3. Mutation of tyrosine into alanine

In this section, we will calculate the free energy change involved in the point mutation of the N- and C-terminally blocked la–Tyr– la tripeptide. Our calculation will be based on the thermodynamic cycle of Figure 8.[24] We will calculate the quantities $G^1_{\text{alch.}}$ and $G^2_{\text{alch.}}$ by running two FEP simulations of the mutation, one *in vacuo* and the other in bulk water. This case is a computationally affordable example of how point mutations in more complex protein systems may be studied.

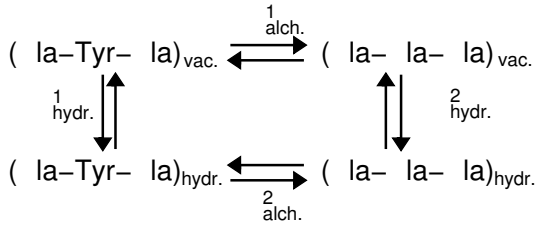

Figure 8: Thermodynamic cycle used in the la–Tyr– la → (la)$_3$ alchemical transformation. The vertical arrows correspond to the hydration of the wild-type tripeptide and its mutant. The horizontal arrows correspond to the point mutation in bulk water and *in vacuo*, so that: G^2_{alch} $G_{\text{alch}} = G^2_{\text{hydr}}$ G_{hydr}

3.1. System setup

We will prepare two systems: the isolated tripeptide, and the same solvated in explicit water. You can build the latter by using the `solvate` plugin of VMDon the former. The files required to get started with this setup can be found in the `03.Mutating-tyrosine-into-alanine` subdirectory of the archive.

s in the ethane-ethane transformation example, you will need a topology file describing the hybrid molecule, which is depicted in Figure 9. The provided topology file `tyr2ala.top` is based on the standard CH RMM topologies for alanine and tyrosine. You can also obtain a topology file containing hybrid amino acids for any point mutation (except those involving proline) with the VMD plugin Mutator, available with VMD 1.8.5 and later. Recent releases of VMD include a database of hybrid topologies compatible with the CM P corrections[25] of the CH RMM force field. The Mutator plugin may be used to prepare hybrid protein topologies and coordinates suitable for alchemical FEP. The manual procedure that we follow here, however, is much more flexible, should one want to include specific patches in the structure.

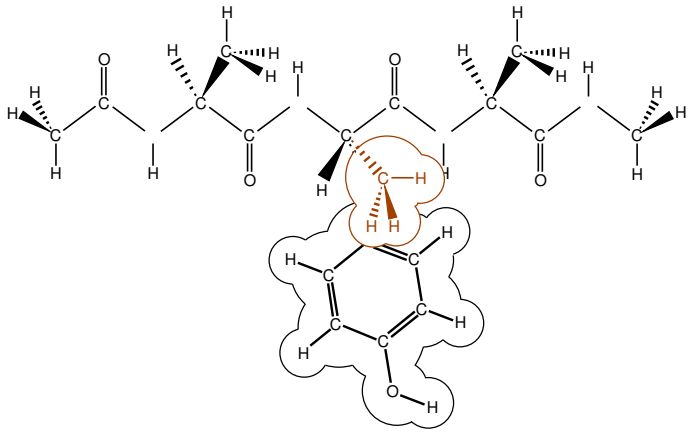

Figure 9: Dual topology hybrid molecule used for the la–Tyr– la $\rightarrow$ (la)$_3$ alchemical transformation. The initial state, *viz.* $\lambda = 0$ (black), and the final state *viz.* $\lambda = 1$ (brown), are defined concurrently. part from the two side chains, the chemical groups of the tripeptide are common to the two topologies.

```
* Topology for tyrosine-to-alanine transformation
27 1    ! Version: pretend we are CH RMM27b1

RESI Y2           0.00
GROUP
 TOM N    NH1    -0.47
 TOM HN   H       0.31
 TOM C    CT1     0.07
 TOM H    HB      0.09
GROUP
 TOM CB   CT2    -0.18
 TOM HB1  H       0.09
 TOM HB2  H       0.09
GROUP
 TOM CG   C       0.00
GROUP
 TOM CD1  C      -0.115
 TOM HD1  HP      0.115
GROUP
 TOM CE1  C      -0.115
 TOM HE1  HP      0.115
GROUP
 TOM CZ   C       0.11
 TOM OH   OH1    -0.54
 TOM HH   H       0.43
GROUP
 TOM CD2  C      -0.115
 TOM HD2  HP      0.115
GROUP
 TOM CE2  C      -0.115
 TOM HE2  HP      0.115
GROUP
 TOM CBB  CT3    -0.27
 TOM HB1B H       0.09
 TOM HB2B H       0.09
 TOM HB3B H       0.09
GROUP
 TOM C    C       0.51
 TOM O    O      -0.51
BOND N    HN
BOND N    C
BOND C    C
BOND C    +N
BOND C    H
BOND O    C
BOND CB   C
BOND CG   CB
```

```
BOND CD2  CG
BOND CE1  CD1
BOND CZ   CE2
BOND OH   CZ
BOND CB   HB1
BOND CB   HB2
BOND CD1  HD1
BOND CD2  HD2
BOND CE1  HE1
BOND CE2  HE2
BOND OH   HH
BOND CD1  CG
BOND CE1  CZ
BOND CE2  CD2
BOND CBB  C
BOND CBB  HB1B
BOND CBB  HB2B
BOND CBB  HB3B
IMPR N    -C    C     HN
IMPR C    C     +N    O
END
```

1 Create the PSF file by making the file `setup.pgn` containing the script below and then running it in the VMD TkConsole using `source setup.pgn`. Note that the first and third residues are standard alanine residues, so you have to use both the topology file from CH RMM27 and the custom hybrid topology.

```
package require psfgen
topology ../common/top_all22_prot.inp
topology tyr2ala.top

# Build the topology of both segments
segment Y2  {
  pdb tyr2ala.pdb
  first  CE
  last CT3
}
# The sequence of this segment is  la-Y2 - la

# Read coordinates from pdb files
coordpdb tyr2ala.pdb Y2

writepsf y2a.psf
writepdb y2a.pdb
```

2 Create the `alchFile` by opening `y2a.pdb` in a text editor and editing the B column to reflect which atoms vanish and appear. Save the file as `y2a.fep`. The modified part should read:

```
...
 TOM     17  N   Y2      2       5.841  -1.926  -3.336  1.00  0.00      YTO  N
 TOM     18  HN  Y2      2       5.362  -2.371  -4.106  1.00  0.00      YTO  H
 TOM     19  C   Y2      2       7.291  -1.926  -3.336  1.00  0.00      YTO  C
 TOM     20  H   Y2      2       7.655  -0.898  -3.336  1.00  0.00      YTO  H
 TOM     21  CB  Y2      2       7.842  -2.640  -2.100  1.00 -1.00      YTO  C
```

```
TOM     22 HB1  Y2      2       7.014  -2.994  -1.485  1.00 -1.00      YTO  H
TOM     23 HB2  Y2      2       8.452  -3.487  -2.411  1.00 -1.00      YTO  H
TOM     24  CG  Y2      2       8.687  -1.679  -1.298  1.00 -1.00      YTO  C
TOM     25 CD1  Y2      2       8.856  -0.360  -1.739  1.00 -1.00      YTO  C
TOM     26 HD1  Y2      2       8.377  -0.028  -2.660  1.00 -1.00      YTO  H
TOM     27 CE1  Y2      2       9.640   0.531  -0.996  1.00 -1.00      YTO  C
TOM     28 HE1  Y2      2       9.771   1.557  -1.339  1.00 -1.00      YTO  H
TOM     29  CZ  Y2      2      10.254   0.104   0.187  1.00 -1.00      YTO  C
TOM     30  OH  Y2      2      11.016   0.969   0.909  1.00 -1.00      YTO  O
TOM     31  HH  Y2      2      11.063   1.844   0.516  1.00 -1.00      YTO  H
TOM     32 CD2  Y2      2       9.302  -2.106  -0.115  1.00 -1.00      YTO  C
TOM     33 HD2  Y2      2       9.170  -3.132   0.227  1.00 -1.00      YTO  H
TOM     34 CE2  Y2      2      10.086  -1.215   0.627  1.00 -1.00      YTO  C
TOM     35 HE2  Y2      2      10.564  -1.547   1.548  1.00 -1.00      YTO  H
TOM     36  CBB Y2      2       7.842  -2.640  -2.100  1.00  1.00      YTO  C
TOM     37 HB1B Y2      2       7.014  -2.994  -1.485  1.00  1.00      YTO  H
TOM     38 HB2B Y2      2       8.452  -3.487  -2.411  1.00  1.00      YTO  H
TOM     39 HB3B Y2      2       8.687  -1.679  -1.298  1.00  1.00      YTO  C
TOM     40  C   Y2      2       7.842  -2.640  -4.572  1.00  0.00      YTO  C
TOM     41  O   Y2      2       7.078  -3.122  -5.407  1.00  0.00      YTO  O
...
```

3 Visualize the system containing the hybrid amino-acid. Run VMD with the following command: `vmd y2a.psf -pdb y2a.fep`.

4 In the Graphics/Representations menu, set the coloring method to `Beta`. The appearing alanine side chain should be colored blue and the tyrosine side chain should be red, while the backbone and the two unperturbed alanine residues should be green. Compare the result with Figure 9.

5 You are now ready to prepare the hydrated system. Load the isolated tripeptide in VMD: `vmd y2a.psf y2a.pdb`.

6 Open the Solvate interface (Extensions/Modeling/ dd Solvation Box).

7 Define a cubic, $30\times30\times30$ $^{\circ}3$ water box: Uncheck the "Use Molecule Dimensions" box and the "Waterbox Only", set the minimum value of x, y and z to 13 and their maximum to $+13$. Run the plugin to create the files `solvate.psf` and `solvate.pdb`.

8 Prepare a new `alchFile` for the solvated structure, by opening `solvate.pdb` in a text editor and editing the B column to reflect the same information you entered in `y2a.fep` for the unsolvated structure. Save the new `alchFile` as `solvate.fep`.

3.2. Running the free energy calculations

You can find the configuration files for the isolated and solvated systems in the `In-vacuo` and `In-aqua` folders respectively.

Traditional MD is configured to run at a constant temperature of 300 K, with cutoff electrostatics and no particular boundary conditions. The FEP sections are configured based on several considerations.

The *in vacuo* transformation requires relatively long sampling times, because there are no solvent fluctuations that could couple to conformational fluctuations of the peptide (see Results). Fortunately, for such a small system, a 0.5-nanosecond trajectory can be generated very quickly on a single processor. Here, we will use 20 contiguous windows, involving 25 ps of MD sampling — interspersed with 2 ps of equilibration (carried out in steps 40 to 42):

```
# FEP P R METERS

source                  ../../tools/fep.tcl

alch                    on
alchType                FEP
alchFile                y2a.fep
alchCol                 B
alchOutFile             forward.fepout
alchOutFreq             10

alchVdwLambdaEnd        1.0
alchElecLambdaStart     0.5
alchVdWShiftCoeff       4.0
alchDecouple            off

alchEquilSteps          4000
set numSteps            50000


runFEP 0.0 1.0 0.05 $numSteps
```

We will use a similar strategy for the solvated system, albeit with a somewhat larger time step. In each window, the system is equilibrated over `alchEquilSteps` MD steps, *viz.* here 100 steps, prior to 400 steps of data collection, making a total of 0.5 ps of MD sampling. ltogether, the alchemical transformation is carried out over 10 ps, and the backward transformation over the same time. (carried out in steps 43 to 45)

```
# FEP P R METERS

source                  ../../tools/fep.tcl

alch                    on
alchType                FEP
alchFile                solvate.fep
alchCol                 B
alchOutFile             forward-off.fepout
alchOutFreq             10

alchVdwLambdaEnd        1.0
alchElecLambdaStart     0.5
alchVdWShiftCoeff       4.0
alchDecouple            off

alchEquilSteps          100
set numSteps            500

runFEP 0.0 1.0 0.05 $numSteps
```

1 Navigate to the `In-vacuo` folder.

2 First, run `equilibration.namd` in the command prompt:
`namd2 equilibration.namd > equilibration.log`

3 Next, repeat the above command for `forward.namd` and `backward.namd` to generate the required data files for analysis.

4 Now navigate to the `In-aqua` folder. Note that the difference between the files suffixed by `-on` and those by `-off` lies in the `alchDecouple` switch. For now you will work with the files with suffix `off`.

The `alchDecouple` switch determines whether or not the vanishing and appearing moieties interact with each other. Switching decoupling off in this example causes the perturbed moieties to interact not only with the environment but also with each other. We discuss the use of this option at greater length in the Results section.

5 Perform the equilibration run.

6 Perform the forward and backward simulations.

3.3. Results

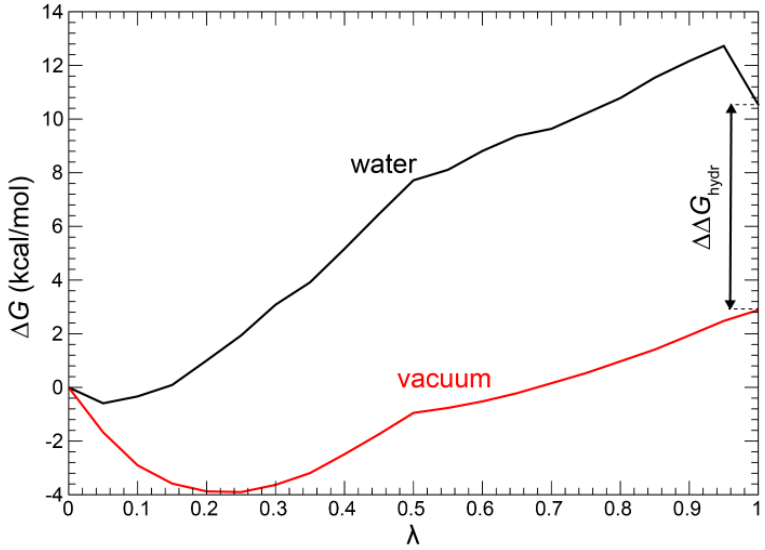

Figure 10: Results for the Tyr → la mutations in water and in vacuum, in the la–Tyr– la blocked tripeptide. The difference between the corresponding free-energy changes yields the relative hydration free energy of la–Tyr– la with respect to (la)$_3$. The transformation in vacuum accounts for intramolecular interactions between the perturbed moieties, *i.e.* `alchDecouple off`.

1 Use ParseFEP to extract the free energy profiles from the `alchOutFiles` generated by your forward and backward simulations.

The $G(\lambda)$ curves for both simulations, as well as an alternate method, are shown in Figure 10. Using an adapted protocol for each of the two mutations, the free energy difference for the hydrated state is $+10.5$ kcal/mol, and $+2.9$ kcal/mol for the isolated state; your values may be slightly different due to the very short simulation times used here. Using the thermodynamic cycle of Figure 8, one may write:

$$G = \quad G^2_{\text{alch.}} \qquad G^1_{\text{alch.}} = \quad G^2_{\text{hydr.}} \qquad G^1_{\text{hydr.}}$$

The net solvation free energy change G for the la–Tyr– la → (la)$_3$ transformation is found to be $+7.6$ kcal/mol. This result may be related to the differential hydration free energy of side–chain analogues, *i.e.* the difference in the hydration free energy of methane and p–cresol, that is, respectively,

1.9 + 6.1 = +8.0 kcal/mol.[26,27] Interestingly enough, Scheraga and coworkers have estimated the side–chain contribution for this mutation to be equal to +8.5 kcal/mol.[28]

This very close agreement with experimental determinations based on side–chain analogues, as well as other computational estimates, may be in part coincidental or due to fortuitous cancellation of errors. Indeed, some deviation could be expected due to environment effects — *viz.* the mutation of a residue embedded in a small peptide chain *versus* that of an isolated, prototypical organic molecule[29] — and, to a lesser extent, the limited accuracy of empirical force fields. The first explanation may be related to the concept of "self solvation" of the side chain. Here, the tyrosyl fragment is not only solvated predominantly by the aqueous environment, but also, to a certain degree, by the peptide chain, which, under certain circumstances, can form hydrogen bonds with the hydroxyl group.

Moreover, it should be noted that even for a small and quickly relaxing system such as the hybrid tripeptide, convergence of the FEP equation requires a significant time. In some cases, very short runs may give better results than moderately longer ones, because the former provide a local sampling around the starting configuration, while the latter start exploring nearby conformations, yet are not long enough to fully sample them.

We note here that the `alchDecouple` switch controls whether or not interactions *within each perturbed group* are scaled or not. If you set `alchDecouple` to `on`, then these interactions will not be scaled. This presumes that the intramolecular interactions are identical in the gas phase and in water. **This is hardly ever true, except for rigid molecules.** In the case of the Tyr → la differential hydration free energy, there is yet another subtlety that necessitates the need for a gas phase calculation, namely the need to account for the interactions between the central Tyr/ la and the flanking la residues. Thus, though tempting as it obviates the need to carry out a separate simulation in the gas phase, the `alchDecouple on` option should be avoided. **s a matter of principle, always set `alch ecouple` to `off` and perform both gas phase and condensed phase transformations.**

In all cases, visualizing MD trajectories is strongly advisable if one wishes to understand the behavior of the system and to solve possible sampling issues. Looking at the present tyrosine-to-alanine trajectories, it appears that the main conformational degree of freedom that has to be sampled is the rotation of the tyrosine hydroxyl group. Convergence is actually faster for the solvated system than for the tripeptide in vacuum, because fluctuations of the solvent help the tyrosine side chain pass the rotational barriers, which does not happen frequently in vacuum.