# Validation of carbohydrate structures in CCP4 6.5

Jon Agirre and Kevin Cowtan
*Department of Chemistry, The University of York, YO10 5DD (UK)*
Correspondence email: kevin.cowtan@york.ac.uk

## Introduction

Pyranose and furanose sugars, as most other cyclic compounds, have strong conformational preferences that are dictated by a minimization of angle, torsional and steric strains. For most of the biologically relevant pyranoses, the preferred conformation is either a $^4C_1$ or a $^1C_4$ chair, and any transitions to higher-energy conformations (*e.g.* half-chair or envelope) are usually a consequence of external factors such as the neighboring presence of catalytic residues from a carbohydrate-active enzyme.

While the set of geometric restraints that crystallographic refinement software impose is usually descriptive enough to reproduce a realistic geometry for amino acids modeled at medium to low resolution, cyclic sugars may end up in a high-energy conformation that, in the absence of clear density supporting it, should be treated as an outlier. Even with the addition of harmonic torsion restraints, any subtle mistakes in the specification of bonding distances – linkages between sugars need to be explicitly declared – or a wrong three-letter code selection (*e.g.* using 'GLC' for β-D-glucopyranose, together with the restraints designated for it) can result in distortion.

## Conformational analysis

The method proposed by Cremer and Pople (1975) has been chosen as primary conformational analysis tool. The algorithm, which is applicable to rings of any cardinality, calculates a minimal set of puckering coordinates that describe each conformation. A total puckering amplitude term (*Q*) is also calculated

$$Q = \sqrt{\sum_{j=1}^{N} (\vec{R}_j \cdot \vec{n})^2} = \sqrt{\sum_{j=1}^{N} Z_j^2}$$

for *N* atoms, with $\vec{R}_j$ being the positional vector of atom *j* in a coordinate system with the origin in the ring's geometrical center, and $\vec{n}$ being the unit vector normal to the ring's mean plane. Therefore, $Z_j$ accounts for the vertical displacement of atom *j*

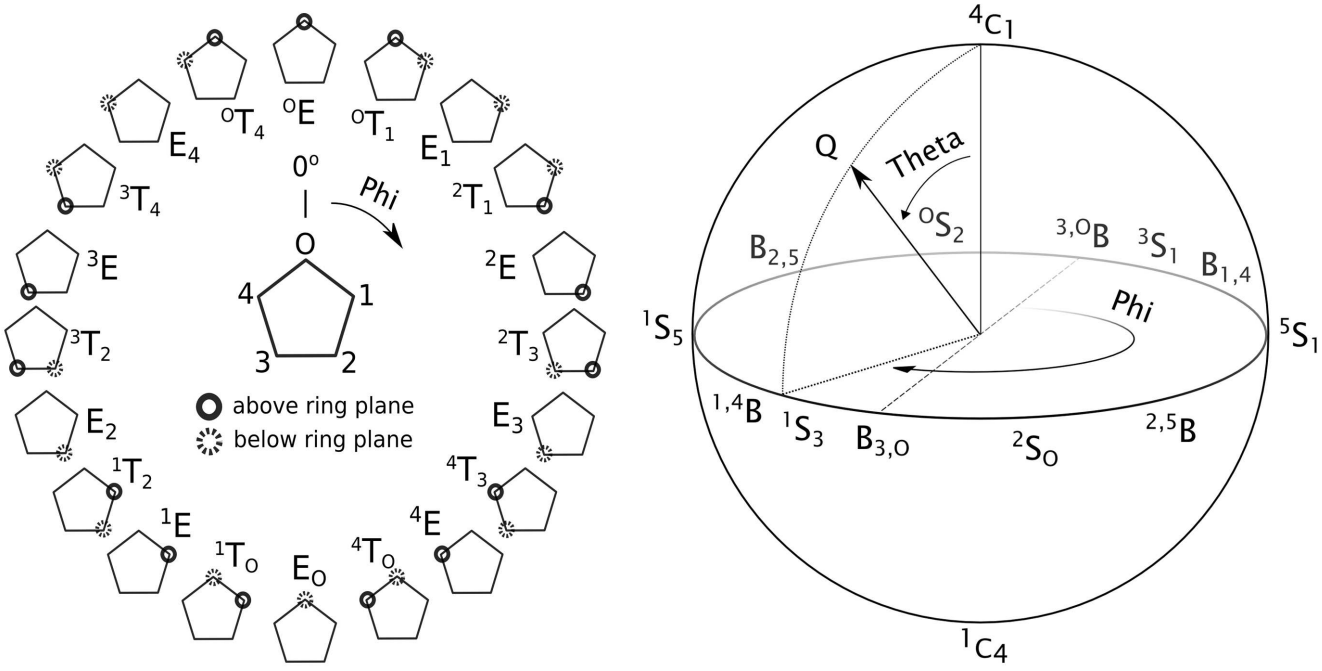


**Figure 1:** Correspondence between the Cremer-Pople angles for furanoses (Φ, image on the left) and pyranoses (Φ, Θ) to the conformation codes defined by IUPAC.

from the mean ring plane. In the case of pyranoside rings, the puckering coordinates are most conveniently expressed in angular form ($\Phi_{[0,2\Pi]}, \Theta_{[0,\Pi]}$) by solving the following set of equations

$$Q \sin\Theta \cos\Phi = \sqrt{\frac{1}{3}} \sum_{j=1}^{6} Z_j \cos\left[\frac{4\Pi(j-1)}{6}\right]$$

$$Q \sin\Theta \sin\Phi = \sqrt{\frac{1}{3}} \sum_{j=1}^{6} Z_j \sin\left[\frac{4\Pi(j-1)}{6}\right]$$

$$Q \cos\Theta = \sqrt{\frac{1}{6}} \sum_{j=1}^{6} (-1)^{j-1} Z_j$$

so that they can be graphically represented on the surface of a sphere of radius $Q$. This sphere, with lowest energy $^4C_1$ and $^1C_4$ ($\Theta = 0$ and $\Theta = \Pi$) chair conformations on the North and South pole respectively, is able to depict every conformational itinerary followed by pyranose sugars in their transition from their low-energy chair conformation to a more distorted boat or skew-boat intermediate ($\Theta = {}^{\Pi}/_{4}$) during catalysis (Davies *et al.*, 2011). For convenience, additional vertical displacements akin to the $Z_j$ ones are calculated for those atoms implicated in the anomer and handedness detection.

A similar calculation is performed for furanose rings, but producing just $Q$ and $\Phi$ .

## Characteristics

Privateer-validate relies on a small database of three-letter codes for which the anomer, handedness and lowest-energy conformation have been calculated. By comparison to these values, the program is able to determine, for instance, if a modeled carbohydrate has been distorted from its initial conformation. When run within CCP4i2 (currently in alpha test phase), an HTML report is displayed with the IUPAC-compliant conformation code, the Cremer-Pople parameters and diagnostics for each sugar. The equivalence between Cremer-Pople angles and conformations can be visualized in Figure 1.

In addition to chemical correctness checks, a real space correlation coefficient is calculated for each sugar against an mFo-DFc map computed omitting all sugar models from the phase calculation. The resulting map coefficients can also be output to an MTZ file for later use.

Whenever glycosylation is present in the input structure, the program will produce linear descriptions of the detected trees. Here is an example of the nomenclature used:

Man←$\alpha_2$–Man←$\alpha_6$–Man←$\alpha_6$–(Man←$\alpha_3$–)Man←
$\beta_4$–GlcNac←$\beta_4$–GlcNac ← β–$ASN_{523}$

Coot script files (Emsley *et al.*, 2010) are also produced with a guided tour of the detected issues. These scripts can be used manually outside CCP4i2 or by simply selecting 'Manual model rebuilding' within the aforementioned graphical interface. The omit mFo-DFc map is presented in pink color while 2mFo-DFc density is displayed in blue. Each button contains a description of the issue or issues detected by privateer-validate, as it can be seen in Figure 2.

The produced startup scripts also activate torsion angle restraints by default; using them in combination with the 'sphere refinement' function (hotkey: 'R') makes most of the issues exposed by privateer-validate easily fixable in Coot. Torsion angle restraints may be subsequently required by refinement software in order to avoid further distortions.

## Availability

The Privateer software package can be obtained as part of the CCP4 distribution (http://www.ccp4.ac.uk). The validation software presented here serves as the prelude to a sugar detection and modeling tool that will be distributed in the forthcoming weeks as an update to CCP4 6.5. Privateer uses the Clipper libraries (Cowtan, 2003) and is distributed under the terms of the GNU Lesser General Public License.

## Acknowledgements

The authors would like to thank Professors Eleanor J Dodson, Keith S Wilson and especially Gideon J Davies for many stimulating discussions.

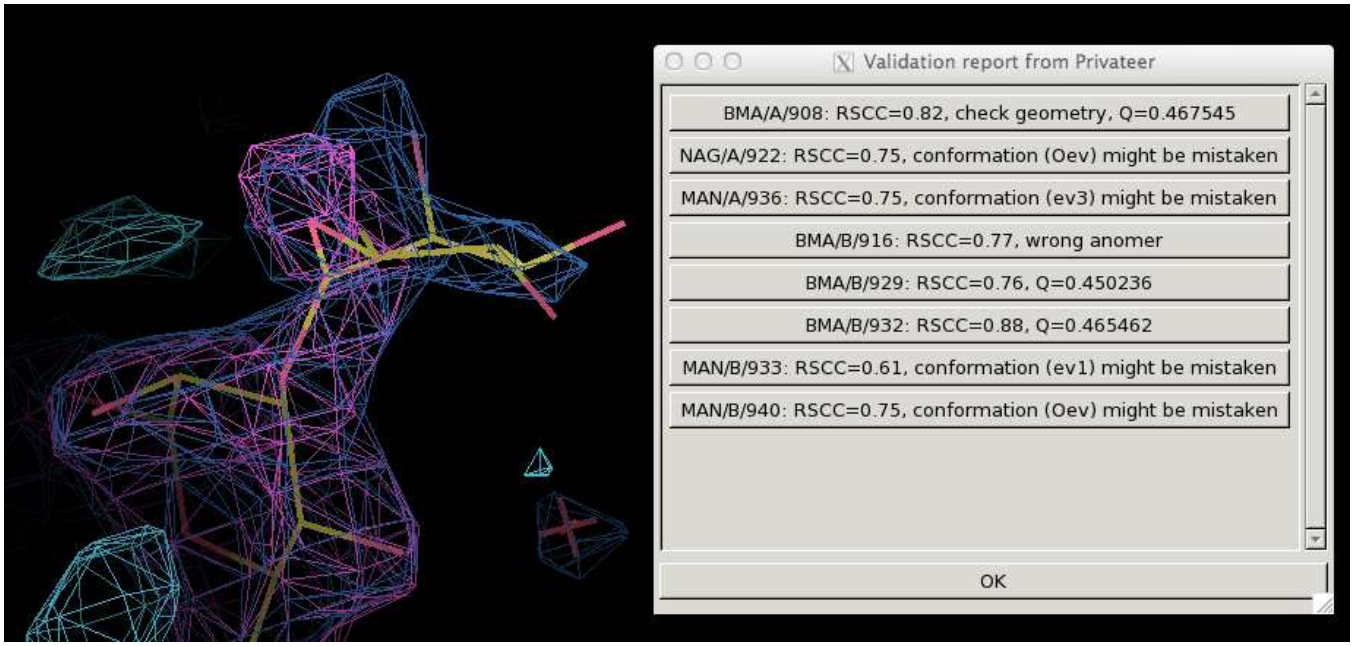


**Figure 2:** Validation of a glycoprotein (PDB code: 4IID). A number of terminal sugars display a high-energy conformation (envelope) as a consequence of the weak density they have been modeled in and the absence of torsion restraints in the original refinement.

## References

Cowtan, K. (2003). The Clipper C++ libraries for x-ray crystallography. IUCr Computing Commission Newsletter, 2:4–9.

Cremer, D. t. and Pople, J. (1975). General definition of ring puckering coordinates. Journal of the American Chemical Society, 97(6):1354–1358.

Davies, G. J., Planas, A., and Rovira, C. (2011). Conformational analyses of the reaction coordinate of glycosidases. Accounts of chemical research, 45(2):308– 316.

Emsley, P., Lohkamp, B., Scott, W., and Cowtan, K. (2010). Features and development of coot. Acta Crystallographica Section D: Biological Crystallography, 66(4):486–501.