Reaction-based De Novo Design

Val Gillet

University of Sheffield

De novo design refers to the design of novel chemical entities from scratch to fit a set of constraints. In drug discovery these constraints are typically modelled properties such as bioactivity, measured as protein ligand binding or using a QSAR (Quantitative Structure Activity Relationship) model; drug-likeness measured, for example, using a model such as QED (Quantitative Estimate of Drug-likeness); selectivity to avoid off-target effects; and so on. De novo drug design techniques were first proposed more than 30 years ago, however, given the huge size of chemical space and the multiobjective nature of drug design, it remains a very challenging area of research. A key issue in de novo design, in addition to the design constraints mentioned above, is ensuring that the designed compounds are synthetically accessible. Early approaches were agnostic of synthesis and consequently their application was limited. More recently, synthetic accessibility has been addressed in two main ways. One is to embed synthesis directly within the de novo design algorithm itself, in what is known as reaction-based de novo design. The other approach is post-generation filtering of the generated molecules, using either simple models trained on molecular descriptors, or more sophisticated retrosynthetic methods. Post-generation filtering is the typical approach taken by modern deep learning methods. This talk will briefly review the evolution of de novo design methods and will then focus on our work on reaction-based de novo design, concluding with a case study as proof-of-concept.