

Computing absolute binding affinities by Streamlined Alchemical Free Energy Perturbation

Mina Ebrahimi^{1,2,3}, Grace Brannigan⁴, Jérôme Hénin^{2,3}

¹Department of Physical Chemistry, School of Chemistry, College of Science, University of Tehran, Tehran, Iran.; ²CNRS, Université de Paris, UPR 9080, Laboratoire de Biochimie Théorique, 13 rue Pierre et Marie Curie, 75005, Paris, France; ³Institut de Biologie Physico-Chimique – Fondation Edmond de Rothschild, PSL Research University, Paris, France; ⁴Rutgers University, Camden, New Jersey, 08102

This version dated
December 20, 2021

Abstract This tutorial describes the double decoupling approach using alchemical free energy perturbation (AFEP) to reproduce the experimental ligand affinity. A wide range of in silico ligand affinity predictions are performed using the AFEP method, this tutorial is built around the Streamlined AFEP (SAFEP) approach, which uses distance-from-bound-configuration (DBC) restraints in the Colvars Module. DBC restraints limit the ligand's conformational changes as well as rototranslational movement relative to the protein to their unbiased fluctuations in the bound state. In the following, the binding free energy of phenol to a mutant lysozyme (L99A/M102H) is estimated using this innovative sampling framework.

*For correspondence:
henin@ibpc.fr (JH)

1 Introduction

2 Scope

This tutorial covers the calculation of absolute binding affinity through computational alchemy, traditionally known as Free Energy Perturbation (FEP). Here, we use a simple and site-centered approach, which relies on a single recently-introduced collective variable to define the binding site and achieve robust convergence. This flavor of FEP is known as Streamlined Alchemical Free Energy Perturbation (SAFEP) [1].

3 Prerequisites

3.1 Background knowledge

It is assumed that users are familiar with standard molecular dynamics simulation using NAMD package. Otherwise, please explore these two tutorials: "NAMD Tutorial" [2] and "In silico alchemy: A tutorial for alchemical free-energy perturbation calculations with NAMD" [3]. Also, basic knowledge of using VMD is necessary for visualization, input preparation, and post-processing. While running the steps in the "Cheat Sheet" requires only superficial understanding, adapting these steps to your own system will require more familiarity with TCL. Furthermore, output analysis concerning with thermodynamic integration method is proceeded by Grace (xmGrace), but many other tools (Python, Matlab, R, Mathematica) can be used alternatively.

3.2 Software requirements

NAMD 2.14 or later [4] will be used to perform simulations using both CPU or GPU. Note that alchemical transformations are only implemented in the CPU versions of NAMD 2, but NAMD 3.0 introduces alchemy on the GPU [5]. In order to monitor, analyze, and visualize the simulations and AFEP outputs, VMD 1.9.4 or later [6] is needed. In the "cheat sheet", the output data post-processing is conducted by a Jupyter notebook (ipynb file) requiring Python 3.0 and the alchemlyb library. In summary, the minimal required software is:

- NAMD 2.14 or later[4]
- VMD 1.9.4 or later [6]
- Python 3.0
- Alchemlyb library [7, 8]

4 Content and Links

4.1 Cheat Sheet

Follow the steps below for a basic calculation of the binding affinity of phenol to lysozyme. More detail on the rationale for each step, as well as alternative options, can be found in the hyperlinked sections. All supporting files are provided in the [SAFEP_tutorial repository](#). In commands of this material, replace the `user_dir` with the directory that you unzip the Supp-Files contents.

The steps below assume that you already have a solvated system prepared for basic MD simulation. We provide these files for the tutorial, and details of how they were prepared are in [System Preparation](#).

1. Defining the occupied state

- (a) Load the "solv-prot.psf" and "solv-prot.pdb" into vmd.
- (b) Open the Colvars Dashboard through VMD main menu→Extensions→Analysis→Colvars Dashboard, then click "Edit" to edit a new collective variable. Load the "DBC ligand RMSD" template from the left panel drop-down menu: "Templates/colvar templates".
- (c) Define the atom selection for the ligand atoms: Delete the "atomNumbers 1 2 3 4" line following the comment "#Define ligand atoms". In its place, insert the atom numbers of all heavy atoms of the ligand by entering "resname PHEN and noh" into the text box "Editing helpers/atoms from selection text" in the left panel, then press Enter to insert the new selection into the configuration text.
- (d) Similarly, define the binding site atoms: Replace the "atomNumbers" line within the "fittingGroup" block with the atom numbers corresponding to the selection text "alpha and within 14 of resname PHEN".

- (e) Create the file containing the reference coordinates for the ligand and binding site (rest-ref.pdb, which is distinct from solv-prot.pdb), by entering the following lines in VMD TkConsole:

```
[atomselect top all] set occupancy 0
[atomselect top "(alpha within 14 of resname PHEN)
or (resname PHEN and noh)"] set occupancy 1
[atomselect top all] writepdb rest-ref.pdb
```

- (f) Edit the "DBC ligand RMSD" template within the Colvars dashboard editor to point to the reference pdb file that you wrote in the previous step, by changing the default line

```
refpositionsfile reference.pdb
```

to

```
refpositionsfile rest-ref.pdb
```

Note that you need to do this for both instances of the keyword (one in the atom group, and one in the RMSD block) and press "Apply[Ctrl-s]". In the Colvar dashboard main menu, click on "Save colvars" at the top, change to the directory and save the "DBC ligand RMSD" as the "DBC-unbiased.colvars". Optional: You can set the output files frequencies by using "colvarsTrajFrequency" and "colvarsRestartFrequency" keywords in the "DBC ligand RMSD" input file.

- (g) Open the "DBC-unbiased.colvars" in a text editor, add a separate block for the histogram feature, with the following content:

```
histogram {
    colvars DBC
    outputFile DBC.dat
    histogramGrid {
        widths          0.1
        lowerboundaries 0.0
        upperboundaries 10.0
    }
}
```

- (h) Run unbiased, traditional MD of the occupied state using the configuration files provided. Look out! The unbiased simulation is a long equilibration (at least 50 ns in this material) for exploratory analysis, predicting the tolerance of the DBC ligand for all bound modes. The short simulation might lead to a false evaluation.

```
user_dir/Supp-Files/DBC-Unbiased/equ/namd2
equ.namd
```

Note: the equilibration included in the unbiased simulation is specifically necessary for this material and

can be skipped for other examples. The rationale is provided in section 6.4.

```
user_dir/Supp-Files/DBC-Unbiased/namd2
    run.namd
```

- (i) The output of the histogram will be in the DBC.dat file. View the histogram using your own preferred tool, or run

```
user_dir/Supp-Files/DBC-Unbiased/outputs/python
    1-Plot-histogram.py
```

Bear in mind that Colvar-grid.py will be called as a class by "1-Plot-histogram.py" script.

- (j) From the histogram, determine the DBC tolerance such that it is larger than at least 95% of the RMSD values (95 percentile or greater).
- (k) From the [Colvars Reference manual](#), add "Harmonic wall restraints" into the "DBC ligand RMSD" input file and modify the `upperWalls` parameter corresponds to the DBC tolerance. In the `harmonicWalls` block, set the `colvars` keyword the DBC, the name of "DBC ligand RMSD" colvar, and choose an appropriate value for the `forceConstant` keyword. Please, find the example in section 7.1.

If it is necessary to change atom group definitions after running the equilibrium MD, you can also analyze through post-processing in Section 6.3.

2. Protein Decoupling

- (a) The atoms which will be decoupled are indicated via the solv-prot-charmm.fep file. These are most conveniently written using VMD. Reload the same solv-prot.psf and solv-prot.pdb (or VMD session) from step 1, and then open the Tkconsole:

```
[atomselect top all] set beta 0
[atomselect top "resname PHEN"] set beta -1
[atomselect top all] writepdb solv-prot-charmm.fep
```

- (b) Run the protein decoupling step. The NAMD configuration file will use the colvar inputs that you created in Step 1.

```
user_dir/Supp-Files/AFEP-Bound-Decoupling/equ/
    namd2 equ.namd
```

After equilibration, then perform the production run:

```
user_dir/Supp-Files/AFEP-Bound-Decoupling/namd2
    run.namd
```

- (c) Open the included ipython notebook `BAR_NAMD_alchemlyb.ipynb`, and run each cell in the workbook using the following stepwise instruction. In the directory containing `.fepout`

and `BAR_NAMD_alchemlyb.ipynb` files, open the terminal (in Linux and MacOS) or Anaconda Prompt (in Windows), executing the command `jupyter notebook`. From the directory tree in the jupyter notebook showing up in the web browser, open the `BAR_NAMD_alchemlyb.ipynb` and run all cells from the menu bar through Cell→Run All. Note: Python 3.0 and alchemlyb library are necessary for executing `BAR_NAMD_alchemlyb.ipynb` script.

- (d) Examine the plot titled "Distribution of fwd-bwd discrepancies". In a perfect calculation, this distribution would be symmetric and centered around 0. If your distribution is highly skewed or has a mean that is more than one standard deviation away from 0, please see "Diagnosing and addressing convergence problems". Otherwise, please move on to the next step.
- (e) Extract the value of ΔG_{site}^* from the first output list beneath the fifth cell in the ipython notebook, the ΔG concerned with the last sampling at $\lambda=1$ exhibits the ΔG_{site}^* in units of kcal/mol.

3. DBC Restraint Free Energy Correction

Run a thermodynamic integration calculation on the ligand in vacuum, releasing the DBC restraint within a spherical restraint of volume V_0 .

- (a) Choose the radius R of the spherical restraint. It should be much larger (at least twice as large) as the DBC tolerance you identified in Step 1j. Calculate the volume $V_0 = 4/3\pi R^3$.
- (b) Load the "DBC ligand RMSD" input file into the Colvar dashboard. In the Colvar config editor, insert the "basic colvar" from the drop-down menu: Templates/colvar templates. First, set the "distance" as "name" in the colvar block. Define the atom selection of ligand atoms for the "group1" in the distance block, same as step 1c. Then, in the "group2", replace "atomNumbers 2" with "dummyAtom (-2.168, 5.235, 9.367)", the ligand center position. In the "harmonicWalls" block, change the "DBC" name following the `colvars` keyword into the "distance", and set the "upperWalls" parameter of the spherical restraint to the radius R to retain the spherical restraint. The completed input file, `DBC-Restraint-RFEP.colvars`, is provided in `/Supp-Files/RFEP`. Furthermore, you can find more detail in section 6.3.3.
- (c) Release the DBC restraint gradually using TI, while retaining the spherical restraint. Practically, the DBC restraint is removed slowly by using a changing flat-bottom harmonic restraint in the `DBC-Restraint-`

RFEP.colvars file. To add the changing harmonic restraint in the "DBC ligand RMSD" input file, please see section 8.2.

```
user_dir\Supp-Files\RFEP\namd2 run.namd
```

- (d) Parse the TI output and get the ΔG_{DBC} :

```
grep "Lambda=" rfep.log | awk '{print $5, $7*0.025}' > RFEP.dat
awk -F' ' '{sum+=$2;}END{print -sum;}' RFEP.dat
```

4. Decoupling from solvent

- (a) Equilibrate the ligand in solvent by running

```
user_dir\Supp-Files\AFEP-Hydration\equ\
namd2 equ.namd
```

- (b) Decouple ligand from solvent

```
user_dir\Supp-Files\AFEP-Hydration\namd2
run.namd
```

- (c) Parse the ".fepout" file using the ipython notebook BAR_NAMD_alchemlyb.ipynb, as in Step 2c.

- (d) Run at least four more replicas to get individual values of ΔG_{bulk}^* , and average them together.

5. Calculate corrections and combine quantities

- (a) Calculate the ideal gas contribution for your desired concentration [L]:

$$\Delta G_V[L] = -k_B T \ln V_o[L]$$

where $\Delta G_V^\circ \equiv \Delta G_V[1M] = -k_B T \ln V_o/1660 \text{ \AA}^3$.

- (b) Use Eq. 1 to determine ΔG_{bind}° , where ΔG_{site}^* was calculated in Step 2, ΔG_{DBC} was calculated in Step 3, ΔG_{bulk}^* was calculated in Step 4.

5 Overview

GB: I renamed this section from "Practical considerations: ligand restraints"

5.1 SAFEP

In the conventional DDM, conformational, rotational, and translational degrees of freedom of the ligand relative to the receptor are individually restrained using biasing potentials. All these restraints affect binding, and their contribution must be calculated to predict the absolute binding free energy [9–14].

The SAFEP procedure that we will use here introduces two changes:[1]

- the number of restraints is reduced to two, including the collective *distance from bound configuration* (DBC);
- both restraints are flat-bottom potentials, which makes all free energy contributions but one trivial to calculate.

The DBC coordinate is the root-mean-square deviation (RMSD) of ligand coordinates from a typical bound pose, *in the frame of reference of the binding site*. This single scalar coordinate captures any relative motion of the ligand with respect to the binding site: changes in external coordinates (global translation and rotation) as well as internal coordinates (conformation). See Ref. 1 for details.

As with all double decoupling methods, SAFEP introduces intermediate stages into the binding process. The simplest calculation requires rewriting the binding process as a series of 4 stages, as illustrated by the thermodynamic cycle of Figure (1): **GB: Decision needed: either a) present cycle in a more abstract framework to explain how SAFEP works, and then divide tutorial into how we calculate each step for our system of choice OR b) tightly integrate introduction of cycle with the specific binding event in question.**

1. decouple the ligand from the solution, into a gas phase;
2. confine the ligand, nominally at standard bulk concentration, in an enclosing volume through a center of mass restraint;
3. restrain translational, rotational and conformational ligand to an ensemble of configurations that resembled the bound state, by applying a single DBC (distance-from-bound-configuration) restraint;
4. couple the ligand to the binding site, under a DBC restraint throughout.

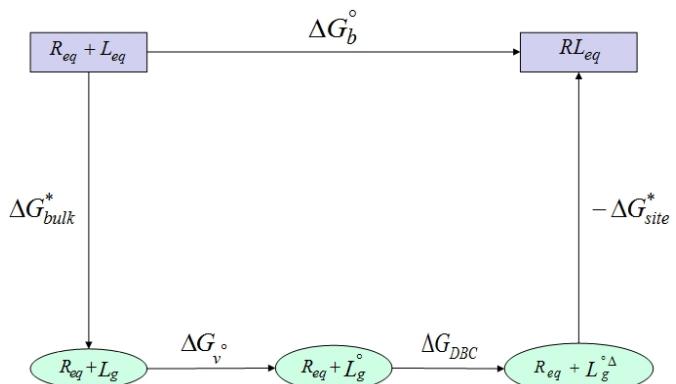


Figure 1. The schematic thermodynamic cycle used in the Streamlined AFEP (SAFEP) approach. ΔG_{bind}° is the standard free energy of ligand binding to the site. ΔG_{bulk}^* is the decoupling free energy of the ligand from the aqueous bulk solvent (minus the excess free energy of solvation). ΔG_V° is the free energy of applying the COM restraint on the ligand, with respect to the standard state in gas phase. ΔG_{DBC} is the free energy of applying the DBC restraint on the ligand. ΔG_{site}^* is the free energy of decoupling the restrained ligand from the binding site.

The sum of the free energies for all stages is the absolute binding affinity of the ligand-protein complex:

$$\Delta G_{bind}^\circ = \Delta G_{bulk}^* + \Delta G_V^\circ + \Delta G_{DBC} - \Delta G_{site}^* \quad (1)$$

Importantly, note the absence of a stage where restraints are removed in the receptor binding site. This is made unnecessary by the use of well-calibrated flat-bottom potentials on the DBC coordinate and the center of mass, which both have zero impact on the coupled, bound state. This is the major practical benefit of SAFEP.

Of the 4 remaining terms in the right-hand side of Eq. (1), two are the results of AFEP decoupling calculations (in bulk water and in the binding site), one (ΔG_V°) has a simple analytical expression ($-RT \ln(V^{\text{Sphere}}/V^\circ)$), and one (ΔG_{DBC}) is estimated by a specific RFEP simulation in vacuum. That is, the process requires only **two condensed-phase simulations**, of which one can be done in a small water box. Importantly, convergence of the AFEP decoupling calculation in the binding site is helped by the DBC restraints, which limits the configurational space that must be sampled.[1]

In practice, we will determine and apply adequate restraints using the Collective Variables Module[15], here in its NAMD and VMD versions. Note that the Gromacs interface could also be used to perform all simulations using Gromacs.

5.2 Tutorial Overview: Application to the mutant lysozyme-phenol complex

In this tutorial we will study binding to lysozyme, which has a buried hydrophobic cavity that can trap various small molecules without large conformational changes [16]. To be precise, we will quantify binding of phenol to the L99A/M102H mutated site introduced by Merski et. al. [17]. While phenol does not bind significantly to the apolar cavity in wild-type or L99 mutant T4 lysozyme[14, 18–20], it does bind the site created by the L99A/M102H double mutation, with an experimental binding affinity $\Delta G_{\text{bind}}^{\circ,\text{exp}}$ of -5.44 kcal/mol [17], to be compared with the result we find for Eq. (1).

6 Preliminary simulation: exploring the bound state

In the alchemical decoupling simulation, we decouple the phenol from lysozyme's binding site to the gas phase while restraining the ligand bound to the bound-state ensemble. In this alchemical transformation, the ligand's non-bonded interactions with the environment are progressively decreased to zero. The flat-bottom DBC restraint should perform two functions:

- limit fluctuations of the decoupled ligand as much as possible;
- perturb fluctuations of the coupled ligand as little as possible.

If we meet the second of these goals, the restraint will leave the bound ensemble unaffected, and therefore have zero free energy contribution to the bound state. In other words, the DBC-restrained bound state should be essentially the same as the unrestrained bound state.

Thus we need a preliminary simulation sampling the bound state to choose the correct DBC boundary value.

Note that such a preliminary simulation is always recommended, whatever the free energy computation scheme, to equilibrate the model of the receptor-ligand complex and characterize it within the chosen modeling parameters (initial structure, force field parameters, solvent model and ionic strength, temperature...). We may find out that the ligand adopts a different binding pose than expected, say, from an experimental structure of the complex, or perhaps binding is unstable and the ligand escapes the site rapidly. This should call into question our modeling assumptions, and that is best done before running an expensive AFEP simulation.

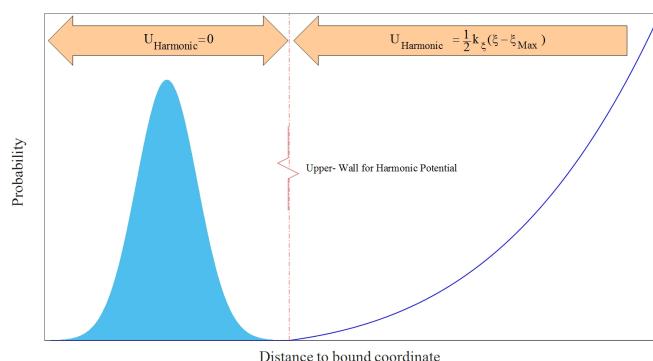


Figure 2. The schematic representation of the half-well harmonic potential mechanism in the DBC restraint, the blue curve indicates the distance evolution of the half-well harmonic potential.

6.1 System preparation

Preparing a protein system for a simulation takes a little work, and is not the focus of this tutorial. Therefore, we provide the necessary files describing the solvated lysozyme-phenol complex. Below, we describe briefly how these files were prepared.

The initial structure of the phenol-mutant lysozyme complex is modeled based on the crystal structure with accession code 4I7L in the PDB. The structure includes two copies of the complex: we retain only chain A. We use two tricks to limit the size of the final simulation box:

1. We truncate the structure's N- and C-terminal segments that are not tightly associated to the rest of the bundle (residues -11 to 1 and 158 to 164);

2. We orient lysozyme along the z axis and solvate it in an elongated water box. Then we need to prevent the protein from rotating during the simulation: we will add a rotational restraint, as explained in section 6.3.4.

We solvate lysozyme in a box of TIP3P water with a physiological NaCl concentration of about 0.15 mol/L. This can be performed directly in VMD using the psfgen, solvate and ionize plugins (see NAMD tutorial for details[2]), or using the online tool CHARMM-GUI[21, 22]. The initial structure of the solvated ligand-protein complex is rendered in Figure (3).

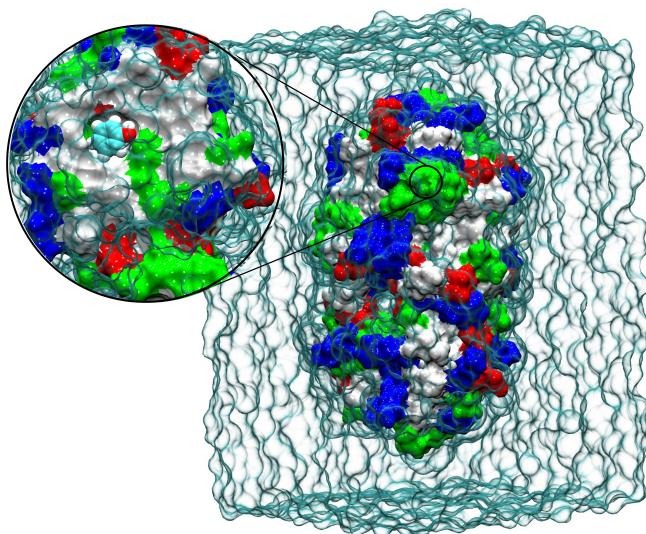


Figure 3. Rendering of the solvated lysozyme-phenol complex, with a zoom in the host-guest site.

6.2 Equilibration

6.2.1 Simulation setup

The scripts for these simulations are provided as examples.

We use NAMD 2.14[4] to minimize, then equilibrate the system at constant temperature of 300 K and constant pressure of 1 bar, controlled by a Langevin thermostat and a Langevin piston algorithm[23]. To prevent anisotropic fluctuations of the periodic cell (preserve its aspect ratio), `useFlexibleCell` is set to no in the pressure control section. The periodic boundary condition (PBC) are applied with particle mesh Ewald (PME) long-range electrostatics[24], with parameters tailored to the size of the simulation box. The timestep is set to 2 fs, with constraints on all bonds to hydrogen atoms (`rigidBonds a11`). The force field is CHARMM36[25], and the water model is TIP3P[26].

The multiple-timestep rRESPA method is used, with fast and slow time steps of 2 and 4 fs, respectively. `WrapAll` should be set to off. `wrapWater` can be set to on optionally. After the simulation, you can wrap water around the protein

in VMD, using the PbcTools package. The center of mass drift of the system is canceled by enabling `zeroMomentum` and retaining `COMmotion` at the default value.

The NAMD configuration file also specifies Colvars parameters. Note that we use "colvar" as shorthand for collective variable, and "Colvars" with a capital "C" for the Collective Variables Module.

Look into `user_dir/Supp-File/AFEP-Bound-Decoupling/equ.namd` configuration file. In the Colvars section, `Colvars` is enabled and its input file is specified:

```
# COLVARS
Colvars          on
ColvarsConfig    DBC-unbiased.colvars
```

This tells the Colvars Module to get its configuration from a file `DBC-unbiased.colvars`. For more information on Colvars configuration, check out the Colvars manual here: <https://colvars.github.io/colvars-refman-namd/colvars-refman-namd.html>.

6.2.2 Definition of the DBC coordinate

During the simulation, lysozyme may move frequently altering the binding site atoms conformations. By enabling the `centerReference` and `rotateReference` in the `rmsd` context, the `fittingGroup` option provides adjusting the phenols' roto-translational movements to the binding site, the moving reference. In this case, the roto-translational and conformational alterations of phenol are adapted to the proteins' alpha carbons within the 14 Å of phenol.

PDB files can be used for two different roles:

1. specify atom groups to use in a collective variable (by setting flags in e.g. the occupancy column);
2. provide reference coordinates when defining RMSD variables, or variables in moving frames of reference that depend on a fit; the DBC coordinate uses both.

```
# DBC colvar
colvar {
  name DBC
  width 0.1
  lowerboundary 0.0
  upperboundary 5.0
  rmsd {
    # Reference coordinates
    # (for ligand RMSD computation)
    refpositionsfile rest-ref.pdb
  }
}
```

In the subsequent, we prolong inspecting the rest of commands in the colvars script.

```
atoms {
```

```
# Define ligand atoms used for RMSD calculation
atomNumbers {11 9 5 1 3 7 12}
```

Use the serial numbers of objective atoms or molecules to tag accurately in atomNumbers sections. As an example, you can use get serial [atomselect top "resname PHEN and noh"] command in VMD TkConsole, defining phenol as ligand atoms;

```
# Moving frame of reference is defined below
centerReference yes
rotateReference yes
fittingGroup {
    # Define binding site atoms used for fitting
    atomNumbers {32 103 1100 1112 1128 1140 1150
    ...
    2448 2472 2486}
}
```

6.3 Defining the Bound State in Post-processing

- Run unbiased, vanilla MD of bound state using the configuration files provided:

```
user_dir\Supp-Files\DBC-Unbiased\equ\
namd2 equ.namd
```

then run:

```
user_dir\Supp-Files\DBC-Unbiased\namd2
run.namd
```

and load the resulting trajectory into VMD.

- Choose the protein atoms whose positions define the protein site. In this case, we'll select all alpha carbons that are contacting the bound ligand.

```
set proteinSel [atomselect top "alpha and
within 14 of resname PHEN"]
```

- Align the trajectory based on these protein atoms, using either the RMSDTT (RMSD Trajectory Tool) GUI in VMD or the command line.
- Choose atoms that define the ligand bound state. In this case we'll use all heavy atoms of the ligand:

```
set ligandSel [atomselect top "resname PHEN and
noh"]
```

- Calculate RMSD for ligand atoms.
- Determine your DBC tolerance: the 95th percentile (the value of the RMSD that is higher than 95% of the equilibrium values) is the lowest DBC tolerance you can use.
- Modify DBC colvar in configuration file to specify (a) protein site atoms (b) ligand site atoms (c) DBC tolerance

6.3.1 Collecting a DBC histogram

We will collect a histogram of the DBC coordinate. Since we do not know beforehand the relevant range of values, we choose a large enough upper boundary value for all colvars. We will use the histograms to predict the optimal value of the upper wall for the DBC and center-of-mass (COM) distance variables.

The histogram block makes the data processing simpler and collects data at every time step, although the histogram could be built based on the colvars trajectory file, or even from a posteriori analysis of DCD trajectories using the Colvars Dashboard in VMD.

```
histogram {
    colvars DBC
    outputFile DBC.dat
    histogramGrid {
        widths          0.1
        lowerboundaries 0.0
        upperboundaries 6.0
    }
}
```

6.3.2 Symmetric DBC

Phenol rotating around C2 imaginary axis, an axis aligned to the aromatic plane and passing through oxygen, changes the position of carbon atoms in the backbone without conformational transformation. The free energy contribution of DBC colvar is invariant concerning atom position changes. Accordingly, the atomPermutation keyword is defined to eliminate the impact of the mentioned permutation. Figure (4) graphically clarifies the atoms permutation in phenol. It is recommended to use the DBC-symmetry Colvar instead of the conventional DBC Colvar for symmetric and pseudo-symmetric ligands.

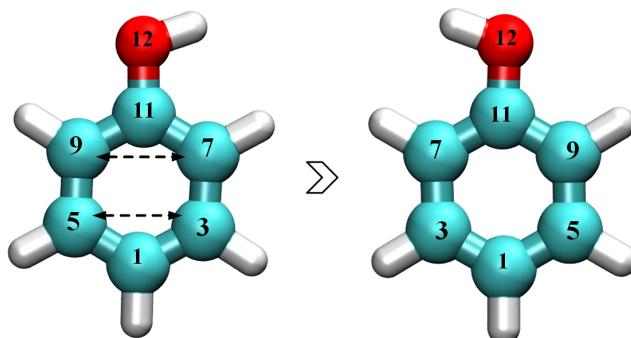


Figure 4. Four carbon atoms may exchange their positions through phenol rotation without any impact on the conformation of phenol.

```
# DBC-symmetry Colvar
```

```

colvar {
    name DBC_sym
    rmsd {
        atomPermutation {11 7 3 1 5 9 12}
        atoms {
            atomNumbers {11 9 5 1 3 7 12}
        }
    }
}
```

6.3.3 Center-of-mass restraint

The distance colvar reflects the phenol center of mass distance with respect to center of the updated binding site in the protein, moving reference. A dummy atom with a given coordinate, center of mass of phenol in the bound-configuration, is used to restrain phenol translational movement in the distance colvar.

```

# Center of mass distance coordinate Colvar
colvar {
    name distance
    distance {
        group2 { # Reference ligand COM position
            dummyAtom (-2.168, 5.235, 9.367)
        # obtained in VMD :
        # > measure center [atomselect top "serial 1 3
        5 7 9 11 12"]
```

6.3.4 Restraint on protein orientation

To keep the water box size at a minimum, lysozyme should be kept aligned along the z axis. By imposing a harmonic force on rotation components, we preserve the initial orientation of lysozyme. Hence, the closestToQuaternion keyword is set to identity rotation. The rotation act is expressed in quaternion, a four-dimensional vector (q_1, q_2, q_3, q_4), furnishing $\sum_i q_i^2 = 1$ [27, 28]. As quaternion components have a magnitude less than 1, so a high enough harmonic bias force constant is required, providing the unit quaternion.

```

colvar {
    name rotation
    orientation {
        # Reference coordinates (for protein)
        refpositionsfile prot-ref.pdb
        atoms {
            # Define protein atoms used for optimal rotation
            atomNumbers {18 32 51 71 86 103 122 146 165 177
            .... 2493}
        }
        # Define reference rotation to null orientation
        closestToQuaternion (1.0, 0.0, 0.0, 0.0)
    }
}
# Harmonic restraint on rotation
harmonic {
```

```

colvars      rotation
forceConstant 10000.0
centers       (1.0, 0.0, 0.0, 0.0)
}
```

6.4 Production run

After completing the short equilibration, the production simulation is carried out starting from equilibration restart files. The configuration file for the production run is the same as the equilibration, only the set value for `margin` keyword should be removed. The `margin` parameter extends the system in all dimensions avoiding the "patch grid error" at the beginning of the simulation. So, if you did not face this error, you can jump over the equilibration step in the unbiased simulation. Furthermore, the production run is executed long enough to collect an ensemble describing the fluctuations of the bound state.

```
user_dir\Supp-Files\DBC-Unbiased\namd2 run.namd
```

6.5 Result assessment

Inspect the trajectory visually in VMD and examine the binding mode of the ligand, its motion with respect to the protein, and its interactions with protein residues. Plot the histogram for the DBC and distance variables to decide the best maximum value for the flat-bottom harmonic restraint. Using the Colvars Dashboard, plot these variables along the trajectory to interpret each peak in their distribution in terms of phenol's position and interactions with the protein.

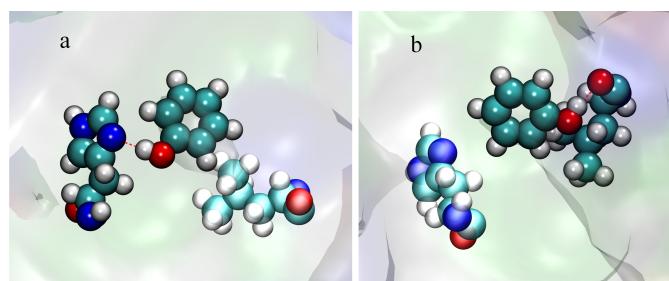


Figure 5. Two detected bound configurations for phenol in the binding pocket. In the snapshots phenol forms a hydrogen-bond with (a) a histidine and (b) a leucine residue.

7 Decoupling the ligand from the binding site

7.1 Simulation under DBC restraint

It may require a long enough equilibration simulation until the Rmsd of lysozyme behaves time independent below 3 Å. Furthermore, the smooth time evolution of DBC colvar trajectories could be another evidence of the equilibration. We

use the `equ.namd` configuration file the same as the `run.namd` file without any modification for this section. In the colvar input file, the measured maximum DBC from the previous subsection (unbiased simulation) should be replaced with the values for `upperboundary` in DBC colvar and histogram blocks. we also utilize this DBC maximum value as the upper wall of the harmonic restraint to implement harmonic bias outside of this range. For applying DBC restraint on the coupled complex, we only added harmonic restraint components to the colvar input script. New keywords in the `DBC-Restraint.colvars` input file are presented in the following:

```
# Flat-bottom harmonic restraint on DBC
harmonicWalls {
    colvars          DBC_sym
    upperWalls      1.5
    forceConstant   200
}
```

7.2 Production run

The fep file format is required to introduce outgoing and incoming atoms for the AFEP evaluation. In this case, the phenol is annihilated in the binding site and the rest of the system is remained without alterations, so beta-column for the phenol should be set to -1 in the fep file format. The `fep.tcl` script enclosed as supplementary Files is required for the alchemical free energy (AFEP) calculations.

To perform the AFEP run, we start from the equilibration restart files with the same conventional MD simulation details discussed before in the unbiased simulation. Now we have a look into the `run.namd` file in the "AFEP-Bound-Decoupling" folder to define sampling strategy for the alchemical simulation.

```
# FEP PARAMETERS
source          ./fep.tcl
alch            on
alchType        FEP
alchFile        solv-prot-charmm.fep
alchCol         B
alchOutFreq    2
alchOutFile    AFEP2-02.fepout
```

The accurate estimation of free energy differences (ΔF) provided by BAR or SOS methods is highly dependent on a sufficiently large number of samples. Hence, the `alchOutFreq` should be assigned small enough to improve the evaluation of the free energy differences by decreasing error bars. It is recommended to define `alchOutFreq` as either 1 or a multiple of `fullElectFreq`. For more information, please check the [Bug advisory and Workaround](#).

```
alchElecLambdaStart 0.5
alchVdwLambdaEnd   1.0
alchVdwShiftCoeff  5.0
```

soft-core potential parameters are utilized to prevent the "end-point catastrophe" by gradual scaling of nonbonded interactions of disappearing atoms with the environment. In this case, the default values fulfill aimed results properly.

```
alchdecouple       on
```

By setting the `alchdecouple` keyword on, in the alchemical transformation, the interatomic nonbonded interactions of changing atoms are not contributed to the free energy estimation.

```
alchEquilSteps 500
set numSteps   50000
set dLambda   0.05
runFEP   0.0 1.0 $dLambda $numSteps true
```

The alchemical FEP sampling is performed by 20 windows with equal intervals i.e. $\Delta\lambda=0.05$, each window is run for 50000 steps interleaved with 500 equilibration steps. The true value at the end of `runFEP` command enables interleaved double-wide sampling (IDWS), as a result of the utilization of this feature, the forward and backward free energy calculation is performed concomitantly. The IDWS provides a fast and an accurate estimation by the frequent interchange of λ between forward and backward values. If the `alchLambdaIDWS` keyword is defined by denoting a given value, the backward free energy will be estimated in the certain range of λ .

7.3 Result assessment

The AFEP simulation output can be analyzed with either the VMD or the alchemlyb library to assess the free energy result. The interleaved double-wide sampling (IDWS) interlaces forward and backward data in the output file, which is extracted by post-processing the `fepout` file. We compressed folders possessing large output files with `tar` command in the subdirectories of `/Supp-Files/` (provided as supplementary files). To follow the tutorial in the Result section, you need extracting tar files, navigate linux or MacOS terminal to the directory of tar files and use the following command:

```
tar -xzvf output.tar.gz
```

In Windows, you can also extract tar files by opening them with WinZip and clicking Unzip.

7.3.1 Ligand binding free energy: ParseFEP plugin in VMD

Before performing the FEP analysis with ParseFEP plugin [29] with VMD, a python script is used to deinterlace the forward

and backward data. Execute the deinterleave-idws.py file by the subsequent command.

```
user_dir/Supp-Files/AFEP-Bound-Decoupling/output/
VMD-ParseFEP/python deinterleave-idws.py
./AFEP2-02.fepout
```

Following the execution of the Python script, two output files, AFEP2-02_bwd.fepout and AFEP2-02_fwd.fepout, are generated in the VMD-ParseFEP folder, representing backward and forward free energy change from sampling of each window. The VMD 1.9.4 is needed to proceed with the free energy estimation, using the ParseFEP plugin [29]. In the main menu of VMD, click the Extensions and from the drop down menu, select Analysis and then Analyze FEP Simulation to open the ParseFEP plugin. In the parameter section of the ParseFEP plugin, set the Gram-Charlier order to zero and enable pictorial representation by checking the disp option. The forward and the backward outfep files are chosen as FEP and FEP (backward) output files and run FEP Parsing, employing the BAR estimator. The ParseFEP graphical interface provides free energy estimation combining forward and backward results with the help of the SOS or the BAR estimator.

7.3.2 Ligand binding free energy: Alchemlyb library

An alternative to the AFEP calculation is the alchemlyb library that recently supports the NAMD's IDWS feature [7, 8], providing the BAR and the MBAR estimators. A jupyter notebook documented by Dr. Grace Brannigan's Lab, /Supp-Files/Jupyter-notebook/BAR_NAMD_alchemlyb.ipynb, conveniently sets up a broad series of built-in functions and tools of the alchemlyb library in a straightforward framework, including parsing the fepout file, AFEP calculation, as well as qualitatively comparing the free energy results in different plot types. Launch a jupyter notebook in the directory of BAR_NAMD_alchemlyb.ipynb file:

```
user_dir/Supp-Files/AFEP-Bound-Decoupling/output/
Alchemlyb/jupyter notebook
```

In the tree view of the jupyter notebook interface, click the BAR_NAMD_alchemlyb.ipynb. In Figure (6), the snapshot of the first cell in the the BAR_NAMD_alchemlyb.ipynb is rendered. Replace the texts in the blue and yellow highlighted boxes with the fepout filename and path. To analyze the fepout file, click the Cell in the top navigation bar of the jupyter notebook and then select the Run All.

For the forward and reverse sampling, the alchemlyb library uses the Bennett acceptance ratio (BAR) estimator to calculate the difference free energy for each window and estimate the overall free energy at $\lambda=1$ by ensemble averaging over all λ -windows. Correspondingly, the evolution of the phenol binding free energy relative to λ is obtained from

```
In [7]: from glob import glob
path = './'
filename = AFEP2-02.fepout
fepoutFiles = glob(path+filename)
temperature = 300
RT = 0.00198720650096 * temperature
print(f'Will process {len(fepoutFiles)} fepout files.')
```

Figure 6. The rendering of the first cell in the BAR_NAMD_alchemlyb.ipynb jupyter notebook.

parsing the AFEP2-02.fepout with the jupyter notebook and illustrated in Figure (7). According to Figure (7) and the screenshot of the printed data in the notebook represented in Figure (8), the phenol decoupling affinity from the binding site is estimated at about (14.65 ± 0.38) kcal/mol.

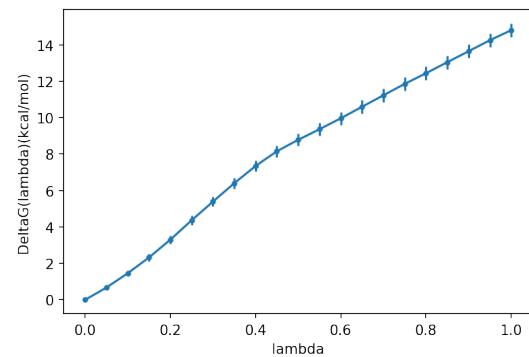


Figure 7. The λ -evolution of the phenol decoupling free energy, from the lysozyme binding site, evaluated by the Alchemlyb library.

```
In [42]: # Extract data for plotting
states = bar.states_
f = bar.delta_f .iloc[0,:] # dataframe
l = np.array([float(s) for s in states]) # lambda midpoints for each window
l_mid = 0.5*(l[1:] + l[:-1])

# FE differences are off diagonal
df = np.array([bar.delta_f .iloc[i, i+1] for i in range(len(states)-1)])
#print(df.cumsum() * RT) #in unit of KT
print("Overall free energy", df.cumsum() * 0.59) #in unit of kcal/mol

# error estimates are too small because we use correlated data
tau = 5e2 # expected correlation length of series

# error estimates are off diagonal
ddf = np.array([bar.delta_f .iloc[i, i+1] for i in range(len(states)-1)]) * np.sqrt(tau)

# Accumulate errors as sum of squares
errors = np.array([np.sqrt((ddf[i]**2).sum()) for i in range(len(states))])
print("Errors", errors)

Overall free energy 0.66442647 1.44231771 2.30032032 3.26590364 4.31841194 5.3336419
0.3336072 7.38561893 8.66824102 8.69183802 9.2664827 9.86107995
19.48001433 11.1899359 11.73835027 12.30762486 12.9992668 13.5046052
14.182725 14.64696263
Errors [0.10463192 0.15648747 0.19370493 0.2197967 0.24404679
0.26418581 0.2842285 0.30285715 0.32128725 0.33510329 0.34170229
0.3472708 0.35241983 0.35675542 0.36137652 0.36592104 0.36971411
0.37312546 0.37655833 0.37915457]
```

Figure 8. The rendering of a patch of the BAR_NAMD_alchemlyb.ipynb notebook. The free energy data with statistical errors for each window is printed in two separate arrays.

The time evolution of phenol decoupling affinity from the binding site for forward sampling and the reverse is well converged in Figure (9). Figure (9), for example, compares the free energy in the first and the last 10% of the sampling, collecting data from the forward pathway and the reverse, and extending the sampled data to the total as the fraction of

time increases. The convergence of the forward and reverse sampling is crucial to optimize the AFEP simulation benefiting from efficient and affordable simulations. The maximum overlap in the free energy estimated from the forward and backward sampling leads to fast convergence. As well as the ParseFEP plugin, the alchemicaly library visualizes the convergence and discrepancy of the forward and backward difference free energy. For decoupling phenol from the bound-state, the forward and backward discrepancy in the binding affinity with respect to λ and its' distribution over the difference free energy are illustrated in Figure (10).

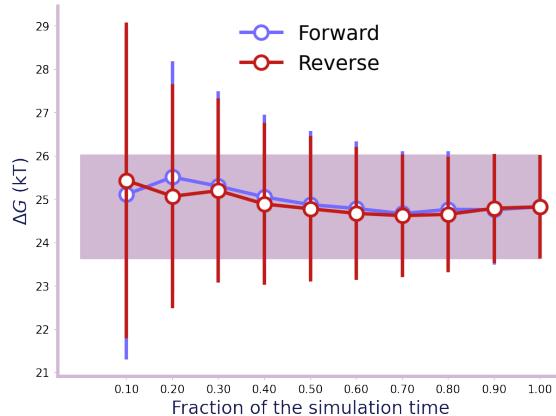


Figure 9. The convergence of the free energy of phenol decoupling from the binding site in the given fractions of simulation time is represented for the forward and backward transformations.

AFEP simulations with appropriate λ intervals, number of sampling, and, in particular, soft-core potential parameters are expected to either exhibit small fwd-bwd affinity discrepancies over the last λs or, in other words, to exhibit a unimodal distribution of fwd-bwd free energy discrepancies around zero.

8 Restraint free energy calculation

Imposing bias potentials on degrees of freedom of the ligand definitely affects the sampled statistical distribution, leading systematic error in the estimation of absolute binding free energy. To remedy this, we calculate restraint energies and consider their contributions to the overall free energy [30, 31].

8.1 Distance from bound-configuration restraint

Although gradual removal of the DBC restraint helps the phenol to retain the bound-conformation during decoupling, the contribution of the DBC restraint must be deducted from the overall free energy estimation. A common approach to handle this issue is establishing the DBC restraint on the phenol

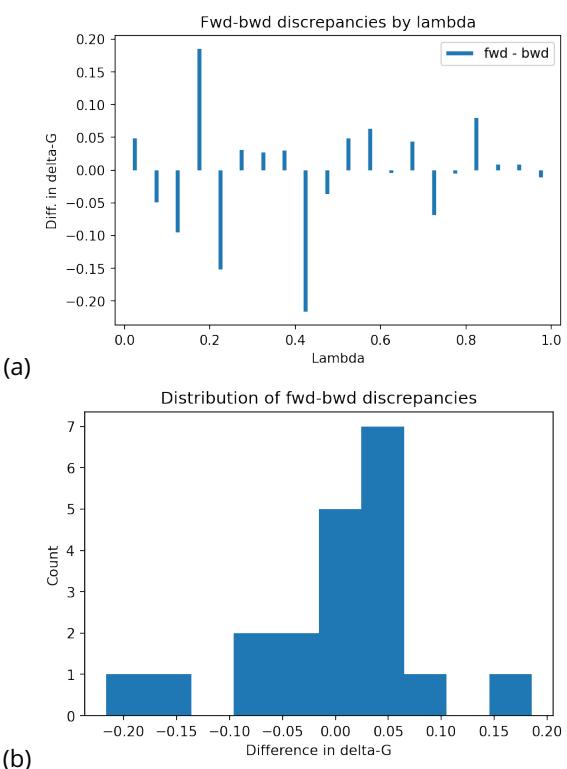


Figure 10. For decoupling phenol from the binding site (a) shows the forward-backward free energy discrepancy relative to λ and (b) represents the distribution of the fwd-bwd binding affinity discrepancy.

progressively, whereas it is confined in an enclosed volume. Employing the harmonic bias with changing force constant in the Colvars module allows us to launch the DBC restraint on the phenol step by step in the RFEP simulation.

8.2 Implementation of the RFEP simulation

The molecular dynamics simulation of the volume enclosed phenol under DBC restraint is practiced through a conventional NVT MD at the constant temperature of about 300 k, controlled by Langevin thermostat. In the configuration file, both the PME and the cell basis vectors as subcommands of the PBC are disabled. The wrapWater and the wrapAll keywords are set to off, followed by disabling the PBC. Here, we investigate the colvars input, DBC-Restraint-RFEP.colvars, and utilized commands for executing RFEP simulation. All keywords used to tailor the phenol roto-translational and conformational changes to the moving reference, including fittingGroup block, are omitted in the colvar context, because the phenol is confined within a discrete volume in the absence of the receptor. In this regard, the centerReference and the rotateReference options are set to off because we do not need defining the optimal ligand roto-translational alignment upon the bound-configuration. To enclose the phenol inside the determined volume, a harmonic bias must

be put on the distance colvar during the whole simulation time.

The RFEP output data files are assessed using the thermodynamic integrating method. We introduce a harmonic bias with the force constant evolving nonlinearly relative to λ , avoiding singularity or integration problems. Thus, it would be sensible to set the value of the `targetForceExponent` higher than 1, chosen by trial and error for this system. Intending to apply the DBC restraint progressively, the force constant of the bias potential starts from zero and increases as a function of the coupling parameter to provide the target force constant.

```
# Changing flat-bottom harmonic restraint on DBC
harmonicWalls {
    colvars          DBC_sym
    targetForceConstant 200.0
    targetForceExponent   6.0
    upperWalls        1.5
    forceConstant      0.0
    targetEquilSteps   500
    lambdaSchedule     1.00 0.95 .... 0.05 0.00
    targetNumSteps     500000
}
```

The RFEP simulation is performed using an arbitrary series of `lambdaSchedule` with a specific interval ($\Delta\lambda$) of about 0.05. Each window of the desired `lambdaSchedule` is sampled over `targetNumSteps` steps with `targetEquilSteps` prior equilibration steps.

8.3 Result assessment

The RFEP simulation does not generate any individual output file reflecting the calculated free energy. We need to extract the primary data from the general NAMD log file (`rfep.log`), written in front of colvars rows with `dA/dLambda` notation. To have a clear vision, a cut of colvars rows in the `rfep.log` file is displayed in Figure (11).

```
ENERGY: 499000      5.1850      2.7640      2.3409      0.0000      -8.6483
WRITING EXTENDED SYSTEM TO RESTART FILE AT STEP 499000
WRITING COORDINATES TO DCD FILE PHEN-RFEP-final.dcd AT STEP 499000
WRITING COORDINATES TO RESTART FILE AT STEP 499000
FINISHED WRITING RESTART COORDINATES
WRITING VELOCITIES TO RESTART FILE AT STEP 499000
FINISHED WRITING RESTART VELOCITIES
colvars: Restrict harmonicwalls2 Lambda= 1 dA/dLambda= 0.476655
colvars: Restrict harmonicwalls2, stage 1: lambda = 0.95, k = 147.018
colvars: Synchronizing (emptying the buffer of) trajectory file "PHEN-RFEP-final.colvars.traj".
colvars: Saving collective variables state to "PHEN-RFEP-final.restart.colvars.state".
```

Figure 11. A sample cut of the `rfep.log` file by highlighting the primary data for restraint free energy calculation.

The derivative of free energy as a function of λ (`dA/dLambda`) are collected from the NAMD log file by Using the following Unix command and saving to the `RFEP.dat` file.

```
grep "Lambda=" rfep.log | awk '{print $5, $7}' > RFEP.dat
```

In the former subsection, the implementation of the RFEP simulation, the restraint free energy was defined using a non-linear changing harmonic potential with respect to the coupling parameter (λ). Here, we estimate the restraint free energy difference between the reference and the target states through the integral of the free energy derivative over the λ , referred to as the thermodynamics integrating method [32, 33]. Eventually, the RFEP free energy is obtained, using `xmgrace` for integrating the $dA/d\lambda$, by typing `xmgrace rfep.dat` command in the Unix terminal, and selecting `cumulative sum` from `Data → transformation → Integration` on the top navigation bar. The $dA/d\lambda$ and its' integrating relative to the λ is illustrated in Figure (12). The restraint free energy is measured of about -1.97 kcal/mol by employing the thermodynamic integration method.

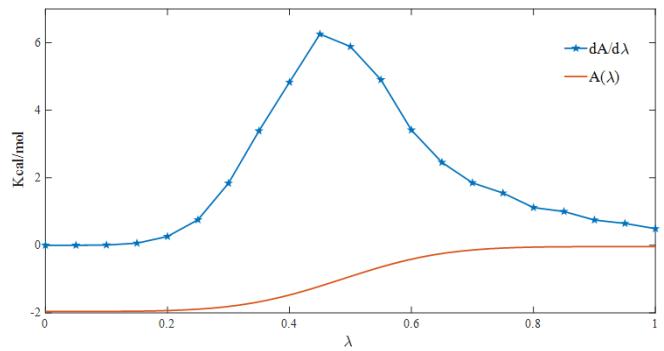


Figure 12. Restraint free energy ($A(\lambda)$, red line) and its derivative with respect to the coupling parameter ($dA/d\lambda$, blue line), as a function of λ .

8.4 Isotropic center-of-mass restraint

The DBC restraint was released while restraining the center of mass of phenol to a spherical enclosing volume. In the next stage, the geometry confined phenol is exchanged between the isotropic and the anisotropic, consistent with the bulk, interaction environments. The free energy cost of this transmission between two environments relies on the ratio of their volumes, which can be calculated by Eq. (2).

$$\Delta F_{iso} = -RT \ln \frac{V_{bulk}}{V_{confined}} \quad (2)$$

The $V_{confined}$, geometry confined, is obtained by $\frac{4}{3}\pi r^3$ and r is the upper boundary of the applied harmonic bias on the distance colvar. The V_{bulk} notation displays the volume of bulk with standard concentration [1]. According to Eqn(1), the contribution of the switch to the isotropic restraint in the overall free energy is calculated about 2.84 kcal/mol.

9 Decoupling the ligand from bulk solution

In this section, phenol is decoupled from the solution in the absence of lysozyme. We calculate the free energy of phenol decoupling from solution based on the alchemical free energy method, using the Bennett acceptance ratio (BAR) estimator for better convergence. Accordingly, the same sampling framework used in decoupling phenol from lysozyme is applied with a little modification. The time step of 1 fs is chosen for the solvated phenol, in view of the fact that the small size of the system satisfies a cheap simulation. The `Colvars` block is disabled in the configuration file as the phenol is not restrained in the solution. In decoupling phenol from the solution, the value of `alchVdwShiftCoeff` is set to 6.8 in the soft-core potential parameters.

```
alchVdwShiftCoeff      6.8
alchEquilSteps         50000
set  numSteps          450000
set  dLambda           0.05
runFEP    0.0   1.0  $dLambda   $numSteps  true
```

The alchemical FEP sampling for decoupling phenol from the solution is performed by 20 windows with equal intervals i.e. $\Delta\lambda=0.05$, each window is run for 450000 steps interleaved with 50000 equilibration steps. In this case, the equilibration steps and number steps for each window are increased.

9.1 Result assessment

We analyze the AFEP simulation output by `alchemyb` library through the procedure as mentioned earlier. The free energy difference of the ligand decoupling from aqueous solution is measured of about (4.18 ± 0.21) kcal/mol and its' evolution with respect to λ is illustrated in Figure(14).

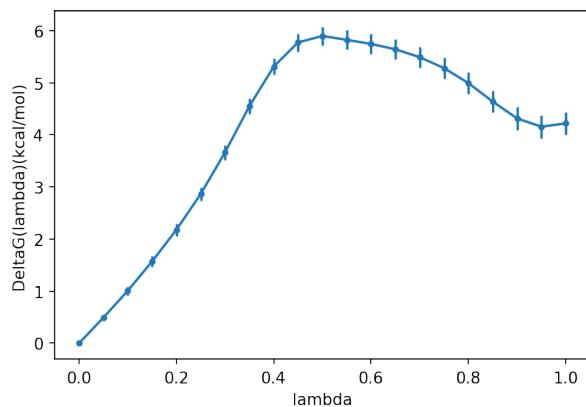


Figure 13. The λ -evolution of phenol decoupling affinity, from aqueous bulk solution, evaluated using `alchemyb` library.

To put the convergence analysis on the phenol decoupling free energy from the aqueous solution or called phenol hydration affinity, we provided Figure (14) visualizing the time-evolution of the forward and reverse sampling. Similar to the ligand binding free energy section, you might find the forward and backward discrepancy of phenol hydration affinity changing with λ and its' distribution over the free energy in Figure (15).

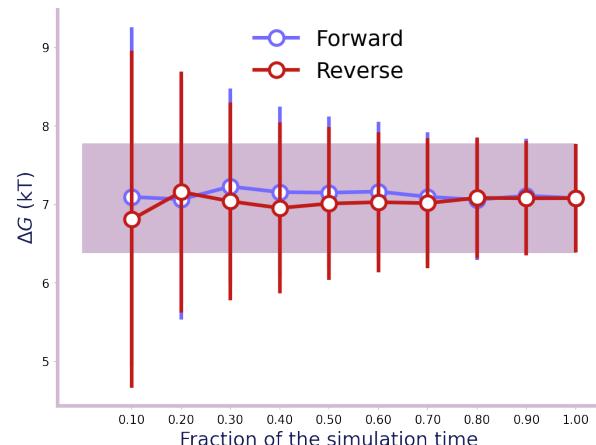


Figure 14. The phenol hydration affinity estimated from the forward and backward sampling changes as a function of fractions of simulation time.

The small fwd-bwd discrepancy of phenol hydration affinity may indicate the negligible hysteresis between the forward pathway and the reverse, which enhances the accuracy of the estimation.

10 Calculating the absolute binding free energy

Ultimately, all obtained free energies for non-physical intermediate states are added up to estimate the absolute binding free energy. In this tutorial, the absolute binding affinity of the phenol-lysozyme is calculated of about -5.67 kcal/mol, employing the streamlined AFEP method. Our assessment concerning the phenol-lysozyme binding affinity is in good accordance with the experimental data, -5.44 kcal/mol reported by Merski et. al. [17]. The exact agreement with the experimental study exhibits the accuracy and reliability of the estimation using the new sampling method. In addition, results converge perfectly without performing the long-run and costly simulations. According to results of the previous study [1] and this tutorial, the streamlined AFEP approach provides more accurate and efficient binding affinity predictions.

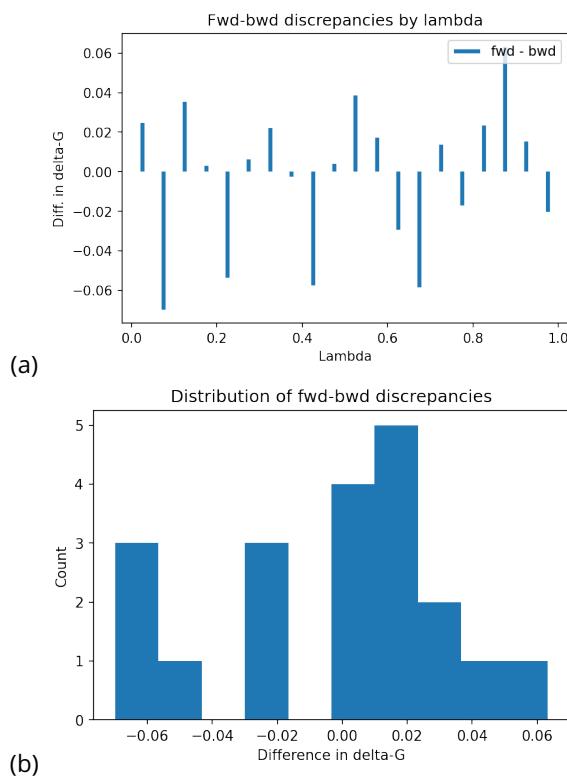


Figure 15. (a) The fwd-bwd discrepancy of the phenol hydration free energy over λ window, and (b) the distribution of the fwd-bwd discrepancy of the phenol hydration affinity.

11 Conclusion

In this instruction, we introduced a new sampling protocol based on the alchemical free energy perturbation method. In order to evaluate the performance of the streamlined AFEP (SAFEP) method, the framework is established to reproduce the experimental binding affinity of the phenol-lysozyme complex. The distance from bound-configuration (DBC) restraint is the remarkable feature of this approach to benefit of decreasing the number of restraining potentials. The predicted absolute binding affinity is in good agreement with the experimental data which furnishes the accuracy of the sampling framework. Furthermore, the convergence of the binding free energy over simulation time, and the low difference free energy discrepancy between the forward and the reverse sampling, underlies the exact prediction, employing the streamlined AFEP approach. This streamlined sampling scaffold, can be readily applied to challenging macromolecules by providing optimal simulations.

12 Author contributions

13 Funding information

LABEX, Iranian travel grant.

References

- [1] Salari R, Joseph T, Lohia R, Hénin J, Brannigan G. A streamlined, general approach for computing ligand binding free energies and its application to GPCR-bound cholesterol. *Journal of chemical theory and computation*. 2018; 14(12):6560–6573.
- [2] Phillips J, Isgro T, Sotomayor M, Villa E, NAMD TUTORIAL. NIH Center for Macromolecular Modeling and Bioinformatics, Beckman Institute; 2003. <http://www.ks.uiuc.edu/Training/TutorialsOverview/namd/namd-tutorial-unix.pdf>.
- [3] Hénin J, Gumbart J, Chipot C. In silico alchemy: A tutorial for alchemical free-energy perturbation calculations with NAMD. see www.ks.uiuc.edu/Training/Tutorials/namd/FEP/tutorial-FEP.pdf for the NAMD software. 2017; .
- [4] Phillips J, Hardy D, Maia J, Stone J, Ribeiro J, Bernardi R, Buch R, Fiorin G, Hénin J, Jiang W, McGreevy R, Melo MCdR, Radak B, Skeel R, Singharoy A, Wang Y, Roux B, Aksimentiev A, Luthey-Schulzen Z, Kale L, et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *The Journal of Chemical Physics*. 2020; 153:044130.
- [5] Chen H, Maia JDC, Radak BK, Hardy DJ, Cai W, Chipot C, Tajkhoshid E. Boosting Free-Energy Perturbation Calculations with GPU-Accelerated NAMD. *Journal of Chemical Information and Modeling*. 2020; <https://doi.org/10.1021/acs.jcim.0c00745>.
- [6] Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *Journal of molecular graphics*. 1996; 14(1):33–38.
- [7] Shirts MR, Chodera JD. Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of chemical physics*. 2008; 129(12):124105.
- [8] Chodera JD. A simple method for automated equilibration detection in molecular simulations. *Journal of chemical theory and computation*. 2016; 12(4):1799–1805.
- [9] Hermans J, Wang LU. Inclusion of loss of translational and rotational freedom in theoretical estimates of free energies of binding. Application to a complex of benzene and mutant T4 lysozyme. *Journal of the American Chemical Society*. 1997; 119(11):2707–2714.
- [10] Gilson MK, Given JA, Bush BL, McCammon JA. The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophysical journal*. 1997; 72(3):1047–1069.
- [11] Boresch S, Tettinger F, Leitgeb M, Karplus M. Absolute binding free energies: a quantitative approach for their calculation. *The Journal of Physical Chemistry B*. 2003; 107(35):9535–9551.
- [12] Hamelberg D, McCammon JA. Standard free energy of releasing a localized water molecule from the binding pockets of proteins: double-decoupling method. *Journal of the American Chemical Society*. 2004; 126(24):7683–7689.
- [13] Woo HJ, Roux B. Calculation of absolute protein-ligand binding free energy from computer simulations. *Proceedings of the National Academy of Sciences*. 2005; 102(19):6825–6830.
- [14] Deng Y, Roux B. Calculation of standard binding free energies: Aromatic molecules in the T4 lysozyme L99A mutant. *Journal of Chemical Theory and Computation*. 2006; 2(5):1255–1273.

- [15] **Fiorin G**, Klein ML, Hénin J. Using collective variables to drive molecular dynamics simulations. *Molecular Physics*. 2013; 111(22-23):3345–3362.
- [16] **Mannhold R**, Kubinyi H, Folkers G. Protein-ligand interactions, vol. 53. John Wiley & Sons; 2012.
- [17] **Merski M**, Shoichet BK. The impact of introducing a histidine into an apolar cavity site on docking and ligand recognition. *Journal of medicinal chemistry*. 2013; 56(7):2874–2884.
- [18] **Wei BQ**, Baase WA, Weaver LH, Matthews BW, Shoichet BK. A model binding site for testing scoring functions in molecular docking. *Journal of molecular biology*. 2002; 322(2):339–355.
- [19] **Morton A**, Baase WA, Matthews BW. Energetic origins of specificity of ligand binding in an interior nonpolar cavity of T4 lysozyme. *Biochemistry*. 1995; 34(27):8564–8575.
- [20] **Graves AP**, Brenk R, Shoichet BK. Decoys for docking. *Journal of medicinal chemistry*. 2005; 48(11):3714–3728.
- [21] **Jo S**, Kim T, Iyer VG, Im W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry*. 2008; 29(11):1859–1865. <https://doi.org/10.1002/jcc.20945>.
- [22] **Lee J**, Cheng X, Swails JM, Yeom MS, Eastman PK, Lemkul JA, Wei S, Buckner J, Jeong JC, Qi Y, Jo S, Pande VS, Case DA, Brooks CL, MacKerell AD, Klauda JB, Im W. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM-M/OpenMM Simulations Using the CHARMM36 Additive Force Field. *Journal of Chemical Theory and Computation*. 2016; 12(1):405–413. <https://doi.org/10.1021/acs.jctc.5b00935>.
- [23] **JC**, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel L R D Kalé, Schulten K. Scalable molecular dynamics with Namd. *J Comput Chem*. 2005; 26:1781–1802.
- [24] **Darden T**, York D, Pedersen L. Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems. *The Journal of chemical physics*. 1993; 98(12):10089–10092.
- [25] **Best RB**, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, MacKerell AD. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ , ψ and Side-Chain χ_1 and χ_2 Dihedral Angles. *Journal of Chemical Theory and Computation*. 2012; 8(9):3257–3273. <https://doi.org/10.1021/ct300400x>.
- [26] **Jorgensen WL**, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 1983; 79:926–935.
- [27] **Bar-Itzhack IY**. New method for extracting the quaternion from a rotation matrix. *Journal of guidance, control, and dynamics*. 2000; 23(6):1085–1087.
- [28] **Eberly D**. Rotation representations and performance issues. Magic Software: Chapel Hill, NC, USA. 2002; .
- [29] **Liu P**, Dehez F, Cai W, Chipot C. A toolkit for the analysis of free-energy perturbation calculations. *Journal of chemical theory and computation*. 2012; 8(8):2606–2616.
- [30] **Hénin J**, Baaden M, Taly A. Foundations of biomolecular simulations: a critical introduction to homology modeling, molecular dynamics simulations, and free energy calculations of membrane proteins. In: *Membrane Proteins Production for Structural Analysis* Springer; 2014.p. 347–392.
- [31] **Fu H**, Cai W, Hénin J, Roux B, Chipot C. New coarse variables for the accurate determination of standard binding free energies. *Journal of Chemical Theory and Computation*. 2017; 13(11):5173–5178.
- [32] **Kirkwood JG**. Statistical mechanics of fluid mixtures. *The Journal of chemical physics*. 1935; 3(5):300–313.
- [33] **Frenkel D**, Smit B. Understanding molecular simulation: from algorithms to applications, vol. 1. Elsevier; 2001.