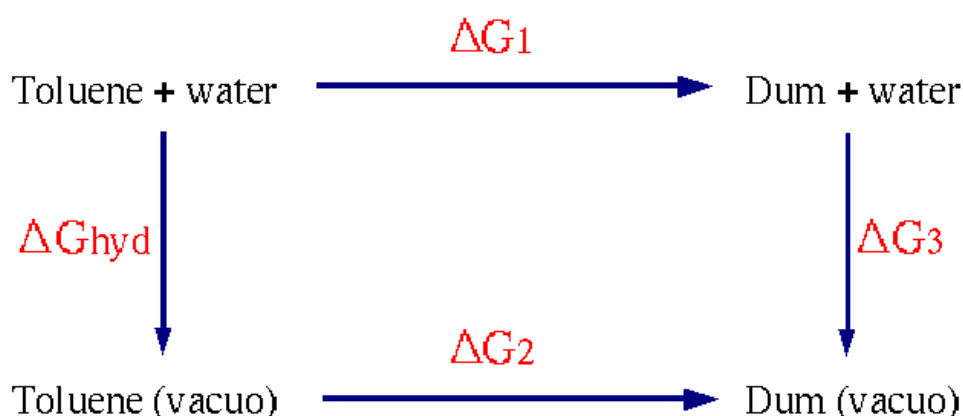# HYDRATION FREE ENERGY OF TOLUENE

Before we start a great follow up resource before and after this prac is the alchemistry.org wiki

http://www.alchemistry.org/wiki

The hydration free energy can be calculated using the following thermodynamic cycle:

The atoms of the molecule of toluene are mutated into dummy atoms along the processes ΔG_1 and ΔG_3 of the cycle (which are therefore nonphysical processes).

Dummy atoms are atoms that do not havelack non-bonded interactions, which basically means that they; they therefore interact only with directly bonded atoms and not by the other molecules around them do not interact with their environment. In practice these atoms are not charged and do not have any Van der Waals interactions with their environment.

> ΔG_{hyd}: hydration free energy.
> ΔG_1: free energy associated to with the mutation of toluene into dummy atoms in water.
> ΔG_2: free energy associated with the mutation of toluene into dummy atoms *in same in vacuo*.
> ΔG_3: can be seen as the hydration free energy of dummy atoms. This term is equal

to 0 since dummies do not have non-bonded interactions and bonded interactions remain the same.

$\Delta G_3 = 0$

$\Delta G_{hyd} = \Delta G_1 +- \Delta G_3 - \Delta G_2$
$\Delta G_{hyd} = \Delta G_1 - \Delta G_3 \Delta G_2$

As dummy atoms are nonphysical, ΔG_1, ΔG_2 and ΔG_3 are purely theoretical constructs. The 'physics' behind these nonphysical processes is that the internal non-bonded potentials of the molecule of toluene are switched off in ΔG_1 , and that you have to account for the energy associated to with that phenomena (which is done by subtracting ΔG_2). Since we're using a computer, we can follow these nonphysical processes by sampling from several simulations in which we mix toluene and dummy-toluene to varying degrees. This is called alchemic mutation.

HYDRATION FREE ENERGY CALCULATIONS USING GROMACS:

In this first example, we will calculate the hydration free energy of a molecule of toluene using the MD software package GROMACS 5.1.x (http://www.gromacs.org). GROMACS is a widely used molecular dynamics simulation package developed at the University of Groningen.

An outline of the procedure you will have to follow to perform MD simulations is given by the GROMACS flowchart: http://manual.gromacs.org/documentation/5.1/user-guide/flow.html

Any simulation involves these steps:

1. Convert a structure file of a molecule into a GROMACS-compatible structure file and a GROMACS topology file
2. Edit the topology file to make it suit your needs
3. Put the molecule in a simulation box
4. Solvate it (if necessary)
5. Perform an energy minimization of the system
6. Perform an equilibration of the system

7. Perform the simulation

8. Analyze the results

To obtain the free energy hydration, you will need to carry out the simulations both in vacuo and in water.

PROCESSING TOLUENE

We now need a coordinate file for toluene. Note, GROMACS is orientated toward simulation of biomolecular systems and there is no building block for toluene in the default distribution. We could build toluene by hand. However, toluene is nothing more than the side-chain of the Phenylalanine amino-acid where the C alpha is replaced by a H atom. Knowing this you can run the pdb2gmx command using Phenylalanine as a template and easily convert the side chain into toluene by hand.

We have given you a .pdb file of phenylalanine, use Vim to delete all the atoms that aren't part of toluene, and delete all the hydrogens. GROMACS will regenerate any hydrogens anyway. (Also delete all the stuff that isn't atoms, as they're not important)

Step 1. **Use pdb2gmx to convert your toluene PDB to a GROMACS structure file (.gro) and generate a GROMACS topology file (.top).**

We start with a pdb file of toluene that has been made by truncating phenylalanine. The program will complain about missing atoms (the backbone atoms of the Phe residue are missing since the pdb file only contains the side chain) but you can discard them (option **-missing**) as toluene is not supposed to contain them. There are many options within this program, which you can see with **gmx help pdb2gmx** or by googling the online manual. Select the Gromos96 54a7 force field for protein simulations in water (14 in the

default distribution), and the SPC water model (1). The default program will continue and provide information on the variety of items being automatically set such as protonation states for certain residues such as the N- and C-termini. You should ideally name your output files with understandable names by using the option flags such as **–p** and **–o** so you don't mix up your files or overwrite them accidentally.

How many atoms does your output GRO file have? How many did your input PDB? This is because GROMACS (re)generates hydrogen positions.

Step 3. Edit the topology file to fit toluene, the A state.

Look at your topology file to get an idea of the atoms in your toluene system. The column names are given in comments, which are lines that start with ';'. GROMACS ignores everything after a semi-colon and comments are there for human benefit. As you can see, default parameters for each atom and bonded interaction are given, but the typeB, charge and massB columns are empty – the 'qtot' entries are after a semi-colon, so they are ignored.

Recall that the GROMOS forcefield is a united atom force field, where nonpolar hydrogens are combined into a single atom with their parent carbon. As GROMACS thinks this toluene is a phenylalanine molecule, it assumes that the CB carbon will bond to the alpha-carbon of phenylalanine and therefore only has 2 hydrogens attached. Toluene, however, just has a terminal methyl group there. Change the **type** of this atom to a methyl (CH3) and add the mass of hydrogen to its current mass.

Once this first modification is done, the topology file corresponds to the molecule of toluene in state A.

Step 4. Create a topology file for the B state, i.e. toluene with a dummy atom.

Define the B state that will be used for the free energy calculation. For this, you need to specify the new atom type (typeB), charge (chargeB) and mass (massB) of each mutated atom, as well as the B-state non-bonded interaction parameters. The dummy atom should be modelled with the DUM type, a 0 charge, and the same mass. The easiest way to create this topology is to copy your toluene topology and edit the copy.

GROMACS infers other non-bonded parameters, such as Lennard-Jones parameters, from the atom type, so they are not explicitly stated in the topology. For the bonded interactions, most parameters are given as keywords that refer back to the forcefield itself. You just need to add a second (identical) entry for each of them to define the B state. Example entries for the `[ bonds ]` directives are provided below, but the `[ angles ]` and `[ dihedrals ]` directives also need to be edited.

```
[ bonds ]
  1     2     2     gb_26     gb_26
```

Step 5. Put your toluene in a simulation box using gmx editconf.

The size of the box must be big enough for the distance between the molecule of toluene and its periodic image to be larger than the cut-off radius, but in this case it should not be a problem given the little size of the molecule of toluene. –c centers the molecule in the box, and –box N N N creates a box where each side is N N N nm long along each unitcell vector. Create a box 3.0 nm long on each side.

Gmx editconf -f phe.pdb -o phe_box.pdb -box 3 3 3

Step 6. Solvate your boxed molecule with gmx solvate.

At this point it might be a good idea to **separate your files into folders for your vacuum and water simulations.** If you don't, beware of accidentally overwriting a file or mixing them up.

Whenever we add molecules to our system, they get an entry in our topology (.top) file. When we solvate our system, therefore, it won't match up with our vacuum simulations. Make a copy of your current (vacuum) topology and name it something sensible and solvent-related – this will be the topology that we modify during solvation.

gmx solvate takes –cp as its input structure flag. The default water model is the simple point charge (SPC) water model, which represents a water molecule using three points. You should have also picked this in pdb2gmx.

gmx solvate –cp tol.pdb -cs -p XXX.top –o XXX_solvated.gro

You can check the result by visualising your output GRO file with VMD.

At this point you should have two simulation systems set up:

1. In vacuum – with your vacuum toluene (pre-solvation) in a box and your vacuum topology
2. In solvent – with your solvated toluene and your solvated topology

You've finished the primary processing of each system. The energy minimisation and equilibration steps after this are required for both, and indeed pretty much all of molecular dynamics. And of course, you should simulate both systems at the end.

Step 7. Perform an energy minimization of the system

Now you have to perform an energy minimization of the structure to remove the local strain in the protein (due to generation of hydrogen positions) and to remove bad Van der Waals contacts (particles that are too close). This can be done with the mdrun command, which is the main computation program. Although it is called mdrun, it is also able to perform energy minimisation and a few other techniques. Before you can use mdrun however, you have to pre-process the topology file, the structure file, and a molecular dynamics parameters file (.mdp) using the GROMacs PreProcessor (grompp).

Copy steep.mdp to your working directory and check the contents of this file. Create a run topology file (.tpr) with this mdp and grompp.

gmx grompp –f XXX.mdp -c XXX.gro -p XXX.top -o XXX.tpr

Then run the minimisation with mdrun. Setting the –deffnm option in mdrun will name your files consistently, which will help with organisation.

gmx mdrun -v –deffnm XXX

Which files were generated by mdrun? (i.e. were not there before mdrun, and are there after). One of the generated files should be a .gro file. What does this structure represent?

You can check the evolution of the potential energy using gmx energy. The program can use the energy file (.edr) of the simulation to create a file (.xvg) containing the values you are interested in (potential energy, kinetic energy, temperature, pressure and many others).

gmx energy –f XXX.edr -o XXX.xvg

You will see something like this:

```
Opened water_enermin.edr as single precision energy file

Select the terms you want from the following list by
selecting either (part of) the name or the number or a combination.
End your selection with an empty line or a zero.
-------------------------------------------------------------------
 1  G96Angle        2  Improper-Dih.    3  LJ-(SR)          4  Coulomb-(SR)
 5  Potential       6  Pressure         7  Constr.-rmsd     8  Vir-XX
 9  Vir-XY         10  Vir-XZ          11  Vir-YX          12  Vir-YY
13  Vir-YZ         14  Vir-ZX          15  Vir-ZY          16  Vir-ZZ
17  Pres-XX        18  Pres-XY         19  Pres-XZ         20  Pres-YX
21  Pres-YY        22  Pres-YZ         23  Pres-ZX         24  Pres-ZY
25  Pres-ZZ                            26  #Surf*SurfTen
27  Coul-SR:Protein-Protein            28  LJ-SR:Protein-Protein
29  Coul-SR:Protein-SOL                30  LJ-SR:Protein-SOL
31  Coul-SR:SOL-SOL                    32  LJ-SR:SOL-SOL
33  T-rest
```

Select 5 (potential energy). You examine the values of various energy and other terms calculated during the simulation. Currently, our main interest is the total value of the potential energy.

Then type 0 (zero) to exit this option.

Now you can make a graph with the software qtgrace. Simply type:

`qtgrace absolute/path/to/your/xvg`

Note, this graph has an error: The label for the horizontal axis must be *Step* instead of Time (ps), as energy minimisation has no time associated with a step.
(To change it go to Plot (in the upper part of the window), select *Tick labels/tick marks* and in *Axis label* delete *Time (ps)* and write *Step*. Then click *Accept* and *Close*.)

To exit qtgrace go to *File* and select *Exit*.

Step 8. Perform equilibration

The equilibration process mainly aims at getting a correct velocity distribution for the molecules in order to simulate a valid statistical ensemble and of course to put the system at equilibrium. As we don't have much time this is all done for you.

Step 9. Free energy calculation

To get the mutation free energy of the toluene in dummy atoms, you need to run several simulations, one for each lambda value. For each simulation at a given λ_i, you will get $\frac{dG}{d\lambda}\lambda_i$.

Depending on the transformation, you usually need 15 and 25 points to get a smooth curve. Here we will only run the simulation for a couple of lambda points.

Get the parameter files for the equilibration (equ-l0.00.mdp) and the simulation (data-l0.00.mdp) with $\lambda = 0$. Remember that your system equilibrates to whatever the mdp settings are, so that each lambda value requires you to equilibrate it again before simulating.

Take a look at one of the parameter files, equ-l0.00.mdp for instance:

You can see that free energy is enabled and also for which lambda value the simulation is performed (`init-lambda  = 0.00`).

As usual, you **would** create a run topology file with gmx grompp. This **would** come up with a warning and fail without producing the output .tpr. Read the warning. In these simulations, soft-core interactions are important to avoid numerical instabilities, and the increased sampling required is not a problem as we have a very small system.

To run the simulation at a different lambda value, you **would** edit them to change the lambda value to the one you want (`init_lambda =  0.00`; `init_lambda =  0.02` etc.). As starting structure you will use the output configuration of the simulation performed at the previous lambda value.

LUCKY FOR YOU

We have done all of this for you, as we don't have much time to work with. You would need to run equilibrations and production simulations at each lambda value to obtain the requisite data. The data files can be found in the folder called demo.

Step 10. Analysis.

You can can see that a file data-l0.00.xvg , one of many similar files was created during the simulation. It contains $\frac{dG}{d\lambda}\lambda_i$. values calculated at each step of the simulation. To calculate the average value and the error associated to it you can use gmx analyze:

```
gmx analyze -f data-l0.00.xvg –ee errest-l0.00.xvg
```

You should see the average value below `average` and the error next to `err.est`.

```
                                    std. dev.     relative deviation of
                        standard    ---------     cumulants from those of
set       average       deviation   sqrt(n-1)     a Gaussian distribition
                                                  cum. 3    cum. 4
SS1   4.278153e-01   5.881016e+00   1.859741e-01   -5.726    36.298
```

```
Back Off! I just backed up errest.xvg to ./#errest.xvg.1#
Data set 1 has strange time correlations:
the std. error using single points is larger than that of blocks of 2 points
The error estimate might be inaccurate, check the fit
a fitted parameter is negative
invalid fit:  e.e. 0.116461  a 1.04389  tau1 0.0723081  tau2 1.27262
Will fix tau2 at the total time: 100
a fitted parameter is negative
invalid fit:  e.e. nan  a 1.00443  tau1 0.0554554  tau2 100
Will use a single exponential fit for set 1
Set   1:  err.est. 0.174381  a 1  tau1 0.0440045  tau2 0
```

These values are also saved in the output file errest-l0.00.xvg. Open the errest-l0.00.xvg in less or vim to find the average as "`av 0.427815`" and error as "`ee 0.174381`"

You can then create a **(dhdl-X.xvg)** file containing all the results according to the following format:

```
lambda_value    dgdl_average    err.est
    0.00           0.427815      0.174381
```

You **should** do this for each lambda value.

Luckily again we have done this for you

`qtgrace -settype xydy absolute/path/to/dHdl.xvg`

To integrate the plot using the identity below, go to the data menu with the mouse and select Transformations/integration. Choose "sum only" and accept.

$$\Delta E = \int_0^1 \left\langle \frac{dU(\lambda, \vec{q})}{d\lambda} \right\rangle_\lambda d\lambda$$

What value do you get? Remember that you still need the value in vacuo to compare the result to the experimental value.

RESULTS

Once you have the results for both in solvent and in vacuo, you can finally calculate the hydration free energy of toluene. Compare it to the experimental value:

$\Delta G^{hyd,exp}$(toluene) = -3.1 kJ.mol^{-1}