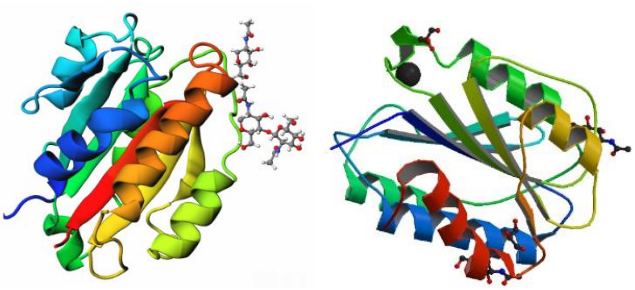
PRATHYSHA KOTHARE, Parkland High School

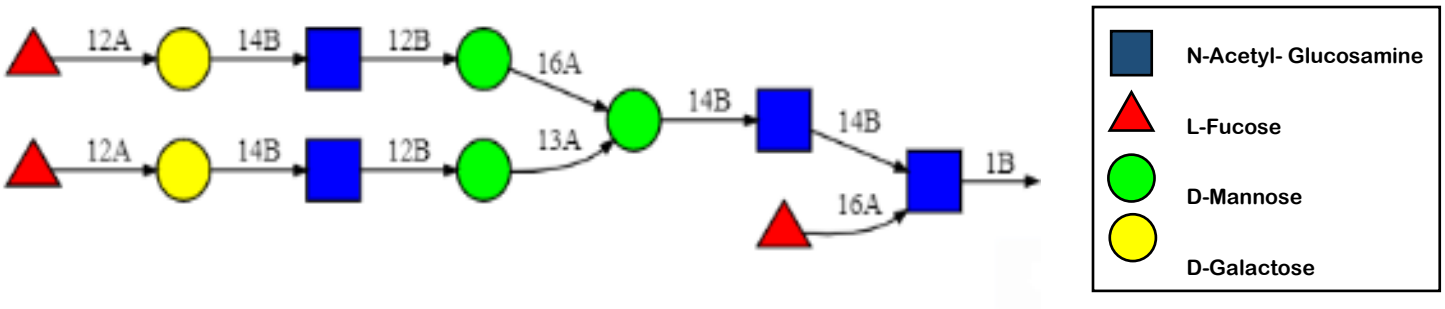
SIMULATION PROCEDURE

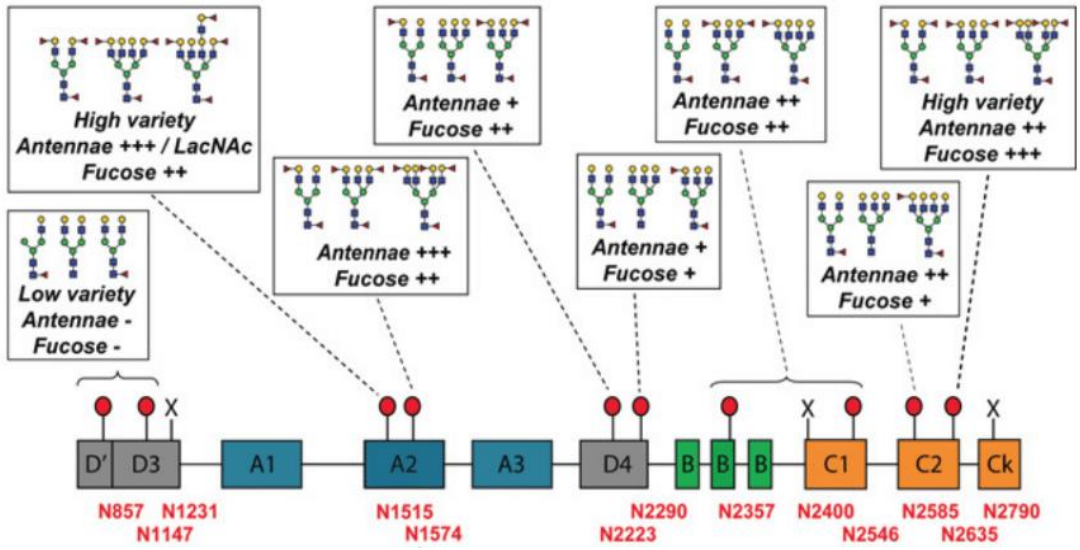


1. Select base protein pdb file from Protein Data Bank:
    a. 3GXB = Wildtype A2 dimer
    b. 3ZQK = Wildtype A2 multimer with calcium
2. Go to charm-gui.org

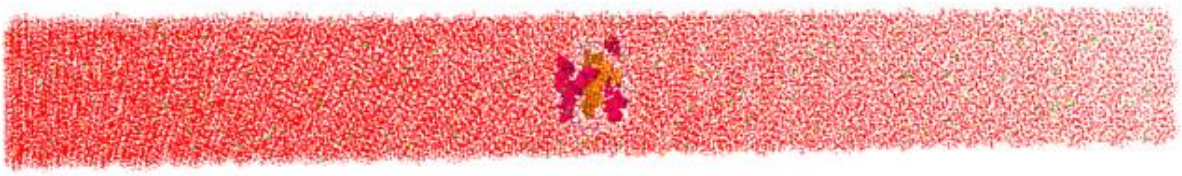
Charmm-gui: web-based graphical user interface used to create various molecular simulation systems and input files to facilitate advanced force simulations.

3. Select Quick MD simulations
4. Select Input Generator
5. Enter in pdb file name and click "upload"
6. STEP1: For monomer unfolding analysis, remove half of dimer
    a. For complex n-glycan simulations, leave 1515 and 1574 selected
    b. For simple A2 unfolding analysis, decode the two n-glycan models
7. STEP 2: Modify monomer protein composition
    a. For WT and SimCAL: no modifications needed
    b. For Sim1541: create amino acid mutation; 1541 Glutamine to Arginine
    c. For building n-glycans, construct based on reference paper: 1515 and 1574
    *For 1574, construct same as 1515, but remove one end fucose monosaccharide
8. STEP 3: Define the waterbox boundaries in which the force simulation will run:
    X: 800 Y: 80 Z: 80 (units=angstroms)





9. STEP 3a: Construct/Add NaCl neutralizing ions to the waterbox in order to closely represent a physiological blood flow environment (use Monte Carlo distribution)
10. STEP 4: Define simulation environment at temperature: 300 K



11. In order to conduct externally induced force simulations, install NAMD sub-directory into the protein tgz file prior to download onto computer.

    Nanoscale Molecular Dynamics (NAMD): Molecular modeling software simulation written using the Charm++ (language developed from C++) parallel programming model. NAMD is one of the only parallel molecular dynamics code that enables interactive simulation by linking to the visualization code VMD

12. "Download" tgz file and save to computer as Sim__.tgz
13. Download CHARMM software package onto PC

    CHARMM (Chemistry at Harvard Macromolecular Mechanics): A molecular simulation program consisting of physical chemistry force fields and data analysis capabilities for understanding molecular interactions and dynamics.

14. Download MobaXTerm application onto PC

    MobaXTerm: Windows application for remote computing; Enables simultaneous and secure connections to SSH and SFTP sessions

15. Create SSH session to remotely connect PC to personal SOL account

SOL: Parallel computing platform with remote login capability via MobaXTerm. The super-computer consists of 1300 computing cores, 6.5625 TB RAM and 50 nVIDIA GTX 1080 GPUs. SOL will be used to conduct protein force simulations in a large waterbox environment.

SSH (Secure Shell): An encrypted remote login session enabling secure client connection to a server. Applications in this project involve running initial protein and waterbox conditions in CHARMM and exporting these files to conduct force simulations in SOL.

16. Initiate parallel SFTP session to transfer downloaded tgz protein file to personal SOL account

SFTP (Secure File Transfer Protocol): a network protocol that allows file transfer and modification over any reliable data stream between client and server. SFTP will be used to transfer downloaded tgz and modeling files from PC to SOL account for running simulations.

17. In SFTP, open SOL directory and create new folder entitled Sim___
18. Transfer tgz into new Simulation folder
19. Transfer downloaded CHARMM software package to SOL account
20. Go to SSH SOL session and cd new directory (Sim__)
21. Untar tgz file (tar –xvzf Sim__.tgz); A charmm-gui directory will be generated containing initial protein and waterbox parameters and NAMD software directory
22. Untar CHARMM software package tgz file using same command
23. Run step2_solvator.inp file from protein tgz in charmm-gui using CHARMM: Step2 solvates defined waterbox and establishes neutralizing ions with Monte Carlo distribution
    /home/mvk2/charmm/exec/gnu/charmm <step2_solvator.inp
24. Using same command line, run step3_pbcsetup.inp to situate monomer within the waterbox and enable protein relaxation
25. Run step3_input.inp to convert pbcsetup output data into alternative python coding to be used as input for step4
26. Go to SFTP and upload original run_step4.slurm, colvars.conf, com.ref, COM.tcl, and run.slurm code files into namd directory under charmm-gui
27. In SSH, cd namd and run run_step4.slurm code: “sbatch run_step4.slurm”. Step 4 thermally equilibrates the protein in the waterbox and enables energy minimization of the molecule
28. “squeue –u mvk2” (SOL account name) to see submitted jobs and run time. Step 4 should run for approximately 2-4 hours

```
[mvk2@sol ~]$ squeue -u mvk2
             JOBID PARTITION     NAME     USER ST       TIME  NODES NODELIST(REASON)
            328116     imlab run.slur     mvk2  R   14:37:56      1 sol-b504
```

29. Once job partition is complete, module load VMD to interpret output data from step 4

VMD (Visual Molecular Dynamics): Computing program utilized to display, animate, and analyze biomolecular systems using 3-D graphics and built-in scripting.

30. Run COM.tcl in vmd to collect end coordinates of thermally equilibrated protein in the waterbox (xyz coordinates)
    "vmd <COM.tcl"
31. "vi com.ref" file in namd to view coordinates after COM.tcl has run

```
14.76758098602295 -2.153451681137085 6.526384353637695
11.428208351135254 0.41509583592414856 18.04734230041504
```

32. Copy new coordinates into colvars.conf file in namd under dummy atoms command line
    vi colvars.conf
    "i" to modify file
    esc :wq to save new modifications and exit file

Collective Variables Module: A software module for molecular simulation and analysis that provides a high-performance implementation of sampling algorithms defined on a reduced space of continuously differentiable functions (aka collective variables).

N-terminal coordinates

C-terminal coordinates

Code to pull C-terminal and fix N-terminal

```
colvar {
 name pull_nter
 distanceZ {
   main {
     psfSegID PROA
     atomNumbersRange 1-19
   }
   ref {
     dummyAtom (14.76758098602295, -2.153451681137085, 6.526384353637695)
   }
   axis (1.0, 0.0, 0.0)
 }
}

colvar {
 name pull_cter
 distanceZ {
   main {
     psfSegID PROA
     atomNumbersRange 2763-2777
   }
   ref {
     dummyAtom (11.428208351135254, 0.41509583592414856, 18.04734230041504)
   }
   axis (1.0, 0.0, 0.0)
 }
}

harmonic {
   colvars pull_nter pull_cter
   centers -0.0 0.0
   targetCenters -0.0 5.0
   forceConstant 5.
   targetNumSteps 500000
}
```

33. If desired to conduct different force induced simulations change colvars command line (Pull N vs C) For this experiment, set both target centers to 50 ang/s
34. vi step5_production.inp auto-generated file in namd and make the following modifications:
    a. Add colvars.conf command line under parameters to incorporate colvars inputs into simulation run

```
colvars on
colvarsConfig colvars.conf
```

   b. Change output name from step5_production to OUTPUT in order to correspond to run.slurm commands. This is crucial to ensure proper rerun of the simulation after 48 hours
   c. Change input name step4_equilibration to INPUT

35. vi run.slurm and change all 10 javlues to 50. This is done to define the pulling speeds for both termini. For this experiment in which both termini are being subject to pulling forces, disregard jval3 and change jval3 in sed –I command line to jval1 and jval2. This establishes the target centers for each termini to correspond to those defined in the jvalue command lines


```
set file = step5_production
set file2 = colvars.conf


while ( ${cnt} <= ${cntmax} )

    if ( ! -e step5_${cnt}.coor ) then
        @ pcnt = ${cnt} - 1

                set ival1 = `sed -n "/centers/"p ${file2}`
        set ival2 = `sed -n "/targetCenters/"p ${file2}`

        set jval1 = `awk -v cnt=${cnt} 'BEGIN{printf "%.1f", (cnt-1)*5.0}'`
        set jval2 = `awk -v cnt=${cnt} 'BEGIN{printf "%.1f", (cnt-1)*5.0 + 5.0}'`
        set jval3 = `awk -v cnt=${cnt} 'BEGIN{printf "%.1f", (0)*5.0}'`

        if ( ${cnt} == 1 ) then
                set pfile = step4_equilibration
        else
                set pfile = step5_${pcnt}
        endif

        sed -e "s/OUTPUT/step5_${cnt}/g" -e "s/INPUT/${pfile}/g" ${file}.inp > step5_${cnt}.inp
                sed -i -e "s/${ival1}/centers -${jval3} ${jval1}/g" -e "s/${ival2}/targetCenters -${jval3} ${jval2}/g" ${file2}


        sleep 60
        namd2 +p ${SLURM_NTASKS} step5_${cnt}.inp > step5_${cnt}.out


        if (! -e step5_${cnt}.coor ) exit

    endif
    @ cnt += 1

end
```

36. submit run.slurm simulation job in namd to conduct force pulling simulation of the C and N termini

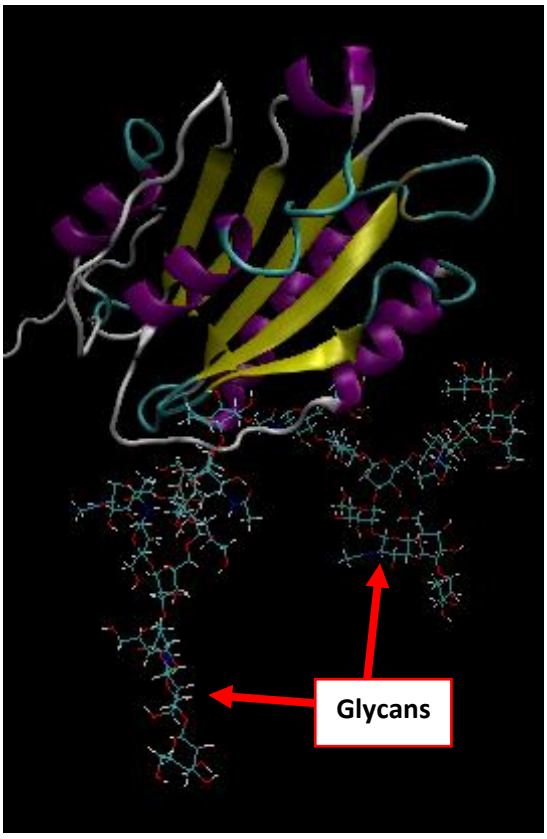
    “sbatch run.slurm”

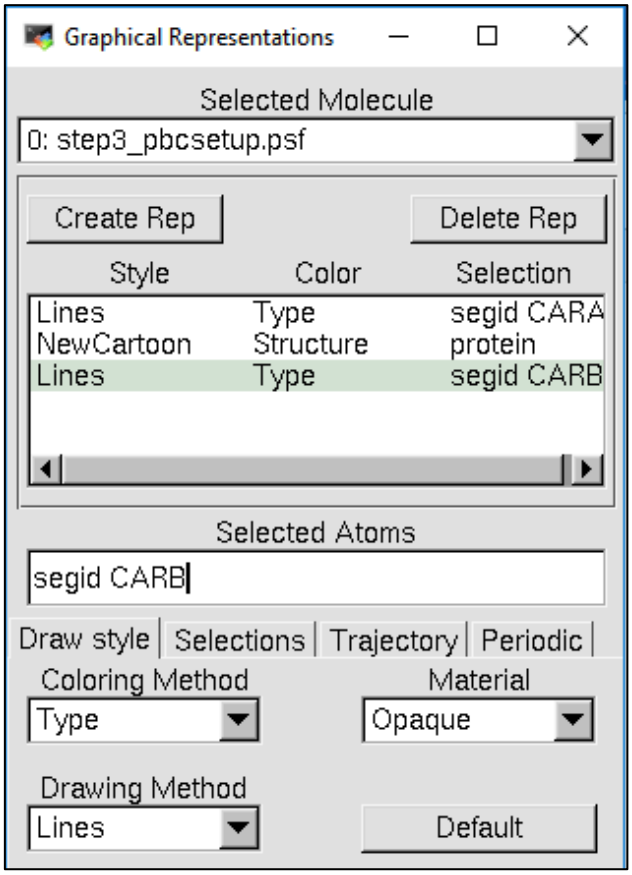
37. Once the job is submitted, it will run for 48 hours, since this is the maximum processing time capability for SOL. In this time, the simulation will generate approximately 0.5 nanoseconds of pulling simulation. To conduct longer simulations, submit a new job (sbatch run.slurm). Input files will all automatically adjust to new coordinates and data based on simulation results.
38. Renew jobs for about two weeks in order to observe complete beta unfolding of the protein monomer

FRAME VIEWING

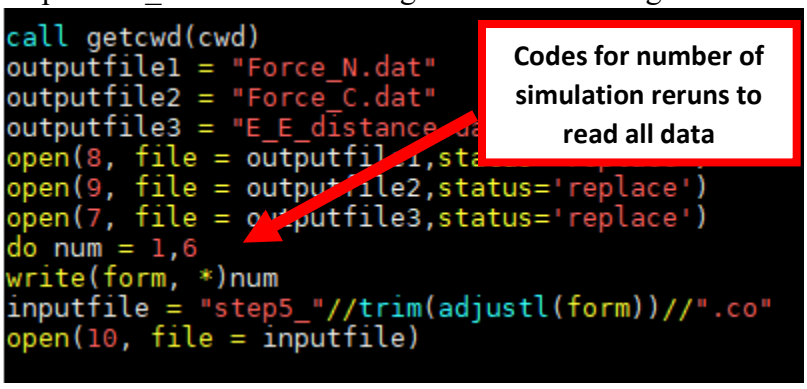
1. To view the VMD frames generated for the entire simulation which illustrate the qualitative unfolding of the domain, use the following commands
   a. Module load vmd
   b. Vmd ../step3_pbcsetup.psf step5_#.out (# refers to the number of times the job has been renewed after 48 hours. Use the latest # in order to view ALL simulation frames

2. A 3D image of the protein in a waterbox will appear. To enhance protein visualization, go to vmd command center and:
    a. Go to Graphics
    b. Representations…
    c. Selections
    d. Reset to eliminate waterbox visualization
    e. Type “protein” in selected atoms and APPLY
    f. Go to Draw Style
    g. Colors: Secondary Structure to view alpha and beta sheets
    h. Lines: New Cartoon
    i. For viewing glycans:
        i. Representations
        ii. Selections
        iii. Create New Rep
        iv. In Selected atoms type “segid CARA” (this selects 1515 glycan)
        v. Go to Draw Style
        vi. In Coloring Method select “Type”
        vii. In Drawing Method select “Lines”
        viii. Repeat for CARB (1574)





FORCE BY TIME ANALYSIS (C &N)

1. Import original position.tcl to namd in monomer subfolder using SFTP session
2. In SSH Module Load VMD
3. Make Analysis directory to generate output data "dat" files
   "mkdir analysis"
4. Run position.tcl in VMD to collect Force by Time data outputs for N and C terminus, and End to End distance "vmd <position.tcl"
5. Once complete, cd analysis and ls to view newly generated files. Should see step_#.co
6. Import original position_force.f90 text file using SFTP into analysis sub-directory
7. Vi position_force.f90 and change "do num" to range of simulation reruns (i.e. 1,6)
8. 

9. Load intel processor to generate "dat" files\
   a. Module load intel
   b. ifort position_force.f90
10. ./a.out to view new "dat" files
11. In SFTP, download SFTP files:
    a. Force_C.dat
    b. Force_N.dat
    c. E_Edistance.dat
12. Open and select all in "dat" file
13. In Excel, Paste Special using Text Wizard (2 columns)
14. Create XY plot and select data from each column as inputs
    =TrialSim!$A$1:$A$2748
15. For smoothing Force by Time curves to get better trend interpretation, go to "Format Trendline"
16. Click "Moving Average" and set to 20
17. Modify X and Y axis and label chart

SURFACE AREA BY TIME ANALYSIS

*Follow same process as for Force by Time, but:

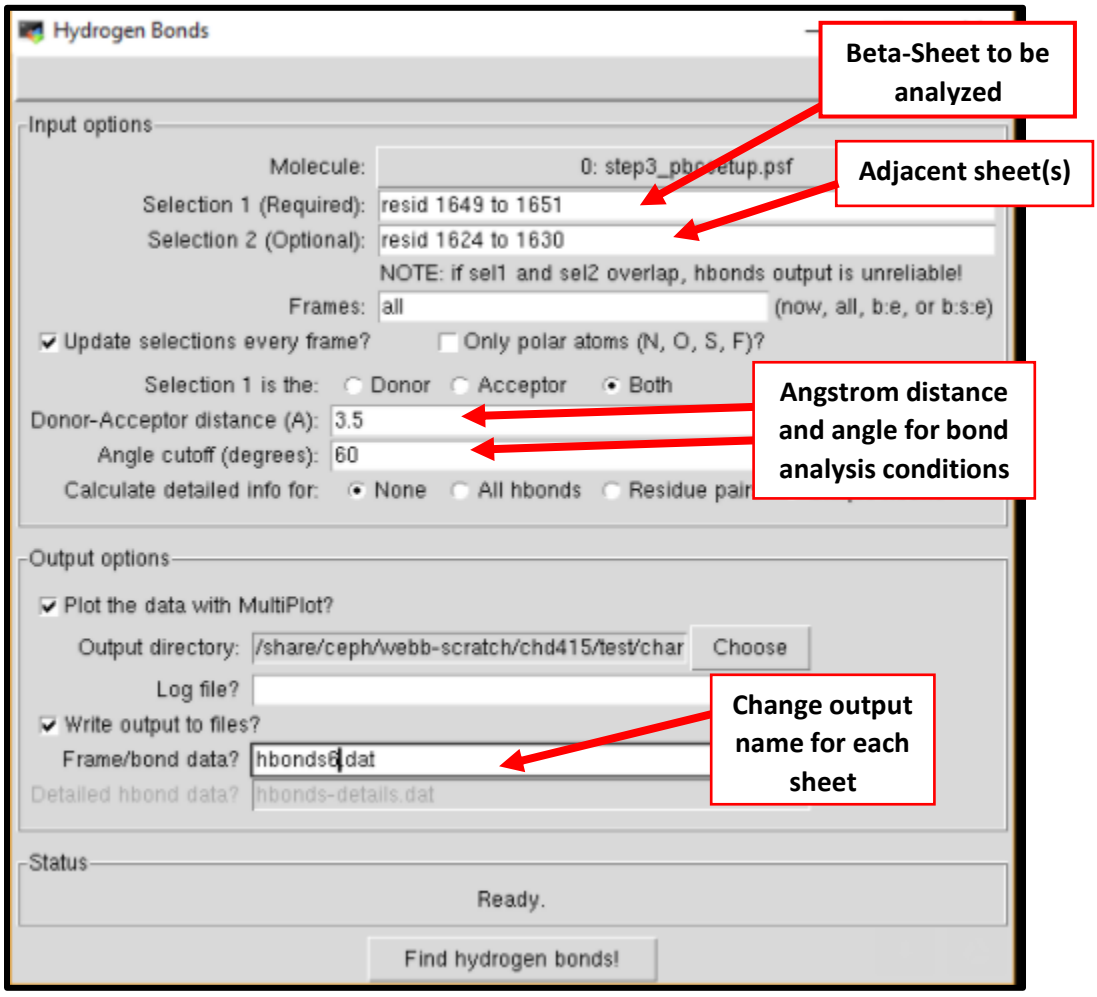
- SASA.tcl and SASA_rewrite.f90
- Change "do num" in SASA_rewrite.f90 as in position_force.f90
- Download SASA_rewrite.dat output file from SFTP to graph in Excel

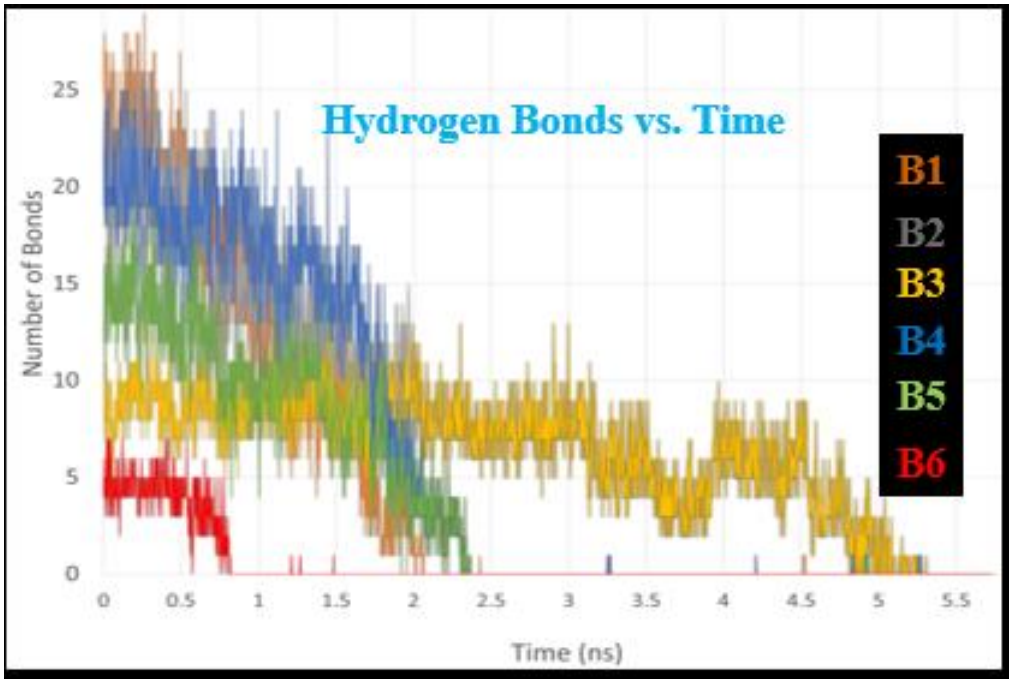
- Surface area by time gives extension rate of monomer

HYDROGEN BOND & BETA SHEET ANALYSIS

1. In charmm-gui directory of monomer tgz file, module load vmd
2. Vmd ..step3_pbcsetup.psf step5_#...
3. In VMD pop-up Main Analysis bar, go to Extensions
4. In Extensions, select Analysis and then "Hydrogen Bonds"
5. A second pop-up VMD browser will open
6. Change Donor-Acceptance Distance to 3.5
7. Change Angle Cut-off to 60 degrees
8. Write Output file to namd subdirectory
9. Using residue IDs (amino acid numbers in each beta-sheet) conduct following process
    a. In Selection 1, type resid ____ to _____ in order to define the beta-sheet being analyzed
    b. In Selection 2, use same formatting to define the beta-sheets adjacent to the sheet being analyzed. This will also the bonds between adjacent sheets to be measured
    c. Select "Find bonds"… Wait until pop-up generate Hbond graph
    d. Go back to residue input page and continue with analysis for each beta sheet
10. After hydrogen bond data is calculated for each sheet, download hbond#.dat files from SFTP in namd
11. Copy and Paste data to Excel and conduct joint plotting to see beta-sheet unfolding in one graphs

*Reference monomer image to determine adjacent beta-sheets


Hydrogen Bonds
Input options
Molecule:
0: step3_pbcsetup.psf
Selection 1 (Required):
resid 1649 to 1651
Selection 2 (Optional):
resid 1624 to 1630
NOTE: if sel1 and sel2 overlap, hbonds output is unreliable!
Frames:
all
(now, all, b:e, or b:s:e)
Update selections every frame?
Only polar atoms (N, O, S, F)?
Selection 1 is the:
Donor
Acceptor
Both
Donor-Acceptor distance (A):
3.5
Angle cutoff (degrees):
60
Calculate detailed info for:
None
All hbonds
Residue pair
Output options
Plot the data with MultiPlot?
Output directory:
/share/ceph/webb-scratch/chd415/test/char
Choose
Log file?
Write output to files?
Frame/bond data?
hbonds6.dat
Detailed hbond data?
hbonds-details.dat
Status
Ready.
Find hydrogen bonds!
Beta-Sheet to be analyzed
Adjacent sheet(s)
Angstrom distance and angle for bond analysis conditions
Change output name for each sheet


Hydrogen Bonds vs. Time
Number of Bonds
Time (ns)
0
5
10
15
20
25
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
B1
B2
B3
B4
B5
B6

PRATHYSHA KOTHARE, Parkland High School

Beta Sheet	Residue Range
Beta 1	1497 to 1504
Beta 2	1535 to 1542
Beta 3	1546 to 1550
Beta 4	1602 to 1608
Beta 5	1624 to 1630
Beta 6	1649 to 1651