# A unique and continuous code of all periodic crystals

O. Anosova, D. Widdowson, V. Kurlin

University of Liverpool, Liverpool L69 3BX, United Kingdom, vitaliy.kurlin@liverpool.ac.uk

A conventional representation of a periodic crystal by its primitive unit cell and motif is well-known to be ambiguous. Indeed, any crystal can be generated from infinitely many primitive unit cells and motifs containing differently located atoms. Niggli's reduced cell is unique but discontinuous under perturbations. Continuity of crystal representations is important for filtering out near duplicates in big datasets [1, Fig. 2d] of simulated crystals in Crystal Structure Prediction (CSP). Symmetry groups and many other descriptors discontinuously change under perturbations. So CSP landscapes are plotted only by two coordinates: the lattice energy and density.

We describe a new geometric approach to generating a unique code (called a *crystal isoset*) of any periodic crystal, which continuously changes under perturbations of atoms [2-3]. This isoset is a *material genome* or a *DNA-type code* that allows an inverse design of new periodic crystals. Using these complete isosets, one can compute invariants via density functions [4] and interatomic distances [5]. For any crystal dataset irrespective of symmetries or chemical compositions, invariants of crystals can be joined in a minimum spanning tree via continuous distances that quantify crystal similarities. Our Python code of distance-based invariants produced a map of all 229K organic crystals in the Cambridge Structural Database overnight on a modest desktop [6 (appendix D), 7].
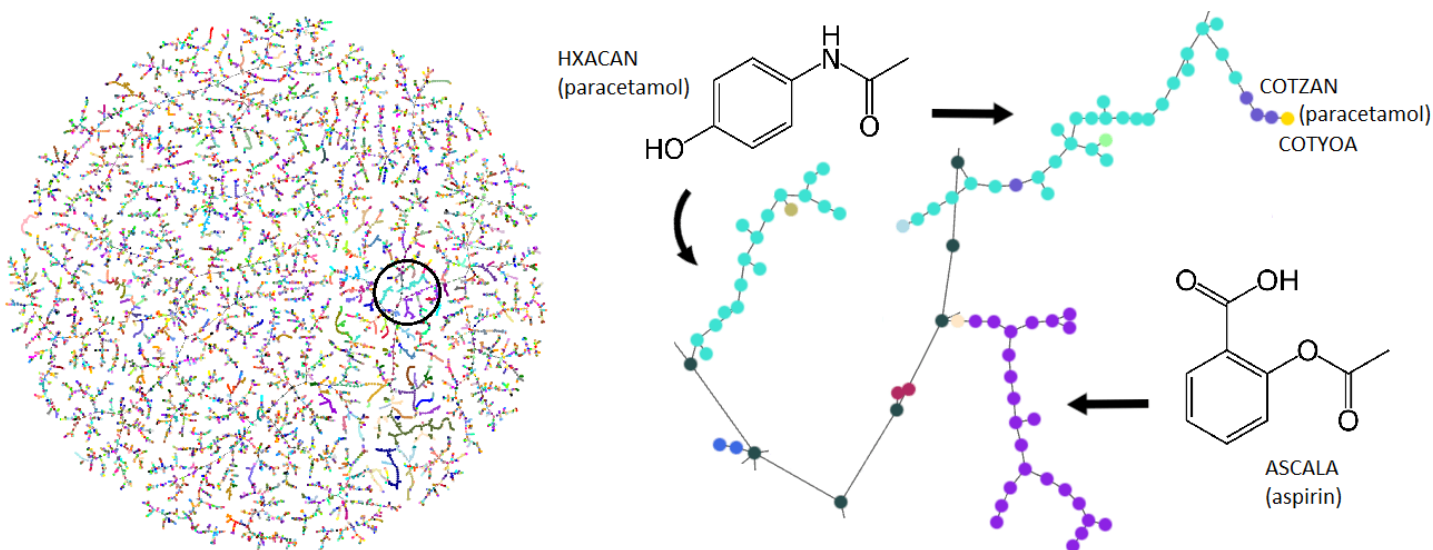
Figure 1. A new invariant-based visualization is illustrated on the CSD Drug Subset of 12,576 structures colored by CSD ref codes. **Left**: A tree joins those structures that have close values of invariants extracted from isosets. **Right**: the interactive invariant-based map is zoomed to show families of chemically different aspirin and paracetamol, which are in close branches as in a pharmacy.

[1] Pulido, A. et al, Functional materials discovery using energy–structure–function maps. *Nature*, *543*(7647), pp.657-664.

[2] Anosova, O., Kurlin, V. (2021). An isometry classification of periodic point sets. Peer-reviewed proceedings of *Discrete Geometry and Mathematical Morphology*. Available at http://kurlin.org/research-papers.php#DGMM2021.

[3] Anosova, O., Kurlin, V. (2021). Introduction to Periodic Geometry and Topology. Available at https://arxiv.org/abs/2103.02749.

[4] Edelsbrunner, H., Heiss, T., Kurlin, V., Smith, P, Wintraecken, M. (2021). The density fingerprint of a periodic point set. Peer-reviewed proceedings of *Symposium on Computational Geometry*. Available at https://arxiv.org/abs/2104.11046.

[5] Widdowson, D., Mosca, M.M., Pulido, A., Kurlin, V., Cooper, A.I. Average Minimum Distances of periodic point sets. To appear in MATCH Communications in Mathematical and in Computer Chemistry (2022). Available at https://arxiv.org/abs/2009.02488.

[6] Widdowson, D., Kurlin, V. Pointwise Distance Distributions of periodic sets. Available at http://arxiv.org/abs/2108.04798.

[7] Ropers, J., Mosca, M.M., Anosova, O., Kurlin, V., Cooper, A.I. Fast predictions of lattice energies by continuous isometry invariants of periodic crystals. Peer-reviewed proceedings of DACOMSIN (Data and Computation for Materials Science and Innovation) 2021.

Keywords: crystal similarities; maps of crystal datasets; crystal structure prediction; continuous classification of crystals