Ideal bond lengths and angles in nucleic acid structures: an update for the 2020s

Andrew J. Howard*

Biological Sciences Department, Illinois Institute of Technology, Chicago, IL 60616, USA. *Correspondence e-mail: howard@iit.edu

In this issue of *Acta Crystallographica* Section B, a toolkit for analysis of nucleic acid structures is reported by Mariusz Jaskolski and co-workers (Gilski *et al.*, 2019). This paper is a valuable addition to the toolkit for analysis of nucleic acid structures. It alters the specific optimal values for bond-length and bond-angle restraint values in Watson–Crick base pairs. It also reduces the variability in those values beyond what is currently available, which are based on a study by Parkinson *et al* (1996). The approach taken here, incorporating Cambridge Structural Database (CSD; Groom *et al*., 2016) structures, quantum-mechanical calculations, and two ultra-high resolution nucleic acid crystal structures in the Protein Data Bank (PDB; Berman *et al*., 2000), provides a robust basis for both the ideal bond-length and bond-angle values and for the estimated variability in those values. The Parkinson study was based only on CSD structures available in 1996; the Gilski study has ten times as many CSD structures from which to draw, and the authors are combining the CSD analyses with these other two sources of information. They provide tables of bond lengths and angles not only for CG and AT pairs, but also for the isolated bases, for an isocytosine–isoguanine base pair, and for isocytosine and isoguanine alone. The authors subject the data to an elaborate statistical analysis that provides further justification for their values. They also inquire into whether the bond lengths and angles found in isolated bases differ significantly from those found in base pairs: the answer is that the differences are small but discernable.

The authors' quantum-mechanical calculations are based on nucleobases attached to methyl groups rather than on actual nucleosides or nucleotides. This simplifies the QM calculations significantly as compared to real nucleosides, and one can hope that that simplification will have little effect on the correctness or applicability of the data. The fact that the QM results are, for the most part, consistent with the CSD and PDB results provides additional confidence that their QM models are adequate. The accuracy of those QM results is poorer than the experimental results, but it remains reliable enough to be usable in situations where, as they state, 'no reliable experimental structural information is available.'

The authors base their analysis in part on two oligonucleotide structures available in the Protein Data Bank, namely 3p4j, which is a CG-only Z-DNA structure determined at 0.55 Å resolution; and 1d8g, which is a B-DNA decamer containing all four bases, determined at 0.74 Å resolution. Both of these structures were determined without applying restraints on the bond lengths and angles in the final refinements, which justifies their scrutiny in a study of this sort. Neither of these structures incorporated charge-density analyses. The authors note that the bond lengths and angles in 3p4j are considerably more uniform than in 1d8g. This higher degree of regularity in 3p4j may arise because the structure was determined at somewhat higher resolution, as the authors suggest; but it may also show the influence of the narrower range of conformational heterogeneity available in a CG-only structure, or even because Z-DNA is perhaps more conformationally constrained than B-DNA. The authors' efforts here could inspire further analysis of other ultra-high resolution nucleic acid structures, to see whether the variability (or lack thereof) discussed in this study carries over into those other structures. There are now at least eight DNA structures and three RNA structures in the PDB that have been determined at better than 0.85 Å resolution. Several of these were refined with
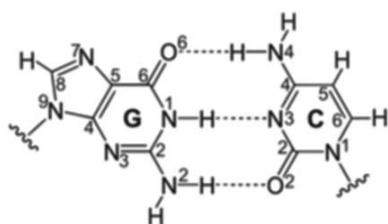
built-in bond length and angle restraints, which render them unsuitable for comparisons of this sort; but it would be straightforward to reanalyze those structures in the absence of those restraints, enabling this kind of analysis.

In general, though, this paper by Gilski *et al* represents a significant step forward in providing statistically justifiable estimates for bond lengths and angles in nucleic acid structures, and should be consulted in the development of the next generation of software packages in encoding restraints.

References

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

Gilski, M., Zhao, J., Kowiel, M., Brzezinski, D., Turner, D. H. & Jaskolski, M. (2019). *Acta Cryst.* B**75**, 235–245.

Groom, C. R., Bruno, I. J., Lightfoot, M. P. & Ward, S. C. (2016). *Acta Cryst.* B**72**, 171–179.

Parkinson, G., Vojtechovsky, J., Clowney, L., Brünger, A. T. & Berman, H. M. (1996). *Acta Cryst.* D**52**, 57–64.