→ **RETROSYNTHESIS**

Computer says yes

> " There were a few steps that I did not really believe would work! That they did makes me wonder about the advantages offered by the ice cold objectivity of an algorithm "

Retrosynthetic analysis is the way that organic chemists draw an imaginary line from a target molecule to available precursors. In addition to functional group interconversions, bonds may be disconnected or rejoined in an attempt to find simpler starting materials. Disconnecting a bond results in two synthons — either a pair of oppositely charged species or two radicals. We then try to identify synthetic equivalents that can stand in for these hypothetical species in a forward reaction. This process is repeated until readily available starting materials are identified. Despite only being formalized in the past half-century — Elias J. Corey was awarded the 1990 Nobel Prize in Chemistry for "his development of the theory and methodology of organic synthesis" — the retrosynthesis process has become an integral part of the practice of organic chemistry. Whether the process might be automated has been debated ever since its earliest applications: can a computer armed with sufficiently large databases of organic reactions and starting materials design a synthesis more effectively than a human? Writing in *Chem*, Bartosz Grzybowski and co-workers describe eight syntheses planned by computer and then executed by the team in the laboratory.

Grzybowski's group have been developing their retrosynthesis planning software — Chematica — for over 10 years. "By 2012, we were quite proficient in using network algorithms to query synthetic paths comprised of published reactions, but we were nowhere near tackling the de novo retrosynthesis problem," he says. "It's easy to imagine that this is simply a question of how many reaction rules we can input, but more rules means more options. With our current knowledge base of some 60,000 reactions, there are, on average, 100 legitimate options to choose from in each step and a rapidly intractable 100^n options for an n-step synthesis."

It took another 5 years and joining forces with chemists at MilliporeSigma/Merck KGaA to tackle real retrosynthesis. "Our collaboration was important because we were keen to get experimental validation of the system and it was essential that this included target molecules selected by a third party," says Grzybowski. Eight targets were selected with different objectives in mind for the new retrosyntheses. Six compounds were selected by the team at MilliporeSigma, each with a market value of >US\$100/mg and for which synthetic attempts yielded little to no product or were not scalable. A seventh compound was an anti-arrhythmia drug (dronedarone, Sanofi-Aventis) whose synthesis was patented, and the eighth a natural product (engelheptanoxide C) for which a synthesis had not, at the time, been reported.

Chematica planned syntheses of each of these eight molecules, needing only 15–20 min for all except dronedarone, though this retrosynthesis was performed using an older and slower version of the software. "We are quite proud that our algorithms can find syntheses improving on previous approaches and on a timescale that would not irritate a practicing chemist," says Grzybowski. The software produces and scores the proposed routes, which were then tested in the lab with provision for only minor adjustments in reaction conditions (for example, to the temperature, solvent or a specific reagent). In all eight syntheses, the results exceeded those of previously reported routes. "There were a few steps that I did not really believe would work! That they did makes me wonder about the advantages offered by the ice cold objectivity of an algorithm over a human brain."

There are challenges ahead, however. "The molecules tackled so far are industrially relevant but certainly not of the highest synthetic complexity," says Grzybowski. "It will take a few years for us to to add more rules, accelerate the algorithms, fine-tune our scoring functions and use more CPU horsepower." He also doesn't believe that the use of such methods will limit creativity in designing completely new chemistry. "The program will not use, for example, a Wittig reaction, if we do not first teach it the Wittig reaction," states Grzybowski.

Stephen G. Davey
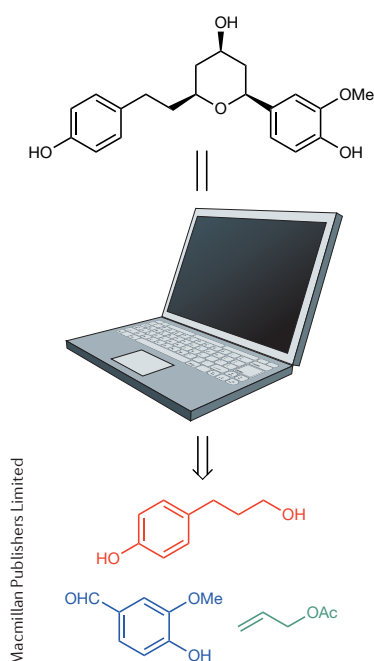
ORIGINAL ARTICLE Klucznik T. et al. Efficient syntheses of diverse, medicinally relevant targets planned by computer and executed in the laboratory. *Chem* **4**, 522–532 (2018)

Macmillan Publishers Limited