New plan for hbond analysis

July 2, 2013

Here's what I plan to do.

1. Divide all the treated data into 150 equally sized bins running from 5 to 35 Angstroms. These lengths will represent the terminal backbone N-to-N distance.
    1. I need this so that all the bins across all the data will be uniform.
2. First, get the bin information, per each runset, for the overall N-N distances for all simulations in that runset. Save this info.
    1. This tells me the distribution of N-N lengths for all frames of all simulations.
3. For each unique type of H-bond found in each runset (over all simulations), bin the N-N distances. Save this info. Then, divide each of these bins by the corresponding (non-zero) value from the whole-run N-N distance bins.
    1. The ratio of HBond N-N distances to overall N-N distances will allow me to more easily find HBonds that correlate with longer or shorter distances.
4. Normalize the ratio bins for each H-bond so that the heights of the bins all add to one.
    1. Because this simplifies making sense of the math.
5. For each H-bond, sum the normalized and ratio'd bins from 5-15, 15-25 and 25-30.
    1. This will give me a sense of…. er…. "enrichment" for each H-bond at short, medium and long lengths.
6. Find the H-bonds for which the sum from 25-30 is the largest, particularly if it is lager than the other sums.
    1. Because these are probably the more important ones for ensuring longer chain length.

It might be interesting to, say, turn off the glycan charges altogether in the Man and GalNAc moieties to see how much their respective bulks/shapes matter irrespective of hydrogen bonding.

I also need to come up with a good way to account for hbond occupancy, but one thing at a time.