1 Implementation of Multiple Walkers Approach

ABF method provides estimate of free energy in the form of PMF and this estimate is used to bias the system [4, 20]. In MWA, N independent simulations (walkers) are run

simultaneously. All of them provides PMF estimate, which is combined with PMF estimates of other walkers and this total PMF estimate is used to bias individual walkers. Because all walkers are independent, the total PMF estimate converged almost *N* times faster than it would be in the case of simple ABF simulation (or single walker). The scheme of our implementation using this strategy is shown in Fig. 2.

Each walker has two accumulators. The first one (I) contains total PMF estimate used to bias molecular dynamics simulation, whereas the second one (II) holds PMF estimate acquired only by particular walker in some defined period. After this period, the walker sends intermediate data to MWA server. MWA server adds them into total accumulator (III). Updated total accumulator is then sent back to the walker. The walker substitute its total accumulator (I) with data received from MWA server and accumulator II is reset. This is repeated until adequate sampling is achieved in the entire space of reaction coordinate.

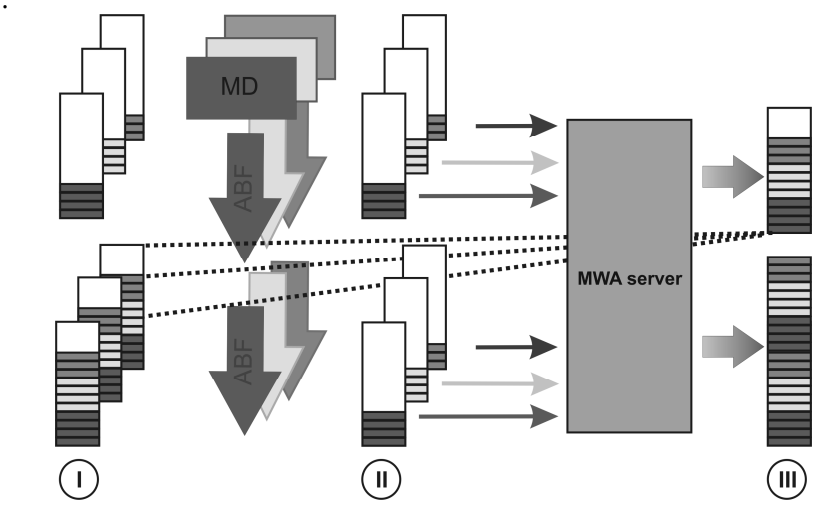

Fig. 2. Scheme of implemented ABF/MWA method

Since the pmemd program is written in Fortran 90 language, ABF method was also implemented in the same language. Client site and MWA server were then implemented in C++ language. Communications among clients and server was implemented over simple TCP/IP connection with client authorization using plain text passphrases. In our tests, the MWA server was always run in MetaCentrum because it requires very long execution time. On contrary, clients were repeatedly run as short jobs in both grid environments.

3.2 Molecular Shuttle in Pseudorotaxane

To test the implemented ABF/MWA method, the molecular shuttle in pseudorotaxane complex was studied. The free energy was calculated as a function of two reaction coordinates. The first coordinate (ξ_1) was specified as the distance between the central plane of the cucurbit[7]uril molecule and the centre of mass of the axle. The second

coordinate (ξ_2) was added to prevent unfavorable axle bending. It is defined as the angle between carboxyl group on the axle terminus and two centers of both pyridinium rings. Obtained results for both studied states are shown in Fig. 3.

Two minima with nearly same free energy separated with moderate size free energy barrier about 8 kcal/mol were found in the state I. This indicates that the wheel can freely move along the axle. On the other hand, only one minimum was found in the state II. This means that the wheel is located in the middle of the axle with limited ability to move to the axle terminals. This computationally predicted pseudorotaxane behavior is in good agreement with known experimental data [10] thus it can be concluded that chosen methodology provides reasonable results at least on qualitative level.

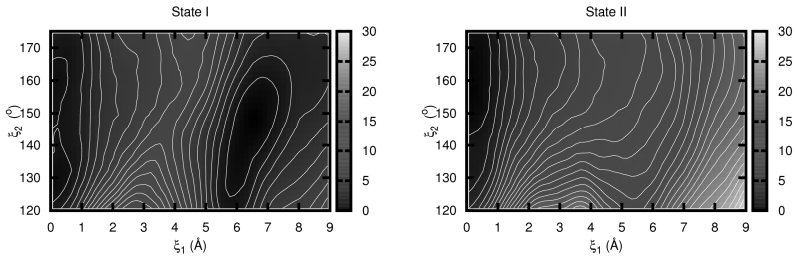

Fig. 3. Free energy profiles of molecular shuttle (in kcal/mol) for state I and II in NaCl solution (contour spacing is 1 kcal/mol)

3.3 Water Molecule Transit through CB[n] Cavity

Simulation protocols validated in the previous section were applied in the study of water molecule transit through CB[n] cavity for n from 5 to 8 (Fig. 4.).

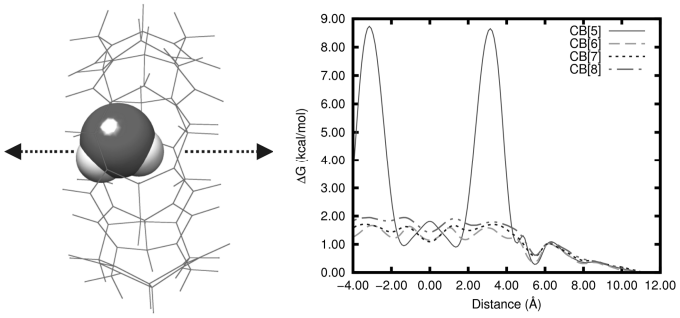

Fig. 4. (left) scheme of water transit through the CB[x] cavity; (right) corresponding free energy profiles

In the free energy calculations, only one-dimensional reaction coordinate was used (Fig. 4. left). It is defined as the distance from the central plane of the CB[n] molecule and the centre of mass of the water molecule. Obtained results are shown in Fig. 4 -

right. In all cases, the free energy increases when the water molecule enters into the cavity. This is in agreement with the fact that the cavity is hydrophobic. The energy increase of only about 2 kcal/mol for CB[6] to CB[8] shows that water molecules are not fully repelled from the cavity. Surprisingly, in the case of CB[5], it was found that the water molecule could possibly enter the cavity because the free energy barrier is only about 8.6 kcal/mol. This was not expected due to very small size of entry portal. Obtained results are important for the understanding of complexation properties of CB[n] with various hosts.

4 Conclusion and Future Directions

We successfully performed the free energy calculations of the molecular shuttle process and single water molecule transit through cucurbit[n]uril cavity using the Adaptive Biasing Force method accelerated by the Multiple Walkers Approach in two different grid computational environments: VOCE and MetaCentrum. In this evaluation study, the free energy profiles were obtained almost 40 times faster than by conventional runs. Our implementation of ABF/MWA method indicates that grid environments are suitable for such kind of scientific calculations.

Several improvements of our current implementation are possible. We plan to use an SSL encrypted communication with authentication via X509 user proxy certificate. Also utilization of several interconnected MWA servers would prevent data lost if one of them crashes due to work node malfunction or communication problems.

Acknowledgement. The presented work is supported by following grants: The Ministry of Education of the Czech Republic (contracts LC06030, MSM0021622413 and MSM6383917201) and by EGEE III project funded by European Commission (contract number INFSO-RI-222667).

References

1. Mulders, T., Kruger, P., Swegat, W., Schlitter, J.: Free energy as the potential of mean constraint force, Journal of Chemical Physics, vol. 104, Mar. 1996, pp. 4869-4870.
2. Torrie, G., Valleau, J.: Non-physical sampling distributions in monte-carlo free-energy estimation - umbrella sampling, Journal of Computational Physics, vol. 23, 1977, pp. 187-199.
3. Laio, A., Parrinello, M.: Escaping free-energy minima, Proceedings of the National Academy of Sciences of the United States of America, vol. 99, Oct. 2002, pp. 12562-12566.
4. Darve, E., Pohorille, A.: Calculating free energies using average force, Journal of Chemical Physics, vol. 115, Nov. 2001, pp. 9169-9183.
5. Kirkwood, J.: Statistical mechanics of fluid mixtures, Journal of Chemical Physics, 1935, pp. 300–313.
6. Sugita, Y., Okamoto, Y.: Replica-exchange molecular dynamics method for protein folding, Chemical Physics Letters, vol. 314, Nov. 1999, pp. 141-151.
7. E, W., Ren, W., Vanden-Eijnden, E.: String method for the study of rare events, Physical Review B, vol. 66, Aug. 2002, p. 052301.

8. Raiteri, P., Laio, A., Gervasio, F., Micheletti, C., Parrinello, M.: Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics, Journal of Physical Chemistry B, vol. 110, Mar. 2006, pp. 3533-3539.
9. Kim, K.: Mechanically interlocked molecules incorporating cucurbituril and their supramolecular assemblies, Chemical Society Reviews, vol. 31, Mar. 2002, pp. 96-107.
10. Sindelar, V., Silvi, S., Parker, S.E., Sobransingh, D., Kaifer, A.E.: Proton and electron transfer control of the position of cucurbit[n]uril wheels in pseudorotaxanes, Advanced Functional Materials, vol. 17, Mar. 2007, pp. 694–701.
11. Sindelar, V., Silvi, S., Kaifer, A.E.: Switching a molecular shuttle on and off: simple, pH-controlled pseudorotaxanes based on cucurbit[7]uril, Chemical Communications, May. 2006, pp. 2185–2187.
12. Frisch, M., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J., Montgomery, J., Vreven, T., Kudin, K.N., Burant, J.C., Millam, J.M., Iyengar, S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, Y., Rega, N., Petersson, G.A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J., Hratchian, H., Cross, J., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R.E., Yazyev, O., Austin, A.J., Cammi, R., Pomelli, C., Ochterski, J., Ayala, P.Y., Morokuma, K., Voth, G., Salvador, P., Dannenberg, J.J., Zakrzewski, V., Dapprich, S., Daniels, A.D., Strain, M.C., Farkas, O., Malick, D.K., Rabuck, A.D., Raghavachari, J., Foresman, J.B., Ortiz, J.V., Cui, Q., Baboul, A.G., Clifford, S., Cioslowski, J., Stefanov, B.B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R.L., Fox, D.J., Keith, T., Al-Laham, M.A., Peng, C.Y., Nanayakkara, A., Challacombe, M., Gill, P.M.W., Johnson, B., Chen, W., Wong, M.W., Gonzalez, C., Pople, J.: Gaussian 03, Revision D.1, Gaussian, Inc., Wallingford, CT (2003).
13. Dupradeau, F., Zaffran, T., Cieplak, P., Pigache, A.: R.E.D. II., Faculté de Pharmacie, Amiens (2005).
14. Case, D., Darden, T., Cheatham, T., Simmerling, C., Wang, J., Duke, R., Luo, R., Merz, K., Pearlman, D., Crowley, M., Walker, R., Zhang, W., B. Wang, Hayik, S., Roitberg, A., Seabra, G., Wong, K., Paesani, F., Wu, X., Brozell, S., Tsui, V., Gohlke, H., Yang, L., Tan, C., Mongan, J., Hornak, V., Cui, G., Beroza, P., Mathews, D., Schafmeister, C., Ross, W., Kollman, P.A.: Amber 9, University of California, San Francisco (2006).
15. Wang, J.M., Wolf, R.M., Caldwell, J.W., Kollman, P.A., Case, D.A.: Development and testing of a general Amber force field, Journal of Computational Chemistry, vol. 25, Jul. 2004, pp. 1157–1174.
16. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., Simmerling, C.: Comparison of multiple Amber force fields and development of improved protein backbone parameters, Proteins-Structure Function and Bioinformatics, vol. 65, Nov. 2006, pp. 712–725.
17. Jorgensen, W., Chandrasekhar, J., Madura, J., Impey, R., Klein, M.: Comparison of simple potential functions for simulating liquid water, Journal of Chemical Physics, vol. 79, 1983, pp. 926-935.
18. Darden, T., York, D., Pedersen, L.: Particle mesh Ewald - an N.log(N) method for Ewald sums in large systems, Journal of Chemical Physics, vol. 98, Jun. 1993, pp. 10089-10092.
19. Ryckaert, J.P., Ciccotti, G., Berendsen, H.J.C.: Numerical-integration of cartesian equations of motion of a system with constraints - molecular-dynamics of n-alkanes, Journal of Computational Physics, vol. 23, 1977, pp. 327–341.
20. Darve, E., Rodriguez-Gomez, D., Pohorille, A.: Adaptive biasing force method for scalar and vector free energy calculations, Journal of Chemical Physics, vol. 128, Apr. 2008, p. 144120.