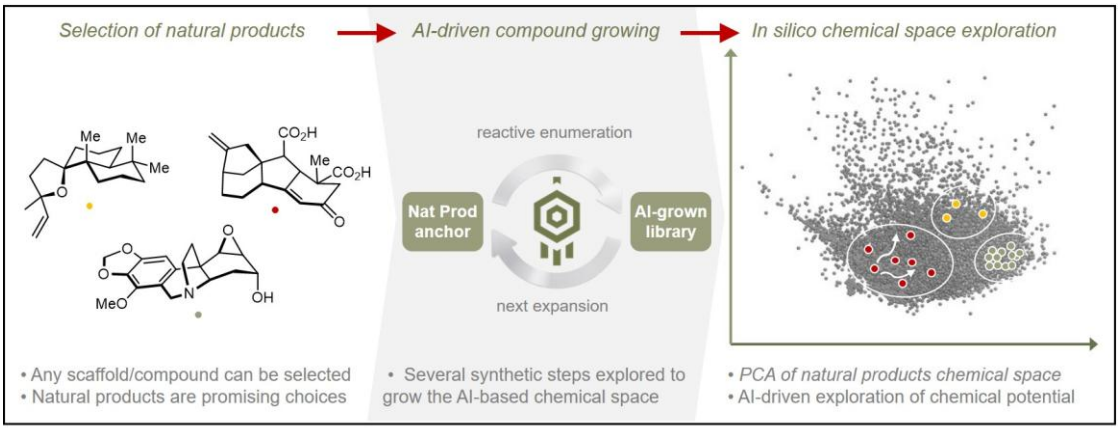
Deep Learning Applied To Ligand-Based De Novo Design: A Real-Life Lead Optimization Case Study

Perron, Q., *[1,2] Hoffmann, B.,[2] Tajmouati, H.,[2] Fourcade, R.,[2] Do Huu, N.,[2] Skiredj, A.,[2] Gaston-Mathé, Y.,[2]
[1]quentin.perron@iktos.com
[2] *Iktos, 65 rue de Prony 75017, Paris, France*



Natural products (NPs) are a valued source of biologically active molecules.[1] However, they often prove to be highly difficult to optimize by medicinal chemists. Indeed, NPs may be sensitive to reaction conditions or not versatile enough to enable derivatization for designing a robust drug candidate. That hurdle prevents their use in most medicinal chemistry projects. This work proposes to address this challenge by a purely data-driven approach. Recent papers have reported successful implementation of data-driven AI approaches to retrosynthetic analysis.[2] We extracted synthetic rules from the USPTO database.[3] We curated & selected several hundreds of NPs from the ChEBI database.[4] Then, applying data-driven retrosynthetic rules in the forward direction, we generated a library of synthetically accessible NP-derived molecules. Those compounds are designed to be druggable, and accessible *via* a limited number of steps, from the NP starting point and commercially available starting materials. Moreover, our method allows to assess the synthetic potential of NPs and to generate the accessible chemical space for a chosen scaffold. To our knowledge, such data-driven *in silico* synthetic approach is unprecedented. Moreover, its application to overcome the synthetic challenge of NP-related compounds is of particular interest for the medicinal chemists community.

Bibliography :

[1] Newman D.J. and Cragg G. M. Natural Products as Sources of New Drugs from 1981 to 2014. *J Nat Prod* **2016**, *79*, 629-661.

[2] a) Coley C.W., Rogers L., Green W.H. and Jensen K.F. Computer-Assisted Retrosynthesis Based on Molecular Similarity. *ACS Cent Sci.* **2017**, *3*, 1237-1245. b) Coley C.W., Green W.H. and Jensen K.F. Machine Learning in Computer-Aided Synthesis Planning. *Acc. Chem. Res.* **2018**, *51*, 1281–1289. c) Plehiers P.P., Marin G.B., Stevens C.V. and Van Geem K.M. Automated reaction database and reaction network analysis: extraction of reaction templates using cheminformatics. *J Cheminformatics* **2018**, 10:11.

[3] Segler M.H., Preuss M. and Waller M.P. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* **2018**, *555*, 604-610.

[4] Hastings J., Owen G., Dekker A, Ennis M, Kale N, Muthukrishnan V, Turner S, Swainston N, Mendes P and Steinbeck C. ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res*. **2016**, *44*, D1214–D1219.