# Crystal structure validation with the use of the DFT method

**F. Fňukal**

*University of Chemistry and Technology, Prague, Technická 5, 166 28 Praha 6 – Dejvice*

*fnukalf@vscht.cz*

## Introduction

Current crystal structure validation procedures rely primarily on diffraction data analysis, as implemented in tools such as PLATON/checkCIF [1]. A complementary approach – comparison of the geometry of experimental structures with their DFT-optimized geometries – was introduced by van de Streek & Neumann [2, 3], but did not gain widespread adoption due to high computational demands and reliance on commercial software. Recent advances in computing power and DFT methodology now make this approach practically viable for routine use on standard desktop hardware even for organic molecular crystals with large unit cells.

## Method

To facilitate practical application of this approach, we implemented the validation workflow in the program checkCIF-DFT, a freely available program that automates parsing CIF files, management of DFT geometry optimizations via the use of external DFT programs and evaluation of structural discrepancies using a set of comparison descriptors. In our testing so far, geometry optimizations were performed using CASTEP [4] with the r2SCAN [5] meta-GGA functional and Many-Body Dispersion (MBD) correction [6]. Our descriptor set extends the previously proposed r.m.s. Cartesian displacement (RMSCD) [2, 3] with maximum changes in bond lengths, bond angles and torsion angles, maximum Cartesian displacement and a temperature-corrected volume difference – the latter accounting for the formal 0 K conditions of DFT optimization using thermal expansion coefficients derived from Cambridge Structural Database (CSD) [7] data [8].

## Results

To set reference values of comparison descriptors, the method was benchmarked against two reference sets: five structures determined by neutron diffraction, representing correct high-quality experimental determinations, and five incorrect structures known to be obtained by data manipulation and fraud [9]. The use of our extended comparison descriptor set is justified as the use of the original RMSCD criterion fails to flag three of the five fraudulent structures as problematic, which was already noted by van de Streek & Neumann in their work. However, the extended comparison descriptor set – particularly maximum bond length difference and maximum bond angle difference – clearly separated all five fraudulent structures from the neutron structures supporting the idea of the need for a multi-parameter validation approach.

The validation procedure was next applied to a representative set of 100 organic crystal structures drawn from the CSD. Of these, 90 calculations completed and the large majority showed deviations consistent with correct experimental determinations (see Fig. 1). However, a subset showed elevated deviations revealing several recurring categories of structural problems: unmodelled solvent-accessible voids, missing atoms, incorrect space group assignment etc. Notably, several calculation failures were also associated with structures showing some of these problems, which offers a possible explanation to the calculation difficulties.

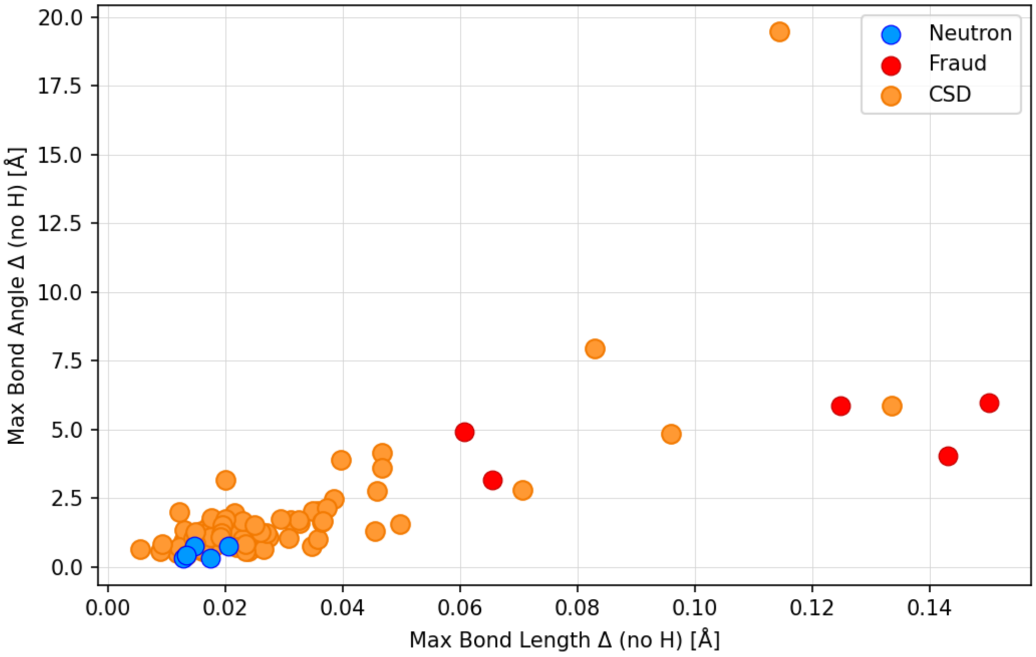

**Figure 1**. Maximum bond length difference versus maximum bond angle difference for the neutron structures (blue circles), fraudulent structures (red circles) and 90 successfully calculated CSD structures (orange circles) calculated excluding hydrogen atoms.

## Conclusions

Our results suggest that DFT-based crystal structure validation may serve as a physically grounded complementary validation method to conventional validation checks. This independent validation is of great use especially in cases where determination errors were made but are difficult to spot or in cases where determination details are poorly documented. Furthermore, the results suggest that the CSD is not free of structures with serious structural issues, even among those that have passed conventional validation filters. Our implementation of the method – the program checkCIF-DFT – is planned to be made freely available and is intended to make this DFT validation workflow accessible to the broad crystallographic community. Additionally, it is planned to extend checkCIF-DFT with pre-calculation checks to reveal most common structural issues that are identifiable prior to DFT optimization calculations. Furthermore, support for the use of Machine Learning Interatomic Potentials (MLIPs) is being developed as a significantly faster alternative to DFT calculations.

1. A. L. Spek, *Acta Cryst.*, **D65**, (2009), 148.
2. J. van de Streek, M. A. Neumann, *Acta. Cryst.*, **B66**, (2010), 544.
3. J. van de Streek, M. A. Neumann, *Acta Cryst.*, **B70**, (2014), 1020.
4. S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. I. J. Probert, K. Refson, M. C. Payne, *Z. Kristallogr.*, **220**, (2005), 567.
5. J. W. Furness, A. D. Kaplan, J. Ning, J. P. Perdew, J. Sun, *J. Phys. Chem. Lett.*, **11**, (2020), 8208.

6. A. Tkatchenko, R. A. DiStasio Jr., R. Car, M. Scheffler, *Phys. Rev. Lett.*, **108**, (2012), 236402.

7. A. van der Lee, D. G. Dumitrescu, *Chem. Sci.*, **12**, (2021), 8537.

8. F. H. Allen, *Acta Cryst.*, **B58**, (2002), 380.

9. W. T. Harrison, J. Simpson, M. Weil, *Acta Cryst.*, **E66**, (2009), e1.

*Computational resources were provided by the e-INFRA CZ project (ID:90254), supported by the Ministry of Education, Youth and Sports of the Czech Republic.*