MS28-1-11 Integrating machine learning in crystal structure prediction for pharmaceutical compounds #MS28-1-11

A. Anelli [1], H. Dietrich [2], P. Ectors [1], F. Stowasser [1], T. Bereau [3], M. Neumann [2], J. Van Den Ende [4]
[1]Hoffmann La Roche - Basel (Switzerland), [2]Avant-garde Materials Simulation - Merzhausen (Germany), [3]University of Amsterdam - Amsterdam (Netherlands), [4]Hoffman La Roche - Basel (Switzerland)

Abstract
Organic molecules can crystallize into various solid forms, which generates a polymorphic degree of freedom[1]. Since the different solid forms have distinct physico-chemical behaviour (e.g. solubility, compressibility etc…), it is paramount to know which crystalline packings an active pharmaceutical ingredient (API) can aggregate into. A promising route to characterize the crystalline landscape of APIs is to perform quantum mechanical atomistic simulations of each potential polymorph, so as to obtain a much needed ranking of the observable crystals[2]. While this routine is fully general and applicable to any class of compounds, it requires a substantial computational investment, allowing its use mostly for molecules in the later drug development stages. To decrease the computational overhead of such approaches, we propose an integrated machine learning crystal structure prediction (ML-CSP) framework. We combine on the fly machine learning potentials[3] with a Bayesian selection scheme to accurately sample the crystal structure landscape of a target API. In Figure 1, we show how the accuracy of a model trained on PBE-DFT energies for fentanyl molecular crystals improves as we progress our ML-CSP routine. By using our method, we report a 3 to 5 fold increase in performances, paving the way for crystal structure prediction to penetrate in the earlier development stages or target more complex compounds.

References
[1] : Cruz-Cabeza, A.J., Feeder, N. and Davey, R.J., 2020. Open questions in organic crystal polymorphism. Communications Chemistry, 3(1) DOI : https://doi.org/10.1038/s42004-020-00388-9
[2] : Hoja, J., Ko, H.Y., Neumann, M.A., Car, R., DiStasio Jr, R.A. and Tkatchenko, A., 2019. Reliable and practical computational description of molecular crystal polymorphs. Science advances, 5(1) DOI : 10.1126/sciadv.aau3338
[3] Sivaraman, G., Krishnamoorthy, A.N., Baur, M., Holm, C., Stan, M., Csányi, G., Benmore, C. and Vázquez-Mayagoitia, Á., 2020. Machine-learned interatomic potentials by active learning: amorphous and liquid hafnium dioxide. npj Computational Materials, 6(1) DOI : https://doi.org/10.1038/s41524-020-00367-7