Invited Lecture

Exploring the cocrystal network in the Cambridge Structural Database

T.E. de Vries[1], H. Meekes[1], P. Tinnemans[1], E. Vlieg[1], R. de Gelder[1]

[1]Radboud University, Institute for Molecules and Materials, Solid State Chemistry, Heyendaalseweg 135, 6525 AJ Nijmegen, The Netherlands

r.degelder@science.ru.nl

The physico-chemical properties of active pharmaceutical ingredients, agrochemicals, and other useful chemicals can be altered and optimised by forming multi-component crystals, such as salts, solvates and cocrystals. Discovering suitable multi-component crystals, especially cocrystals, is time-consuming and labour-intensive. The Cambridge Structural Database (CSD) contains a massive number of cocrystals that can be used to construct an extensive network of coformers [1]. This network contains predictive power for discovering new cocrystals by applying link-prediction methods [2,3] or artificial neural networks [4].

It is possible to bipartise the entire CSD cocrystal network, which shows the remarkable global complementarity that is present in the network structure [5]. Bipartisation also reveals 'universal' coformers, that do not conform to the bipartite nature of the network, and makes identification of anticommunities of coformers possible. The accuracy of network-based cocrystal prediction can be improved by applying bipartisation to the CSD cocrystal network prior to link prediction. Moreover, combining bipartisation with the Multi- Steps Resource Allocation (MSRA) measure [6] outperforms the earlier applied scoring functions for link prediction [7].

Iteratively applying link prediction, cocrystal screening and network updating demonstrates that saturation of the local network around a target compound eventually diminishes the returns of the method. At the same time, it demonstrates that a steadily expanding CSD database is of great value to the discovery of new multi-component crystals.

[1] Devogelaer, J.J., Meekes, H., Vlieg, E. & de Gelder, R. (2019). *Acta Cryst. B,* **75**, 371-383.

[2] Devogelaer, J.J., Brugman, S.J.T., Meekes, H., Tinnemans, P., Vlieg, E. & de Gelder, R. (2019). *Crystengcomm,* **21**, 6875-6885.

[3] 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. (2021). *Cryst. Growth Des.,* **21**, 3428−3437.

[4] Devogelaer, J.J., Meekes, H., Tinnemans, P., Vlieg, E. & de Gelder, R. (2020). *Angew. Chem. Int. Ed.,* **59**, 21711-21718.

[5] De Vries, T.E., Vlieg, E. & de Gelder, R. (2024). *CrystEngComm*, **26**, 192-202.

[6] Wu, Z. & Li, Y. (2014). In: *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, **1**, 355-360.

[7] De Vries, T.E., van Eert, E., Weevers, L., Tinnemans, P., Vlieg, E., Meekes, H. & de Gelder, R. (2024). Cryst. Growth Des., 24, 5200–5210.