User guide for AFFPEL.py – v1.0

1. Introduction

The Automated Protein-Protein Free Energy tooL (APPFEL.py) is an automated tool designed to computationally determine the affinity between two polypeptide chains. Examples of this type of system are the complex between two large proteins, or a protein-peptide complex. Starting only from the coordinates of the bound system, APPFEL performs all the necessary steps in the absolute binding free energy (ABFE) calculation: assigning the needed parameters, building and equilibrating the simulation boxes, and performing/analyzing each of the the free energy components. The associated Molecular Dynamics (MD) simulations are performed using the NAMD software, which combines high performance with a set of collective variables that is suitable for large molecules. For ABFE calculations on smaller systems, such as protein-ligand or host-guest complexes, the user is invited to try APPFEL's cousin programs BAT.py and GHOAT.py, which are freely available at https://github.com/GHeinzelmann/BAT.py and https://github.com/Gheinzelmann/GHOAT.py.

In this user guide we will first describe the theory and the methods behind the calculations, in which the binding free energy is determined by pulling the two molecules apart in the presence of restraints. We then go through the practical aspects of the program, explaining how the equilibration and free energy stages are carried out, and detailing each of the parameters to be used in the APPFEL.py input file. Finally, we show how to add a new system to the APPFEL workflow, allowing the calculations to be extended to several other systems with minimal effort.

2. Theory and methods

2.1 Absolute binding free energy

We can relate the value of the dissociation constant K_d, between a protein receptor and a single ligand, to their absolute (or standard) binding free energy ΔG^o_b:

$$\Delta G^o_b = RT \ln\left(\frac{K_d}{C^o}\right) \tag{1}$$

where R is the gas constant and C^o is the standard concentration of 1 M. In the APPFEL program, the calculation of ΔG^o_b is done through a series of MD simulations along an artificial path that connects the bound and unbound states (Fig. 1). This path starts with the application of a set of restraints to the two bound molecules, followed by separating them along a physical path, and then removing the applied restraints. By calculating the free energy variation at every step, we can obtain the value of ΔG^o_b that is valid for the spontaneous process as well (Eq. 1), since G is a state function and thus is path-independent.

Following the cycle from Fig. 1, the value of the calculated binding free energy will be written as a sum of seven components:

$$-\Delta G^o_{bind} = \Delta G_{r,cf,att} + \Delta G_{l,cf,att} + \Delta G_{l,TR,att} + \Delta G_{pull} + \Delta G_{l,TR,rel} + \Delta G_{l,cf,rel} + \Delta G_{r,cf,rel} \tag{2}$$

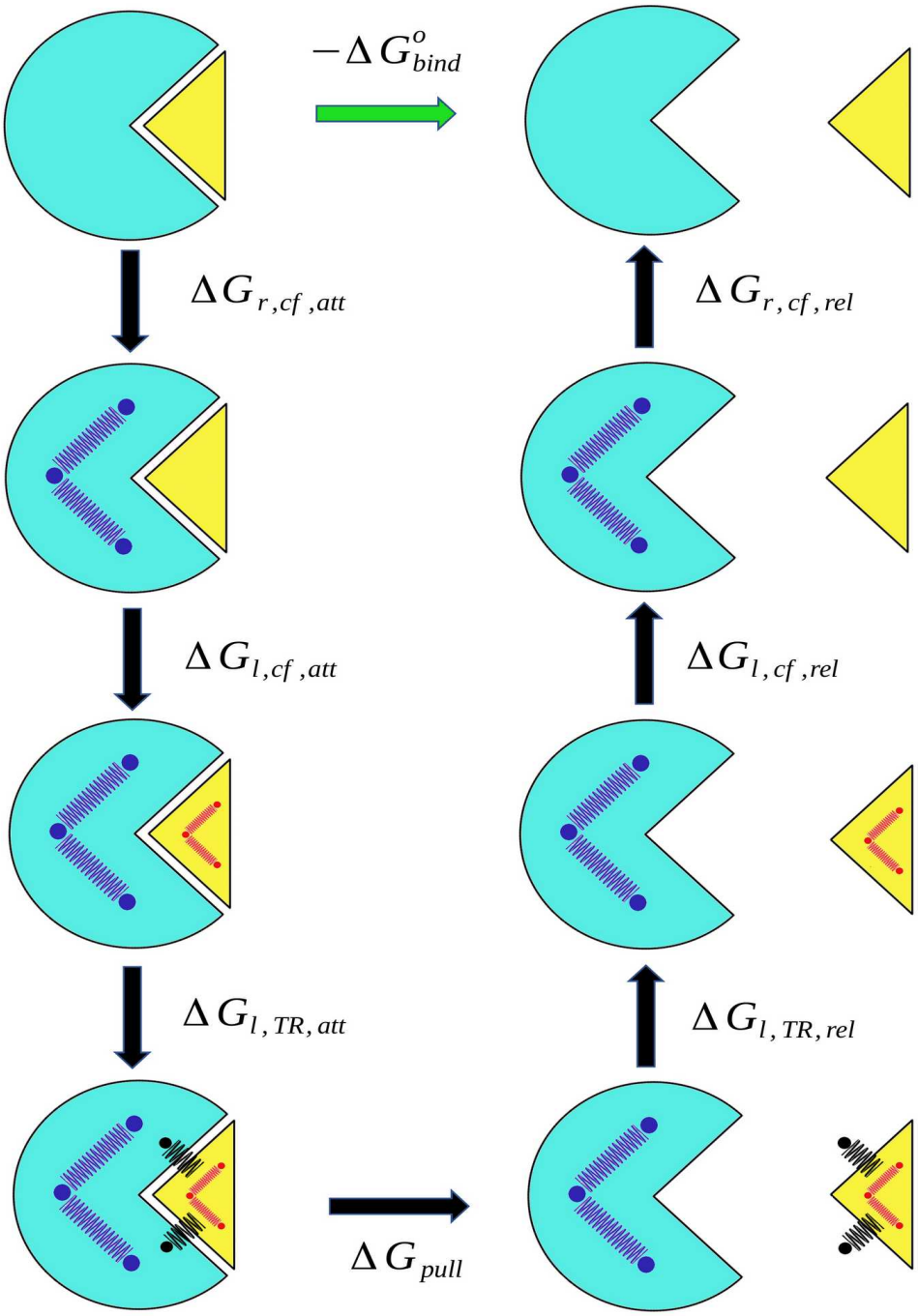

Figure 1: Thermodynamic cycle showing all the steps in the binding free energy calculation between the receptor (blue) and the ligand (yellow). The conformational (*cf*) restraints applied to the receptor and the ligand are shown as the blue and red springs, respectively, and the black springs denote the ligand translational/rotational (*TR*) restraints.

The first three terms on the right side of Eq. 2 are the free energy contributions of attaching (index *att*) restraints to the receptor (index *r*) and the ligand (index *l*), when the system is in the bound state. The nature of the restraints can be either conformational (index *cf*), or translational/rotational (index *TR*), with the former restricting the internal degrees of freedom of the molecule, and the latter used to maintain its position and overall orientation.

The ΔG_{pull} term is the free energy change of bringing the ligand from the receptor binding site to a point in which they do not interact anymore, with all restraints applied to both. Once the two species are separated and each considered free in bulk solvent, the last three free energy terms on the right side of the Eq. 2 are calculated (index *rel*), by releasing each of the restraining potentials used in the pulling step.

2.2 Restraint setup

The restraint setup employed here makes use of the collective variables module from NAMD, which can apply harmonic potentials to several groups of atoms during the simulation. As noted in the previous subsection, the restraints applied to the ligand and receptor are divided into conformational (*cf*) and translational/rotational (*TR*) components.

The conformational restraints use the root mean square displacement (RMSD) of a group of *n* atoms throughout the simulation, calculated relative to a reference set of *n* atom coordinates. The restraining potential applied to this RMSD collective variable has the expression:

$$u_c = \frac{k_c}{2n} \sum_{i=1}^{n} (\vec{x}_i - \vec{x}_{0i})^2 \tag{3}$$

with k_c being the chosen force constant, $\vec{x}_i$ the position of atom *i* at a given MD-generated state and $\vec{x}_{0i}$ its position in the reference structure. The $(\vec{x}_i - \vec{x}_{0i})$ distances are computed after the set of coordinates $\vec{x}_i$ has its overall position and orientation aligned relative to $\vec{x}_{0i}$, by first centering their centers of geometry and then applying the rotation that best superimposes the two structures.

The translational/rotational restraints are present in both the receptor and the ligand during the pulling stage. Like the conformational restraints, the TR restraints are applied to a group of atoms and determined relative to a reference structure. Here, the collective variables are the distance between the centers of mass of the MD-generated and reference coordinates, and the relative rotation between the current and reference atom groups. For the receptor, these restraints only maintain the position and orientation of this molecule relative to the simulation box reference frame, and are not computed in the calculated free energies.

For the ligand, the applied TR potentials will have the expressions:

$$u_t = \frac{k_t}{2}(\vec{r} - \vec{r}_0) = \frac{k_t}{2}[x^2 + y^2 + (z - z_0)^2] \tag{4}$$

$$u_o = \frac{k_o}{2}\Omega^2 \tag{5}$$

Eq. 4 shows the translational component, with $\vec{r}$ being the current position of the chosen ligand atoms center of mass, $\vec{r}_0$ the reference position, and k_t the translational spring constant (Fig. 2). We place the origin so that $\vec{r}_0 = (0, 0, z_0)$, and thus we can write this equation in terms of the *x*, *y*,

and *z* coordinates, as well as the chosen value of z_0. Eq. 5 corresponds to the rotational component, with k_o as the rotational spring constant and $\Omega = \cos^{-1}(\vec{q}\cdot\vec{q}_r)$. The vectors **q** and $\mathbf{q}_r$ are the MD-generated and the reference quaternions, respectively, each made of four components $\mathbf{q} = (q_0, q_1, q_2, q_3)$. They represent rotations of a rigid body in three dimensions, and are an elegant alternative to the Euler rotation angles.

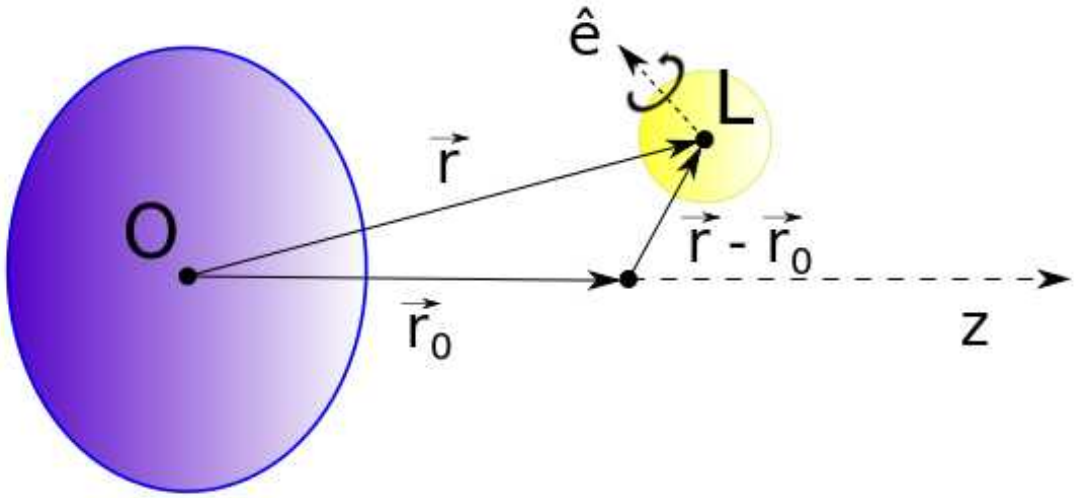

Figure 2: Scheme showing the applied restraints during the binding free energy calculations, with the receptor in blue and the ligand in yellow. Points O and L represent the origin and the center of mass of the chosen ligand atoms, respectively. The **r** vector is the position of the ligand atoms center of mass relative to the origin, and $\mathbf{r}_0$ vector the reference position. The ê unit vector is the Euler axis of rotation, according to the quaternion representation.

To obtain the free energy contributions related to the application or removal of the conformational and TR restraints $\Delta G_{r,cf,att}$, $\Delta G_{l,cf,att}$, $\Delta G_{l,TR,att}$, $\Delta G_{l,cf,rel}$ and $\Delta G_{r,cf,rel}$, a set of simulation windows with intermediate values of the associated force constants is used, ranging between 0 and the final chosen value (k_c, k_t or k_o). The energy output of these windows are then combined using the Multistate Bennett Acceptance Ratio (MBAR), providing the free energy difference of the process. The exception is the $\Delta G_{l,TR,rel}$ term, which is computed analytically according to the expressions:

$$\Delta G_{l,TR,rel} = k_B T \ln\left[C^o\left(\frac{2\pi k_B T}{k_t}\right)^{3/2} + \frac{1}{8\pi^2}\left(\frac{8\pi k_B T}{k_o}\right)^{3/2}\right] , \qquad (6)$$

where k_B is the Boltzmann constant and *T* is the temperature. A detailed demonstration of Eq. 6 can be found in the Appendix.

2.3 Pulling stage

The pulling stage is performed with all restraints attached, as demonstrated in Fig. 1, by varying only the value of z_0 from Eq. 4, between the bound state and a state in which the receptor and the ligand are far away from each other (Figs. 1 and 2). The free energy of this process is calculated by using a series of windows with different values of z_0, and using MBAR to combine the data from the windows and extract the free energy of the process.

The initial states for each of the pulling windows are prepared through an steered molecular dynamics process, which the value of z0 is slowly increased from its initial to the final value of the pulling distance.