

Carbohydrate anomalies in the PDB

To the Editor: The importance of carbohydrates both to fundamental cellular biology and as integral parts of therapeutics (including antibodies) continues to grow. The presence of the correct glycans is important for the beneficial effects of therapeutic glycoproteins and is likely to be increasingly required by regulatory agencies. However, carbohydrates (and other small molecules) are handled poorly in macromolecular structural biology. When such small molecules are present in macromolecule structures, they are often reported with stereo- and regiochemical errors and in unlikely conformations. Stereo- and regiochemistry should always be correct, and although conformational distortions may reflect interactions taking place in a complex¹, most are also likely to be erroneous—resulting from poor chemical understanding and lack of appropriate stereochemical restraints in refinement, often against low-resolution data².

Pyranoside sugars have clear conformational preferences dictated by a minimization of angle, torsional and trans-annular strains, resulting in a favored chair (C) conformation. While carbohydrate-active enzymes can force a distortion from this minimal-energy conformation to a higher-energy one in order to enable catalysis³, this is a specialized event; of all the sugars deposited in the PDB, 65% sit undisturbed on N-glycosylation trees.

Using the Privateer⁴ software, we computed the real-space correlation coefficient (RSCC) against positive omit electron density (as a bias-minimized fit to observed density) for all N-glycan-forming D-pyranosides from the PDB as well as calculating their conformation-determining Cremer-Pople puckering parameters⁵. This subgroup was chosen because all are expected to be in the preferred ⁴C₁ chair conformation. A number of anomalies appeared from this analysis. First, 64% of all N-glycan D-pyranosides have RSCCs of <0.8, reflecting poor fit to the electron density. Indeed 12% have RSCC <0.5. Reflecting comments by others⁶, it is clear that a number of models have been built incorrectly from the start: approximately 5% of the N-acetyl-D-glucosamine (GlcNAc) moieties attached to asparagine residues show an incorrect α -linkage, and some monosaccharides are then built 'upside down', resulting in a ¹C₄ conformation. In addition, there are more subtle errors reflecting incorrect refinement of the deposited sugar that could nevertheless have enormous implications

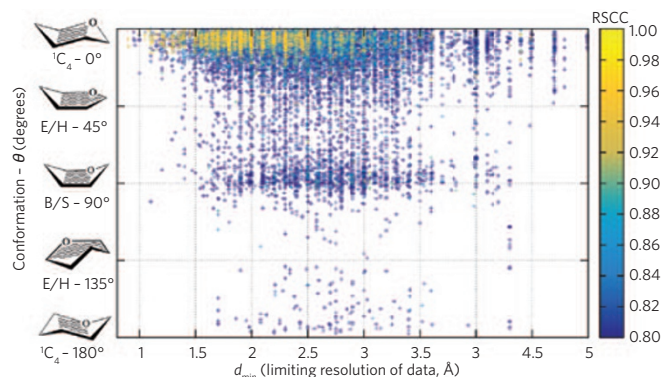


Figure 1 | Distribution of D-pyranoside ring conformations as a function of resolution for all N-linked sugars (at distance <2.0 Å) in the PDB as of January 2015, identified by their Chemical Component Dictionary IDs: NAG, NDG, MAN, BMA, BGC, GLC, GAL and GLA. E/H, envelopes and half-chairs; B/S, boats and skew-boats; wavy lines denote the main ring plane. For clarity, an envelope is depicted at $\theta = 45^\circ$ and a half-chair at $\theta = 135^\circ$, and skew-boat is omitted from the equator.

when interpreted in a biological context. A plot of ring conformation against resolution of the X-ray analysis is shown in **Figure 1**. As expected, at atomic resolutions (<1.2 Å, model precision better than 0.01 Å), all sugars showing high RSCC are ⁴C₁ chairs (yellow cluster). However, as the resolution gets lower and model precision poorer, unexpected higher-energy conformations start to appear. Most of these models also show low RSCC (<0.8; blue entries).

Although energetically unfavorable models may reflect a poor knowledge of glycochemistry and 'optimistic' density interpretation (reflected in low RSCCs), it is nevertheless clear that in many cases macromolecular crystallographers are failing to apply appropriate conformational restraints to encourage chemically sensible models at lower resolutions (>1.6 Å). Although community re-refinement efforts such as PDB_REDO⁷ have led to substantial improvement in protein models, many sugars are still in high-energy conformations as a result of re-refinement without dihedral restraints. Torsion restraints, which approximate the eponymous energy barriers, can be used to penalize models with eclipsed conformations, encouraging a particular ring puckering for sugars. However, the perceived difficulty of modeling torsional preferences often results in these restraints being tacitly turned off, regardless of the resolution, in many refinement and model-building programs. This creates a vicious circle: publication and deposition of incorrect structures informs subsequent statistical

analyses that suggest the deposited structures are 'normal'.

Problems with the refinement of protein structures in the 1980s led to the rise of standard dictionaries, consistent refinement strategies, better graphics programs and community-accepted best practices. Ligands generally, and carbohydrates especially, got left behind. The fundamental roles of carbohydrates in cell biology and medicine, the extraordinary experimental advances in carbohydrate synthesis and the large increase in eukaryotic expression systems now demand improved refinement protocols for these key biological species, too.

Acknowledgments

This work was supported by Biotechnology and Biological Sciences Research Council grant BB/K008153/1.

Competing financial interests

The authors declare no competing financial interests.

References

1. Davies, G.J. *et al. Acc. Chem. Res.* **45**, 308–316 (2012).
2. Reynolds, C.H. *ACS Med. Chem. Lett.* **5**, 727–729 (2014).
3. Speciale, G. *et al. Curr. Opin. Struct. Biol.* **28**, 1–13 (2014).
4. Agirre, J. & Cowtan, K. *Comput. Crystallogr. Newslett.* **Jan**, 10–12 (2015).
5. Cremer, D. & Pople, J.A. *J. Am. Chem. Soc.* **97**, 1354–1358 (1975).
6. Lütke, T. *et al. Carbohydr. Res.* **339**, 1015–1020 (2004).
7. Joosten, R.P. *et al. Bioinformatics* **27**, 3392–3398 (2011).

Jon Agirre, Gideon Davies, Keith Wilson & Kevin Cowtan

York Structural Biology Laboratory, Department of Chemistry, The University of York, UK.
e-mail: kevin.cowtan@york.ac.uk or jon.agirre@york.ac.uk

ERRATUM

Carbohydrate anomalies in the PDB

Jon Agirre, Gideon Davies, Keith Wilson & Kevin Cowtan

Nat. Chem. Biol. **11**, 303 (2015); published online 17 April 2015; corrected after print 17 June 2015

In the version of this article initially printed, the shorthand label in Figure 1 for the preferred chair conformation (top structure shown) was incorrectly shown as 1C_4 instead of 4C_1 . The figure is correct in the HTML and PDF versions of the article.