Bioinformatics, 32(14), 2016, 2216–2218
doi: 10.1093/bioinformatics/btw215
Advance Access Publication Date: 22 April 2016
Applications Note

OXFORD

Structural bioinformatics

CaFE: a tool for binding affinity prediction using end-point free energy methods

Hui Liu and Tingjun Hou*

College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, Zhejiang 310058, China

*To whom correspondence should be addressed.

Associate Editor: Burkhard Rost

Received on January 21, 2016; revised on March 23, 2016; accepted on April 14, 2016

Abstract

Summary: Accurate prediction of binding free energy is of particular importance to computational biology and structure-based drug design. Among those methods for binding affinity predictions, the end-point approaches, such as MM/PBSA and LIE, have been widely used because they can achieve a good balance between prediction accuracy and computational cost. Here we present an easy-to-use pipeline tool named Calculation of Free Energy (CaFE) to conduct MM/PBSA and LIE calculations. Powered by the VMD and NAMD programs, CaFE is able to handle numerous static coordinate and molecular dynamics trajectory file formats generated by different molecular simulation packages and supports various force field parameters.

Availability and implementation: CaFE source code and documentation are freely available under the GNU General Public License via GitHub at https://github.com/huiliucode/cafe_plugin. It is a VMD plugin written in Tcl and the usage is platform-independent.

Contact: tingjunhou@zju.edu.cn

1 Introduction

Accurate estimation of binding affinity (alternatively, binding free energy) is among the most attractive and challenging topics in both computational biology and computer-aided drug design (CADD) (Chipot, 2014). On one hand, noncovalent association is ubiquitous and essential in the cell machinery. It is involved in the processes of enzymatic catalysis, signal transduction, molecular recognition and receptor activation. Knowledge of the thermodynamic basis of association at the molecular level provides insights into the underlying mechanisms of various biological phenomena. On the other hand, from a pharmaceutical point of view, the major goal of drug design and discovery is to identify new chemical entities with high affinity to a certain pharmaceutical target. Obviously, these research fields would benefit from reliable methods for estimating binding affinity.

A number of computational methods have been proposed to predict binding free energy. There are some theoretically rigorous ones based on statistical mechanics; however, these methods are time-consuming and thus their applications in real-life CADD are limited. Another kind of approaches frequently used is so-called scoring functions. They are often embedded in molecular docking programs, for which simplicity and speed are generally required; however, it usually comes with the loss of accuracy. The third type of free energy prediction methods is referred to as end-point methods, which sample conformations only with the free and bound states and get binding free energy by taking a difference between these states. These methods can reach a good balance between accuracy and efficiency (Hou *et al.*, 2011a, b).

2 Overview of CaFE

We developed a program named Calculation of Free Energy (CaFE) to automatically conduct binding affinity prediction with end-point methods. In CaFE, two most popular end-point methods are incorporated, namely, the molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) (Kollman *et al.*, 2000) and the linear interaction energy (LIE) (Åqvist *et al.*, 1994) methods. Some tools are already available for conducting these two types of calculations (Case *et al.*, 2005; Homeyer and Gohlke, 2013; Kumari *et al.*, 2014; Lill and Danielson, 2011; Miller *et al.*, 2012). However, most of them are specifically designed for processing the topology and

trajectory files that are used by the AMBER or GROMACS packages and are focused on the usage of AMBER-related force field parameters. To date, there are no freely available pipeline tools that can be directly used to deal with the files (PSF and DCD file formats) generated by the CHARMM/NAMD software as well as others using the same file system. CaFE is designed to fill the gap. Compared with them, CaFE is capable of handling a variety of coordinate and trajectory files produced by different molecular simulation programs. Although files produced by different programs can be interconverted by using additional conversion tools, the CaFE program provides a simple solution to facilitate the setup of end-point free energy calculations by rendering such a file format conversion unnecessary. All of the post-processing and calculations are automated by a simple command without additional procedures, from which users in the field of CADD who often have quite a few samples to be evaluated would benefit a lot. Different types of force field parameters can be used in CaFE. This is powered by NAMD (Phillips *et al.*, 2005), which is employed by CaFE as the back-end molecular mechanical engine. In addition, some optimized atom radius sets (Banavali and Roux, 2002; Nina *et al.*, 1997; Sitkoff *et al.*, 1994) are embedded for MM/PBSA calculations.

By using the Tcl/Tk scripting interface of the versatile VMD package (Humphrey *et al.*, 1996), CaFE is implemented as a set of Tcl scripts, aiming to shorten and automate the process of binding affinity prediction while taking advantage of the powerful scripting interface to VMD's built-in functions. Several internal procedures of CaFE are implemented as Tcl wrappers around external programs for evaluating the single point gas-phase energy and solving the Poisson–Boltzmann (PB) equation. CaFE features are accessible as command-line procedures. By importing the *cafe* package in analysis scripts, two procedures, namely, *mmpbsa* and *lie*, can be invoked for the corresponding calculations. The internal work flow is depicted in Figure 1A and B.

3 Application of CaFE

We performed the calculations of binding affinity between avidin (Fig. 1C) and biotin as well as its analogues (Green, 1975) to exemplify the usage of CaFE. Avidin, found in egg white of oviparous vertebrates, is a naturally occurring protein that specifically binds biotin. The avidin-biotin system forms one of the strongest noncovalent association pair in nature and is widely used in immunological and diagnostic assays. Also, it is frequently employed for assessing the performance of theoretical methods for estimating binding affinity (Hou *et al.*, 2011a).

3.1 MM/PBSA

MM/PBSA is a method combining molecular mechanics and continuum solvents (Kollman *et al.*, 2000). Hypothetically, a receptor-ligand binding event is split into two stages: association in the gas phase and solvation in the aqueous phase. While the single point gas-phase energy is evaluated with a classical force field, the solvation term is quantified by using implicit solvent models, which is the core of this method.

Step 1: Conformational sampling. The end-point methods are actually post-processing methods. Therefore, prior to calculations, trajectory files are usually generated by common molecular dynamics (MD) simulation packages. For MM/PBSA ligand binding energy calculations, one trajectory can be used for each complex. In this case, parameters of CHARMM and its general force fields were used for the sample calculations.

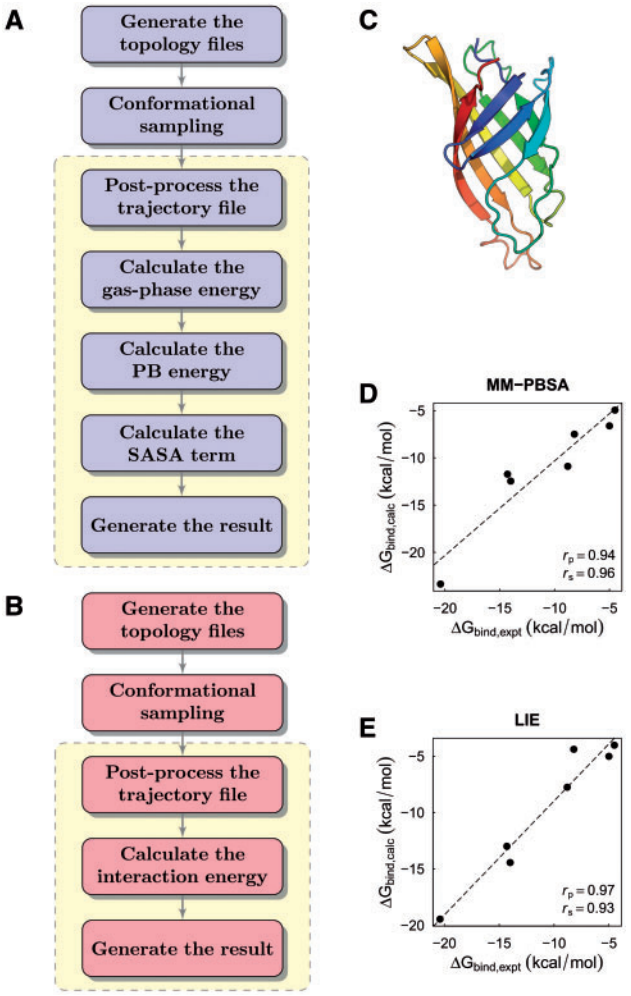

Fig. 1 (**A**) Work flow for the typical MM/PBSA and (**B**) LIE calculations. Procedures with yellow background are automated by CaFE. (**C**) Structure of avidin monomer (PDB accession code: 1AVD). The correlations between the experimental binding affinities and the predicted values by the (**D**) MM/PBSA and (**E**) LIE methods (Color version of this figure is available at *Bioinformatics* online.)

Step 2: Post-processing of trajectory. For MM/PBSA, the complex, receptor and ligand conformations of interest are extracted from the produced MD trajectories, according to the one-trajectory protocol (Kollman *et al.*, 2000). Unlike some other tools, additional topology files for neither the receptor nor the ligand is required. Instead, a human-readable selection language, which is used by VMD, is given by the users to select the subgroup of atoms involved in the complex, receptor and ligand. In addition, the intrinsic implicit solvent radii can be assigned without modification of the input file.

Step 3: Energy calculations and output. Three energetic components are calculated in the MM/PBSA method. At first, the gas-phase energy difference between the complex and the separated receptor and ligand is obtained by calling NAMD. Then, the polar solvation free energy is calculated by numerically solving the PB equation implemented in APBS (Baker *et al.*, 2001). Subsequently, the difference of solvent-accessible surface area (SASA) is measured and the nonpolar solvation free energy is estimated by its approximate linear relation with SASA. At last, the binding free energy is summed and averaged throughout an ensemble of conformations as follows:

$$\Delta G_{\text{bind}} = \Delta H - T\Delta S = \langle \Delta E_{\text{gas}} + \Delta G_{\text{sol}}^{\text{polar}} + \Delta G_{\text{sol}}^{\text{nonpolar}} - T\Delta S \rangle \quad (1)$$

It is worth noting that the entropic term is ignorable for a group of structurally similar compounds as in this case. Due to high computational cost and inaccuracy of current methods for entropy calculations, we generally ignored the entropic term in CaFE.

3.2 LIE

LIE is based on the linear response approximation (Åqvist *et al.*, 1994). An external PB solver is not needed, and the preorganization of the ligand is explicitly taken into account. However, it is more expensive in conformational sampling and predefined scaling factors are required.

Step 1: Conformational sampling. Similar to MM/PBSA, trajectories have to be generated before free energy calculations; however, an additional trajectory for the solvated ligand is required for the LIE method.

Step 2: Post-processing of trajectory. Water molecules are not removed as in the MM/PBSA case. Nevertheless, trajectory or coordinate files are converted to the DCD file format which is used for later energy evaluations.

Step 3: Energy calculations and output. For LIE, only two energetic differences between the ligand and its surrounding environments, in both the bound and free aqueous states, are calculated. They are the electrostatic and van der Waals interactions on behalf of the polar and nonpolar binding free energy terms, respectively. Then, the total binding free energy is obtained by

$$\begin{aligned}\Delta G_{\text{bind}} &= \Delta G_{\text{bind}}^{\text{polar}} + \Delta G_{\text{bind}}^{\text{nonpolar}} \\ &= \beta(\langle E_{\text{elec}}^{\text{bound}}\rangle - \langle E_{\text{elec}}^{\text{free}}\rangle) + \alpha(\langle E_{vdW}^{\text{bound}}\rangle - \langle E_{vdW}^{\text{free}}\rangle) + \gamma \qquad (2)\end{aligned}$$

where α, β and γ are predefined coefficients and γ is ignored for cognate ligands.

3.3 Results

The results are shown in Figure 1D and E. Both methods show high correlation coefficients with experimental values in the showcased example, which is in line with our previous evaluations using different force field parameters (Hou *et al.*, 2011a).

4 Conclusion

To facilitate the prediction of binding affinity of protein–protein/ligand, we present an automated tool CaFE, which provides a user-friendly choice for researchers who want to perform a post-MD energetic analysis using the end-point methods.

Funding

This work was funded by the National Science Foundation of China through its grant 21575128.

Conflict of Interest: none declared.

References

Åqvist,J. *et al.* (1994) A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.*, **7**, 385–391.

Baker,N.A. *et al.* (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA*, **98**, 10037–10041.

Banavali,N.K. and Roux,B. (2002) Atomic radii for continuum electrostatics calculations on nucleic acids. *J. Phys. Chem. B*, **106**, 11026–11035.

Case,D.A. *et al.* (2005) The amber biomolecular simulation programs. *J. Comput. Chem.*, **26**, 1668–1688.

Chipot,C. (2014) Frontiers in free-energy calculations of biological systems. *WIREs. Comput. Mol. Sci.*, **4**, 71–89.

Green,N.M. (1975) Avidin. *Adv. Protein Chem.*, **29**, 85–133.

Homeyer,N. and Gohlke,H. (2013) FEW: a workflow tool for free energy calculations of ligand binding. *J. Comput. Chem.*, **34**, 965–973.

Hou,T. *et al.* (2011a) Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model*, **51**, 69–82.

Hou,T. *et al.* (2011b) Assessing the performance of the molecular mechanics/poisson boltzmann surface area and molecular mechanics/generalized born surface area methods. II. The accuracy of ranking poses generated from docking. *J. Comput. Chem*, **32**, 866–877.

Humphrey,W. *et al.* (1996) VMD: visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38.

Kollman,P.A. *et al.* (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.*, **33**, 889–897.

Kumari,R. *et al.* (2014) G_mmpbsa—a GROMACS tool for high-throughput MM-PBSA calculations. *J. Chem. Inf. Model.*, **54**, 1951–1962.

Lill,M.A. and Danielson,M.L. (2011) Computer-aided drug design platform using pymol. *J. Comput. Aided Mol. Des.*, **25**, 13–19.

Miller,B.R. *et al.* (2012) MMPBSA.py: an efficient program for end-state free energy calculations. *J. Chem. Theory Comput.*, **8**, 3314–3321.

Nina,M. *et al.* (1997) Atomic radii for continuum electrostatics calculations based on molecular dynamics free energy simulations. *J. Phys. Chem. B*, **101**, 5239–5248.

Phillips,J.C. *et al.* (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, **26**, 1781–1802.

Sitkoff,D. *et al.* (1994) Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.*, **98**, 1978–1988.