MS35 Artificial intelligence in photon and neutron crystallography, data mining, machine learning

MS35-03
Cocrystal Discovery with Network Science and Machine Learning
J.J. Devogelaer [1], H. Meekes [1], P. Tinnemans [1], E. Vlieg [1], R. De Gelder [1]
[1]Radboud University - Nijmegen (Netherlands)

Abstract
Physicochemical and biopharmaceutical properties of pure compounds can be modified by the formulation of multicomponent crystals, such as salts, solvates and cocrystals, which are crystalline aggregates containing multiple ionic and/or neutral species in the crystal lattice [1]. The formation of salts is rather straightforward but the design of cocrystals remains challenging. Experimental screening of cocrystals is time-consuming and therefore computational tools to understand and predict cocrystals can be crucial for an efficient and successful screening process.
Recently we constructed a network of coformers from the information present in the Cambridge Structural Database (CSD). Analysis of this network shows that the coformers in the network behave primarily in a bipartite manner, demonstrating the importance of combining complementary coformers [2].
Using link-prediction algorithms, missing links in the network can be identified: cocrystals that are not yet present in the CSD but are likely to exist based on the information in the network [3]. For a target compound unknown to the CSD, in-house data on cocrystallization can be combined with the entire CSD coformer network and link-prediction may quite accurately predict new combinations for the target compound [4]. Moreover, using link prediction an invalid cocrystal set can be generated that can serve as negative input data for an approach based on artificial intelligence. The cocrystals present in the CSD, together with the results from link prediction applied to the coformer network, were used as input to a machine learning approach. We trained artificial neural networks (ANNs) to recognize fruitful combinations of coformers. Validation of the ANN model showed that high accuracies are obtained and that the combination of network science and machine learning delivers a powerful prediction tool for discovering new cocrystals [5].

References
[1] Grothe, E., Meekes, H., Vlieg, E., ter Horst, J. H., de Gelder, R. *Cryst. Growth Des.*, 2016, **16**, 3237–3243.
[2] Devogelaer, J.J., Meekes, H., Vlieg, E., de Gelder, R. *Acta Cryst. B*, 2019, **75**, 371-383.
[3] Devogelaer, J.J., Brugman, S.J.T., Meekes, H., Tinnemans, P., Vlieg, E., de Gelder, R. *Crystengcomm*, 2019, **21**, 6875-6885.
[4] Devogelaer, J.J., Charpentier, M.D., Tijink, A., Dupray, V., Coquerel, G., Johnston, K., Meekes, H., Tinnemans, P., Vlieg, E., ter Horst, J.H., de Gelder, R. *Cryst. Growth Des.*, 2021, **21**, 3428-3437.
[5] Devogelaer, J.J., Meekes, H., Tinnemans, P., Vlieg, E., de Gelder, R. *Angew. Chem. Int. Ed.*, 2020, **59**, 21711-21718.