Biochemical and Biophysical Research Communications 483 (2017) 203–208

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Computing the binding affinity of a ligand buried deep inside a protein with the hybrid steered molecular dynamics

Oscar D. Villarreal [a, 1], Lili Yu [a, b, 1], Roberto A. Rodriguez [a], Liao Y. Chen [a, *]

[a] Department of Physics, University of Texas at San Antonio, San Antonio, TX 78249, USA
[b] Department of Laboratory Medicine, Yancheng Vocational Institute of Health Sciences, Yancheng, Jiangsu 224006, PR China

ARTICLE INFO

Article history:
Received 15 December 2016
Accepted 24 December 2016
Available online 26 December 2016

Keywords:
Binding affinity
Retinol-binding protein
Ligand-protein interaction
T4-lysozyme mutants
Molecular dynamics

ABSTRACT

Computing the ligand-protein binding affinity (or the Gibbs free energy) with chemical accuracy has long been a challenge for which many methods/approaches have been developed and refined with various successful applications. False positives and, even more harmful, false negatives have been and still are a common occurrence in practical applications. Inevitable in all approaches are the errors in the force field parameters we obtain from quantum mechanical computation and/or empirical fittings for the intra- and inter-molecular interactions. These errors propagate to the final results of the computed binding affinities even if we were able to perfectly implement the statistical mechanics of all the processes relevant to a given problem. And they are actually amplified to various degrees even in the mature, sophisticated computational approaches. In particular, the free energy perturbation (alchemical) approaches amplify the errors in the force field parameters because they rely on extracting the small differences between similarly large numbers. In this paper, we develop a hybrid steered molecular dynamics (hSMD) approach to the difficult binding problems of a ligand buried deep inside a protein. Sampling the transition along a physical (not alchemical) dissociation path of opening up the binding cavity—pulling out the ligand—closing back the cavity, we can avoid the problem of error amplifications by not relying on small differences between similar numbers. We tested this new form of hSMD on retinol inside cellular retinol-binding protein 1 and three cases of a ligand (a benzylacetate, a 2-nitrothiophene, and a benzene) inside a T4 lysozyme L99A/M102Q(H) double mutant. In all cases, we obtained binding free energies in close agreement with the experimentally measured values. This indicates that the force field parameters we employed are accurate and that hSMD (a brute force, unsophisticated approach) is free from the problem of error amplification suffered by many sophisticated approaches in the literature.

1. Introduction

Accurately computing the free-energy of binding a ligand to a protein is a task of essential importance in biochemical and biophysical studies that still presents us considerable difficulty to overcome [1–19]. For example, the free energy perturbation (FEP) alchemical methods have been widely applied with many successes but they have also been found to have their share of producing false positives/negatives [20]. Examining the sources of the computation errors, the first is the errors in the force field parameters used for intra- and inter-molecular interactions, which can be regarded as intrinsic or systematic errors. Improvements of the force fields led to higher accuracy nearing the range of chemical accuracy in recent years. The second source of errors is extrinsic, which can simply be an amplification of the intrinsic errors. Errors also arise due to insufficient sampling of the relevant events for a given process. These extrinsic errors are dependent upon the computational approach used in a given study. For example, in an alchemical approach that involves thermodynamic (alchemical) cycles illustrated in Fig. 1, the absolute binding energy is computed as the difference $\Delta G_{binding} = \Delta G^{apo}_{L=>0} - \Delta G^{holo}_{L=>0}$, between the free energy of annihilating the ligand in the holo-protein state $\Delta G^{holo}_{L=>0}$, and the same in the apo-protein state, $\Delta G^{apo}_{L=>0}$. If the two free energies of annihilation are large and similar, computing the free energy of binding is to extract the small difference between two large numbers, which typically amplifies the intrinsic errors. Hypothetically, if the free energy of annihilation in the apo state is

* Corresponding author.
E-mail address: Liao.Chen@utsa.edu (L.Y. Chen).
[1] Equal contributions.

http://dx.doi.org/10.1016/j.bbrc.2016.12.165

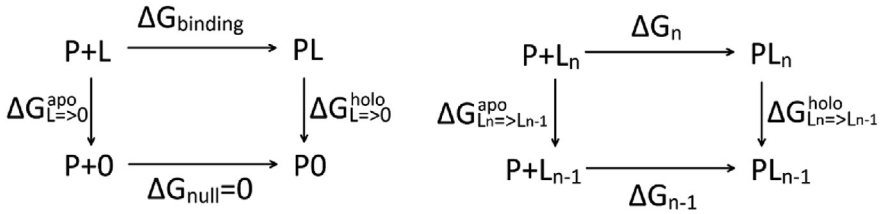

Fig. 1. Alchemical route/thermodynamic cycles for binding free energy. In the left panel, the ligand L is totally annihilated in the apo and the holo states respectively. In the right panel, only one fragment of the ligand is annihilated in the apo and the holo states respectively.

200 ± 10 kcal/mol and the free energy of annihilation in the holo state is 190 ± 10 kcal/mol, both with errors around 5%, then the computational result of the absolute binding energy is −10 ± 14 kcal/mol with an uncertainty of 140%. It should be noted that other methods including docking and MM/GBSA/PBSA can have similar problems of error amplification when they involve small differences between large numbers.

In the current literature, many efforts have been put forward to improve the accuracy of the force fields and to eliminate the extrinsic errors. In particular, various delicate approaches have been devised and refined to avoid the afore-illustrated amplification of errors, *e.g.,* to compute the absolute binding energy as the sum of multiple relative binding energies (Fig. 1, right panel). For each of the N steps of annihilating a ligand, L, by part, $L = L_N \Rightarrow L_{N-1} \Rightarrow \cdots L_n \Rightarrow L_{n-1} \Rightarrow \cdots L_1 \Rightarrow 0$, we can compute the relative binding energy and then assemble the relative binding free energies together for the absolute binding free energy of the entire ligand, L, $\Delta G_{binding} = \sum_n (\Delta G^{apo}_{L_n => L_{n-1}} - \Delta G^{holo}_{L_n => L_{n-1}})$. This assembly of a series of relative binding energies will be as accurate as the force fields we use if the major contributors to the sum are not small differences between similarly large numbers. Depending on how we choose the intermediates, L_n, we might be able to avoid the problem of error amplifications.

Approaches without invoking alchemical cycles can be free from the error amplification inherent in the process of extracting small differences between similarly large numbers, in which the potential of mean force (PMF) [21–25] is computed along a physical (not alchemical) dissociation path of the ligand leading from its binding site to a place far away from the protein. The PMF difference between the ends of the path (holo and apo states) is obtained not by subtracting the apo state PMF from the holo state PMF but instead by accumulating small PMF differences along the dissociation path. Each small PMF difference is computed not by subtraction but by conducting the statistical average of similar numbers from the equilibrium samplings with designed biases and constraints or from the nonequilibrium samplings with steered molecular dynamics (SMD). (Note that the brute force SMD has not been used reliably for free-energy calculations with quantitative accuracy without the specially designed correction factors [16–19,26,27]. The hybrid SMD (hSMD) method [28,29], also brute force in nature, has been shown to produce accurate results.) These PMF-based approaches, delicate equilibrium or brute force nonequilibrium, have proven to be effective in cases where a small molecule (or protein) adheres onto the surface of a protein or resides in an open binding site and, therefore, can be removed from the protein along an unhindered path [12,15,16,28–37]. However, are they applicable to the cases where a ligand is completely buried in a deep binding site such as the complex of retinol (RTL) bound inside the human cellular retinol-binding protein 1 (CRBP1) [38] illustrated in Fig. 2?

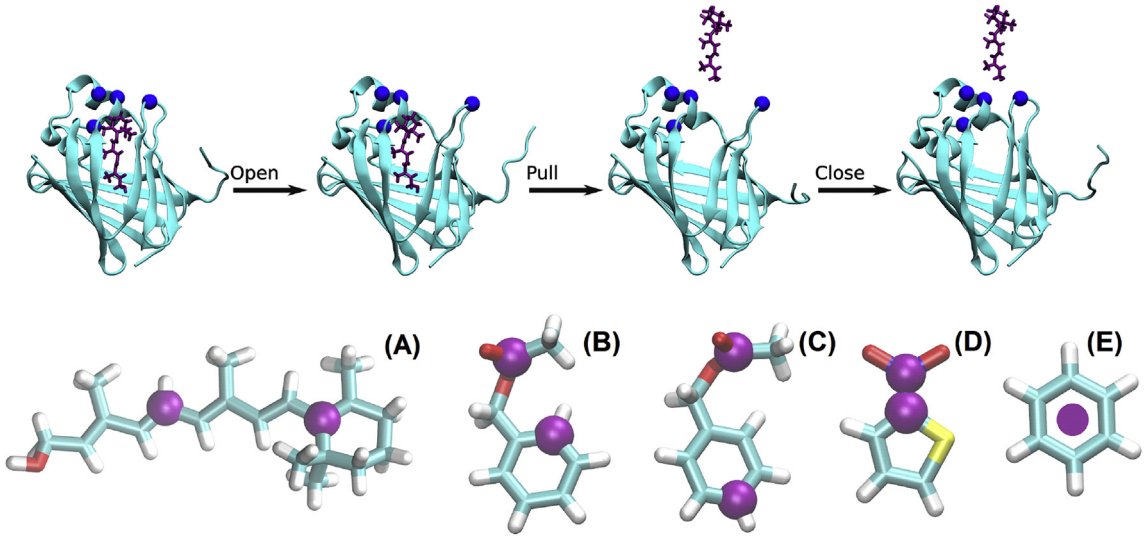

Fig. 2. An example of the open-pull-close pulling path and the pulling centers (purple spheres/disks) on the ligands. In the top panel, the protein (CRBP1) is shown as cyan ribbons and the ligand (RTL) in purple licorice. The blues spheres indicate the four pulling centers on CRBP1. In the bottom panel, the ligands are represented with licorices (colored by atom names: carbon, cyan; hydrogen, white; nitrogen, blue; oxygen, red; sulfur, yellow). (A) Retinol (RTL). The pulling centers on this ligand are atoms C6 and C11. (B) and (C) Benzylacetate (J0Z). The pulling centers on J0Z are, respectively, atoms CAJ and CAG in simulation Set I and atoms CAJ and CAC in Set II. (D) 2-Nitrothiophene (265). The pulling centers on this ligand are atoms CAG and NAH. (E) Benzene (BNZ). The pulling center is the center of mass of this ligand. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)