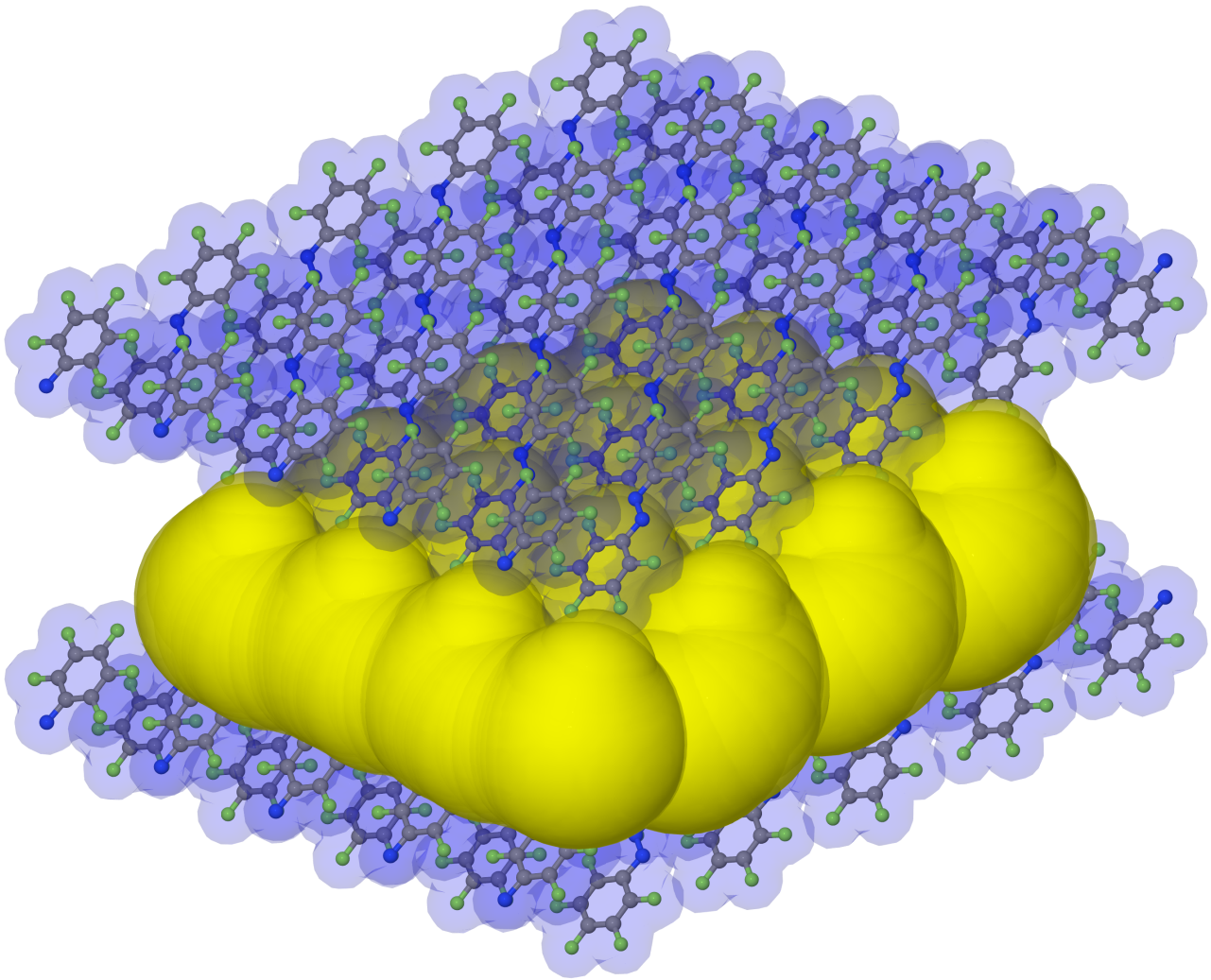

Spotting voids in the crystals of small molecules

Andrius Merkys
Doctoral assistant – Theory and Simulation of Materials (THEOS) and National Center for Computational Design and Discovery of Novel Materials (MARVEL)

Detection of incomplete data is an easy task for a human being, but a complex one for a computer. Usually, checking a few data items manually is not a burden for the researcher, however manual checks become a serious bottleneck for the database mining and analysis of big data. Crystallography Open Database (COD) [1], the largest up to date open resource of experimentally determined small molecule crystal structures, aims to automatically detect incomplete crystal structures by scanning it's data for voids and either fixing or marking such entries to be treated with caution in automated analyses. A program, based on *voronota* software tool [2], is harnessed in such scans.

The yellow balls are fitted inside voids (only balls larger than 3 Å in diameter are shown) in transdecafluoroazobenzene crystal structure (van der Waals surface shown in blue), originating from COD entry 2001917. The largest ball has a diameter larger than 5 Å. The voids in this structure were apparently introduced by the lack of 6 symmetry operators in the description of crystal structure. However, all voids have disappeared after adding missing symmetry operators, as suggested by the symmetry space group symbol. Image is generated using *Jmol* [3] and *POV-ray* [4].

[1] Gražulis et al. *Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration*, Nucleic Acids Research, 2012

[2] Olechnovič & Venclovas. *Voronota: A Fast and Reliable Tool for Computing the Vertices of the Voronoi Diagram of Atomic Balls*, Journal of Computational Chemistry, 2014

[3] Hanson. *Jmol – a paradigm shift in crystallographic visualization*, Journal of Applied Crystallography, 2010

[4] Persistence of Vision Pty. Ltd. Persistence of vision raytracer (version 3.6), 2004.