HHS Public Access

Author manuscript

J Chem Inf Model. Author manuscript; available in PMC 2019 March 26.

Published in final edited form as:

J Chem Inf Model. 2018 March 26; 58(3): 556–560. doi:10.1021/acs.jcim.7b00695.

BFEE: A User-Friendly Graphical Interface Facilitating Absolute Binding Free-energy Calculations

Haohao Fu[†], **James C. Gumbart**[▽], **Haochuan Chen**[†], **Xueguang Shao**[†,‡,#], **Wensheng Cai**[*,†,‡], and **Christophe Chipot**[*,⊥,$,§]

[†]Research Center for Analytical Sciences, College of Chemistry, Tianjin Key Laboratory of Biosensing and Molecular Recognition, Nankai University, Tianjin 300071, China

[▽]School of Physics and School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332

[‡]Collaborative Innovation Center of Chemical Science and Engineering (Tianjin), Tianjin 300071, China

[#]State Key Laboratory of Medicinal Chemical Biology, Tianjin 300071, China

[⊥]Laboratoire International Associé Centre National de la Recherche Scientifique et University of Illinois at Urbana-Champaign, Unité Mixte de Recherche No. 7565, Université de Lorraine, B.P. 70239, 54506 Vandoeuvre-lès-Nancy cedex, France

[$]Theoretical and Computational Biophysics Group, Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

[§]Department of Physics, University of Illinois at Urbana−Champaign, 1110 West Green Street, Urbana, Illinois 61801, United States

Abstract

Quantifying protein-ligand binding has attracted the attention of both theorists and experimentalists for decades. Many methods for estimating binding free energies *in silico* have been reported in recent years. Proper use of the proposed strategies requires, however, adequate knowledge of the protein-ligand complex, the mathematical background for deriving the underlying theory, and time for setting up the simulations, bookkeeping and post-processing. Here, to minimize human intervention, we propose a toolkit aimed at facilitating the accurate estimation of standard binding free energies using a geometrical route, coined binding free-energy estimator (BFEE), and introduced as a plug-in of the popular visualization program VMD. Benefit from recent developments in new collective variables, BFEE can be used to generate the simulation input files, based solely on the structure of the complex. Once the simulations are completed, BFEE can also be utilized to perform the post-treatment of the free-energy calculations, allowing

[*] wscai@nankai.edu.cn (W.C.); chipot@ks.uiuc.edu (C.C.).

Supporting Information.

The following files are available free of charge.

Source code of BFEE and example files (ZIP)

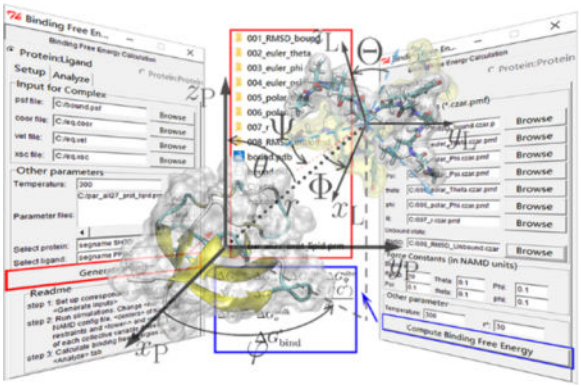
Readme of example files. Difference between pmf, UI.pmf and czar.pmf files (PDF)

Notes

The authors declare no competing financial interest.

the absolute binding free energy to be estimated directly from the one dimensional potentials of mean force in simulation outputs. The minimal amount of human intervention required during the whole process combined with the ergonomic graphical interface makes BFEE a very effective and practical tool for the end-user.

Graphical abstract



INTRODUCTION

The importance of accurate binding free-energy calculations of protein-ligand complexes is a truism.[1] The most serious difficulty of *in silico* estimation of the binding affinity resides in capturing the change in configurational entropy associated to protein-ligand association, which requires adequate sampling of all the relevant movements of the ligand with respect to the protein.[2,3] Approximate methods, on the one hand, such as molecular mechanics/Poisson-Boltzmann surface area,[4] usually ignore part of the entropic contribution to the binding free energy. Strategies for the accurate evaluation of this contribution, on the other hand, involve a rather elaborate workflow. For example, the confine-and-release method[5] uses extensive potential of mean force (PMF) calculations to search for the metastable states of the complex and add corrections arising from geometrical restraints to the binding affinity obtained by an alchemical route. The attach-pull-release strategy[6] decomposes the absolute binding free energy into the reversible work of connecting an artificial spring to the host, adjusting the spring to extract the host from the guest, and ultimately releasing the spring. We also purposed a geometrical route, in which protein-ligand binding is decomposed into several independent subprocesses, each of which describes the sampling of one degree of freedom at a time. Restraints are added to these degrees of freedom, the contributions of which are evaluated in PMF calculations (Figure 1).[7–9]

As one can easily expect, setting up an absolute binding free-energy calculation using the latter strategy can require substantial human intervention, i.e., from building the simulation assays to choosing the proper simulation parameters. Attempts to automate the setup have been made, e.g., using the CharmmGUI server.[10,11] This approach, however, selects groups of atoms to define the relative orientation and position of the ligand with respect to the protein, which can be problematic for unusual geometries of the host or the guest. Moreover, this tool only helps users to generate input files. The more taxing aspect of the stepwise

strategy of the geometric route is without a doubt the bookkeeping of the different PMF calculations, in particular their post-treatment and the evaluation of the configurational integrals that appear in the expression of the binding constant, K_{eq} (Equation 1). Some other codes, like mmpbsa.py[12] in Ambertools,[13] analyze the results of a binding free energy estimation using the approximate MM-PBSA method. Development of an automated tool for the setup and the post-treatment of accurate free-energy calculations is, therefore, highly desirable.

With the objective of minimizing human intervention in the estimation of protein-ligand binding affinities, we present in this contribution a plug-in for the visualization program VMD[14] coined binding free-energy estimator (BFEE). This plug-in can automatically set up and analyze absolute binding free-energy calculations carried out with the popular molecular dynamics engine NAMD,[15] eliminating tedious and repetitive file preparation and bookkeeping, as shown in Figure 1.

THEORETICAL BACKGROUND

The detail of the theoretical background of the binding free-energy calculation can be found elsewhere.[9,16] Here, we simply recall the expression of K_{eq}, utilized in the post-treatment of the PMF calculations,

$$\begin{aligned} K_{\mathrm{eq}} &= \frac{\int_{\mathrm{site}} \mathrm{d}\mathbf{1} \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta U}}{\int_{bulk} \mathrm{d}\mathbf{1}\, \delta\left(\mathbf{x}_1 - \mathbf{x}_1^*\right) \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta U}} \\ &= \frac{\int_{\mathrm{site}} \mathrm{d}\mathbf{1} \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta U}}{\int_{\mathrm{site}} \mathrm{d}\mathbf{1} \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta (U+u_c)}} \\ &\times \frac{\int_{\mathrm{site}} \mathrm{d}\mathbf{1} \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta (U+u_c)}}{\int_{\mathrm{site}} \mathrm{d}\mathbf{1} \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta (U+u_c+u_o)}} \\ &\times \frac{\int_{\mathrm{site}} \mathrm{d}\mathbf{1} \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta (U+u_c+u_o)}}{\int_{\mathrm{site}} \mathrm{d}\mathbf{1} \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta (U+u_c+u_o+u_a)}} \\ &\times \frac{\int_{\mathrm{site}} \mathrm{d}\mathbf{1} \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta (U+u_c+u_o+u_a)}}{\int_{\mathrm{bulk}} \mathrm{d}\mathbf{1}\, \delta\left(\mathbf{x}_1 - \mathbf{x}_1^*\right) \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta (U+u_c+u_o)}} \\ &\times \frac{\int_{\mathrm{bulk}} \mathrm{d}\mathbf{1}\, \delta\left(\mathbf{x}_1 - \mathbf{x}_1^*\right) \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta (U+u_c+u_o)}}{\int_{\mathrm{bulk}} \mathrm{d}\mathbf{1}\, \delta\left(\mathbf{x}_1 - \mathbf{x}_1^*\right) \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta (U+u_c)}} \\ &\times \frac{\int_{\mathrm{bulk}} \mathrm{d}\mathbf{1}\, \delta\left(\mathbf{x}_1 - \mathbf{x}_1^*\right) \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta (U+u_c)}}{\int_{\mathrm{bulk}} \mathrm{d}\mathbf{1}\, \delta\left(\mathbf{x}_1 - \mathbf{x}_1^*\right) \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta U}} \\ &= \mathrm{e}^{-\beta(\Delta G_c^{\mathrm{site}} + \Delta G_o^{\mathrm{site}} + \Delta G_a^{\mathrm{site}} - \frac{1}{\beta}\ln(S*I*C^\circ) + \Delta G_o^{\mathrm{bulk}} + \Delta G_c^{\mathrm{bulk}})} \end{aligned} \quad (1)$$

where **1** denotes the ligand, $\mathbf{x_1}$ the position of its center of mass, and $\mathbf{x_1}^*$ an arbitrary location in the solution, sufficiently far from the binding site. U represents the potential energy of the whole protein-ligand assembly, $u_o = u_\Theta + u_\Phi + u_\Psi$ is the restraining potential for the three Euler angles, Θ, Φ, and Ψ, and $u_a = u_\theta + u_\phi$ is that for the spherical angles, θ and ϕ (see Figure 1). Under these premises, the standard binding free energy is given by

$$\Delta G^{\circ}_{\mathrm{bind}} = k_{\mathrm{B}} T \times \ln K_{\mathrm{eq}} C^{\circ}$$

As made clear by Equation 1, one needs to run eight individual one-dimensional PMF calculations, seven of which involve the protein-ligand complex, using in a sequential order the distance root-mean-square deviation (RMSD) with respect to the native conformation of the ligand in the bound state, the three Euler angles, Θ, Φ, and Ψ, the two spherical angles, θ and ϕ, and the distance between the centers of mass of the protein and the ligand. The remaining PMF calculation describes the change in the distance RMSD of the ligand in the unbound state. The corresponding restraints, namely u_c, u_o and u_a, are introduced as described in Equation 1.

METHOD

Implementation Details

BFEE is implemented as a Tcl plug-in in VMD,[14] which is available for Microsoft Windows, Linux and macOS operating systems. Although BFEE is designed for PMF calculations applying the extended adaptive biasing force (eABF) method with an unbiased estimator[17,18] provided in the Colvars[19] module of NAMD[15], there is no technical barrier preventing one from porting this toolkit to other molecular dynamics engines. For example, by integrating TopoTools[20] or charmm2lammps.pl,[21] one can easily generate LAMMPS[22] input files with BFEE. The source code of this plug-in is provided in the Supporting Information. The most updated version of BFEE will be always released together with VMD.[14]

Functional Demonstration

BFEE is shown in Figure 2. We now detail the usage of the plug-in.

Preparing the input files for the PMF calculations—Users are expected to be familiarized with NAMD[15] and have successfully equilibrated the protein-ligand complex in bulk water. In order to set up the different simulations involved in the geometric route for standard binding free-energy determination, BFEE requires the structure file of the complex in bulk water (`psf` file), the binary coordinates (`coor` file), the binary velocities (`vel` file), the periodic-cell dimensions (xsc file) and the force-field parameters (usually, `prm` or `str` files) as inputs. All of these files should already exist or be generated through an equilibrium simulation. The input files required for the eight individual PMF calculations are then automatically generated by clicking the “Generate Inputs” button, as shown in Figure 2.

Running the simulations—Ideally, one can submit the configuration files directly to the NAMD program without any modification. It might be, however, desirable to tailor the range of values sampled in the PMF calculations, as well as the center of the harmonic restraints as a function of the protein-ligand complex at hand. For the simulations characterizing the change in the RMSDs of the ligand (both in the bound and in the unbound states), in the Euler angles and in the spherical angles, we suggest choosing a range of values that satisfies $\Delta G(\mathrm{max})$-$\Delta G(\mathrm{min}) \geq 5$ kcal/mol (by default, BFEE sets the center of the angular harmonic

restraints to their equilibrium value in the native complex, and samples a range of ±10° around this value in the PMF calculations). A short preliminary simulation for each angle can be performed to refine the range of values that ought to be covered in the different free-energy calculations. In addition, the center of each angular harmonic restraint acting on Θ, Φ, Ψ, θ, or ϕ may need to be adapted to reflect the corresponding free-energy minimum. These fine-tunings constitute the only human intervention needed during the entire binding free-energy calculation workflow. In practice, due to the inherent complexity of protein-ligand binding, one may turn to use a staging, or stratification strategy in PMF calculations rather than a single-window simulation (by default) for an improved convergence rate (see the Supporting Information for examples).

Post-treatment—BFEE calculates the binding free energy as well as the contribution of each degree of freedom automatically. As shown in Figure 2, only the free-energy profile at each step of the workflow (usually `czar.pmf`. See the Supporting Information for the difference between `pmf` and `czar.pmf` files) is needed as input if default force constants (listed below in the same window) are used for the geometric restraints. The results of the free-energy calculation will be shown after clicking the button “Compute Binding Free Energy”, as depicted in Figure 3, which shows the affinity of ligand p41 binding to SH3 domain of Abl kinase. Although to reach this degree of convergence required a stratification strategy, an example using single-window simulations produced a result within ~1 kcal/mol. This example and others are provided in the Supporting Information.

Error Estimation—Error estimation is extremely important in binding free-energy calculations. The recommended way to estimate the error using BFEE is to run parallel simulations independently, with different seeds, and then calculate the standard deviation over the different binding free energies. Since BFEE handles all the generation of inputs and analysis of outputs, almost no additional human intervention is needed in the error calculation, unlike the determination of the precision of the calculation, which requires an estimation of the correlation length of the time series.[23]

Additionally, it is crucial to guarantee that each step of the PMF calculation is converged when evaluating the reliability of binding free-energy estimates. This evaluation can be done by monitoring the time evolution of individual free-energy landscape (`hist.czar.pmf` file). If a PMF calculation is converged, the free-energy and gradient profiles should remain unchanged with respect to the simulation time. We have an in-house script to perform this task, and we will make it available with the formal release of BFEE. (Currently, this script can be found at https://github.com/fhh2626/eABF-analyzer-for-NAMD)

CONCLUSION AND PERSPECTIVE

Motivated by the idea that to be truly useful beyond the academic walls, accurate binding free-energy estimation ought to be automated to reduce human intervention as much as possible, in a continuing effort we have designed a “parameterless” computational workflow[16] and an ergonomic graphical user-friendly interface for the setup and post-treatment of the different simulations. With these objectives in mind, we introduce a new plug-in, BFEE, for accurate standard binding free-energy calculations, based on a strategy

recently put forth,[9,16] making use of geometric transformations. The rapid setup and post-treatment of the simulations by BFEE make the latter a very attractive option for the non-expert in the field of advanced statistical-mechanics simulations.

At the design level, BFEE offers a sufficiently general and easy way to investigate any recognition-and-association process relevant to chemistry and biology. In practice, however, there are at least two cases for which usage of BFEE is ill-advised, namely, (i) when the guest is deeply buried in the host molecule, rendering the PMF calculation of the separation nearly intractable over common timescales, and (ii) in protein-protein complexes, where the interaction network at the interface is intricate, necessitating the introduction of additional restraints.[24] Possible workarounds consist of (i) turning to an alchemical approach as an alternative to the geometric route, with the corresponding input files generated automatically by BFEE (which is currently under development as an extension) and (ii) generalizing the plug-in to protein-protein binding to account for the complexity of the interfacial interactions through harmonic restraints acting on the participating amino-acid side chains. In its present version, BFEE provides a very convenient, user-friendly framework for investigating protein-ligand binding. Future developments include an extension of the plug-in to address virtually any problem relevant to recognition and association applied to arbitrary host-guest complexes, such as protein-protein and protein-membrane molecular assemblies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work is supported by National Natural Science Foundation of China (Nos. 21373117, 21773125), and special Program for Applied Research on Super Computation of the NSFC-Guangdong Joint Fund (the second phase) under Grant No. U1501501. The GENCI and CINES, Montpellier, France are gratefully acknowledged for provision of generous amounts of CPU time. C. C. is indebted to the Centre National de la Recherche Scientifique for the support of his joint research program (PICS) with the People's Republic of China. JCG acknowledges support from the US National Institutes of Health (R01-GM123169), US National Science Foundation (1452464), and the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by NSF grant number OCI-1053575, for additional computational resources.

References

1. Gilson MK, Zhou H-X. Calculation of protein-ligand binding affinities. Annu Rev Biophys Biomol Struct. 2007; 36:21–42. [PubMed: 17201676]
2. Chipot C. Frontiers in free-energy calculations of biological systems. WIRES: Comput Mol Sci. 2014; 4:71–89.
3. Chodera JD, Mobley DL. Entropy-enthalpy compensation: role and ramifications in biomolecular ligand recognition and design. Annu Rev Biophys. 2013; 42:121–142. [PubMed: 23654303]
4. Srinivasan J, Cheatham TE, Cieplak P, Kollman PA, Case DA. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate- DNA helices. J Am Chem Soc. 1998; 120:9401–9409.
5. Mobley DL, Chodera JD, Dill KA. The confine-and-release method: Obtaining correct binding fee energies in the presence of protein conformational change. J Chem Theory Comput. 2007; 3:1231–1235. [PubMed: 18843379]

6. Velez-vega C, Gilson MK. Overcoming dissipation in the calculation of standard binding free energies by ligand extraction. J Comput Chem. 2013; 34:2360–2371. [PubMed: 24038118]
7. Woo H-J, Roux B. Calculation of absolute protein–ligand binding free energy from computer simulations. Proc Natl Acad Sci US A. 2005; 102:6825–6830.
8. Wang J, Deng Y, Roux B. Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. Biophys J. 2006; 91:2798–2814. [PubMed: 16844742]
9. Gumbart JC, Roux B, Chipot C. Standard binding free energies from computer simulations: What is the best strategy? J Chem Theory Comput. 2013; 9:794–802. [PubMed: 23794960]
10. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: A web-based graphical user interface for CHARMM. J Comput Chem. 2008; 29:1859–1865. [PubMed: 18351591]
11. Jo S, Jiang W, Lee HS, Roux B, Im W. CHARMM-GUI ligand binder for binding free energy calculations and its application. J Chem Inf Model. 2013; 53:267–277. [PubMed: 23205773]
12. Miller BR III, McGee TD Jr, Swails JM, Homeyer N, Gohlke H, Roitberg AE. MMPBSA.py: an efficient program for end-state free energy calculations. J Chem Theory Comput. 2012; 8:3314–3321. [PubMed: 26605738]
13. Case, DA., Cerutti, DS., Cheatham, TE., III, Darden, TA., Duke, RE., Giese, TJ., Gohlke, H., Goetz, AW., Greene, D., Homeyer, N., Izadi, S., Kovalenko, A., Lee, TS., LeGrand, S., Li, P., Lin, C., Liu, J., Luchko, T., Luo, R., Mermelstein, D., Merz, KM., Monard, G., Nguyen, H., Omelyan, I., Onufriev, A., Pan, F., Qi, R., Roe, DR., Roitberg, A., Sagui, C., Simmerling, CL., Botello-Smith, WM., Swails, J., Walker, RC., Wang, J., Wolf, RM., Wu, X., Xiao, L., York, DM., Kollman, PA. AMBER 2017. University of California; San Francisco: 2017.
14. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996; 14:33–38. [PubMed: 8744570]
15. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable molecular dynamics with NAMD. J Comput Chem. 2005; 26:1781–1802. [PubMed: 16222654]
16. Fu H, Cai W, Hénin J, Roux B, Chipot C. New coarse variables for the accurate determination of standard binding free energies. J Chem Theory Comput. 2017; 13:5173–5178. [PubMed: 28965398]
17. Fu H, Shao X, Chipot C, Cai W. Extended adaptive biasing force algorithm. An on-the-fly implementation for accurate free-energy calculations. J Chem Theory Comput. 2016; 12:3506–3513. [PubMed: 27398726]
18. Lesage A, Lelièvre T, Stoltz G, Hénin J. Smoothed biasing forces yield unbiased free energies with the extended-system adaptive biasing force method. J Phys Chem B. 2017; 121:3676–3685. [PubMed: 27959559]
19. Fiorin G, Klein ML, Hénin J. Using collective variables to drive molecular dynamics simulations. Mol Phys. 2013; 111:3345–3362.
20. Kohlmeyer, Axel. TopoTools. 2017. DOI: http://dx.doi.org/10.5281/zenodo.598373
21. charmm2lammps.pl. https://github.com/lammps/lammps/tree/master/tools/ch2lmp (accessed Oct 25, 2017)
22. Plimpton S. Fast parallel algorithms for short-range molecular dynamics. J Comput Phys. 1995; 117:1–19.
23. Rodriguez-Gómez D, Darve E, Pohorille A. Assessing the efficiency of free energy calculation methods. J Chem Phys. 2004; 120:3563–3578. [PubMed: 15268518]
24. Gumbart JC, Roux B, Chipot C. Efficient determination of protein–protein standard binding free energies from first principles. J Chem Theory Comput. 2013; 9:3789–3798.

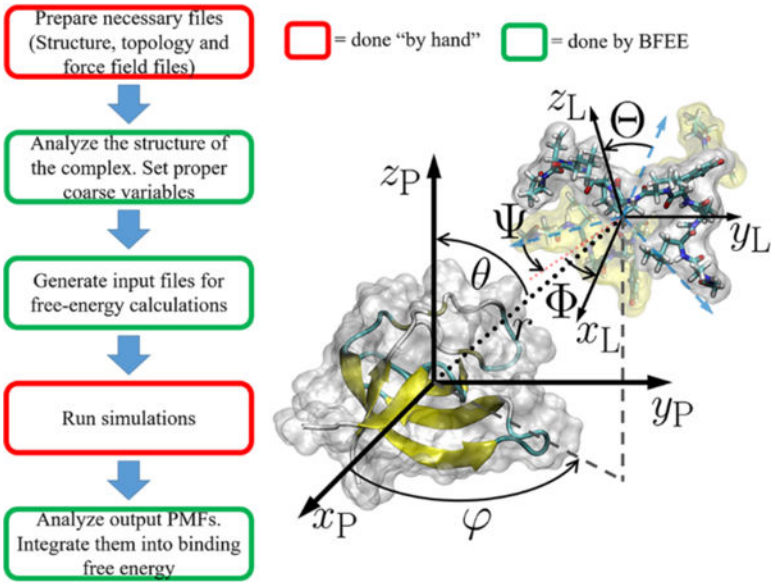


Figure 1.

Left: Workflow of the BFEE plug-in. Right: Degrees of freedom considered in the binding free energy calculation strategy. The isomerization of the ligand is considered by characterizing its RMSD with respect to the conformation of the ligand in the bound state, but this variable is not shown in the figure for clarity.

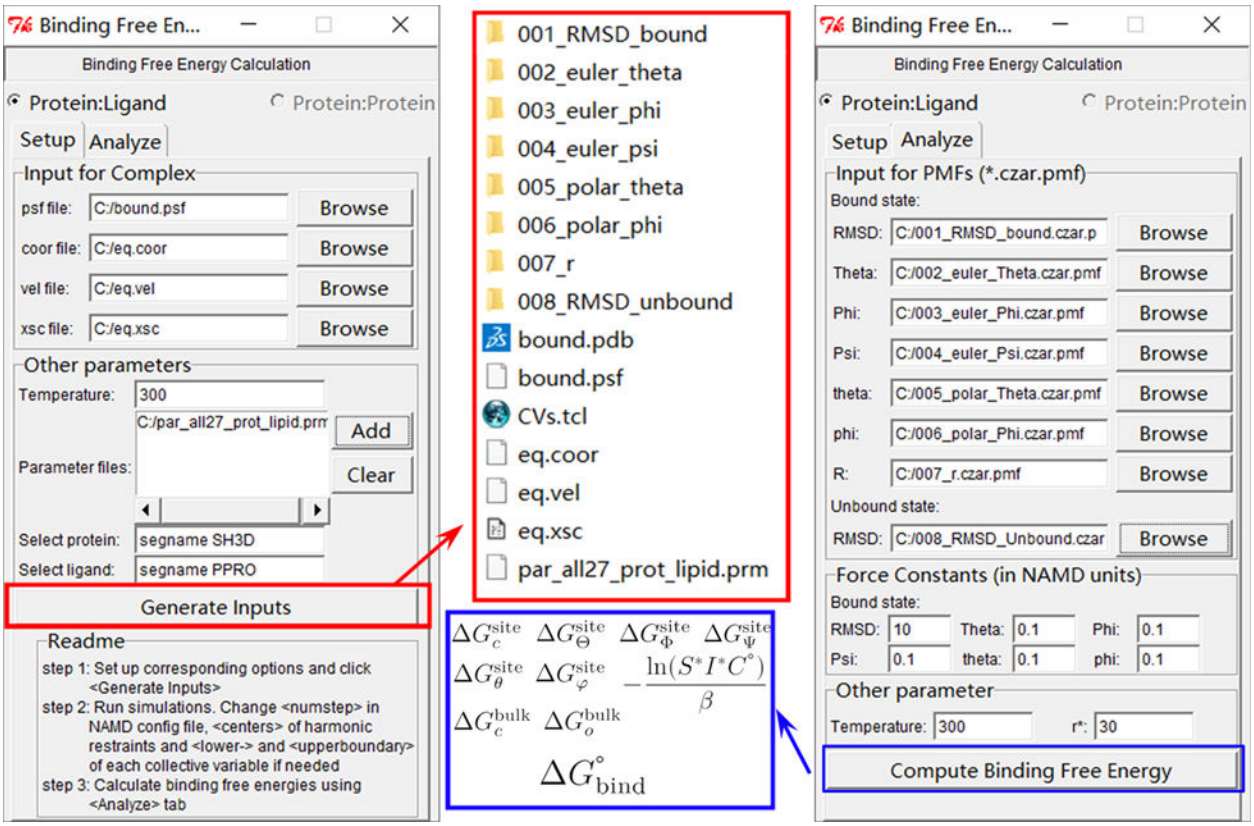


Figure 2.
Graphical user interface and functional demonstration of the BFEE plug-in.

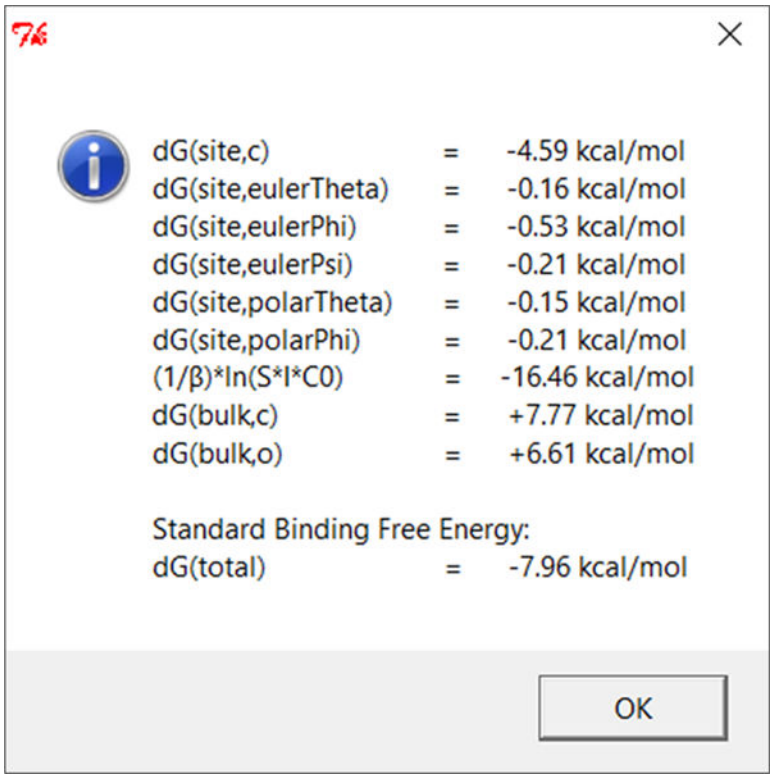


Figure 3.

Results shown in a message box of BFEE. The data were obtained from the case of ligand p41 binding to SH3 domain of Abl kinase. The initial coor, vel, xsc and force field files of the molecular system are provided in the Supporting Information as a test example of the BFEE plug-in. See ref. [12] for more information about the molecular assembly.