TO-DO list for Mucin analysis

August 22, 2013

I've got most of the hard parts done. Now, to tie up a few things and start making final images.

List items are noted by:

- Regular type if I only need to do some make-pretty with plots. Initial, draft gnuplot input files are attached.
- *Italic type* if the data are prepared, but I still need to generate plots, some or all. The data are typically too large to attach here, so I won't do that.
- **Bold type** if I still need to do some data generation (plus generate plots, of course).

My highest priority right now is anything in **bold type**. The next priority is anything in *italic type*. Of course, I reserve the right to get frustrated and do them in any order I please.

Most plots below will include separated plots for all four linkages.

Contour plots:

- Phi-Psi plots
  - Bin all-run data.
    - Files: Bin_PhiPsi_for_plots.c makefile
    - There's also this control file, run from the ANALYSIS/PhiPsi_Bins directory: finding_a_good_scale.bash
  - *Backbone*
  - *Glycosidic linkage* (runsets d4g and d4m only)
- N — O-back versus N — H — O-back *w/NMR*
  Files: Bak_contour Bak_contour_1 Bak_contour_2 Bak_contour_3 Bak_contour_4 Bak_contour_multi
- *N — O-glyc versus N — H — O-glyc w/NMR*
  Files: Gly_contour
- (N — O-back + N — O-glyc) versus O — N — O
  Files: SUMvsONO_contour
- H — O-back versus O — H — O
  Files: HObvsOHO_contour HObvsOHO_contour_1 HObvsOHO_contour_2 HObvsOHO_contour_3 HObvsOHO_contour_4 HObvsOHO_multi
- H — O glyc versus O — H — O
  Files: HOgvsOHO_contour

- (H — O-back + H — O-glyc) versus O — H — O
  Files: SUMvsOHO_contour Sum_contour_1 Sum_contour_2 Sum_contour_3 Sum_contour_4 Sum_contour_multi
- *H — O-back versus H — O — O*

Histogram plots:

- *Distribution of Phi angles, showing NMR average and simulated average*
  - All three runsets
  - Also separated out by linkage
- *Distribution of Psi angles, showing NMR average and simulated average*
  - All three runsets
  - Also separated out by linkage
- *N — O-back distances w/NMR*
- *N — O-glyc distances w/NMR*

Scatter Plots:

RMSD's for all frames (all sims) versus each initial structure
Plot in sets of four initial structures

Images:

- O-H-O from N — O-back < 4.5 for all frames from two angles
  Files: Gal3_set4_shared
- Previous, plus the backbone H
  Files: see previous for now. Will separate when making pretty.
- ~~*O-H-O*~~ *H-O-H-O* ***(find representative structure somehow)*** *with surface from two angles*
  - *Do the H-O-H-O 4-way thing. It's more relevant and makes a nicer case.*
  - **To determine representative structure:**
    Find structure with backbone angles closest to NMR-derived value
- Structure with RMSD closest to average RMSD from NMR-best overlayed with NMR-best, show backbone only.
  (I did this once… must find the files)

- **Check equilibration runs for evidence of water-mediated H-bond**
    - Find reasonable number of frames that are equilibrated (use energy)
      They are all well-equilibrated 100 frames in. This leaves about 900 frames (about 1.8 ns) from each simulation.
    - Write program to search for N-H … O … H-N bond.
        - This is done, and seems to work well Current version: do_equil_water_check.c
        - **Update program to do these things:**
            - **Write the distances info. Also binned.**
            - **Calculate the H-H <1/r^6>^1/6 averages for comparison to earlier work.**

Other Data/Information:

- Average RMSD, with fluctuations, for each initial structure versus all simulations
  Files: Fluctuations_all_bak Fluctuations_all_sug RMSD_all_bak RMSD_all_sug
- *Previous, also split out per simulation*