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

The spontaneous process of ligand binding/unbinding to a protein receptor involves a large configurational space, which generally cannot be sufficiently sampled when employing regular MD simulations. To overcome this limitation, we can relate the value of the dissociation constant between the two species (K_d) to their absolute binding free energy ΔG^o_b:

$$\Delta G_b^o = RT \ln\left(\frac{K_d}{C_o}\right) \quad (1)$$

where R is the gas constant and C^o is the standard concentration of 1 M. The value of ΔG^o_b can be obtained using MD simulations, by creating a path to transfer the ligand from the protein binding site to bulk solvent in the presence of restraints. By computing the free energy of this process, and the free energies of applying and removing these restraints on both ends, we create an alternative route to calculate ΔG^o_b using a thermodynamic cycle (Fig. 1).

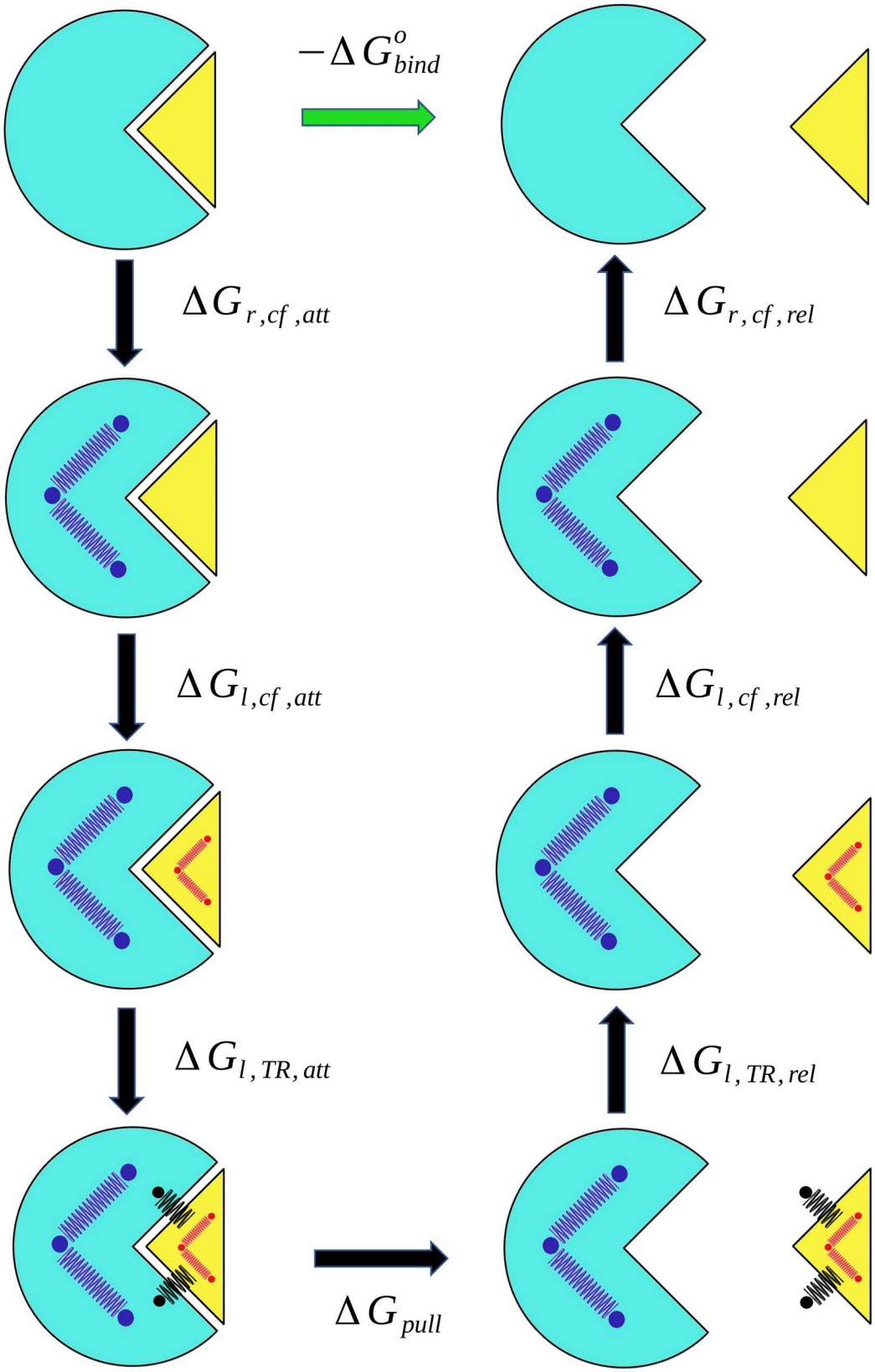

Figure 1: Thermodynamic cycle showing all the steps in the binding free energy calculation between the receptor (blue) and the ligand (yellow). The conformational restraints applied to the receptor and the ligand are shown as the blue and red springs, respectively, and the black springs denote the ligand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

The value of the binding free energy is, therefore, written as a sum of seven components:

$$-\Delta G^{o}_{bind} = \Delta G_{r,cf,att} + \Delta G_{l,cf,att} + \Delta G_{l,TR,att} + \Delta G_{pull} + \Delta G_{l,TR,rel} + \Delta G_{l,cf,rel} + \Delta G_{r,cf,rel} \quad (2)$$

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. We consider the receptor as the polypeptide chain that is kept fixed in the laboratory frame, and the ligand as the polypeptide that is pulled from the receptor towards bulk solvent. The nature of the applied 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 orientation. ΔG_{pull} is the free energy change of bringing the ligand from the receptor binding site to a point in bulk solvent, with all restraints applied to both. Once the two species are separated and considered free in bulk solvent, the last three free energy terms on the right side of the Eq. 2 are calculated, by releasing (index *rel*) each of the restraining potentials used in the pulling step.

2.2 Restraint setup

The restraint setup employed here employs the collective variables module from NAMD, which applies 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:

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

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. As shown in Fig. 1, the conformational restraints are attached to the receptor first, and then to the ligand, yielding the $\Delta G_{r,cf,att}$ and the $\Delta G_{l,cf,att}$ free energy terms. Their release is performed with the two molecules in separate boxes, since at this point they no longer interact with each other, giving the $\Delta G_{l,cf,rel}$ and $\Delta G_{r,cf,rel}$ free energies.

To obtain the free energies associated with the attachment/release of the conformational restraints, a set of simulation windows with intermediate values of the k_c force constant is used, between 0 and the final chosen value. The energy output of these windows are then combined using the Multistate Bennett Acceptance Ratio (MBAR), thus providing the free energy of the process.