

BCH222 - Sidechain Rotamer-plot Data

Reading

Lovell et al. (2000) Proteins: Struct Funct Genomics **40**, 389 on the "penultimate library" of sidechain rotamers.

What was the B-factor cutoff used in this dataset? _____ How many PDB files were included? _____

What are two of the advantages of defining rotamer positions by modal (highest occurrence) rather than mean (average) chi values? _____

(In Lys section) What is the order of preference for χ_1 bins (called p,t,m)? ____ > ____ > ____

What is considered the strongest piece of evidence that the backward-fit "decoy" rotamers of Leu (tt* & mp*) are actually incorrect? _____

[Later work (Headd 2009 JSFG 10:83) showed that the decoy examples could be successfully rebuilt.]

Does the B<40 filter used here get rid of the decoys? _____

What is the conclusion about allowable ring puckers for cis Pro? _____

The m-30 rotamer of Phe or Tyr clashes with its backbone if built in standard geometry; how much was the $C\alpha-C\beta-C\gamma$ angle observed to be widened out in the examples, to prevent that clash? about _____°

Graphics etc. assignment

To work in kinemages of rotamer data, most including also dimensions for ϕ, ψ , and study individual examples. *** Make an initial choice of project topic, and of whether to work on your own or as a pair. We'll sort these ideas out in class next week. Note that the start of a set of project-relevant resources are now on the class website. ***

1. Gln: [rota4400-gln_ptm.kin.gz](#) (832KB)

Open the Gln rotamer kinemage in KiNG. In the initial view, note that the overall ϕ, ψ plot for Gln looks a lot like the general case. Turn on the p, t, and m chi1 bins one at a time; do they each individually look like the general-case? _____

Which one has a sizable cluster in the $L\alpha$ region? _____

Does the order of relative occurrence frequencies for p, t, and m χ_1 values match the order in Lovell 2000? _____

Go to view2, to see the neat pattern of peaks in χ_1, χ_2 . There is still a scatter of some datapoints through the eclipsed values, however. Turn on markers, and do a Find point search for 1n83 321; which angle is eclipsed for that Gln 321? _____ [It may be helpful to use Tools/Show XYZ coordinates.]

Go to view3, to see how spread out the χ_3 values are; this is because of the nearly flat torsion-angle preference around a bond where one end is tetrahedral and the other end is planar.

However, choose view4 to appreciate the complex interactions between χ_2 and χ_3 .

To see the eclipsed-angle residue you identified above, open the 1n83FH_Gln321_ecl_chi2.kin.gz kinemage in KiNG. Measure the χ_2 angle: _____° Are the sidechain atoms in strong, well-shaped density? _____ Turn on the 4σ contours; is the Oe in stronger density than the Ne? _____ How many good H-bonds does this sidechain form? _____

Overall, do you think it's established that this Gln has a genuinely eclipsed χ ? _____

2. Leu: [rota4400-leu_byB.kin.gz](#) (1.2MB)

Open the Leu rotamer kin in KiNG. [Press the w keyboard key to make the ptm labels larger.] This χ_1, χ_2 plot includes all the data, so it has prominent peaks for the misfit "decoy" rotamers tt* and mp* (as described in Lovell 2000). Try filtering on all-atom clashes by choosing view2 ("noclash"); does that get rid of the decoy peaks? _____

Go back to view1, and turn off B-factor bins one at a time from the high values down, watching the decoy peaks. At what B-factor do both the tt* and the mp* peaks go away? B= _____ So with 4400 PDB files the filtering can be much more stringent than was possible for the 240 files of the penultimate library; this will let us get rid of decoy rotamers naturally rather than by fiat.

Open kin file 1xwlfH-Leu473.kin.gz in KiNG, showing Leu 473 fairly well fit into density. However, it is a 0.1% rotamer outlier, with χ_1 _____° and χ_2 _____°. Turn on Tools/Structural biology/Sidechain rotator, choose the PDB file 1xwlfH.pdb, and middle-click on an atom in Leu 473. Turn off "dots" on the button panel, and turn on "Probe dots" in the model-manager box. Try each of the rotamer alternatives; most have dire clashes - which one is best to start with? _____ Use the dials to move χ_1 and χ_2 slightly, to optimize the contacts and the fit to density. What are your final χ_1 _____° and χ_2 _____°? What is the % score for this conformation? _____%

Is there any aspect for which the new version is not better than the original? _____ Close without saving.

3. Pro: [rota4400-Pro_cis-trans.kin.gz](#) (1MB)

Open the Pro rotamer kin in KiNG. In the ϕ, ψ starting view, animate between trans and cis

Pro, to see the slight overall shift of peaks in ϕ and the absence of the small central peak in ψ for the cis form. Has that previously-noted absence held up well, somewhat, or not at all?

Choose view3 to see the χ_1, χ_2 plot, showing the strong anti-correlation enforced by the ring closure. χ_1 is most populated at about $+30^\circ$ and _____ $^\circ$, corresponding to C γ -endo and C γ -exo pucker states, with a trail of points in between with flatter pucker. [We believe the flat versions should actually be fit as a mixture of the two puckers, but that is not yet conclusively proven.] Animate trans to cis.

Choose view2 to see ψ vs χ_1 . This view surprised us by showing that as well as cis-trans, ring pucker also affects the ϕ, ψ plot. Lovell 2000 says that cis Pro only has the C γ -endo pucker ($+30^\circ$); has that restriction held up well, somewhat, or not at all? _____

Open the ribonuclease kin file 1dy5-cisTransPro.kin.gz to see cis-trans and both pucker examples in a high-resolution map. Pro 117 is trans, with ω _____ $^\circ$. Its χ_1 is _____ $^\circ$ [make sure to measure from the N atom], or C γ -exo pucker (curled away from its CO), and is well determined by the density.

Pro 93 is cis, with ω _____ $^\circ$, and is at the end of a β -hairpin turn. Its χ_1 is _____ $^\circ$, or C γ -endo pucker (also well determined by the density). Note the distinctive appearance of the two flanking CO groups pointing away from each other, which can't happen for a trans peptide. Turn off the N atom balls, and make sure you can still tell the N from the C α on each Pro.

4. Ser/Thr/Val: [rota4400-SerThrVal_3D.kin.gz](#) (3.1MB)

Open the Ser/Thr/Val rotamer kin. Animate between the 3 ϕ, ψ plots, noting that each is in Z layers by χ_1 so that the front layer of points show much blacker. This is most striking for Thr, where you can see that the peaks in its β region come from distinct layers - turn the view a bit, to see that the two peaks lower in ψ are mainly from $\chi_1 =$ _____, and the upper two are mainly from $\chi_1 =$ _____.

Choose view2 to see χ_1 , and hit the "w" key to enlarge the p,t,m labels. Animate among the 3 amino-acid types, to see that there are complex shape differences among p,t,m bins, but the shapes and relative populations are not the same for the 3 types. Both visually and from the datapoint counts given on the button panel, fill in the matrix below for relative frequencies (that is, 1>2>3).

... χ_1 p ... χ_1 t ... χ_1 m .

Ser _____

Thr _____

Val _____

There are (at least) two different things going on here.

One is that the Ser & Thr O γ atoms are H-bond donor/acceptors and can tuck in tightly, for instance to H-bond with a CO of the preceding turn of α -helix; that gives both of them a

favorable interaction in χ_1 p, whereas a methyl group in that position would not fit at all.

The second thing is that dihedral angles are defined according to generic chemical conventions, which don't necessarily match up well for specific macromolecular parts. For Ser and Thr, the O atom is dominant and defines the "1" branch to which χ_1 is measured (that's the left-hand branch for Thr). But since the two branches are equivalent for Val, the default assignment of the right-hand branch as "1" takes over. Therefore the p,t,m labels do not mean the same thing for Val as for Thr.

[Note that since χ_1 is measured from the backbone N thru C α and C β to the next atom on the "1" branch, a trans (t) χ_1 means that the "1" branch atom lies opposite (trans) to the N.

Similarly, for minus (m) χ_1 the "1" branch atom is opposite the CO, and for plus (p) χ_1 it is opposite the C α H. Since the H is smallest and the CO largest, being opposite H is unfavorable in general (as we saw for Gln), since the branch then lies between the larger N and CO atoms.]

The above is one example of the many Catch 22 "gotchas" that you need to be constantly alert for. If you do your data-mining blindly by the numbers and labels, you'll make many mistakes. You undoubtedly already know that it's easy for a person to confuse "O" with "0" or "1" with "l" in PDB file names. A more insidious problem is that if you read PDB codes into Excel in the default mode for either .csv or .txt, it will try to be too smart, and will change 1mar to a date (1-Mar) and 1e12 to an exponential number (1.00E+12). [The fix for that problem, once you notice it, is to read in as .txt and specify that the PDBid column is read as text, not general, format. If you're curious, try a simple text read into Excel for the

VTF_noFlip100104_nDate_filtProt.txt.gz file in the project resource section (unzip it first). It seems OK, but sort on PDBid to see what values then show up at the top.] In general, just be vigilant and expect some silly problems.