Tutorial: Accelerating Simulations of Large Systems Using Virtual Sites

Abhishek Acharya

June 28, 2019

I. Introduction

The aim of this tutorial is to give a demonstration of how virtual sites can be used for accelerating all-atom MD simulations, particularly helpful for systems with large proteins. Of course, there are other applications of virtual sites such as construction of complex toplogies for molecules or in hybrid all-atom/coarse-grained simulations. The topologies for TIP4P or TIP5P water models (included with the GROMACS distribution) would be instructive in this regard. Interested readers may also look at the MARTINI tutorial for running hybrid simulations using virtual sites (http://cgmartini.nl/index.php/tutorials-general-introduction/tutorial-hybrid-model-using-virtual-sites).

The key idea is to increase the largest possible timestep used for integrating the equations of motion with reasonable accuracy. As we know, the timestep allowed is limited by the fastest degrees of freedom present in the system. In usual all-atom MD setups, the fastest degrees of freedom are the bond stretching vibrations involving H-atoms (10 fs). These are usually treated using constrainsts. The next fastest DOF is that of bond-angle vibrations involving H-atoms (13 fs), which currently limits the largest timestep to 2 fs. Although, constrainst can be used to set bond-lengths and angles as constant, and increase the limit to about 5 fs, a nice way to run stable simulations with larger timesteps is to altogether remove these high-frequency motions involving hydrogen from the system. In this alternative approach, hydrogen atoms are treated as Virtual Interaction Sites (simply, dummy atoms) that are NOT connected to the parent atom with a bond. Instead, the positions of the dummy atoms are calculated at every MD step from the positions of nearby connected heavy atoms using geometrical rules. The force acting on a dummy atom is redistributed to the heavy atoms, and the positions of these heavy atoms evolve under the action of these forces. The hydrogen atoms treated as dummy particles do not have an associated mass; therefore, to keep the momentum of the system umchanged, the mass is added to the bonded heavy atoms. The bond, angle and dihedral terms describing the natural connection of the hydrogen atom is removed. But, in case of specific functional groups requiring rotational freedom (amine, hydroxyl) care must be taken to keep the rotational degree of freedom intact. We will go over the finer details of these aspects in sections below. For more details about virtual sites, the reader is referred to the GROMACS manual and the orginal paper on virtual sites [Feenstra, Hess and Berendsen.(1999).JCC.Vol. 20, No. 8, 786-7989].

Here on, we will give a brief overview of the types of virtual sites implemented in GROMACS. We will look at an example protein (trp-cage) to see how the topology is constructed when treating hydrogens as virtual sites. Note that for proteins, pdb2gmx can automatically generate the virtual site topology file using the -vsite flag. But, it would be instructive to look at the protein topology first to get an understanding of the toplogy construction. Eventually, it would be helpful in the second example where we construct the topology of a ligand with dummy hydrogen atoms by hand.

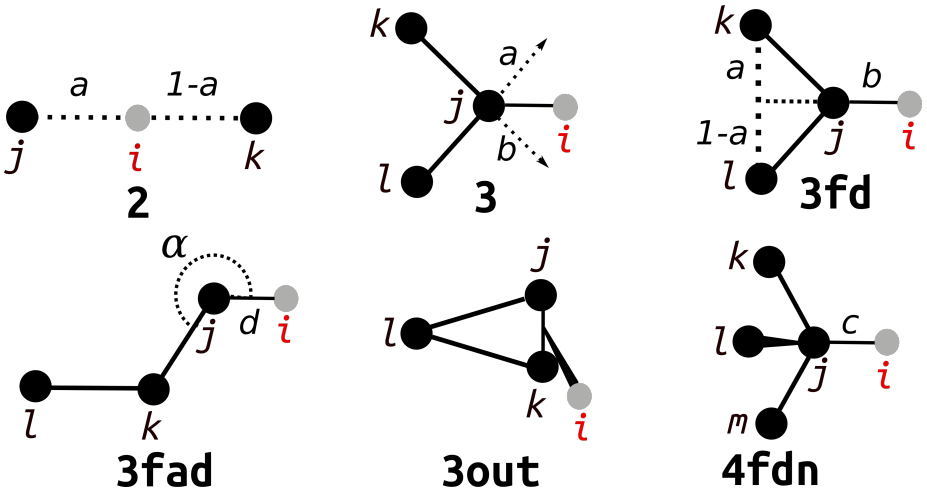

Figure 1: Types of virtual sites in GROMACS.The gray circles are dummy atoms and the black ones are the constructing atoms.

II. Types of Virtual Sites in GROMACS

There are six types of virtual site construction possible in GROMACS. For particular types of hydrogen atoms (polar connected to oxygen, non-polar connected to sp3 carbon etc.) particular constuctions can be used. This will be more clear with appropriate examples for each type given below.

Figure 1 below shows the different types of virtual site construction available in GROMACS.

a. Type 2 Virtual Site

This type is constructed as a linear combination of two atoms. As shown in Figure 1, the virtual site lies in the line connecting the two constructing atoms. In the topology file, this construction is specified under the topology directive *[virtual_sites2]*. The first column is the index of the virtual atom and the next two columns specify the two constructing atoms. Finally, the last column specifies the function to be used, which in this case is 1.

This type is not used for construction of hydrogen atoms in proteins but may be useful in construction of virtual sites for other molecules. See the virtual site tutorial by J. Lemkul for an example of this type used for a Carbon-dioxide molecule (http://www.mdtutorials.com/gmx/vsites/index.html).

b. Type 3 Virtual Site

This type is constructed as a linear combination of three atoms (Refer to Figure 1). The virtual site in this case lies in the plane containing the three constructing atoms (i,j and k). In the topology file, this construction is specified under the topology directive *[virtual_sites3]*. The first column is the index for the virtual atom, while the next three columns specify the index of the three constructing atoms. Again, the final column here is the function type 1.

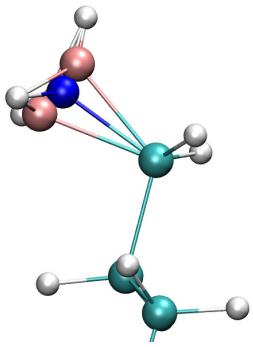

Figure 2: Construction of a NH3 group in Lys

In proteins, this type of construction can be used in case of umbrella-NH2 (freely-rotating amines), -NH3 and -CH3 groups.

c. Type 3fd Virtual Site

This type has the virtual site constructed in the plane of the three constructing atoms (*j,k* and *l*), but at a fixed distance b from the adjacent bonded atom j (Refer to Figure 1). In the topology file, this construction is specified under the topology directive *[virtual_sites3]*. The first column is the index for the virtual atom, while the next three columns specify the index of the three constructing atoms. Again, the final column here is the function type 2.

Examples of such a construction in proteins can be observed in case of backbone -NH- group, hydrogens of aromatic rings (Phe,Trp,His,Tyr) and Nε hydrogen in Arg.

d. Type 3fad Virtual Site

For this type the virtual site is constructed in the plane of the three constructing atoms (*j,k* and *l*) at a fixed angle α with $r_{j,k}$ and distance d from the adjacent bonded atom j (Refer to Figure 1). Like other three atom constructions, this construction is also specified under the topology directive *[virtual_sites3]*. First four columns are the index for the virtual site and the three constructing atoms, respectively, and the fifth column is the function type 3.

This construction is employed for groups such as planar -NH2 in Asn and Gln and the two -NH2 moieties in the guanidino group of Arg.

e. Type 3out Virtual Site

This type of virtual site is constructed out of the plane containing the three constructing atoms (Refer to Figure 1). In the topology file, this construction is specified under the topology directive *[virtual_sites3]*. The first column is the

index for the virtual atom, while the next three columns specify the index of the three constructing atoms. The final column here is the function type 4.

In proteins, this type of construction can be used in case of -NH3 and -CH3 groups, and Gly CA hydrogens.

f. Type 4fdn Virtual Site

Here, the virtual site is constructed from 4 atoms (*j,k,l* and *m*) at a fixed distance *c* from the adjacent bonded atom *j*. In the topology file, this construction is specified under the topology directive *[virtual_sites4]*. The first column is the index for the virtual atom, while the next four columns specify the index of the four constructing atoms. The final column here is the function type 2 (function type 1 is also available, but it is not recommended as it can be unstable).

This type of construction is observed in case of the Cα hydrogen (exception is Gly Cα hydrogens which are constructed using 3out type construction).

III. Case 1: Simulations of Trpcage Protein

Now that we have looked at the various virtual site types, we will go into the details of how each construction type is implemented in the protein toplogy file, by taking the example of the Trpcage protein.

a. Treatment of different types of hydrogens in protein

a.i. Umbrella-NH2, -NH3 and -CH3 groups

We are going to take the example of -NH3 group of Lysine, to go over the details of topology construction. Note in Figure 2 the two dummy mass centers that are added. In the toplogy file (relevant sections printed below) you would see the following for Lysine.

```
[ atoms ]
...
163         CT      8    LYS     CE    149    -0.0143     14.026
164         HP      8    LYS    HE1    150     0.1135          0
165         HP      8    LYS    HE2    151     0.1135          0
166       MNH3      8    LYS   MNZ1    152          0      8.517
167       MNH3      8    LYS   MNZ2    152          0      8.517
168         N3      8    LYS     NZ    152    -0.3854          0
169          H      8    LYS    HZ1    153       0.34          0
170          H      8    LYS    HZ2    154       0.34          0
171          H      8    LYS    HZ3    155       0.34          0
...

[ bonds ]
...
160   163     1
163   164     1
163   165     1
163   168     1
```

```
168   169     1
168   170     1
168   171     1
...

[ constraints ]
...
163   166     2
163   167     2
166   167     2
...

[ pairs ]
...
162   164     1
162   165     1
162   168     1
164   169     1
164   170     1
164   171     1
165   169     1
165   170     1
165   171     1
...

[ angles ]
...
157   160   163     1
161   160   163     1
162   160   163     1
160   163   164     1
160   163   165     1
160   163   168     1
164   163   165     1
164   163   168     1
165   163   168     1
163   168   169     1
163   168   170     1
163   168   171     1
169   168   170     1
169   168   171     1
170   168   171     1
...

[ dihedrals ]
...
  157   160   163   164     9
  157   160   163   165     9
  157   160   163   168     9
  161   160   163   164     9
```

```
  161   160   163   165     9
  161   160   163   168     9
  162   160   163   164     9
  162   160   163   165     9
  162   160   163   168     9
  160   163   168   169     9
  160   163   168   170     9
  160   163   168   171     9
  164   163   168   169     9
  164   163   168   170     9
  164   163   168   171     9
  165   163   168   169     9
  165   163   168   170     9
  165   163   168   171     9
  ...

 [ virtual_sites3 ]
...
  168   163   166   167     1
  169   163   166   167     1
...
 [ virtual_sites3 ]
...
  170   163   166   167    -4
  171   163   166   167     4
...
```

Here above are the topology directives used to define virtual sites in case of NH3 of Lys.

You will notice that in order to construct the -NH3 group, additional dummy sites and contraints are required in the topology. The two dummy masses are added and constrained with respect to the Cϵ atom in a triangular formation. The total mass on the two dummy sites is just the mass of the NZ atom. Based on the positions of these three points (one Cϵ and the two dummy masses), we can calculate on the fly the position of the NZ atom considering the fact that the new atom NZ must lie on the line passing through and equidistant from both dummy mass points. This is the type 3 virtual site construction and would appear under the topology directive *[virtual_sites3]*, as shown above. Additional virtual site directives are required to provide the rules for generating the positions of the three hydrogens connected to NZ atom. We use Type 3 virtual site construction for the position of the hydrogen (atom 169) in the plane of the three constructing atoms. For the other two hydrogens (atoms 170 and 171) we generate the positions using the Type 3out directive. The sign associated with the function number 4 is used to denote whether the virtual site is constructed above or below the plane containing the constructing atoms.

As an exercise, you may inspect the topology for -CH3 group.

a.ii. Planar-NH2 group

For planar -NH2, we take the example of the amide group of Asn. Relevant sections of the topology are printed below.

```
[ atoms ]
...
    12          C      1    ASN     CG     10     0.5833      12.01
    13          O      1    ASN    OD1     11    -0.5744         16
    14          N      1    ASN    ND2     12    -0.8634     16.026
    15          H      1    ASN   HD21     13     0.4097          0
    16          H      1    ASN   HD22     14     0.4097          0
...

[ bonds ]
...
    9    12     1
   12    13     1
   12    14     1
   14    15     1
   14    16     1
...

[ pairs ]
...
   12    17     1
   13    15     1
   13    16     1
...

[ angles ]
...
   10     9    12     1
   11     9    12     1
    9    12    13     1
    9    12    14     1
   13    12    14     1
   12    14    15     1
   12    14    16     1
   15    14    16     1
...

[ dihedrals ]
...
    7     9    12    14     9    torsion_ASN_CA_CB_CG_ND2_mult1
    7     9    12    14     9    torsion_ASN_CA_CB_CG_ND2_mult2
    7     9    12    14     9    torsion_ASN_CA_CB_CG_ND2_mult3
    7     9    12    14     9    torsion_ASN_CA_CB_CG_ND2_mult4
    7     9    12    14     9    torsion_ASN_CA_CB_CG_ND2_mult5
    7     9    12    14     9    torsion_ASN_CA_CB_CG_ND2_mult6
```

```
    7     9    12    13     9
   10     9    12    13     9
   10     9    12    14     9
   11     9    12    13     9
   11     9    12    14     9
    9    12    14    15     9
    9    12    14    16     9
   13    12    14    15     9
   13    12    14    16     9
...

[ dihedrals ]
...
    9    14    12    13     4
   12    15    14    16     4
...

[ virtual_sites3 ]
...
   15    14    12     9     3
   16    14    12     9    -3
...

Here above are the topology directives used to define virtual sites in
case of NH2 group of Asn.
```

You will notice in the topology that the two hydrogens (15 and 16) are constructed using the Type 3fad virtual site directives. Here, the function has a sign associated which denotes the direction in which the angle is measured.

a.iii. Secondary amine or amide (-NH-) and aromatic -CH groups

This is a very straightforward. The relevant section of the topology is given below. As you can see, the virtual site type used for construction of the hydrogen is Type 3fd.

```
[ atoms ]
...
    17          C      1    ASN      C     15     0.6163      12.01
    18          O      1    ASN      O     16    -0.5722         16
    19          N      2    LEU      N     17    -0.4157     15.018
    20          H      2    LEU      H     18     0.2719          0
    21         CT      2    LEU     CA     19    -0.0518     13.018
...

[ bonds ]
...
   17    18     1
   17    19     1
```

```
   19    20     1
   19    21     1
   21    22     1
   21    23     1
   21    40     1
...

[ pairs ]
...
...

[ angles ]
...
    7    17    18     1
    7    17    19     1
   18    17    19     1
   17    19    20     1
   17    19    21     1
   20    19    21     1
   19    21    22     1
   19    21    23     1
   19    21    40     1
   22    21    23     1
   22    21    40     1
   23    21    40     1
   21    23    24     1
   21    23    25     1
   21    23    26     1

...

[ dihedrals ]
...
    7    17    19    20     9
    7    17    19    21     9
   18    17    19    20     9
   18    17    19    21     9
   17    19    21    22     9
   17    19    21    23     9
   17    19    21    40     9
   20    19    21    22     9
   20    19    21    23     9
   20    19    21    40     9
   40    21    23    26     9    torsion_LEU_C_CA_CB_CG_mult1
   40    21    23    26     9    torsion_LEU_C_CA_CB_CG_mult2
   40    21    23    26     9    torsion_LEU_C_CA_CB_CG_mult3
   19    21    23    24     9
   19    21    23    25     9
   19    21    23    26     9
   22    21    23    24     9
```

```
   22    21    23    25     9
   22    21    23    26     9
   40    21    23    24     9
   40    21    23    25     9
   19    21    40    41     9
   19    21    40    42     9
   22    21    40    41     9
   22    21    40    42     9
   23    21    40    41     9
   23    21    40    42     9

...

[ dihedrals ]
...
   17    21    19    20     4
...

[ virtual_sites3 ]
...
  20    19    17    21     2
...
```

Here above are the topology directives used to define virtual sites in case of -NH group of backbone amide linkages.

a.iv. Hydroxyl (-OH) and sulfhydryl (-SH) groups

The -OH and -SH groups are treated a little differently. As you will notice in the topology printed below, in case of Ser the polar hydrogen is not treated as a dummy atom. Your will find that the mass of the polar hydrogen (atom 222) is not set to zero. In this case, we just add an additional constraint on the C-O-H angle by putting distance constraints on C—H distance. No virtual site directive is needed in this case.

```
[ atoms ]
...
   218         CT     13    SER     CB    202     0.2117     14.026
   219         H1     13    SER    HB1    203     0.0352          0
   220         H1     13    SER    HB2    204     0.0352          0
   221         OH     13    SER     OG    205    -0.6546         16
   222         HO     13    SER     HG    206     0.4275      1.008
...

[ bonds ]
...
  218   219     1
  218   220     1
```

```
  218   221     1
  221   222     1
...

[ constraints ]
...
218   222     2
...

[ pairs ]
...
...

[ angles ]
...
    217   216   218     1
    217   216   223     1
    218   216   223     1
    216   218   219     1
    216   218   220     1
    216   218   221     1
    219   218   220     1
    219   218   221     1
    220   218   221     1
    218   221   222     1
...

[ dihedrals ]
...
  217   216   223   224     9
  217   216   223   225     9
  218   216   223   224     9
  218   216   223   225     9
  216   218   221   222     9
  219   218   221   222     9
  220   218   221   222     9
...
```

b. Setting up the Trpcage simulation

Now once the topology is ready, it is generally trivial to run the simulations. During equilibration steps, it is advised to start the NVT equilibration step using 2fs as the timestep; this can be followed by 2fs of NPT. Thereafter, additional equilibration runs at increasing timestep values or 4fs and finally 5fs, should be run. Finally, production runs can be done using 5fs as a timestep, which leads to at least 2.5 times accelation in simulations.

IV. Case 2: Simulations of a Ligand

Unlike for proteins, virtual site topology for ligand are not generated automatically by gromacs, and need to be built by hand. This is where the undestanding of virtual site topology construction obtained from previous section would be helpful.

a. Treating ligand hydrogens as virtual interaction sites

Building virtual site topology for ligands typically involves modifying a normal full atom topology. So, the following procedure assumes that one already has an optimized topology obtained using methods routinely used for a given forcefield. Once we have a normal working topology for the ligand of interest, we can work through the following steps to edit the file and generate a virtual site topology. For the current demonstration, we take the example of N-Acetyl Glucosamine-1-phosphate (GlcNAc1P).

Step 1: Inspect the ligand structure. We study the ligand structure and make note of the different types of hydrogens that are present in the ligand. This will give us an idea of the different types of virtual site directives we need in the topology file. On inspection, you will find the following types of hydrogen (see Table 1). It is helpful to also note down the number of occurences of each and the gromacs virtual site type that could be used to treat these as virtual sites.

Table 1: Different types of hydrogens in GlcNAc1P.

No.	**Hydrogen type**	**# of occurences**	**Virtual site type**
i.	Polar hydrogen of -OH group	3	None
ii.	Hydrogen on CH3 group	3	Type 3 and 3out
iii.	Amide hydrogen	1	Type 3fd
iv.	Hydrogens on ring sp3-Carbon atoms	5	Type 4fdn
v.	Hydrogen on -CH2- group	2	Type 3out

This table helps to visualize in our mind how to go about constructing the topology.

Step 2: Edit the coordinate and topology to include additional dummy masses and redistribute hydrogen masses to heavy atoms. Since, treating functional groups such as -NH3 and -CH3 requires addition of extra dummy masses which would eventually change the atom numbering, this step should be done first. It involves changing both the coordinate file (addition of extra dummy particle) and the topology (addition of new fields under atomtypes and atoms directives).

In case of GlcNAc1P, we have one -CH3 group. So, we would have to add two additional dummy particles both in the coordinate and the topology file. The coordinate file can be build (using any software of choice such as chimera,

avogadro etc.) trivially by addition of the two dummy atoms such that the -CH3 carbon atom lies in the plane formed by the adjacent heavy atom (bonded to -CH3) and the two dummy atoms, and falls on the line connecting the two dummy atoms, equidistant from both (refer to Figure 2 for -NH3 group, but the same is applicable to -CH3 as well).

Coordinate file for ligand with VS:

```
GN1vs vaccum
     35
      1GN1     C1    1   0.195   6.318   5.501
      1GN1     H1    2   0.228   6.413   5.459
      1GN1     C2    3   0.318   6.227   5.518
      1GN1     H2    4   0.368   6.250   5.613
      1GN1      O    5   0.288   6.086   5.516
      1GN1     C3    6   0.153   6.048   5.550
      1GN1     H3    7   0.134   6.082   5.652
      1GN1     C4    8   0.051   6.116   5.457
      1GN1     H4    9  -0.041   6.135   5.513
      1GN1     C5   10   0.105   6.247   5.400
      1GN1     H5   11   0.162   6.226   5.309
      1GN1     O1   12   0.410   6.256   5.412
      1GN1      P   13   0.541   6.347   5.428
      1GN1     O2   14   0.517   6.485   5.475
      1GN1     O3   15   0.630   6.304   5.538
      1GN1     O4   16   0.624   6.357   5.306
      1GN1     C6   17   0.149   5.895   5.548
      1GN1    H61   18   0.045   5.860   5.541
      1GN1    H62   19   0.202   5.856   5.460
      1GN1     O5   20   0.209   5.846   5.667
      1GN1    HO5   21   0.291   5.897   5.679
      1GN1     O6   22   0.021   6.029   5.347
      1GN1    HO6   23   0.080   6.055   5.275
      1GN1     O7   24  -0.003   6.335   5.366
      1GN1    HO7   25  -0.010   6.399   5.439
      1GN1      N   26   0.138   6.344   5.633
      1GN1     HN   27   0.172   6.289   5.710
      1GN1     C7   28   0.042   6.436   5.655
      1GN1     O8   29  -0.008   6.501   5.566
      1GN1   MCG1   30   0.035   6.448   5.814
      1GN1   MCG2   31  -0.051   6.460   5.782
      1GN1     C8   32  -0.008   6.454   5.796
      1GN1    H71   33   0.076   6.445   5.866
      1GN1    H72   34  -0.053   6.552   5.807
      1GN1    H73   35  -0.082   6.377   5.819
    0.00000   0.00000   0.00000
```

In the coordinate file printed above, you can see that two additional dummy particles (MCG1, MCG2) have been included in the ligand structure. Here,

atoms C8, H71, H72 and H73 constitute the acetyl group. So, following the convention observed in topologies that we observed in case of proteins, we introduce the atoms MCG1 and MCG2 such, C8 lies in the plane formed by atoms C7, MCG1 and MCG2, on the mid-point of the line connecting MCG1 and MCG2. Once the coordinate file is constructed, we next move on to edit the topology file.

You can start by copying over only the contents under the *[atoms]* directive of the normal topology to a new file and introduce two additional lines for the dummy masses, making sure that the atom order is consistent between the coordinate and topology file. Your topology should become like the one printed below.

Atoms section of the new topology:

```
 [ atoms ]
;   nr       type  resnr residue  atom   cgnr    charge       mass  typeB    chargeB
; residue    1 GN1 rtp GN1 q -1.6
   1         c3      1    GN1     C1      1  -0.069900    13.0180
   2         h1      1    GN1     H1      2   0.124100     0.0000
   3         c3      1    GN1     C2      3   0.033200    13.0180
   4         h2      1    GN1     H2      4   0.174300     0.0000
   5         os      1    GN1      O      5  -0.351600    16.0000
   6         c3      1    GN1     C3      6   0.043900    13.0180
   7         h1      1    GN1     H3      7   0.181200     0.0000
   8         c3      1    GN1     C4      8   0.108700    13.0180
   9         h1      1    GN1     H4      9   0.069100     0.0000
  10         c3      1    GN1     C5     10   0.169400    13.0180
  11         h1      1    GN1     H5     11   0.098200     0.0000
  12         os      1    GN1     O1     12  -0.505900    16.0000
  13         p5      1    GN1      P     13   1.351600    30.9700
  14          o      1    GN1     O2     14  -0.901600    16.0000
  15          o      1    GN1     O3     15  -0.901600    16.0000
  16          o      1    GN1     O4     16  -0.901600    16.0000
  17         c3      1    GN1     C6     17   0.257900    14.0260
  18         h1      1    GN1    H61     18  -0.005700     0.0000
  19         h1      1    GN1    H62     19  -0.005700     0.0000
  20         oh      1    GN1     O5     20  -0.742400    16.0000
  21         ho      1    GN1    HO5     21   0.493900     1.0080
  22         oh      1    GN1     O6     22  -0.703400    16.0000
  23         ho      1    GN1    HO6     23   0.444400     1.0080
  24         oh      1    GN1     O7     24  -0.733800    16.0000
  25         ho      1    GN1    HO7     25   0.454400     1.0080
  26          n      1    GN1      N     26  -0.486100    15.0180
  27         hn      1    GN1     HN     27   0.348700     0.0000
  28          c      1    GN1     C7     28   0.759100    12.0100
  29          o      1    GN1     O8     29  -0.700300    16.0000
  30       MCH3      1    GN1   MCG1     30   0.000000     7.5170
  31       MCH3      1    GN1   MCG2     30   0.000000     7.5170
  32         c3      1    GN1     C8     30  -0.412100     0.0000
  33         hc      1    GN1    H71     31   0.103200     0.0000
```

```
  34          hc       1     GN1     H72      32    0.103200      0.0000
  35          hc       1     GN1     H73      33    0.103200      0.0000
```

Please notice in the topology that the masses for the dummy atoms are essentially the mass of the C8 carbon distributed equally between the two, and that the C8 mass has been set to zero. Also, it is important to note that for all the hydrogens (except for polar -OH hydrogens), the masses are set to zero and these masses the added to the adjacent bonded heavy atom. Charges on all atoms are kept unchanged and the charges on the two dummy particles (or more correctly, dummy masses) are set to zero.

Step 3: Add the bonds, constraints, pairs, angles and dihedral sections In the next step, one must build the rest of the sections such as bonds, pairs, angles and dihedrals by hand. Most of the parameters can be copied but one must be careful about the correct numbering (since it changed due to addition of new atoms). For the *[constraints]* section, we must note the contraints that need to be introduced. As observed in case of the example of proteins, distance constraints are added to constrain either the C-O-H angle in polar -OH (or -SH), or the distances between the two dummy masses and the third heavy atom in -CH3 (or -NH3) groups. In Table 1, you will find that our ligand has 3 -OH groups and 1 -CH3 group, so we should include a total of 6 [3(1) + 1(3)] constraints in the topology file (printed below). The first three correspond to the constrain on the C-O-H angle vibrations by constraining the C–H distance. The next three correspond to the -CH3 group with the dummy masses.

```
 Constraints section of the new topology:

 [ constraints ]
;  ai    aj funct            c0            c1
   17    21     2
    8    23     2
   10    25     2
   28    30     2
   28    31     2
   30    31     2
```

Step 4: Add the topology sections of virtual sites. Again, refer to the Table 1 where all the types of VS constructions required are tabulated. From here, one can make a mental note of the number of virtual site directives one must add to the topology file. For the polar hydrogens no VS constructions are required as these are not treated as dummy atoms (note that these hydrogens retain their usual mass in the new toplogy). Next, we have the -CH3 group hydrogens; as we have learned for this group one needs 2 Type 3 directives and 2 Type 3out directives (for details refer to -NH3 construction in subsection *a* under section III). Next type of hydrogens in the list is the amide hydrogen, which can be treated as VS by adding 1 Type 3fd directive. This is followed by the hydrogens on the ring carbon atoms, which again can be defined by 1 Type 4fdn directive for each hydrogen. So, in total there would be 5 Type 4fdn

directives for the ligand as there are 5 such hydrogens. Finally, we have the two hydrogens on the aliphatic -CH2- group, which can be defined as virtual sites by adding 2 Type 3out directives. So then, in total we would need to introduce 2 Type 3, 1 Type 3fd, 4 Type 3out and 5 Type 4fdn definitions. These can be constructed by hand and should look similar to the one printed below. Do make it a habit to finally check the correct numbering of the atoms to ensure correctness of the toplogy.

```
Virtual sites section of the new topology:

 [ virtual_sites3 ]
;  ai    aj    ak    al funct            c0            c1
   32    28    30    31     1
   33    28    30    31     1

[ virtual_sites3 ]
;  ai    aj    ak    al funct            c0            c1
   27    26     1    28     2

[ virtual_sites3 ]
;  ai    aj    ak    al funct            c0            c1
   34    28    30    31    -4
   35    28    30    31     4
   18    17     6    20     4
   19    17     6    20    -4

[ virtual_sites4 ]
;  ai    aj    ak    al    am funct            c0            c1            c2
    2     1     3    10    26     2
    4     3     1     5    12     2
    7     6     5     8    17     2
    9     8     6    10    22     2
   11    10     1     8    24     2
```

b. Final notes

It is important to run some test simulations with the ligand and compare some ligand properties (from simulations with and without VS), to ensure the system is stable and gives expected results. Thereafter, these topologies can be integrated within complex gromacs topologies in the usual manner.

With this we conclude the tutorial. Feel free to contact via email (abhi117acharya@example.com) if your find any mistake in the document or have any other concerns regarding clarity or tips to make this tutorial better.