ISSN 0027-1314, Moscow University Chemistry Bulletin, 2016, Vol. 71, Nos. 5–6, pp. 283–286. © Allerton Press, Inc., 2016.
Original Russian Text © V.A. Mironov, L.A. Lychko, M.G. Khrenova, 2016, published in Vestnik Moskovskogo Universiteta, Seriya 2: Khimiya, 2016, No. 5, pp. 303–306.

Methodological Aspects of the Calculation of the Free Energy Profile of Guanosine Triphosphate Hydrolysis by Ras-GAP Protein Complex

V. A. Mironov*, L. A. Lychko, and M. G. Khrenova

Department of Chemistry, Moscow State University, Moscow, Russia

**e-mail: vmironov@lcc.chem.msu.ru*

Received February 12, 2016

Abstract—Methods of statistical data analysis for umbrella sampling simulations of enzyme catalytic reactions were considered using guanosine triphosphate hydrolysis by the Ras-GAP protein complex as an example.

Keywords: small GTPases, free energy of a reaction, molecular dynamics

DOI: 10.3103/S0027131416050047

Prediction and direct modification of enzyme properties is one of major goals of modern biochemistry. It requires knowledge about all elementary steps, particularly activation energy barriers of enzymatic reactions. Traditional approach for energy barrier prediction implies calculation of minimum energy profile of the reaction. It requires the search for stationary points on the potential energy surface (PES). The minima on the PES correspond to the reagents and products of the reaction, while the saddle point corresponds to the transition state.

Most biological systems have an extremely large number of degrees of freedom. Consequently, numerous stationary points corresponding to reagents, products and transition states exist. They almost do not differ in energy thereby complicating theoretical studies. One of the approaches for solving this problem is calculation of the free energy of a reaction that accounts density of states of the system. One of the possible choices of free energy functions is Helmholtz free energy. In order to obtain accurate results, this approach requires computationally very intensive molecular-dynamic (MD) simulations using quantum mechanical (QM) energy potential. Such simulations became affordable only recently with the growing power of supercomputers and the development of molecular modeling methods. Free energy accuracy is directly related to the amount of gathered statistical data (length of MD trajectory). Nowadays, simulation time for QM-based MD rarely exceeds several tens of picoseconds.

In this work we studied the dependence of calculated free energy profile on the length of the MD trajectory and on the choice of the data analysis method. The process of guanosine triphosphate (GTP) hydrolysis by the Ras-GAP protein complex was selected as a sample reaction. The reason was its biological importance and the fact that its mechanism is not yet completely understood. The Ras protein belongs to the class of small GTPases, which play an important role in signal transduction in cells. They act as molecular switches: the activity of the transduced signal depends on the molecule (guanosine triphosphate or guanosine diphosphate (GDP)) they are bound to. The hydrolysis of GTP to GDP in these proteins is controlled by the regulatory protein (GAP).

Numerous theoretical studies were devoted to the Ras-GAP protein mediated hydrolysis [1–5]. According to one of the suggested mechanisms [2], the reaction proceeds by attacking the terminal phosphorus atom of the GTP molecule via the oxygen atom in the reacting water molecule accompanied by simultaneous cleavage of the bond between the terminal phosphorus atom and bridging oxygen of the GTP molecule (Fig. 1). After that cyclic proton transfer from the water molecule to the terminal phosphate group via Gln61 residue occurs, leading to the formation of the reaction products (GDP and inorganic phosphate as well as the imide form of Gln61 residue). A similar mechanism was suggested in [5]. The authors used the method of empirical valence bonds calibrated on the basis of the reaction of GTP hydrolysis in water in order to reproduce the free energy profile of GTP hydrolysis in protein environment. Two paths were investigated for the second step of the reaction, with one of them corresponding to the abovementioned mechanism and another including participation of two water molecules for the proton transfer.

In this work, coordinates of heavy atoms of the model system were taken from the crystal structure of the Ras-GAP complex containing the GTP analogue with the terminal phosphate group replaced with aluminum trifluoride. Reference structure was taken from the protein structure database PDB (PDB ID

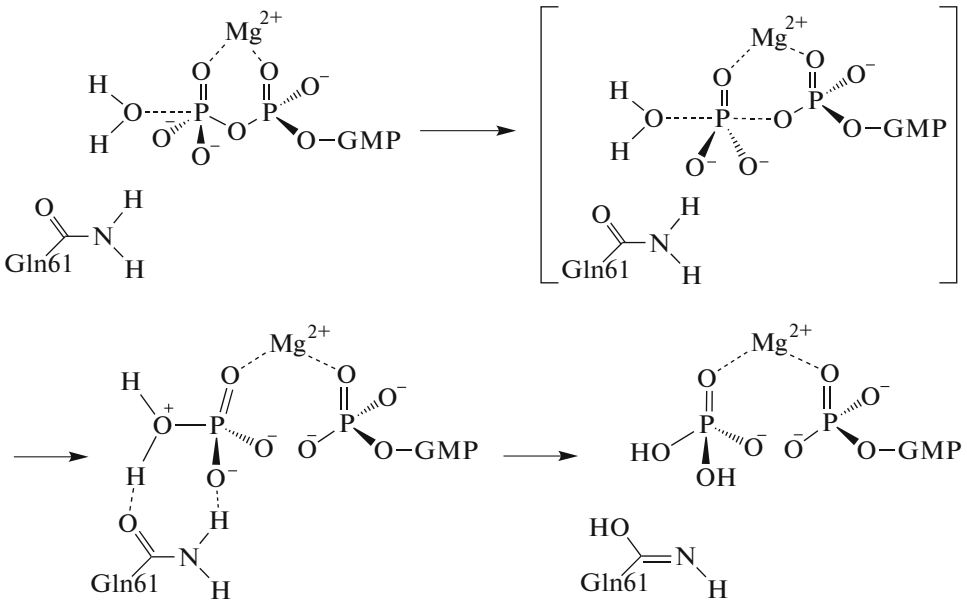

Fig. 1. Schematic representation of the mechanism of GTP hydrolysis in the active center of Ras-GAP enzymatic complex.

1WQ1) [6]. The AlF_3 group was replaced with the corresponding PO_3 group. The missing hydrogen atoms were recovered with the help of the VMD software [7]. The obtained all-atom model was placed into a rectangular box of water molecules in such a way that the distance from the protein atoms to the box boundary being no less than 10 Å. Na^+ and Cl^- ions were added at physiological concentration (~0.15 mol/L) in order to neutralize the system charge. Relaxation of the resulting system was achieved by 20 ns MD simulation using the NAMD program [8]. The Langevin molecular dynamics was used for simulation, with temperature and pressure maintained at 300 K and 1 bar, respectively. The time step of MD simulation was 1 fs. The parameters of atoms in the Ras-GAP protein complex, GTP molecule, and ions were taken from the CHARMM force field [9]; the TIP3P model was used for description of water molecules.

Molecular dynamic modeling was performed using quantum mechanics/molecular mechanics (QM/MM) approach with the help of the CP2K software [10] that uses hybrid basis sets of Gaussian functions and plane waves. The quantum mechanical subsystem included the triphosphate group of GTP, Mg^{2+} ion, water molecules closest to the GTP molecule (including the catalytic one), and the so-called 'arginine finger' of the GAP protein (GAP-Arg789). In addition, the residues of Ras protein (Gln61, Ser17, and Thr35) were included in the QM-subsystem together with the backbone of residues Gly12...Ala18 and Ala59...Gly60. The MM-subsystem was described using CHARMM force field. The forces on the atoms of the QM subsystem were calculated using density functional theory with BLYP functional and empirical dispersion correction DFT-D3 [11]. The DZVP basis optimized for calculation of molecular systems [12] with GHT pseudopotential was used for calculations. The MD simulations were conducted using NTV ensemble (300 K) using the Nosé–Hoover thermostat chains.

The free energy profiles were calculated with the help of the umbrella sampling (US) approach. This method allows collecting statistical data from high-energy regions of configuration space along the reaction coordinate. For this purpose a specially tuned potential is added to the system that compensates the energy change during the reaction. Due to additive nature of this potential the initial profile can be easily recovered from the statistical data obtained in simulation with that potential. In order to simplify the method and to accelerate convergence, the reaction path is usually separated into several parts ('windows'), and the data from all windows are combined using special methods described below. A unique potential is used for each window that provides the best sampling in the preset region; harmonic potential is used as a rule. In this work the reaction path was broken down into 18 regions. For each of the regions, MD-calculation was performed with external potential. The trajectory length obtained for each window was 15–20 ps. The distance from the oxygen atom of the reacting water molecule to the phosphorus atom of the terminal phosphate in GTP was used as a reaction coordinate.

The reaction mechanism suggested in this work is in agreement with the mechanism described in [1]. The first step of the reaction corresponds to formation of the intermediate with water molecule bound to GTP. This step is also the limiting one with the barrier

of ~10 kcal/mol. Next step comprises cyclic proton transfer from catalytic water molecule to the oxygen atom of inorganic phosphate via the Gln61 residue. Imide form of Gln61 is formed on this stage.

The data obtained in different windows along the reaction path were combined with the help of weighted histogram analysis method (WHAM [13]), umbrella integration (UI, [14]), and transition-based reweighting analysis method (TRAM, [15]). These methods differ significantly in the ways they construct the free energy profile. In WHAM approach the probability density of finding the system along the reaction coordinate is calculated on the basis of combined data from multiple windows. The desired value of Helmholtz free energy is proportional to the logarithm of the probability density. In the UI method, the derivative of free energy is calculated from the assumption of normal distribution of the reaction coordinate value in each window, and the final free energy profile is obtained via numerical integration.

A fundamentally different approach is used in the TRAM method. According to this approach each MD-trajectory is considered as a Markov chain. Some states of interest of the chemical system are defined (e.g. regions of configuration space along the reaction path) which are assumed to be in equilibrium between each other. The probability density of finding the system in the selected states and corresponding free energy can be calculated from the matrix of transition constants between these states. The results of this method are fairly good even when the configuration space was poorly sampled. Despite the differences in methodology, all methods used for the analysis of MD with additional external potential result in similar energy profiles (Fig. 2). This indicates that these approaches are equivalent if the amount of statistical data is large enough.

It needs to be noted that the local equilibrium must be reached for each MD window. Hence, some of the data collected at the beginning of each MD window corresponds to equilibration and should not be taken into account when calculating free energy profile. To estimate the actual amount of these data we studied the dependencies of the free energy profiles on the number of discarded frames in each window (Fig. 3). It was found that the equilibrium was reached after ~6 ps. Considering that the preliminary MD simulation (2 ps) was performed for each window, we can state that the time for reaching local equilibrium is approximately 8 ps for the system in our study. It should be expected that this estimate is valid for numerous other proteins specifically for GTPases and ATPases, which have very conservative structure of the active center (for the entire protein series).

Hence, it can be concluded that the relaxation of the system takes a significant part of computing time (30–50%). If the sampling is sufficient, the results obtained by all considered approaches are similar.

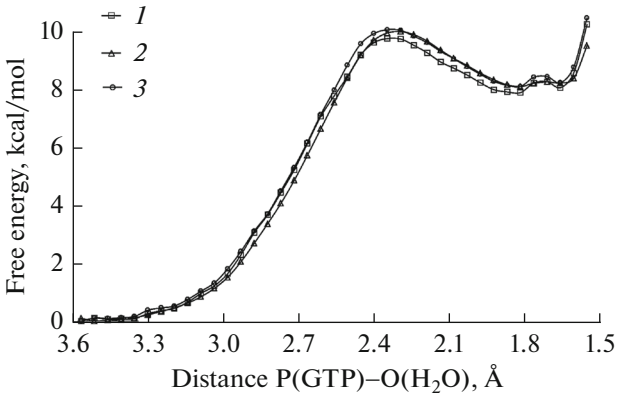

Fig. 2. Comparison of results of calculation of free energy profiles using WHAM (*1*), UI (*2*), and TRAM (*3*) methods.

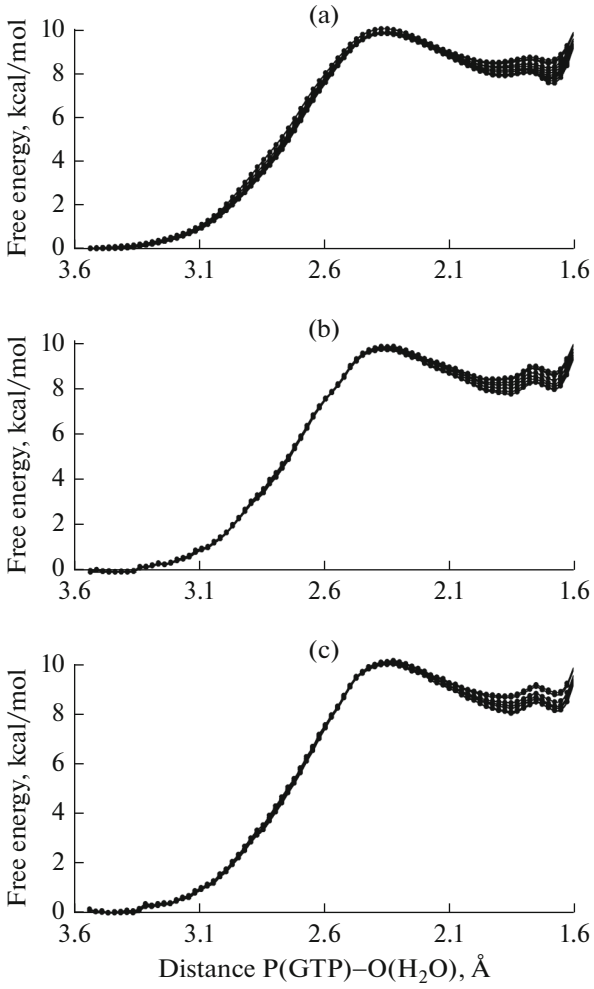

Fig. 3. Convergence of free energy profiles depending on the number of discarded data at the beginning of each US window calculated with methods: (a) UI, (b) WHAM, (c) TRAM.

ACKNOWLEDGEMENTS

This work was supported by the Russian Foundation for Basic Research, project no. 14-03-31898.

REFERENCES

1. Topol, I.A., Cachau, R.E., Nemukhin, A.V., et al., *Biochim. Biophys. Acta*, 2004, vol. 1700, no. 1, p. 125.
2. Grigorenko, B.L., Nemukhin, A.V., Topol, I.A., et al., *Proteins*, 2005, vol. 60, no. 3, p. 495.
3. Lu, Q., Nassar, N., and Wang, J., *Chem. Phys. Lett.*, 2011, vol. 516, nos. 4–6, p. 233.
4. Martín-García, F., Mendieta-Moreno, J.I., López-Viñas, E., et al., *Biophys. J.*, 2012, vol. 102, no. 1, p. 152.
5. Prasad, B.R., Plotnikov, N.V., Lameira, J., et al., *Proc. Natl. Acad. Sci. U. S. A.*, 2013, vol. 110, no. 51, p. 20509.
6. Berman, H.M., Westbrook, J., Feng, Z., et al., *Nucleic Acid Res.*, 2000, vol. 28, no. 1, p. 235.
7. Humphrey, W., *J. Mol. Graphics*, 1996, vol. 14, no. 1, p. 33.
8. Phillips, J.C., Braun, R., Wang, W., et al., *J. Comput. Chem.*, 2005, vol. 26, no. 16, p. 1781.
9. Mackerell, A.D., Feig, M., and Brooks, C.L., *J. Comput. Chem.*, 2004, vol. 25, no. 11, p. 1400.
10. CP2K: A general program to perform molecular dynamics simulations: 2.3, CP2K developers group, 2012.
11. Grimme, S., Antony, J., Ehrlich, S., et al., *J. Chem. Phys.*, 2010, vol. 132, no. 15, 154104. doi 10.1063/1.3382344
12. VandeVondele, J. and Hutter, J., *J. Chem. Phys.*, 2007, vol. 127, no. 11, 114105. doi 10.1063/1.2770708
13. Roux, B., *Comput. Phys. Commun.*, 1995, vol. 91, nos. 1–3, p. 275.
14. Kästner, J. and Thiel, W., *J. Chem. Phys.*, 2005, vol. 123, no. 14, 144104. doi 10.1063/1.2052648
15. Wu, H., Mey, A.S.J.S., Rosta, E., et al., *J. Chem. Phys.*, 2014, vol. 141, no. 21, 214106. doi 10.1063/1.4902240

Translated by L. Brovko