

MATTS2025: Machine Learning-Enhanced Expansion of the MATTS Data Bank for Small Molecule Electron Density Modeling

P. M. Rybicka¹, V. M. Ignat'ev¹, M. Kulik¹, P. M. Dominiak¹

¹University of Warsaw, Faculty of Chemistry, Biological and Chemical Research Centre, ul. Żwirki i Wigury 101, 02-089 Warsaw, Poland

p.rybicka@uw.edu.pl

Accurate electron density modelling is crucial for determining molecular structures from experimental X-ray and 3D electron diffraction data, as well as for understanding chemical bonding, intermolecular interactions, and calculating properties like electrostatic potential. This is significantly enhanced by using aspherical models, such as the Multipole Model [1], supported by transferable electron density databanks like MATTS [2, 3].

Following the release of MATTS2021, several limitations were identified, particularly in representing unusual functional groups and complex chemical topologies. To address this, we introduce **MATTS2025**, a substantially expanded and restructured data bank. The set of model molecules was enlarged by 40%, covering less common functional groups and fused or bridged ring systems, leading to the addition of numerous new atom types and significantly reducing unassigned atoms. Key methodological improvements include lowering the covalent bond detection threshold from 0.40 to 0.23 Å for better handling of small rings, switching from Clementi-Roetti to Su-Coppens-Macchi radial functions, and replacing the 6-31G** basis set with def2-TZVP, resulting in improved refinement accuracy [4]. A systematic analysis of the effect of the symmetry constraints led to implementing refinements without them and the introduction of new local coordinate systems to better represent the apparent pseudosymmetry of atomic electron densities [5]. Furthermore, supervised and unsupervised machine learning techniques were employed to classify, cluster, and redefine atom types based on their electron density descriptors, enhancing transferability and generalization for chemically complex cases.

Together, these developments deliver a more accurate, comprehensive, and versatile databank for aspherical electron density modelling in small-molecule crystallography, expanding the practical applicability of multipole-based methods in structural and property analysis.

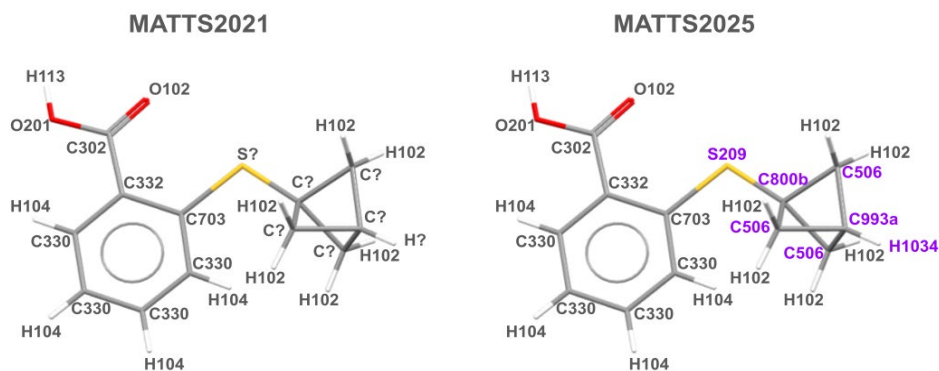


Figure 1. Example of a molecule featuring complex ring systems and functional groups newly covered in the MATTS2025 data bank. Atom types assigned according to MATTS2021 (left) and MATTS2025 (right) data banks are indicated.

[1] Hansen, N.K. *et al.* (1978). *Acta Cryst. A* **34**, 909-921.

[2] Jha, K.K. *et al.* (2022). *J. Chem. Inf. Model.* **62**, 3752-3765.

[3] Rybicka, P.M. *et al.* (2022). *J. Chem. Inf. Model.* **62**, 3766-3783.

[4] Ignat'ev, V.M. *et al.* (2024). *J. Appl. Cryst.* **57**, 1884-1895.

[5] Rybicka, P.M. *et al.*, *in preparation*.

This research was funded by National Science Centre, Poland UMO-2020/39/I/ST4/02904.