Quantifying Structural Novelty via Element Substitutions for AI-generated Crystals

Masahiro Negishi[1], Aron Walsh[1]
[1]Department of Materials, Imperial College London, Exhibition Road, South Kensington, London SW7 2AZ, UK

m.negishi25@imperial.ac.uk

Recent years have witnessed a growing interest in the inverse design of crystals using generative models [1]. These models are capable, in principle, of exploring a wider material space than conventional methods such as genetic algorithms (GA). However, the structural novelty of AI-generated crystals has been questioned, with critics suggesting that their underlying structures are frequently already known despite their novel compositions [2, 3]. This raises the important question of whether generative models truly expand the search space beyond what can be achieved using conventional strategies such as substitution into known structure types. To address this question, this study set out to investigate whether AI-generated crystals could be obtained by applying simple operations to training data crystals, similar to those used in GA. For simplicity, the operations were restricted to atomic substitutions and a single final structural relaxation. For each generated crystal, a training prototype crystal exhibiting high structural and elemental similarity was selected as the starting point. Structural similarity was determined based on atomic coordinates and lattice constants, or space groups and Wyckoff alphabets, while elemental similarity was assessed using the modified Pettifor scale [4]. Our experimental results revealed that over 80% of the metastable and chemically valid crystals generated by MatterGen could be obtained solely through elemental substitution from the training data [5, 6]. Furthermore, 100% and 99% of the generated high-symmetry cubic and hexagonal crystals, respectively, were found to be obtainable via substitution. These results suggest that current generative models fail to significantly broaden the search space compared to substitution strategies, particularly in the high-symmetry region.

References

[1] Pierre-Paul De Breuck, et al., npj Computational Materials, 1 (2025) 370
[2] Masahiro Negishi, et al., arXiv, 2510.12405 (2026).
[3] Nathan J. Szymanski and Christopher J. Bartel, Materials Horizons, 19 (2025) 8000-8011.
[4] Henning Glawe, et al., New Journal of Physics, 9 (2016) 093011
[5] Daniel W. Davies, et al., Chem, 4 (2016) 617-627
[6] Claudio Zeni, et al., Nature, 8055 (2025) 624-632

Figures

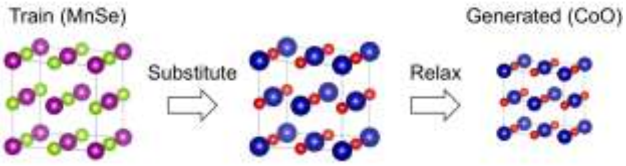

Figure 1. A schema of our novelty evaluation algorithm. An AI-generated structure is not structurally novel if it can be derived from a training structure via element substitutions and one structural relaxation. In this example, the AI-generated CoO crystal is not structurally novel, because it can be obtained from MnSe.

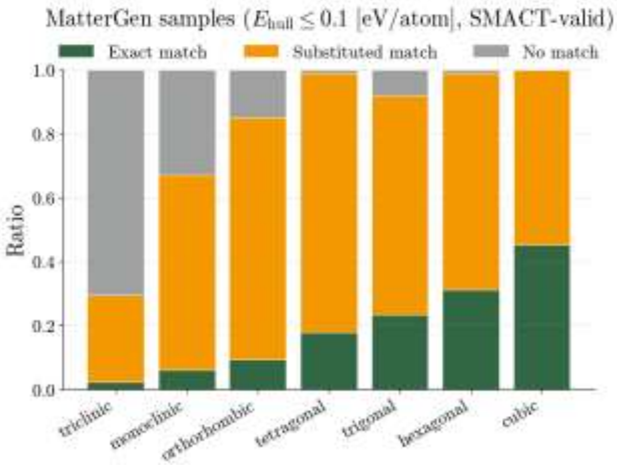

Figure 2. Matching ratios for MatterGen-generated crystal structures (metastable and chemically valid) across different crystal systems. "Exact match" refers to generated crystals that duplicate a training structure. "Substituted match" means that the generated structures can be obtained by applying element substitutions and one structural relaxation to a training sample. Those entries that do not belong to either category are classified as "No match."