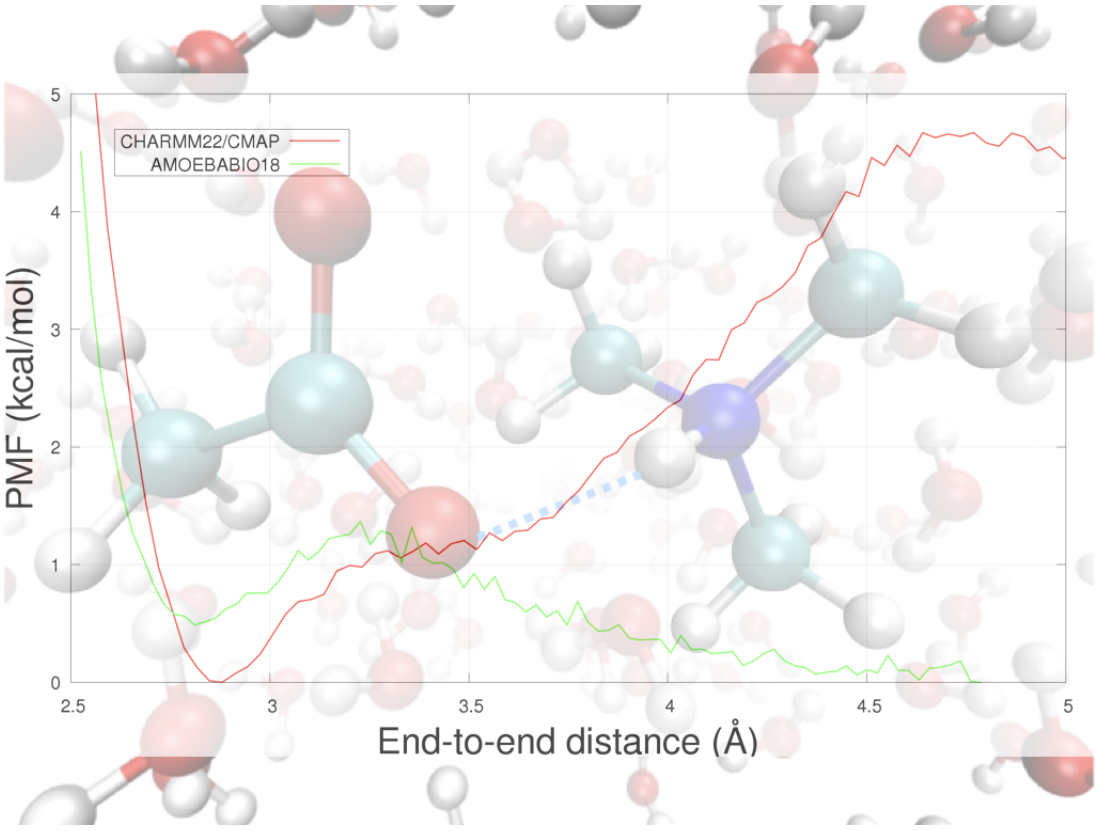
Umbrella Sampling with Tinker–HP TUTORIAL



⚠ Please visit the website: https://github.com/TinkerTools to obtain last updates of this tutorial and informations about other new tutorials.

Contents

1 Introduction

1.1 Presentation of the study

Be able to sample free energy landscape along a specific reaction coordinate (RC) is one of many current challenge. Such a profile is known as the Potential of Mean Focre (PMF), and is really useful in searching transitions states, as well as establishing secondary strcuture stabilities in biological systems. A simple way should be to run a couple of classical molecular dynamics and to identify the different stabilized states. However, often the free energy barrier to pass from one state to the other one being of many times of k_BT, it should take a long time to be able to observe such an event. When MDs are in the frame of nanoseconds, these events are called rare events and remain of a hard task to be well sampled. Several methods have been developed in this specific field until few decades to tackle this huge problem, such as: Free Energy Perturbation (FEP), Thermodynamic Integration (TI), Steered Molecular Dynamics (SMD), Umbrella Sampling (US), Adaptive Bias Forced (ABF), ...
In this tutorial, we will specifically be focused on one of them, the Umbrella Sampling method. Due to its accuracy in terms of free energy results it represents a powerful tool in order to well sample a free energy landscape along a specific reaction coordinate.
Remark: However, although this methodology is powerful, it also represents an important cost in terms of computational time. That is why when the RC looks as well as a straight line and the system does not seem too big (not too important degrees of freedom), it begins to be more interesting to consider the use of the Steered Molecular Dynamics methodology (explained in an other tutorial and currently implemented in the Tinker-HP v1.2). Nevertheless, US stays more useful (to not say essential) than SMD when:

- The chosen RC is not a simple straight line
- The SMD velocity is too high to be able to consider the quasi–equilibrium approach

After a few reminds about how the US method works, we will explain how to perform a US study with the use of the Tinker-HP V1.2 software. The system of interest in this tutorial will be a single ionic interaction between an acetate and a trimethylamonium (See Figure 1), where US will be used to determine the free energy barrier of this interaction, and thus scrutinize the strength of such a salt bridge (SB) by using of polarizable (AMOEBABIO18) and nonpolarizable (CHARMM22/CMAP) force fields.

1.2 Umbrella Sampling methodology

As told before, Umbrella Sampling is a powerful method to sample free energy landscape along a RC when it could be more complicated than a simple straight line (SMD).

- Firstly, we have to choose a RC (a distance, an angle, a dihedral, ...). This is what we call **Collective Variable (CV)** in the literature.
- After having identified the RC, we have to generate several windows. Each window represents a specific configuration of the system according to the chosen reaction coordinate. The choice of the number of windows to select is not easy and is also of a crucial choice (this point will be discussed in the next part of this tutorial).

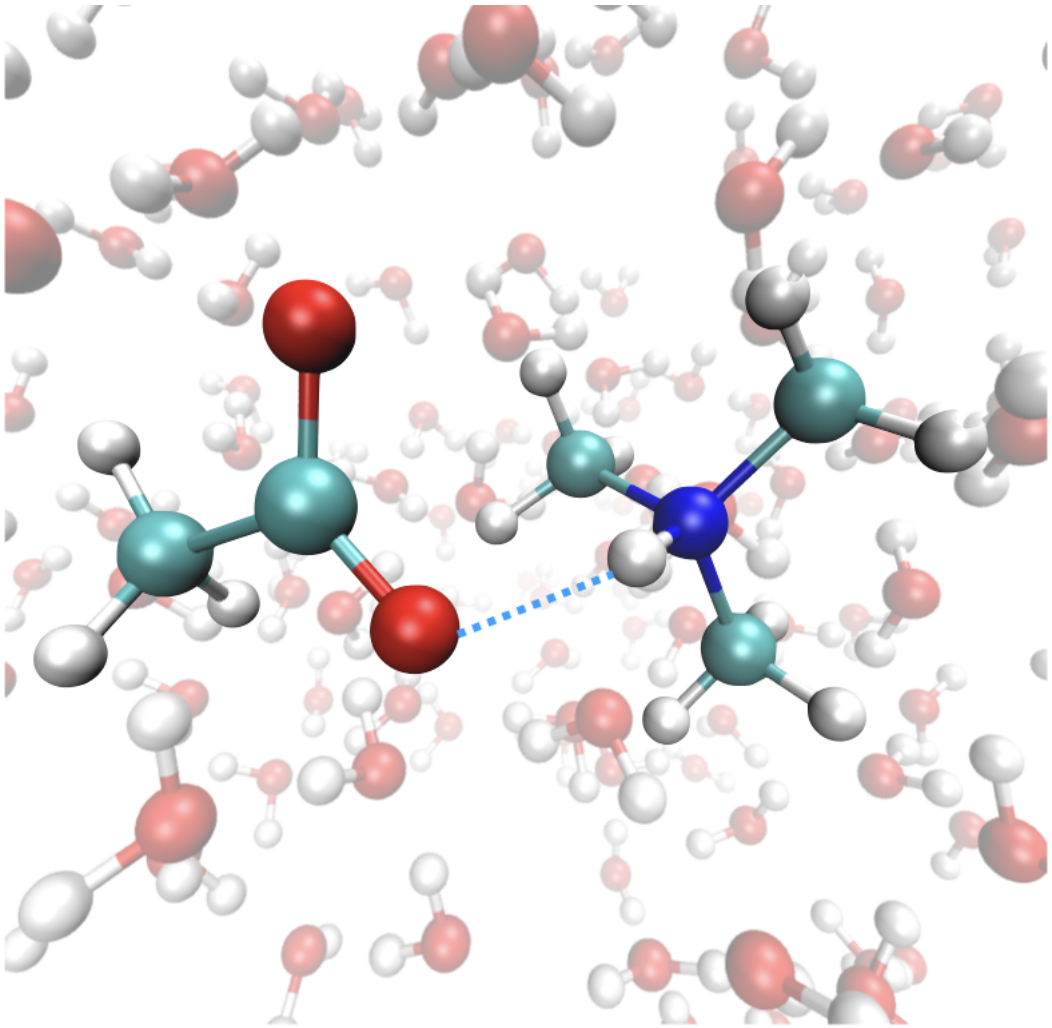


Figure 1: CPK representations of the interaction between an acetate and a trimethylamonium in a water environment. The chosen reaction coordinate (RC) is depcited in blue as it corresponds to the distance between the oxygen of the acetate and the hydrogen of the trimethylamonium.

- We then perform a classical molecular dynamics on each configuration, by adding a restraint with a harmonic form like that:

$$w(\epsilon) = \frac{1}{2}k(\epsilon - \epsilon_0)^2 \tag{1}$$

when k is the spring constant (in kcal/mol.Å^2), and ϵ_0 the initial restraint of the windows.

- Finally, we construct for each window its umbrella, and we combine all of them in order to reconstruct the desired potential energy surface. To perform this final step in the frame of this tutorial, we will limit us to the Weighted Histogram Analaysis Method (WHAM). However, keep in mind that other methods exist and can be used (Umbrella Integration, Bennett Acceptance Ratio, ...) if some troubles are observed in the futur in terms of convergence.

2 Running an US simulation with Tinker and Tinker–HP

Input Files and Executables

- complex.xyz
- dynamic.key
- AMOEBABIO18_modified.prm
- Tinker–HP dynamic

Let us start to explain how to perform an US simulation. As explained in the previous point, we normally first have to prepare each window properly, and so to construct our system, minimize and equilibrate it. The preparation, minimization and dynamic steps being already done, we will directly focus on the preparation of each US window.
Please follow these instructions step by step:

1. Open a new terminal and place in a same direction complex.arc, dynamic.key, AMOEBABIO18_modified.prm files and the Tinker–HP dynamic executable

2. As we decided to sample the distance between the O of the acetate and the N of the trimethylamonium, we have to create windows starting from 2.2 to 5 Å. Each window will be separated by a step of 0.2 Å. If the directory you placed before each file is called "REF_file", please copy like that:

   - *cp REF_file 2.2*
   - *cp REF_file 2.4*
   - *cp REF_file 2.6*
   - *cp REF_file 2.8*
   - *cp REF_file 3.0*

- *cp REF_file 3.2*
- *cp REF_file 3.4*
- *cp REF_file 3.6*
- *cp REF_file 3.8*
- *cp REF_file 4.0*
- *cp REF_file 4.2*
- *cp REF_file 4.4*
- *cp REF_file 4.6*
- *cp REF_file 4.8*
- *cp REF_file 5.0*

3. In each dynamic.key file of each window, adjust the *restrain–distance* line according to the window's restrain. For instance, for the 3.2 window, you have to write:

```
restrain-distance      2 20 10 3.2
```

   2 and 20 represents the index atoms on which the restrain is applied, 10 the spring constant of the harmonic restraint (in kcal/mol.Å^2), and 3.2 is the reference value for the harmonic restraint.

4. Launch each window separately by typing in the terminal:

```
mpi -np 8 ./dynamic complex.xyz 1000000 1.0 1 2 300 > dynamic.out
```

Automatization script !

Due to the fact that this task could be long in a manual way, an automatization script named **umbrella.sh** is available to perform all these tasks, starting from your initial Ref_file repertory.

3D spatial decomposition limitation in US ...

As we are using the 3D spatial decomposition in Tinker–HP, each cores know the x/y/z coordinates of their respective atoms + atoms localized at 7 Å max. In order to ensure the well consistency of our US procedure, we need to be sure that both atoms used in the *restrain–distance* could be read by the same core. In this way, one possibility is to consider the cutoff range equal to the distance between the both atoms. However, you can decrease these distance according to the number of cores you are using in your dynamics. You just then have to enter a keyword labeled **dd-cutoff**.
Example: We have a distance of 40 Å between the both atoms. The 3D spatial decomposition indicates that we have 6 procs in each x/y/z direction of a cubic box of 50 Å length. A dd–cutoff 30 Å should be useful in this case.

3 Umbrella Sampling: Use of the WHAM algorithm to reconstruct the Potential of Mean Force

Input Files and Executables

- size.tcl
- wham
- gnuplot
- PMF.gnu
- Umbrellas.gnu

On each window we have run a dynamic of 1000000 ps. The aim now is to be able to reconstruct the Potential of Mean Force (PMF) profile of the ion pair according to the US sampling. The WHAM algorithm can be used to perform this step. Follow these instructions:

1. Create a file labeled arcfile
2. The aim now is to generate for each window a distance file containing two columns: the first for the step and the second for the distance O(acetate)–N(trimethylamonium). It can be done manually by compiling and applying the distarc.f90 fortran file for each window and move the generated file in the arcfille file. It can also be easily automatized by a little bash script.
3. Once it is done, you have in your arcfile file the following files:
   - wham2.2
   - wham2.4
   - wham2.6
   - wham2.8
   - wham3.0
   - wham3.2
   - wham3.4
   - wham3.6
   - wham3.8
   - wham4.0
   - wham4.2
   - wham4.4
   - wham4.6
   - wham4.8
   - wham5.0

Create a file labeled metadata and fill for each wham file a line like that:

```
/home/celerse/arcfile/wham3.4    3.4    20
```

/home/celerse/arcfile/wham3.4 represents the path of your wham3.4 file, 3.4 is the reference for your restraint and 20 the spring constant. Be careful about the value you enter here ! Indeed, you saw that we used 10 kcal/mol.Å^2 before. Due to the fact that Tinker double count the harmonic interaction at each step, we have to adjust it by double the spring constant value in the metadata file.

4. Apply the wham executable by typing in the terminal:

```
./wham 2.2 5 100 0.0001 300 metadata out
```

Let us detail each components used in the wham's line:

- 2.2: The left boarder of the PMF
- 5: The right boarder of the PMF
- 100: Number of bins
- 0.0001: Threshold criteria
- 300: Temperature (in K)
- metadata: name of the metadata file
- out: name of output file generated by the wham procedure

Finally, a output file labeled out is generated and contain the free energy value for each step of the US. You can easily make the same procedure as before for the CHARMM22/CMAP case. If you open and plot the output files, you should observe this kind of graph:

Interpretation tool !

The ionic interaction is weaker in the case of the polarizable force field AMOEBABIO18 with a barrier of approximatively 1 kcal/mol between the interacted and dissociated forms. The nonpolarizable force field CHARMM22/CMAP depicts on its side a stronger ionic interaction of 4.7 kcal/mol, which seems to be too high for a simple salt bridge. This result goes in the same direction as the literature, as well as it was demonstrated that nonpolarizable force fields tend to overstablize the salt bridges. It thus clearly encourage to use in the future polarizable force fields such as AMOEBA to describe these kinds of interactions.

We can also draw the umbrellas of each windows to have an idea about the distance stabilization of the ionic pair. It could be done by applying the umbrella.py python code on each windows. Finally, still by using Gnuplot, we obtain for AMOEBABIO18 a graph like that:

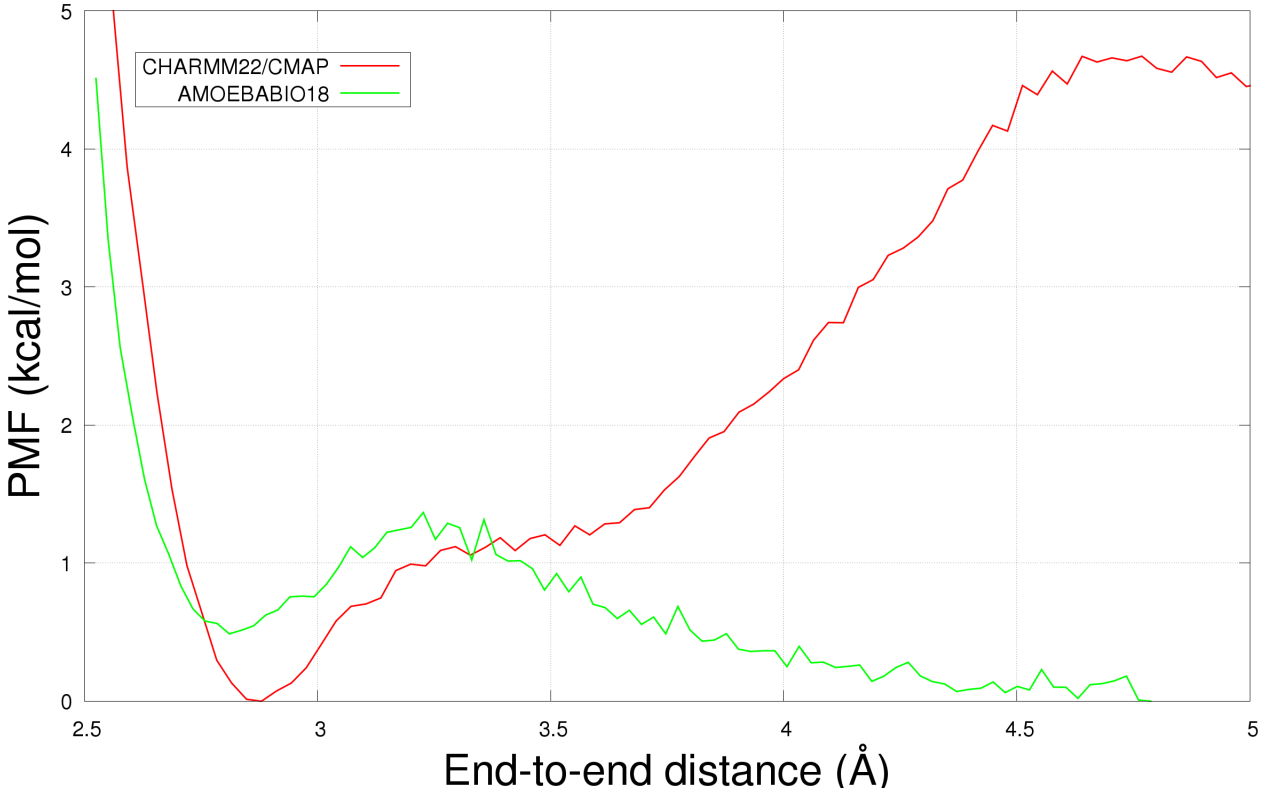


Figure 2: Potential of Mean Force obtained for the interaction of the acetate with the trimethylamonium using polarizable (AMOEBABIO18) and nonpolarizable (CHARMM22/CMAP) force fields in the Tinker–HP package.

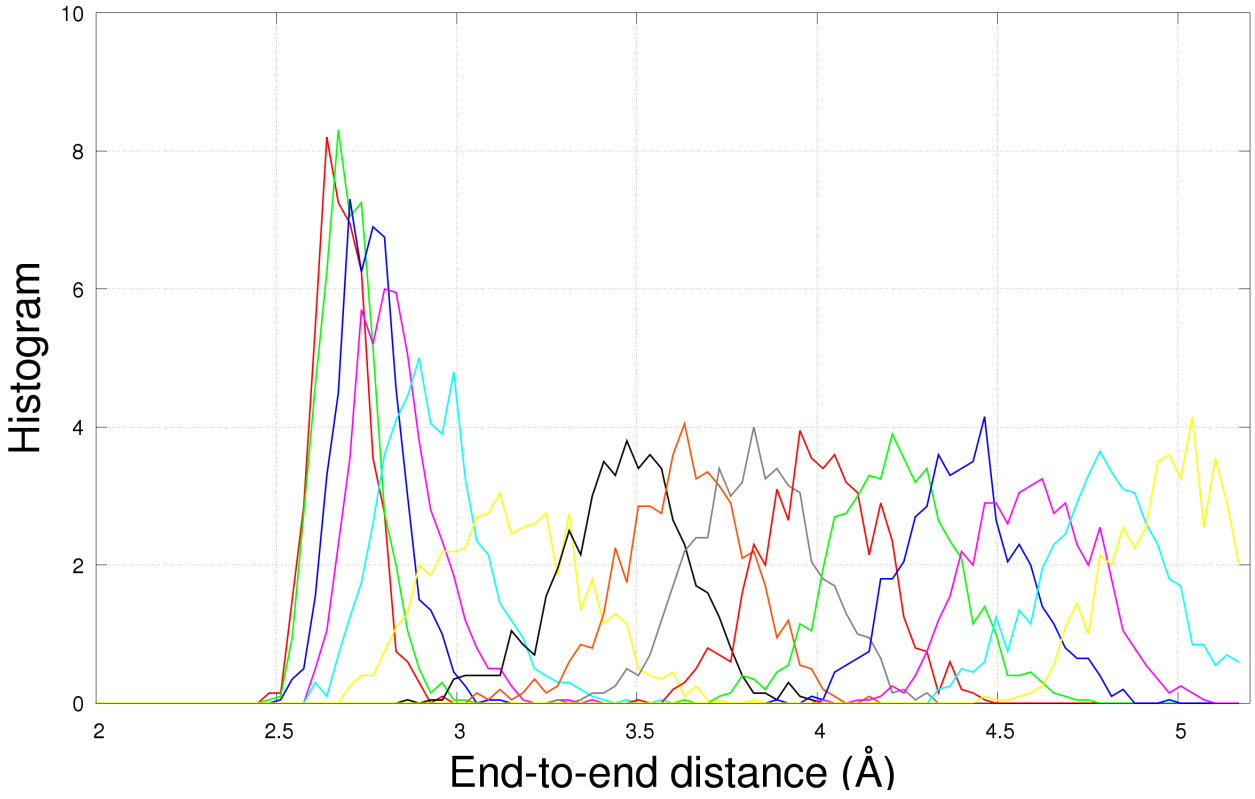


Figure 3: Umbrellas of each windows for AMOEBABIO18.

Interpretation tool !

- We observe a large overlap of the umbrellas around the region of 2.7/2.8 Å, which well corresponds to the stabilization hole observed in Figure 2 for the ionic pair charge.
- The region between 3 and 3.2 Å should not be well sampled. To tackle this problem, we could either increase the simulation time of each window or add additional windows (3.05, 3.1, 3.15) in this region to enhance the sampling
- Finally, the sharked shape of the major part of the umbrellas let think that PMF could be not well converged. One way to ensure that we converged is to double the simulation time (from 1 to 2 ns per window) and overlap PMFs obtained with windows of 1 and 2 ns. If the recovering is correct, we can thus consider that we converged.

Avoid the beginning of each windows

⚠It is advisable to avoid the beginning of each windows as well as it corresponds to the equilibration of the system according to the restraints.

4 Conclusions and perspectives

We saw toward this smooth tutorial how we can use the Umbrella Sampling method within the Tinker–HP software, enabling to use now polarizable force field such as AMOEBA for this free energy method. We also saw the influence of the polarization on the strengthness of a single ionic pair charge in a water environment. Litterature was recovered, with a highest stabilization (4.7 kcal/mol) observed with the nonpolarizable force field CHARMM22/CMAP compared to the polarizable force field AMOEBABIO18 (1 kcal/mol). It thus encourage us to use polarizable force fields in the future to enhance the description's accuracy of such a SB interaction.

5 References

Torrie, G. M.; Valleau, J. P. Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. *Chem. Phys. Lett.* **1974**, 28, 578–581.

Kästner, J. Umbrella Sampling. *Wiley. Interdiscip. Rev. Comput. Mol. Sci.* **2011**, 1, 932–942.

Grossfield, A. WHAM: the weighted histogram analysis method.
http://membrane.urmc.rochester.edu/content/wham.

Lagardère, L.; Jolly, L.H.; Lipparini, F.; Aviat, F.; Stamm, B.; Jing, Z. F.; Harger, M.; Torabifard, H.; Cisneros, G. A.; Schnieders, M. J.; Gresh, N.;Maday, Y.; Ren, P. Y.; Ponder, J.W.; Piquemal, J.-P. Tinker–HP: a massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields. *Chem. Sci.* **2018**, 9, 956–972.

Célerse, F.; Lagardère, L.; Derat, E.; Piquemal, J.-P. Massively parallel implementation of Steered Molecular Dynamics in Tinker-HP: comparisons of polarizable and non-polarizable simulations of realistic systems. *J. Chem. Theory. Comput.*, **2019**, X, XXX–XXX