A tutorial to set up alchemical free energy perturbation calculations in NAMD

Surjit B. Dixit and Christophe Chipot

Equipe de chimie et biochimie théoriques,
UMR CNRS/UHP 7565,
Institut nancéien de chimie moléculaire,
Université Henri Poincaré,
BP 239,
54506 Vandœuvre–lès–Nancy cedex, France

Version: July 23, 2002

1. Introduction

The main goal of this tutorial is to provide a guidance when setting up free energy calculations of alchemical transformations within NAMD.[1] As has been commented on amply, such *in silico* experiments have not reached yet the maturity to be viewed as black–box, routine jobs.[2] Either the set–up, the sampling protocol, or the analysis of the result should be considered with great care. The paradigm chosen in NAMD for performing alchemical transformations is the so–called *dual–topology* approach,[3] wherein both the initial state, *viz.* $\lambda = 0$, and the final state, *viz.* $\lambda = 1$, are defined concurrently. As the molecular dynamics (MD) simulation progresses, the potential energy function characteristic of $\lambda = 0$ is scaled into that representative of $\lambda = 1$. Whereas the initial and the final states do not see each other in the course of the transformation, they, however, interact with the environment. The implication of these conditions is that a list of excluded atoms should be defined in the `psf` topology file. At the present time, PSFGEN does not permit the construction of such a list in an easy and straightforward fashion. The set–up of the two examples described hereafter will, therefore, be done within CHARMM.[4,5]

2. Setting up the system

Perhaps the simplest alchemical transformation one could imagine, the result of which is completely independent of the potential energy function utilized, is the *zero–sum* ethane $\rightarrow$ ethane mutation,[6,3] wherein a methyl group vanishes at one end of the molecule, while another one appears at the other end. The accuracy of the computed free energy depends solely upon the sampling protocol adopted, regardless of the force field employed.

As can be seen in Figure 1, the hybrid defined for this transformation is a propane molecule, consisting of a juxtaposition of two ethane fragments, with a common $–CH_2–$ moiety. To prevent the initial state, *viz.* $\lambda = 0$, from interacting with the final state, *viz.* $\lambda = 1$, the atoms pertaining to the latter should be excluded. In CHARMM. this can be achieved by declaring explicitly, next to the definition of each atom, those atoms of the hybrid molecule that should not interact. In the CHARMM definition below, atoms `CF`, `HF1`, `HF2`, `HF3` and `HF` of the final state will not see

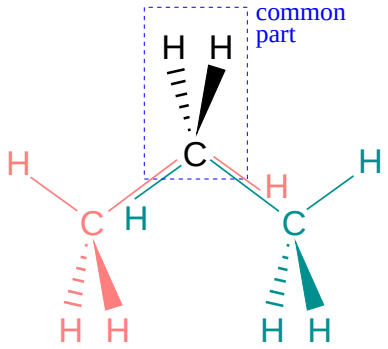

Figure 1: Dual–topology hybrid molecule used for the *zero–sum* ethane → ethane alchemical transformation. The initial state, *viz.* $\lambda = 0$ (pink), and the final state *viz.* $\lambda = 1$ (cyan), are defined concurrently. The central $–CH_2–$ moiety is common to the two topologies.

atoms `CI`, `HI1`, `HI2`, `HI3` and `HI` of the initial state. Alternatively, by uncommenting the list of atoms representative of the final state, and commenting out those of the initial state, the reverse transformation would take place.

```
read rtf card append

RESI ZERO        0.00                          ! ethane -> ethane
GROUP                                          !
ATOM CI   CT3   -0.27      CF  HF1 HF2 HF3 HF  !
ATOM HI1  HA     0.09      CF  HF1 HF2 HF3 HF  !
ATOM HI2  HA     0.09      CF  HF1 HF2 HF3 HF  !
ATOM HI3  HA     0.09      CF  HF1 HF2 HF3 HF  !
GROUP                                              HI1       HM1   HF2  HF3
ATOM CM   CT3   -0.27                          !      \        |     |  /
ATOM HM1  HA     0.09                          !       \HF     |     | /
ATOM HM2  HA     0.09                          !        CI----CM----CF
ATOM HI   HA     0.09      CF  HF1 HF2 HF3 HF  !       / |     |     HI\
ATOM HF   HA     0.09   !  CI  HI1 HI2 HI3 HI  !      /  |     |        \
GROUP                                              HI2  HI3  HM2        HF1
ATOM CF   CT3   -0.27   !  CI  HI1 HI2 HI3 HI  !
ATOM HF1  HA     0.09   !  CI  HI1 HI2 HI3 HI  !
ATOM HF2  HA     0.09   !  CI  HI1 HI2 HI3 HI  !
ATOM HF3  HA     0.09   !  CI  HI1 HI2 HI3 HI  !
BOND     CI  HI1           CI  HI2           CI   HI3           ! ethane 1
BOND     CF  HF1           CF  HF2           CF   HF3           ! ethane 2
BOND     CI  CM            CF  CM                               ! common
BOND     CM  HM1           CM  HM2                              ! common
BOND     CM  HI                                                 ! ethane 1
```

```
BOND     CM  HF                                                    ! ethane 2
ANGLE    HI1 CI  CM        HI2 CI  CM        HI3 CI  CM          ! ethane 1
ANGLE    HI1 CI  HI2       HI1 CI  HI3       HI2 CI  HI3         ! ethane 1
ANGLE    HF1 CF  CM        HF2 CF  CM        HF3 CF  CM          ! ethane 2
ANGLE    HF1 CF  HF2       HF1 CF  HF3       HF2 CF  HF3         ! ethane 2
ANGLE    CF  CM  HF        CI  CM  HI                            ! common
ANGLE    HM1 CM  HM2                                             ! common
ANGLE    HM1 CM  CI        HM1 CM  CF                            ! common
ANGLE    HM1 CM  HF        HM1 CM  HI                            ! common
ANGLE    HM2 CM  CI        HM2 CM  CF                            ! common
ANGLE    HM2 CM  HF        HM2 CM  HI                            ! common
DIHEDRAL HI1 CI  CM  HM1   HI1 CI  CM  HM2   HI1 CI  CM  HI     ! ethane 1
DIHEDRAL HI2 CI  CM  HM1   HI2 CI  CM  HM2   HI2 CI  CM  HI     ! ethane 1
DIHEDRAL HI3 CI  CM  HM1   HI3 CI  CM  HM2   HI3 CI  CM  HI     ! ethane 1
DIHEDRAL HF1 CF  CM  HM1   HF1 CF  CM  HM2   HF1 CF  CM  HF     ! ethane 2
DIHEDRAL HF2 CF  CM  HM1   HF2 CF  CM  HM2   HF2 CF  CM  HF     ! ethane 2
DIHEDRAL HF3 CF  CM  HM1   HF3 CF  CM  HM2   HF3 CF  CM  HF     ! ethane 2

END
```

The second, less trivial application of alchemical free energy calculations consists in mutating in a short peptide the side chain of an amino acid. The example of the terminally blocked Ala–Ser–Ala tripeptide was chosen, in which the L–serine (Ser) residue was transformed into L–alanine (Ala).

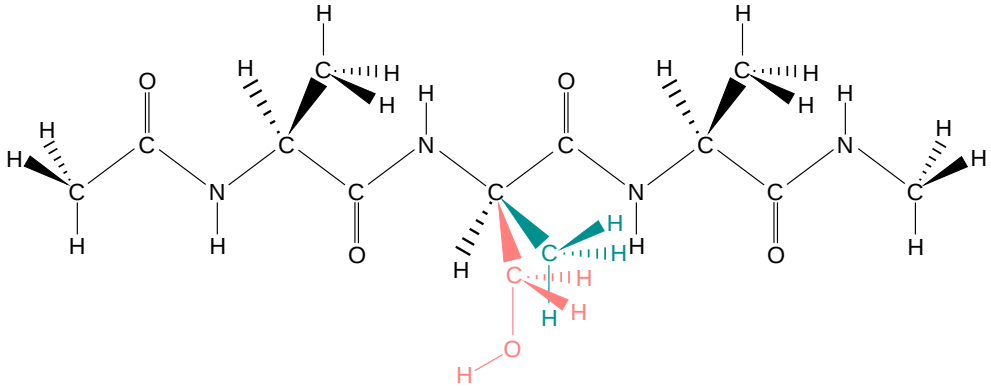

Figure 2: Dual–topology hybrid molecule used for the Ala–Ser–Ala → $(Ala)_3$ alchemical transformation. The initial state, *viz.* $\lambda = 0$ (pink), and the final state *viz.* $\lambda = 1$ (cyan), are defined concurrently. Apart from the two side chains, the chemical groups of the tripeptide are common to the two topologies.

Here again, one must check that the side chain of the wild type does not see that of the mutant. To

achieve this condition, a hybrid amino acid should be defined, wherein those atoms characteristic of the final state, *viz.* $\lambda = 1$, are excluded from the list of atoms representative of the initial state, *viz.* $\lambda = 0$ — see Figure 2. In the CHARMM definition of the topology, atoms CBB, HB1B, HB2B and HB3B pertaining to Ala will not interact with atoms CBA, HB1A, HB2A, OGA and HG1A of Ser.

```
read rtf card append
! serine/alanine hybrid
27   1

DECL -CA
DECL -C
DECL -O
DECL +N
DECL +HN
DECL +CA

DEFA FIRS NTER LAST CTER
AUTO ANGLES DIHE

RESI STA          0.00
GROUP
ATOM N    NH1    -0.47                              !     |
ATOM HN   H       0.31                              !  N--N    HB1B
ATOM CA   CT1     0.07                              !     |    HB1A
ATOM HA   HB      0.09                              !     |    |     HB3
GROUP                                               ! HA--CA---CBA---OGA
ATOM CBA  CT2     0.05   CBB  HB1B HB2B HB3B        !     |    CBB      \
ATOM HB1A HA      0.09   CBB  HB1B HB2B HB3B        !     |    |         HG1A
ATOM HB2A HA      0.09   CBB  HB1B HB2B HB3B        !  O==C    HB2A
ATOM OGA  OH1    -0.66   CBB  HB1B HB2B HB3B        !     |
ATOM HG1A H       0.43   CBB  HB1B HB2B HB3B        !  serine part
GROUP                                               !
ATOM CBB  CT3    -0.27 ! CBA  HB1A HB2A OGA  HG1A !
ATOM HB1B HA      0.09 ! CBA  HB1A HB2A OGA  HG1A !
ATOM HB2B HA      0.09 ! CBA  HB1A HB2A OGA  HG1A !
ATOM HB3B HA      0.09 ! CBA  HB1A HB2A OGA  HG1A !  alanine part
GROUP
ATOM C    C       0.51
ATOM O    O      -0.51
BOND N    HN    N    CA                ! common
BOND C    CA    C    +N    CA   HA     ! common
```

```
BOND CBA  CA    OGA  CBA   OGA  HG1A   ! serine
BOND HB1A CBA   HB2A CBA               ! serine
BOND CBB  CA                           ! alanine
BOND HB1B CBB   HB2B CBB   HB3B CBB    ! alanine
DOUBLE    O  C
IMPR N -C CA HN  C CA +N O
DONOR HN N
DONOR HG1A OGA
ACCEPTOR OGA
ACCEPTOR O C

PRES HYB    0.00
DELE ANGLE  CBA  CA   CBB  ! remove angle between side chains

END
```

Interestingly enough, atoms belonging to the backbone are common to the initial and the final states. Only the side chain of the hybrid residue is affected by the alchemical transformation, and, therefore, requires a dual–topological definition. In addition, considering that only the central amino acid of the blocked tripeptide is modified in the course of the free energy calculation, the standard Ala residues should be invoked in the CHARMM script when building the sequence.

```
read sequence card
* title
*
1
ALA STA ALA
```

It should be mentioned that as the potential energy function is scaled when λ varies from 0 to 1, the side chain of the mutant might flip, thereby resulting in the wrong chirality for the final state of the transformation. To circumvent this undesirable effect, the valence angle formed by the β–carbon atoms of the side chains and the common α–carbon atom should be restrained to zero.

3. Setting up the free energy calculations

Now that the topology of the hybrid molecule is defined, we will detail how the free energy calculation proceeds in NAMD. Execution of the CHARMM script will generate the `psf` topology file used in the MD simulation. One should remember, however, that, on account of the dual–topology paradigm, both the initial state and the final states of the alchemical transformation are present simultaneously. It is, therefore, pivotal that the information about the nature of these states be passed to NAMD, indicating which atoms of the hybrid correspond to $\lambda = 0$, and similarly, which correspond to $\lambda = 1$. This information is given by `fepFile`, a file written in the PDB format, wherein a `-1.00` or `1.00` flag characterizes those atoms of the hybrid molecule that, respectively, vanish or appear in the course of the simulation:

```
ATOM      1  CI  ZERO    1      -1.167   0.224   0.034  1.00 -1.00      ZERO
ATOM      2  HI1 ZERO    1      -2.133  -0.414   0.000  1.00 -1.00      ZERO
ATOM      3  HI2 ZERO    1      -1.260   0.824   0.876  1.00 -1.00      ZERO
ATOM      4  HI3 ZERO    1      -1.258   0.825  -0.874  1.00 -1.00      ZERO
ATOM      5  CM  ZERO    1       0.001  -0.652  -0.002  1.00  0.00      ZERO
ATOM      6  HM1 ZERO    1       0.000  -1.313  -0.890  1.00  0.00      ZERO
ATOM      7  HM2 ZERO    1       0.005  -1.308   0.889  1.00  0.00      ZERO
ATOM      8  HI  ZERO    1       1.234   0.192   0.000  1.00 -1.00      ZERO
ATOM      9  HF  ZERO    1      -1.237   0.190   0.000  1.00  1.00      ZERO
ATOM     10  CF  ZERO    1       1.289   0.150  -0.078  1.00  1.00      ZERO
ATOM     11  HF1 ZERO    1       2.149  -0.425  -0.001  1.00  1.00      ZERO
ATOM     12  HF2 ZERO    1       1.256   0.837  -0.893  1.00  1.00      ZERO
ATOM     13  HF3 ZERO    1       1.131   0.871   0.940  1.00  1.00      ZERO
```

The flag that distinguishes between “growing” and “shrinking” atoms can be declared in either the X, Y, Z, O or B column of `fepFile`. In the case of the *zero–sum* ethane → ethane transformation, atoms `CI`, `HI1`, `HI2`, `HI3` and `HI` of the initial state vanish, as atoms `CF`, `HF1`, `HF2`, `HF3` and `HF` of the final state appear. A `0.00` flag is assigned to those atoms that are left unchanged as λ varies from 0 to 1.

```
# FEP

fep                     on
```

```
fepFile                 zero.fep
fepCol                  B
fepOutFile              zero.fepout
fepOutFreq              5
FepEquilbSteps          3200


# LOOP OVER LAMBDA-STATES -- TCL

set step                0.00
set dstep               0.025
dlambda                 $dstep

while {$step <= 1.00} {
   firsttimestep  0
   lambda $step
   lambda2 [expr $step+$dstep]
   run 6400
   set step [expr $step+$dstep]
}
```

TCL scripts allow to set up the protocol of the free energy calculation in a straightforward, user–friendly fashion. In the above example, the potential energy function of the system is scaled from $\lambda = 0$ to $\lambda = 1$ by increments $\delta\lambda = 0.025$, *i.e.* 40 intermediate λ–states or "windows".[7] In each "window", the system is equilibrated over `FepEquilbSteps` MD steps, *viz.* here 3,200 MD steps, prior to 6,400 MD steps of data collection, from which the ensemble average is evaluated. Using this protocol, the free energy varies smoothly as λ progresses from 0 to 1, and the the net free energy change is +0.03 kcal/mol, altogether suggestive that the convergence of the simulation is appropriate.

In the second example, wherein the terminally blocked Ala–Ser–Ala tripeptide is mutated into $(Ala)_3$, atoms `CBA`, `HB1A`, `HB2A`, `OGA` and `HG1A` vanish, as atoms `CBB`, `HB1B`, `HB2B` and `HB3B` appear:

```
ATOM     13  N   STA     2      -1.035  -1.123   0.248  1.00  0.00      AHA
ATOM     14  HN  STA     2      -1.281  -1.562  -0.608  1.00  0.00      AHA
ATOM     15  CA  STA     2       0.229  -1.546   0.803  1.00  0.00      AHA
ATOM     16  HA  STA     2      -0.005  -1.844   1.799  1.00  0.00      AHA
```

```
ATOM     17  CBA STA     2       0.933  -2.644   0.012  1.00 -1.00      AHA
ATOM     18 HB1A STA     2       1.827  -2.977   0.506  1.00 -1.00      AHA
ATOM     19 HB2A STA     2       0.225  -3.507  -0.122  1.00 -1.00      AHA
ATOM     20  OGA STA     2       1.263  -2.022  -1.274  1.00 -1.00      AHA
ATOM     21 HG1A STA     2       1.853  -1.309  -1.234  1.00 -1.00      AHA
ATOM     22  CBB STA     2       0.937  -2.495  -0.279  1.00  1.00      AHA
ATOM     23 HB1B STA     2       1.814  -2.963   0.480  1.00  1.00      AHA
ATOM     24 HB2B STA     2       0.236  -3.490  -0.132  1.00  1.00      AHA
ATOM     25 HB3B STA     2       1.175  -2.169  -1.095  1.00  1.00      AHA
ATOM     26  C   STA     2       1.207  -0.431   1.069  1.00  0.00      AHA
ATOM     27  O   STA     2       1.752  -0.299   2.140  1.00  0.00      AHA
```

Because the backbone atoms of the hybrid amino acid are left unchanged throughout the alchemical transformation, a `0.00` flag is assigned to them in the B column of `fepFile`. It is crucial, here, that the declarations in `fepCol` coincides with the list of excluded atoms defined in the CHARMM set–up script. In this particular example, the initial state corresponds to Ala–Ser–Ala, in which the atoms of the L–serine side chain, *viz.* λ =0, do not see those of L–alanine, *viz.* λ = 1. Accordingly, atoms `CBA`, `HB1A`, `HB2A`, `OGA` and `HG1A` in `fepFile` should be assigned a `-1.00` flag, while the flag for atoms `CBB`, `HB1B`, `HB2B` and `HB3B` should be `1.00`.

```
# FEP

fep                     on
fepFile                 serine.fep
fepCol                  B
fepOutFile              serine.fepout
fepOutFreq              5
FepEquilbSteps          6400


# LOOP OVER LAMBDA-STATES -- TCL

set step                0.00
set dstep               0.025
dlambda                 $dstep

while {$step <= 1.00} {
   firsttimestep  0
   lambda $step
```

```
   lambda2 [expr $step+$dstep]
   run 12800
   set step [expr $step+$dstep]
}
```

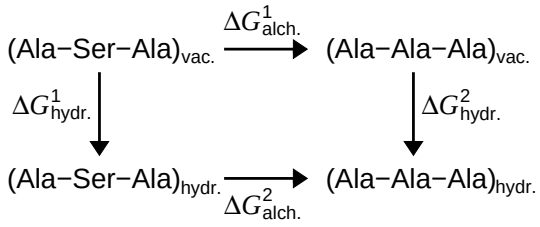

Figure 3: Thermodynamic cycle used in the Ala–Ser–Ala $\rightarrow$ $(Ala)_3$ alchemical transformation. The vertical arrows correspond to the hydration of the wild–type tripeptide and its mutant. The horizontal arrows correspond to the point mutation in bulk water and *in vacuo*, so that: $\Delta G^2_{\rm alch.} - \Delta G^1_{\rm alch.} = \Delta G^2_{\rm hydr.} - \Delta G^1_{\rm hydr.}$.

Just like in the *zero–sum* ethane $\rightarrow$ ethane alchemical transformation, use is made here of 40 "windows" to connect the initial state, *viz.* Ser, to the final state, *viz.* Ala, of the mutation. Here, each individual λ–state consists, however, of 6,400 MD steps of equilibration, followed by 12,800 MD steps of data collection, from which the ensemble average is computed. In order to estimate the free energy change involved in the point mutation of the hydrated tripeptide, the same simulation should be carried out in bulk water and *in vacuo*, to close the thermodynamic cycle of Figure 3.[8] Adopting the same sampling protocol for the two legs of the mutation, the free energy difference for the hydrated state is +1.80 kcal/mol, and -5.57 kcal/mol for the isolated state, thus yielding a net free energy change equal to +7.37 kcal/mol for the overall Ala–Ser–Ala $\rightarrow$ $(Ala)_3$ transformation. In sharp contrast with the preceding example, the accuracy of the free energy difference associated to this alchemical transformation inherently depends upon the quality of the potential energy function utilized. Deconvoluting the error in such calculations in terms of (i) inadequacy of the force field, and, (ii) sampling of inappropriate length, constitutes a daunting task. A close agreement with the experimental value may very well be the fortuitous result of an insufficient sampling and a poorly parameterized potential energy function.[9,2] Here, the smoothly changing free energy and the nice accord with the experimental estimate of +7.02 kcal/mol,[10,11] not only suggest that the simulation has converged, but also that the parameters used are well–adapted to the problem tackled.

References

[1] S. B. Dixit and C. Chipot. Can absolute free energies of association be estimated from molecular mechanical simulations ? The biotin–streptavidin system revisited. *J. Phys. Chem. A*, 105:9795–9799, 2001.

[2] C. Chipot and D. A. Pearlman. Free energy calculations. The long and winding gilded road. *Mol. Sim.*, 28:1–12, 2002.

[3] D. A. Pearlman. A comparison of alternative approaches to free energy calculations. *J. Phys. Chem.*, 98:1487–1493, 1994.

[4] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, 4:187–217, 1983.

[5] A. D. MacKerell Jr., D. Bashford, M. Bellott, R. L. Dunbrack Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus. All–atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, 102:3586–3616, 1998.

[6] D. A. Pearlman and P. A. Kollman. The overlooked bond–stretching contribution in free energy perturbation calculations. *J. Chem. Phys.*, 94:4532–4545, 1991.

[7] D. L. Beveridge and F. M. DiCapua. Free energy via molecular simulation: A primer. In W. F. Van Gunsteren and P. K. Weiner, editors, *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*, pages 1–26. Escom, The Netherlands, 1989.

[8] P. A. Kollman. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.*, 93:2395–2417, 1993.

[9] D. A. Pearlman and P. A. Kollman. Free energy perturbation calculations: Problems and pitfalls along the gilded road. In W. F. Van Gunsteren and P. K. Weiner, editors, *Computer*

Simulation of Biomolecular Systems: Theoretical and Experimental Applications, pages 101–119. Escom, The Netherlands, 1989.

[10] P. A. Bash, U. C. Singh, R. Langridge, and P. A. Kollman. Free energy calculations by computer simulation. *Science*, 236:564–568, 1987.

[11] R. Wolfenden, L. Andersson, P. M. Cullins, and C. C. B. Southgate. Affinities of amino acid side chains for solvent water. *Biochemistry*, 20:849–855, 1981.