Martini Glycosylator

(Tool for glycosylating CG martini protein)

1. Martinize your protein using martinize.py. For tutorial follow (http://cgmartini.nl/index.php/tutorials-general-introduction-gmx5/proteins-gmx5)
   If protein has multiple chains, then resultant itp files should be named according to their chain names.

2. Make sure the protein is in pdb format with chain names specified inside the file even if protein is monomeric.
   a) For monomeric systems (for example 'lysozyme') the files will be named lysozyme.pdb (with chain name for example 'A') and lysozyme_A.itp.
   b) For dimeric systems (for example 'denv') the files will be named as denv.pdb (with chain names A and B) along with denv_A.itp and denv_B.itp.

3. Next an input file specifying the ASN_site, chain and glycan type is required. An example input.dat is available in the files provided. (list of available glycans are on the next page)

```
ASN_site chain  glycan
14         X     m9
144        X     m5
```

4. The martini_glycosylator script is then run as

```
python martini_glycosylator.py -d 1.2 -c protein.pdb -f input.dat
```

   Here, the -d (3.5*d Angstrom) controls the distance that the glycan is placed away from the glycosylation sites. A value of 1 works in most cases, but a higher value between 1 to 2 could be used to avoid any clashes. Note that, currently glycan is placed along the ASN vector (along backbone bead and side chain bead) which points outwards in crystal structures. If glycan is attached to already equilibrated structure, it can result in erroneous placing of glycans. Do a visualisation check if the glycans are placed properly, adjust the positions of glycans using pymol if required. Protein.pdb is the CG protein pdb file generated from the step 1.

5. The output will have itp files (protein_A_gly.itp, protein_B_gly.itp etc) and pdb for each chain (protein_A_gly.pdb, protein_B_gly.pdb) which could be used for visually checking the placement of the glycans.

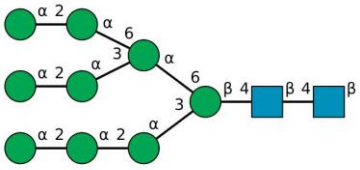
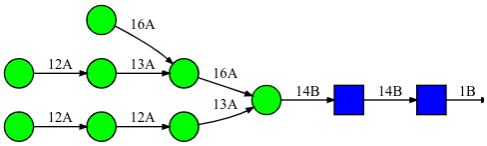
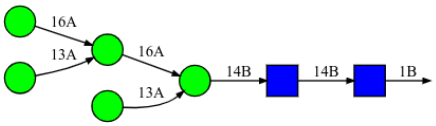
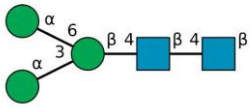
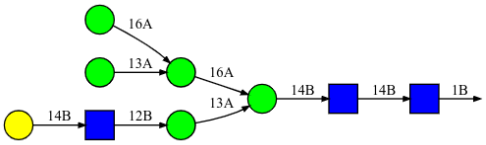
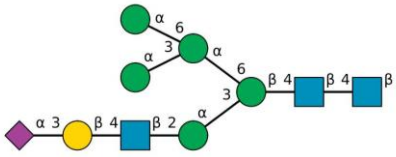
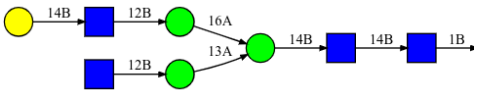
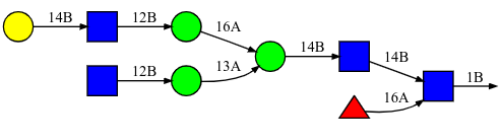
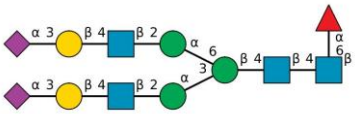
6. The combined pdb file can be generated using cat command. For example,

```
cat protein_A_gly.pdb protein_B_gly.pdb > protein_gly.pdb
```

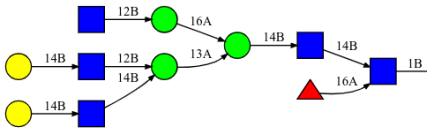
7. The first minimization should be done with restraints on protein (mdp file included). Also, equilibration should be started with 5fs and increase till 15 fs if no LINCS warnings are generated (10 fs was most stable in our studies). In most cases

martini_2.2 glycans parameters scaled by 0.9 should be used (available in the folder) (0.95 if majority of the glycans attached are noncharged).

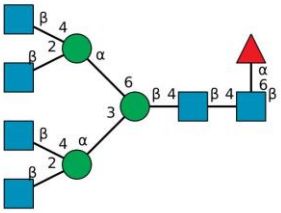
Table 1: Glycans currently available for attachment to the CG protein. The glycan codes are represented by oxford notation. Follow https://www.ludger.com/docs/tables/ludger-n-glycan-nomenclature-table.pdf for more information. Additional, glycan types could be generated using parameters published in https://doi.org/10.1021/acs.jcim.0c00495

Glycan code	Schematic representation
M9	
M8	
M5	
M3	
M5A1G1	
M5A1G1S1	
A2G1	
FA2G1	
FA2G2S2	

FA3G2



FA4



FA4G2S1