The Future of Retrosynthesis and Synthetic Planning: Algorithmic, Humanistic or the Interplay?

Craig M. Williams and Madeleine A. Dallaston

School of Chemistry and Molecular Biosciences, University of Queensland, Brisbane, Qld 4072, Australia.

Corresponding author. Email: c.williams3@uq.edu.au

The practice of deploying and teaching retrosynthesis is on the cusp of considerable change, which in turn forces practitioners and educators to contemplate whether this impending change will advance or erode the efficiency and elegance of organic synthesis in the future. A short treatise is presented herein that covers the concept of retrosynthesis, along with exemplified methods and theories, and an attempt to comprehend the impact of artificial intelligence in an era when freely and commercially available retrosynthetic and forward synthesis planning programs are increasingly prevalent. Will the computer ever compete with human retrosynthetic design and the art of organic synthesis?

Keywords: retrosynthesis, biomimetic synthesis, natural products, computer assisted synthesis, total synthesis, computer assisted retrosynthesis, artificial intelligence, antithetic analysis.

Received 21 December 2020, accepted 25 March 2021, published online 12 May 2021

Introduction

Retrosynthesis (or retrosynthetic analysis or antithetic analysis) is defined as ‘a problem-solving technique for transforming the structure of a synthetic target molecule to a sequence of progressively simpler structures along a pathway which ultimately leads to simple or commercially available starting materials for a chemical synthesis’.[^1] This definition was formulated and underpinned by extensive research endeavours in heuristic chemical synthesis performed by the research group led by Elias James Corey, who won the 1990 Nobel Prize in Chemistry ‘for his development of the theory and methodology of organic synthesis’.[^2] It should be recognised, however, that many talented and famous scientists performing organic synthesis were already utilising some form of innate retrosynthesis before the term was coined. From an educational perspective, Warren masterfully described synthetic planning using the disconnection approach[^3] which is an intuitive method that likely embodied early thinking. Overall, however, it is hard to ascertain whether the very early methods preceding Corey’s antithetic analysis focussed on bond breaking (disconnection), or more on literature-inspired bond formation (i.e. literature knowledge of chemical reaction methodology), or both[^4] but in

Craig M. Williams (C.M.W.) was born in Adelaide, Australia. He received his B.Sc. (Hons) degree in chemistry in 1994 from Flinders University. In 1997, he was awarded his Ph.D. in organic chemistry from the same institution under the supervision of Professor Rolf H. Prager. He undertook post-doctoral studies as an Alexander von Humboldt Fellow working with Professor Armin de Meijere at the Georg-August-Universität, Göttingen, Germany, from 1997 to 1999. In early 1999, he accepted a second post-doctoral fellowship at the Australian National University with Professor Lewis N. Mander. Professor Williams has held an academic position at the University of Queensland since 2000, and during this time has won a number of awards including a Thieme Chemistry Journals Award in 2007, an Australian Research Council Future Fellowship award in 2011, and the Award for Outstanding Contribution to Research (SCMB, UQ, 2019). The Williams research group explores numerous interests within the discipline of organic chemistry (e.g. medicinal chemistry, fundamental molecules, natural product isolation, microelectronics, drug and agrichemical development, impact sensitive molecules) enabled by organic synthesis refined through the construction of biologically active complex natural products (diterpenes, polyketides, alkaloids), and designs synthetic methodology to assist in this endeavour (synthetic transformations and reagents). Professor Williams especially enjoys teaching whole molecule retrosynthesis to undergraduate and post-graduate students.

Madeleine A. Dallaston (M.A.D.) was born in Leicester, England, and grew up in Brisbane, Australia. In 2016, she received her B.Sc. from Griffith University, earning the RACI Queensland Branch Prize for that year. In 2017, she undertook a B.Sc. (Hons) at the University of Queensland in collaboration with Defense Science and Technology, working on energetic materials for countermeasures development. She then commenced her Ph.D. in organic chemistry at the University of Queensland under the supervision of Professor Williams and G. Paul Savage in 2018, focusing on the synthesis of novel heterocycles for application as bioisosteres.
modern times, the situation is rapidly changing. For example, reviews covering synthetic achievements by subject (e.g. drug synthesis\cite{10} and natural product synthesis\cite{11-13}) are generating highly refined retrosynthetic approaches, as are reaction popularity assessments (e.g. medicinal chemistry\cite{14}) and modern methodology-focussed retrosynthetic planning (e.g. C–H activation\cite{15} and photochemistry\cite{16}). The most evolutionary change currently being confronted by the synthetic community, however, is the realisation that computer-aided retrosynthesis,\cite{17} and the accompanying artificial intelligence (AI) that proposes forward syntheses,\cite{18} may very soon become commonly available to educators and practitioners. It is this verve of revolutionary retrosynthesis that has inspired the present review, and as such the authors will attempt to shed light on a subject that has been in motion since the late 1960s, but not in the past widely adopted. To assist in achieving the goal of this perspective, classical and historical methods and practices of retrosynthesis will be briefly overviewed (i.e. using examples and experiences well known to the authors), along with an assessment of currently available retrosynthetic programs. Combined, this assessment will hopefully generate philosophical debate around the pros and cons of AI introduction and contemplation on the future of organic synthesis design.

**Retrosynthetic Methods and Practices**

There are essentially two forms of retrosynthesis, and these are highly integrated. The first is functional group interconversion (FGI) or functional group exchanges, and the second is whole-molecule (or target-orientated) retrosynthesis.

FGI plays a role in both, whereby the first mode is the ability to navigate through a range of functional groups in a linear fashion to arrive at an alternative functional group (this also includes protecting groups). An example to highlight this situation is the stereoselective synthesis of squalene (1) in 1959, which was performed by husband-and-wife team Drs John and Rita Cornforth and, presumably, their joint student K. Mathew.\cite{19} At this time, the term retrosynthesis did not exist; however, conceptual thinking around construction of the target was discussed. Interestingly, the approach presented focussed on the number of carbon subunits required to achieve the 30-carbon framework (i.e. in terms of numerical values entered into an equation\cite{20}), and that a reported olefin synthesis method would be used. This is a clear example of deploying a literature-inspired bond formation strategy, and to achieve this goal a large number of FGIs were undertaken (Scheme 1). To exemplify this point, homogeranyl lithium (2) was required to add to dichloroketone (3) in the final stages of the synthesis, and this was derived from homogeranyl bromide (4). Bromide 4 was in turn obtained from alcohol 5, which was prepared via reduction of homogeranic acid (6) (Scheme 1). Overall, five FGIs were deployed to transform 7 into 2, but the homogeranyl carbon framework went unchanged, i.e. only the end functional group is different.

Although arguably FGI can also be considered target-orientated retrosynthesis, the latter term is really reserved for targets that require as part of their retrosynthetic planning a combination of carbon framework manipulations and FGIs. Examples that fit this category include small, large, simple, and complex molecular targets, and thus are only limited by academic imagination and industrial desires. Some (of the many) areas that have benefitted from these endeavours include flavours, fragrances, pharmaceuticals, agrichemicals, materials, natural products, and environmental and fundamental molecules.

With all these drivers and inspiration to make such molecules, members of the synthetic community have resorted to numerous retrosynthetic approaches to achieve successful completion of their proposed targets (see below).\cite{21} Interestingly, the majority of these approaches were developed through the process of performing natural product total synthesis, mainly because natural products were the first on the scene during the emergence of the discipline of organic chemistry.\cite{22} Following isolation and subsequent elucidation of a natural product, there was an immediate structural challenge, often augmented by intriguing biology activity, that provided a substantial synthetic challenge. This is not to say that non-natural targets have not contributed to the art, but the fact that Mother Nature has for eternity been generating a vast library of molecules has presented a ‘spoiled for choice’ situation for synthetic enthusiasts and educationalists.

The various retrosynthetic approach concepts are outlined below using select examples well known to the authors.

**Biomimetic Synthesis**

Biomimetic, or biogenetically inspired, synthesis continues to be a regularly deployed method to construct complex natural products. In essence, this area evolved from natural product isolation and synthetic chemists postulating how organisms might produce secondary metabolites through biosynthesis and/or accompanying metabolic processes. The retrosynthetic planning process therefore adopts synthetic transformations, and/or key intermediates, that mimic a postulated biological process (e.g. enzymatic oxidation) undertaken by the organism.\cite{23} It should be noted that domino,\cite{24} tandem,\cite{25} and cascade\cite{26} reactions are prevalent, but importantly inspired by nature.
An example that embraces both biomimetic and domino-based retrosynthetic planning is the total synthesis of diterpenes in the vibsane series of natural products.\[22] Interestingly, the isolation of this diterpene class was inspired by reports that the shrub *Viburnum awabuki* (Caprifoliaceae) was used as a fish poison by Japanese villagers (i.e. piscicidal activity).\[23] Since that time, other viburnum species (i.e. *V. odoratissimum* and *V. suspensum*) have yielded a wide variety of different structure types all proposed to be biosynthetically related, and many displaying pronounced neurite outgrowth activity.\[24] Of relevance to this section are the neovibsanins, which Fukuyama\[25] proposed were biosynthetically derived from vibsanin B (8), via key intermediate 9 (Scheme 2). Intermediate 9 was then proposed to undergo a series of cyclisations and rearrangements, the first being tetrahydrofuran formation to give 10, which is poised to undergo further cyclisation to neovibsanin A (11, via path A), or provide carboxylation 12 via path B. Carboxylation 12 can then be attacked either intermolecularly by water to give neovibsanin H (13, via path C), or intramolecularly by the pendant homoisoprenyl alkene to give the bicyclic cation 14 (via path D). Cation 14 is then open to either nucleophilic attack by water, resulting in the formation of neovibsanin F (15, via path E), or elimination via path F to afford neovibsanin G (16) (Scheme 2).

Digestion of this extensive biosynthetic proposal suggested that if a candidate similar to intermediate 9 was entertained during retrosynthetic analysis, then via multiple acid-catalysed steps, it might be possible to access the entire neovibsanin class of natural products. Equally, it could be considered that total synthesis of vibsanin B (8) might be a viable starting point, but approaches to construct the 11-membered ring are daunting in comparison with developing strategies for decorating a six-membered ring of the type seen in 9. With these considerations front of mind, the following retrosynthesis was formulated (Scheme 3). It was opted to install the enol ester side chain last, giving rise to 17, which although carrying the requisite retrosynthetic was better equated as 18 to hold oxygenation at the doubly allylic position. Access to 18 would likely arise from a combination of alkylation and addition to cyclohexone 19, with the overall sequence starting from methylcyclohexene 20 (Scheme 3).

Although commercially available at the time, it was more convenient to synthesise methylcyclohexene 20 in-house on large scale from ethyl acrylate (21).\[26] Addition of homoprenyl magnesium bromide (22) to 20 could be achieved using standard cuprate conditions on reasonable scale in the racemic series (i.e. 23),\[27] but also asymmetrically using the chiral ligand 24, affording the 1R-isomer 25 (Scheme 4). Dehydrogenation using IBX-NMO (26; IBX = o-iodoxybenzoic acid, NMO = *N*-methylmorpholine *N*-oxide)\[28] gave 27, setting the stage for introduction of the important methylene hydroxy group (i.e. 28) via a Morita–Baylis–Hillman reaction. Ultimately, this was achieved by developing a green chemistry method using the common surfactant sodium dodecyl sulfate (SDS) in water, which could be performed at scale.\[29] Silyl protection of the pendant hydroxyl group enabled enolate-mediated alkylation of ketone 28 with ethyl iodoacetate, which gave a mixture of diastereomers 29 and 30. The major diastereomer 30 was then reacted with the acyl anion equivalent,\[30] lithium dithiazide (31)\[31] to give lactone 32. Unmasking the aldehyde was performed using mercury acetate, and the final methyl enone was installed using ylid 33, to afford 34 (Scheme 4).\[32]

Having access to an advanced intermediate (i.e. 34) that most closely matched that proposed biosynthetically (i.e. intermediate 9) enabled the next phase of the biomimetic synthesis to be investigated. An important aspect in moving forward in this regard was the fact that many neovibsanin family members contained methoxy groups, presumably from methylation of the corresponding hydroxyl group via plant methyl transferases.\[33] Therefore, instead of attempting to emulate an aqueous
environment or conditions, methanol could be used as both the solvent and nucleophile. Accordingly, 34 was dissolved in methanol and treated with concentrated sulfuric acid at room temperature overnight. The strongly acidic conditions promoted several cascading reactions, the first of which was deprotection of the t-butyldimethylsilyl (TBS) group to unmask the primary alcohol seen in 35. Oxy-Michael addition formed the tetrahydrofuran ring observed in 36, which underwent epimerisation (i.e. 37), before solvolysis and attack by the solvent (path A) to give carboxylic acid 38. Under the reaction conditions, the acid undergoes Fischer esterification to afford the methyl ester (39).

Global reduction followed by global oxidation provided aldehyde 40, which was immediately subjected to conditions developed by Davies to install the elaborate enol ester sidechain. This final manoeuvre delivered (−)-2-O-methylneovibsanin H (41) in 12 synthetic operations, whereby an acid-catalysed, one-pot, four-step sequence was key to achieving a concise total synthesis based on a biosynthetic proposal (Scheme 5).

Based on these biogenetically inspired results, a similar domino reaction, albeit combining Brønsted and Lewis acid promotion, successfully led to the asymmetric total synthesis of (−)-neovibsanin G (42) starting from 25 (Scheme 5). However, it is worthwhile to note that even the most well-considered retrosynthetic planning can come unstuck, and this occurred when attempting the total synthesis of neovibsanins A and B. It was thought that the proposed advanced intermediate 34 was suitably flexible to follow the proposed biosynthetic route, but one fault resided within 34, which amounted to a single stereocentre that was opposite to that proposed for biosynthetic intermediate 9 (Scheme 5, red highlight). On treatment of 34 with concentrated sulfuric acid, although this time at 4°C overnight (room temperature (r.t.) resulted in decomposition), the resulting alcohol (43) did not undergo solvolysis, but instead acted as a nucleophile, giving rise to ester 44. Ketalisation then afforded the 4,5-bis-epi-esters 45 and 46, which were carried through via the respective aldehydes (47 and 48) to the non-natural isomers 4,5-bis-epi-neovibsanin A (49) and B (50) (Scheme 5). Of note is that the biological activity of 49 and 50 was very similar to that of neovibsanins A and B. Furthermore, a related outcome was experienced when spirovibsanin A was targeted, in that (−)-5,14-bis-epi-spirovibsanin A (51) was obtained (Scheme 5).

Although retrosynthetic planning based on biosynthesis postulates rarely utilises enzymatic methods for performing forward synthesis...
synthesis, such synthetic transformations mediated by enzymatic control have found wide application in all other areas of organic synthesis.\(^3\) That said, they are very substrate-dependent.\(^4\)

Lastly, from a philosophical perspective, adopting a retrosynthetic plan based on naturally occurring biosynthetic considerations could be considered plagiarism from Mother Nature. A similar argument could be mounted for AI, which presents a similar philosophical dilemma (see further below).

### Linear Synthesis

Linear (or consecutive) syntheses are the most common outcomes of retrosynthetic planning, because they are by design iterative, and thus cater for both small and large target molecules. Like all synthetic approaches to a target molecule, linear syntheses have their place; however, generally they are avoided from both an elegance perspective (e.g. ideal syntheses)\(^5\) and efficiency point of view (e.g. step economy),\(^6\) i.e. convergence approaches are preferred when possible (see below).

A landmark contribution of tactical linear retrosynthesis deployment is that of the cubane skeleton (i.e. cubane-1,4-dicarboxylic acid \((52)\) and the corresponding dimethyl ester \((53)\)) reported in 1964 by Philip Eaton, and coworker Thomas Cole.\(^7\) The sequence was later modified by the same authors for conversion into the Platonic solid cubane \((54)\), a seminal contribution to the field of strained hydrocarbons.\(^8\) Although a retrosynthetic approach was not discussed, as was the convention in that era, the clear inspiration for the route taken hinged on the fact that cyclopentadienone \((55)\) underwent regioselective \([4+2]\)-dimerisation.\(^9\) However, at that time the endo stereoselective nature of these reactions was not defined, but was thought to be endo as drawn by Hafner and Goliasch in 1961.\(^10\) Interestingly, Eaton independently confirmed the endo assignment via the known hydrocarbon \((56)\), but the route was not verified with either experimental procedures or cited.\(^11\) Apparently, this was achieved by reduction of dicyclopentadien-1,8-dione \((57)\) to the diketone \((58)\), followed by conversion to the bisthioketal \((59)\), and final reduction via Raney nickel to endo-tetrahydrodicyclopentadiene \((56)\). These key pieces of knowledge regarding the endo-stereochernistry, combined with reports that chlorinated cyclopentadienones also undergo \([4+2]\)-dimerisation,\(^12\) enabled Eaton to consider the following approach. Gaining access to a suitably substituted...
cyclopentadienone (60) would provide a dihalogenated endo-dicyclopentadien-1,8-dione (61). Although, given the photochemically sensitive bridgehead bromide seen in 61, a daring regioselective photolysis was postulated to drive a [2 + 2]-cycloaddition to form the precursor cage (62), a subsequent double ring contraction was then envisaged to give the 1,4-substituted cubane framework (i.e. 63), enabled by the Favorskii rearrangement (Scheme 6).

In the forward direction, the synthesis started by treating cyclopentenone (64) with N-bromosuccinimide, which gave the monobrominated material 65 en route to the tribromide 66. Exposing 66 to ethylamine in diethyl ether at low temperature generated 2-bromo-cyclopentadienone (67), which underwent immediate dimerisation to afford the diketone dimer (68) in 40 % yield over three steps. Subsequent UV irradiation effected the [2 + 2]-cycloaddition to give the cage precursor 69, which was treated with potassium hydroxide to mediate the double Favorskii, and reveal cubane-1,4-dicarboxylic acid (52). The corresponding dimethyl ester (53) could be readily obtained through classical Fischer esterification conditions (i.e. methanol and catalytic amounts of sulfuric acid). Although there are now numerous methods to access cubane itself (i.e. 54), a more expedient route consists of double decarboxylation of the diacid (52)[46] using modified Barton conditions that use chloroform as the H atom donor (Scheme 7).[47]

Given the original questions around endo-selectivity concerning the cyclopentadienone dimerisation, Eaton justified this stereochemical configuration on the grounds that ‘interactions of like dipoles should be minimized in the geometry of the transition state [i.e. 70] for Diels–Alder dimerization’ (Scheme 7).[41]

Since Eaton’s reports, there has been a great deal of activity in terms of both broadening the understanding of and developing synthetic methods for the cubane system.[48] A large-scale industrial process was developed for the synthesis of dimethyl cubane-1,4-dicarboxylate (53),[49] which was found not to be impact-sensitive, unlike the corresponding diacid (52).[50] Furthermore, cubane has more recently become known as a phenyl ring isostere, and has been deployed as a bioisostere in medicinal and agricultural chemistry.[51]

Convergent Synthesis

Convergent syntheses are based on a branched retrosynthetic pathway such that the individual branches lead to smaller fragments (synthons) that are more synthetically tractable, and that can be rapidly assembled in a convergent manner. This approach saves considerable synthetic steps (and manipulations) compared with other methods, in particular linear sequences (see above).[52]

The Australian rainforest polyketide EBC-23 (71), isolated from Cinnamomum laubattii (family Lauraceae) in northern Queensland, is an example of a three-component convergent
The inspiration to undertake this collaborative industry project included encouraging biological activity (i.e. growth inhibitor activity of the androgen-independent prostate tumour cell line DU145), and the necessity to confirm both the proposed relative and absolute stereocentric assignment. Ancillary drivers included undertaking a total synthesis campaign under project management milestone conditions, and being mindful that the synthesis should be designed such that it was process chemistry transferrable for potential large-scale production.

The retrosynthetic approach for EBC-23 started by disconnection of the EE-spiroketone unit, because it was plausible that this was occurring biosynthetically (i.e. selective oxidation of the polyol chain to a single keto moiety facilitating thermodynamic cyclisation via the anomeric effect), and it was also highly conceivable synthetically. The corresponding ketone (72) initiated the first point of convergence, which was facilitated by considering acyl anion equivalent chemistry (i.e. (73), which provided a left- (74) and right- (75) hand fragment. The left-hand fragment (74) could be further dissected using the same acyl anion approach to give epoxides 76 and 77, whereas the right-hand fragment (75) was perfectly poised for a ring-closing metathesis-styled disconnection leading to 78, which could be accessed from a functionalised syn-diol of the type seen in 79. The convergent approach also provided several built-in fallback strategies associated with the left- (74) and right- (75) hand fragments. For example, the left could be accessed via traditional asymmetric aldol methods (e.g. sequentially coupling 80 to aldehydes 81 and 82), whereas the right hand could materialise from the chiral pool (e.g. deoxy sugar 83) (Scheme 8).

In the forward direction, the left-hand fragment 74 was tackled first, as it was a chance to explore and select the appropriate acyl anion chemistry for the final coupling of both halves. In this regard, Tietze-Smith linchpin methodology was chosen, which is founded on the Brook rearrangement of silylated dithiane anions that facilitate an anion relay reaction through to the target. The final sequence proceeded as follows. Addition of the acryloyl group to 74, which was achieved in two steps (via 85) using the method reported by Lepoivitin, which provided the desired material in high enantiomeric excess (ee) and in large quantities, starting from commercial epi-chlorohydrin (86). TMS-dithiane (87) was prepared by treating dithiane (88) with n-butyl lithium (nBuLi) and quenching the resulting anion with trimethylsilyl chloride (TMSCl). Subsequent treatment of TMS-dithiane (87) with nBuLi provided the TMS-dithiane anion (89), which on reaction with epoxide 84, followed by addition of epi-chlorohydrin (86), afforded the desired disubstituted dithiane (90). Interestingly, unmasking the acyl anion equivalent (i.e. dithiane moiety seen in 90) by oxidative removal with mercury perchlorate not only provided the ketone function, but also mediated TMS deprotection, giving the hydroxyketone 91. The unstable hydroxyketone (91) was immediately subjected to a Prasad syn-selective reduction using diethylmethoxymethyl chloride and sodium borohydride and the resulting diol protected as an acetone (i.e. 92). All transformations proceeded in high yields with high diastereomeric (de) selection and enantiomeric excess (Scheme 9).

Completing the synthesis required considerable trouble-shooting in terms of finding ring-closing metathesis (RCM) conditions, and a workable sequence of deprotection and cyclisation through to the target. The final sequence proceeded as follows. Addition of the acryloyl group to 99 proceeded smoothly to afford 100, but the ensuing RCM reaction required both harsh conditions and a catalyst that could maintain sustained activity. In this regard, the second-generation Hoveyda–Grubbs (101) catalyst delivered in spades, in what is believed to be overcoming severe steric
encumbrance and likely counterproductive co-ordination to the dithiane sulfur atoms. The desired product (102) was then stripped of silyl protection using hydrogen fluoride (HF) to give the polyl (103), setting the stage for final spiroketalsation. Cerium ammonium nitrate (CAN) was found to be critical for this final step in that it not only facilitated the unmasking of the keto functionality, but also metal-templated the cyclisation, as observed previously by Evans. This last manoeuvre delivered EBC-23 (71) concisely (i.e. two degrees of convergence of nine total linear steps in 8% overall yield), which enabled confirmation of structure and further biological studies (Scheme 10).

Chemical Degradation

Chemical degradation covers an aspect of retrosynthetic thinking that was only really made possible owing to the then-emerging field of natural product isolation and elucidation (i.e. before modern methods involving NMR spectroscopy). The main approach in the early era was to perform a multitude of chemical degradation studies on the unknown isolate looking for functional group and skeletal clues (e.g. via melting point, boiling point, density, refractive indices, optical rotation) to assist in solving the structure. Elucidation methodology of this type took a great leap forward, however, when in 1948 Woodward combined the evolving mechanistic understanding of chemical reactions with natural product elucidation (i.e. understanding degradation pathways from a mechanistic standpoint). By this time, of course, chemical reactions were becoming better understood through the adoption of reaction arrows; however, although having been introduced 25 years prior, these concepts were slow to be adopted in the natural product isolation field. To highlight the point of mechanistic insight, Woodward chose the santonin (104) and santonic acid (105) story as an exemplar case. For 100 years, many organic chemists had been trying to solve the structure of santonin (104), which had been isolated from wormseed (Artemisia santonica) in 1830. Although the flat structure of santonin (104) was solved at the end of this 100-year period, one of the unexplained products of degradation was santonic acid (105). Interestingly, this Holy Grail of the time was obtained with vigorous boiling of 104 with concentrated alkali. Woodward set about demonstrating that by evaluating the various base-induced degradation products and intermediates from a mechanistic viewpoint, a proposed structure could be rationally determined, and then verified experimentally. Woodward focussed on the functionality that would be likely susceptible to treatment with hydroxide, which first involved ring-opening of the lactone in santonin (104) to give alcohol 106. Enolisation of the conjugated γ-enone in 106, via 107, revealed the possibility of a ketone (i.e. 108) in the C-ring. The C-ring ketone could then undergo enolisation and reach over and attack the enone in the A-ring (i.e. 109). It was this critical thinking that led to the notion that santonic acid (105) was a cage bicyclic system (i.e. quite the novelty system in the day) (Scheme 11). Note it took many more decades for the absolute stereochemistry of both (–)-santonin (110) and santonic acid (111) to be solved.

Although santonic acid (105) was not a natural product in its own right, this endeavour married the emerging field of reaction mechanism understanding with the field of natural product chemistry. The outcome of this merger produced several key discipline shifts: (1) it added a new dimension to the elucidation of the lactone in santonin (104)) to give alcohol (106). Enolisation of the conjugated γ-enone in 106, via 107, revealed the possibility of a ketone (i.e. 108) in the C-ring. The C-ring ketone could then undergo enolisation and reach over and attack the enone in the A-ring (i.e. 109). It was this critical thinking that led to the notion that santonic acid (105) was a cage bicyclic system (i.e. quite the novelty system in the day) (Scheme 11). Note it took many more decades for the absolute stereochemistry of both (–)-santonin (110) and santonic acid (111) to be solved.

Although santonic acid (105) was not a natural product in its own right, this endeavour married the emerging field of reaction mechanism understanding with the field of natural product chemistry. The outcome of this merger produced several key discipline shifts: (1) it added a new dimension to the elucidation of the lactone in santonin (104)) to give alcohol (106). Enolisation of the conjugated γ-enone in 106, via 107, revealed the possibility of a ketone (i.e. 108) in the C-ring. The C-ring ketone could then undergo enolisation and reach over and attack the enone in the A-ring (i.e. 109). It was this critical thinking that led to the notion that santonic acid (105) was a cage bicyclic system (i.e. quite the novelty system in the day) (Scheme 11). Note it took many more decades for the absolute stereochemistry of both (–)-santonin (110) and santonic acid (111) to be solved.

Although santonic acid (105) was not a natural product in its own right, this endeavour married the emerging field of reaction mechanism understanding with the field of natural product chemistry. The outcome of this merger produced several key discipline shifts: (1) it added a new dimension to the elucidation of the lactone in santonin (104)) to give alcohol (106). Enolisation of the conjugated γ-enone in 106, via 107, revealed the possibility of a ketone (i.e. 108) in the C-ring. The C-ring ketone could then undergo enolisation and reach over and attack the enone in the A-ring (i.e. 109). It was this critical thinking that led to the notion that santonic acid (105) was a cage bicyclic system (i.e. quite the novelty system in the day) (Scheme 11). Note it took many more decades for the absolute stereochemistry of both (–)-santonin (110) and santonic acid (111) to be solved.

Although santonic acid (105) was not a natural product in its own right, this endeavour married the emerging field of reaction mechanism understanding with the field of natural product chemistry. The outcome of this merger produced several key discipline shifts: (1) it added a new dimension to the elucidation of the lactone in santonin (104)) to give alcohol (106). Enolisation of the conjugated γ-enone in 106, via 107, revealed the possibility of a ketone (i.e. 108) in the C-ring. The C-ring ketone could then undergo enolisation and reach over and attack the enone in the A-ring (i.e. 109). It was this critical thinking that led to the notion that santonic acid (105) was a cage bicyclic system (i.e. quite the novelty system in the day) (Scheme 11). Note it took many more decades for the absolute stereochemistry of both (–)-santonin (110) and santonic acid (111) to be solved.

Although santonic acid (105) was not a natural product in its own right, this endeavour married the emerging field of reaction mechanism understanding with the field of natural product chemistry. The outcome of this merger produced several key discipline shifts: (1) it added a new dimension to the elucidation of the lactone in santonin (104)) to give alcohol (106). Enolisation of the conjugated γ-enone in 106, via 107, revealed the possibility of a ketone (i.e. 108) in the C-ring. The C-ring ketone could then undergo enolisation and reach over and attack the enone in the A-ring (i.e. 109). It was this critical thinking that led to the notion that santonic acid (105) was a cage bicyclic system (i.e. quite the novelty system in the day) (Scheme 11). Note it took many more decades for the absolute stereochemistry of both (–)-santonin (110) and santonic acid (111) to be solved.

Although santonic acid (105) was not a natural product in its own right, this endeavour married the emerging field of reaction mechanism understanding with the field of natural product chemistry. The outcome of this merger produced several key discipline shifts: (1) it added a new dimension to the elucidation of the lactone in santonin (104)) to give alcohol (106). Enolisation of the conjugated γ-enone in 106, via 107, revealed the possibility of a ketone (i.e. 108) in the C-ring. The C-ring ketone could then undergo enolisation and reach over and attack the enone in the A-ring (i.e. 109). It was this critical thinking that led to the notion that santonic acid (105) was a cage bicyclic system (i.e. quite the novelty system in the day) (Scheme 11). Note it took many more decades for the absolute stereochemistry of both (–)-santonin (110) and santonic acid (111) to be solved.

Although santonic acid (105) was not a natural product in its own right, this endeavour married the emerging field of reaction mechanism understanding with the field of natural product chemistry. The outcome of this merger produced several key discipline shifts: (1) it added a new dimension to the elucidation of the lactone in santonin (104)) to give alcohol (106). Enolisation of the conjugated γ-enone in 106, via 107, revealed the possibility of a ketone (i.e. 108) in the C-ring. The C-ring ketone could then undergo enolisation and reach over and attack the enone in the A-ring (i.e. 109). It was this critical thinking that led to the notion that santonic acid (105) was a cage bicyclic system (i.e. quite the novelty system in the day) (Scheme 11). Note it took many more decades for the absolute stereochemistry of both (–)-santonin (110) and santonic acid (111) to be solved.

Although santonic acid (105) was not a natural product in its own right, this endeavour married the emerging field of reaction mechanism understanding with the field of natural product chemistry. The outcome of this merger produced several key discipline shifts: (1) it added a new dimension to the elucidation of the lactone in santonin (104)) to give alcohol (106). Enolisation of the conjugated γ-enone in 106, via 107, revealed the possibility of a ketone (i.e. 108) in the C-ring. The C-ring ketone could then undergo enolisation and reach over and attack the enone in the A-ring (i.e. 109). It was this critical thinking that led to the notion that santonic acid (105) was a cage bicyclic system (i.e. quite the novelty system in the day) (Scheme 11). Note it took many more decades for the absolute stereochemistry of both (–)-santonin (110) and santonic acid (111) to be solved.
of natural products (and degradation products); (2) it elevated the notion that natural products could be transformed into related natural products and/or systems; (3) it assisted in laying the foundation for partial and relay syntheses (see below); and (4) it demonstrated that new chemical transformations could be incorporated into retrosynthetic thinking.

A further example to highlight these points are the Galbulimima alkaloids, which also integrated biosynthetic understanding. These unique alkaloids, which now number ~35 members in four separate classes, have so far been isolated from the bark of the tropical rainforest trees Galbulimima belgraveana and Galbulimima baccata.\cite{72–75} These evergreen species are found in northern Australia, Indonesia, and Papua New Guinea, and have been used extensively by native tribes owing to their hallucinogenic and medicinal properties.\cite{72–75} Himgaline (112) is an interesting case in point.\cite{72,76,77} It was solved using classical chemical degradation reinforced by semi-synthesis, with the final stages of the structure being proved using oxidative elimination of 112 using 70 % nitric acid to give GB13 (113). The overall conversion was represented in several simplified schemes showing the starting alcohol moiety (i.e. 114) giving the eliminated enone function (i.e. 115) and the fragmented dialkylamine (116).\cite{78,79} Furthermore, it was reported that catalytic hydrogenation of GB13
In light of this information, it was clear that any total synthesis of himgaline would be best achieved via GB13 (Scheme 12).

Mander and McLachlan were the first to report a synthesis of (±)-GB13, via a sequence that involved multiple dearmatization steps, ring contraction/cleavage/anneluation, and a Diels–Alder cycloaddition. The details of this synthesis, which showcases the utility of benzenoid synths, started with a Birch reduction alklylation protocol. The dianion of 2,5-dimethoxybenzoic acid was generated via lithium–ammonia reduction, and then quenched with 3-methoxybenzyl bromide to give the dearmatized bis-enol ether. Double protonolysis and cyclisation of p-nitrobenzenesulfonyl azide gave ketone, setting the stage for a Wolff rearrangement-mediated ring contraction. Irradiation of 122 in the presence of hexamethyldisilazane resulted in the intermediate ketene being intercepted, and on protonolysis afforded the amide function into a cyano group, and then via deprotonation and trapping with diphenylselenide, providing the option to install the bridge double bond.

Additional examples that are recommended to the interested reader include the total synthesis of bleomycin A2 based on a mass spectrometry degradative synthesis using a mixture of biosynthetic, degradation, and FGI-inspired retrosynthetic concepts. This often means that the skeleton is already (or mostly) intact and that FGI chemistry is then deployed to achieve the synthesis of the target. In some cases, the starting point (i.e. starting material) can be non-natural, but ultimately it is also derived from a nature source (e.g. via degradation). A wonderful example where the wide utility of semi-syntheses can contribute to an area of science is that of the gibberellin plant hormones.

Partial Synthesis

Partial synthesis (or semi-synthesis) relates to natural product synthesis using a mixture of biosynthetic, degradation, and FGI-inspired retrosynthetic concepts. This often means that the skeleton is already (or mostly) intact and that FGI chemistry is then deployed to achieve the synthesis of the target. In some cases, the starting point (i.e. starting material) can be non-natural, but ultimately it is also derived from a nature source (e.g. via degradation). A wonderful example where the wide utility of semi-syntheses can contribute to an area of science is that of the gibberellin plant hormones.

**Scheme 12.** Himgaline (112) structure confirmed by degradation and reconstitution from GB13 (113).
A specific example is the conversion of gibberellic acid (139) to gibberellin A19 (GA19, 140) [92] At the time, GA19 had only ever been isolated once in workable analytical quantities from bamboo shoots (i.e. 14 mg from 44 tons), and was required for biosynthetic studies [92]. First, gibberellic acid (139) was esterified with diazomethane followed by protection of the
bridgehead hydroxyl (via MOM) to give 141. Dissolving metal reduction cleaved the lactone bridge, removed the A ring oxygenation, and isomerised the double bond to afford 142, which was converted into the diazoketone (143). Cyclopropanation mediated by copper was the key synthetic manoeuvre to introduce the desired carbon bridge, and as expected, the cyclopropyl ketone (144) was obtained in high yield. A second lithium in ammonia reduction cleaved the cyclopropane bond that made up a six-membered ring, in turn leaving a five-membered ring ketone (145). Oxidative cleavage of the resulting five-membered ring was challenging, but it eventually yielded to a two-step process first involving treatment with potassium hydride (KH) and quenching with methyl iodide to afford an enol ether (146), followed by exposure to ozone. Reductive workup with dimethylsulfide (DMS) provided the desired enol ether, followed by exposure to ozone. Reductive workup with dimethylsulfide (DMS) provided the desired terpenoid carbocation (147), which was converted into the diazoketone (148) enabled by a relay sequence obtained from degradation of 152 (via 157). The ability to inject a vital advanced intermediate into the total synthesis campaign was critical to the completion of the planned route, which involved re-establishment of the furanone ring via 158 and 159 (Scheme 15).

Chemoenzymatic Synthesis

Building on statements made above concerning biomimetic syntheses, it was noted that enzymes are an important set of ‘reagents’ that can be integrated solely or partly into retrosynthetic route design. A prime example is enzyme-mediated hydroxylation of hydrocarbon scaffolds using cytochromes P450 (P450s). To highlight this point, Shen and Renata took commonly available diterpenes and converted them into high-value and rare natural products to demonstrate the power of combining chemical and enzymatic oxidation methods via a chemoenzymatic platform. Among the numerous examples, steviol (160) was first oxidised with PtmO6, which was better overproduced on expression in Escherichia coli, followed by a traditional pyridinium dichromate (PDC) oxidation to give the C7 ketone (161). Not only was the yield high over two steps, but the selectivity was remarkable.

Note: regioselective C7 oxidation of terpenoid substrates is challenging, as exemplified by attempts to construct the ABC ring system (i.e. 162) in this study.

Relay Synthesis

Relay synthesis is a term used for the conversion of one natural product into another so as to intercept and supply an advanced intermediate for a total synthesis campaign. For retrosynthetic thinking, a route that employs a relay overcomes access to advanced intermediates in meagre quantities via chemical synthesis by using partial synthesis. A very early example was provided by Robinson for the synthesis of cholesterol (150), with more recent examples reported by Shi, Tan, and Gin for the synthesis of cholesterol (151) and Deslongchamps et al. for ouabagenin (152). The latter is a particularly instructive example whereby advanced intermediate 155 was achieved via the planned Diels–Alder based retrosynthetic route (i.e. reaction of 153 with 154 to give 155), but additional material (i.e. 156) was produced, reflecting sites of retrosynthetic disconnections.

Pharmaceutically relevant examples include the semi-synthesis of both Taxol® (148) and artemether (149) (Fig. 2).[94]

![Fig. 2. Taxol® (148) and artemether (149). Atoms highlighted in green reflect sites of retrosynthetic disconnections.](image-url)

Scheme 14. Partial synthesis of gibberellin A19 (GA19, 140).
retrosynthetic plan involving advanced intermediate 164. After considerable investigation, the synthesis of 162 was achieved in 17 steps starting from 165 (Scheme 16).\(^{[102]}\)

Continuation of the chemoenzymatic synthesis involved another enzymatic oxidation, but this time using PtmO5, which is a class I P450 that requires a separate reductase partner (e.g. RhFRed) to support function and turnover. The adoption of this enzyme installed a \(\beta\)-OH at C11 with absolute stereo-specificity to give 166. Acylation of 166 and deployment of the Appel reaction gave amide 167, which underwent double hydride reduction to give 168. Subjecting 168 to selenium dioxide gave the bridge allylic alcohol (i.e. C15) that was selectively oxidised with IBX\(^{[103]}\) to afford rosthornin C (169) (Scheme 16).

**Formal Synthesis**

A formal synthesis is achieved when a practitioner performs a synthesis of a known target molecule using a different approach to that reported, but instead of completing the entire synthesis, the sequence stops at an advanced intermediate intersecting the previously disclosed synthetic route. For the purposes of retrosynthetic thinking, it is not the target itself that warrants a full disconnection evaluation, and associated forward planning, but the advanced intermediate of choice becomes the synthetic target. These indirect targets are usually embedded in the final stages of a prior reported synthetic route.

Formal syntheses are numerous and widespread. An instructive example is that of the marine natural product 7,20-diisocyanoadociane (170), which became a popular target because of considerable antimalarial activity shown by this and related systems.\(^{[104]}\) Corey was the first to complete a total synthesis of 170 in 26 steps, via the dione advanced intermediate 171, starting from glutaric anhydride (172).\(^{[105]}\) Mander was the first to establish the stereochemical relationships at C7 and C20, by completing the first formal synthesis of 170 starting from the anisole derivative 173.\(^{[106]}\) The intersecting target in this formal synthesis was the diamine (174), which Simpson and Garson had previously reported could be converted in two steps to 170 (Scheme 17).\(^{[107]}\) However, numerous groups focussed on targeting Corey’s dione (i.e. 171) as the point of establishing their formal syntheses. The first was that by Miyaoka,\(^{[108]}\) which took 29 steps from (S)-4-methyltetrahydro-2H-pyran-2-one (175), followed by Vanderwal in 21 steps from perillaldehyde (176),\(^{[109]}\) and then Thomson with a shattering 12-step synthesis of the dione (171) originating from desmethylcarvone (177).\(^{[110]}\) Vanderwal ended up finishing this entertaining fight with a knockout blow constituting a 10-step total synthesis.
Methodology-Driven Synthesis

Many target-orientated syntheses are retrosynthetically designed around a certain type of synthetic methodology. This choice can be inspired by a multitude of reasons, some of which include: (1) testing newly discovered methodology designed in-house (e.g. C–H activation); (2) long-standing experience with certain reaction types (e.g. Claisen rearrangement); and (3) trust (or even distrust) of certain synthetic methodologies (Suzuki cross-couplings). Overall, a key piece of methodology is identified as the focal point of the synthetic route design, and this in turn impacts the retrosynthetic thinking process, as the prime task then becomes strategic implementation (and designing accommodation) to reach the target.

A synthetic campaign that encapsulated aspects of methodology-driven synthesis concepts was that of vibsanin (Scheme 16). (a) Chemoenzymatic synthesis of rosthornin C (169) from steviol (160); and (b) retrosynthesis of gedunin (163) with construction of the advanced intermediate (162). aKG = a-ketoglutaric acid; NADP = nicotinamide adenine dinucleotide phosphate; Opt13 = thermostable phosphite dehydrogenase variant; Imid = imidazole; TBHP = tert-butyl hydrogenperoxide.

Scheme 16. (a) Chemoenzymatic synthesis of rosthornin C (169) from steviol (160); and (b) retrosynthesis of gedunin (163) with construction of the advanced intermediate (162). aKG = a-ketoglutaric acid; NADP = nicotinamide adenine dinucleotide phosphate; Opt13 = thermostable phosphite dehydrogenase variant; Imid = imidazole; TBHP = tert-butyl hydrogenperoxide.

Scheme 17. Formal and total syntheses of 7,20-diisocyanoadociane (170).

(111) C. M. Williams and M. A. Dallaston
E (179)\(^{[113]}\) and 5-epi-vibsanin (180)\(^{[114]}\) which are two rarely occurring complex cage-containing diterpenes isolated from *Viburnum awabuki* in Japan.\(^{[115]}\) Initially, the synthetic methods of choice combined an acid-catalysed cyclisation of 181, followed by ring expansion of 182, to give an advanced intermediate containing features seen in 183.\(^{[127]}\) However, although this work provided substantial experience for the neovibsanins (see Schemes 3 and 4), it was not possible to convert the advanced intermediate into either of the targets.\(^{[116]}\) Therefore, a more modern methodology was employed. However, it is always a concern when adopting new methods as many often fail when applied to complex systems. In this case, however, the Davies group had shown that donor–acceptor diazo functions (e.g. 184) undergo rhodium-catalysed [4+3]-cycloaddition with dienes (e.g. 185) (a cyclopropanation Cope rearrangement sequence) to give seven-membered rings (e.g. 186). In addition, they had applied this protocol to the total synthesis of (±)-5,10-bis-epi-vibsanin E (187), which was a non-natural product.\(^{[117]}\) Clearly, combining the [4+3] and [4+2]-cycloaddition methodology was very powerful, but the end game was compromised. This opened an opportunity for the Williams and Davies groups to collaborate and pool experiences to complete the synthesis, by adopting the Davies protocol to access an advanced intermediate of type 183, and then deploying Williams end-game know-how (Scheme 18).\(^{[118]}\)

Reducing this approach to practice, the initial seven-membered ring (i.e. 188) was obtained in 65% yield using the chiral catalyst known as PTAD (189), which required deoxygenation using in part Comin’s reagent (190) to afford a suitably substituted ring (i.e. 191) poised for the [4+2] cycloaddition. The Diels–Alder reaction proceeded as planned and the resulting tricyclic ring system (192) was converted to the key advanced intermediate 193 in three steps. Conjugate addition using the lithiated MOM derivative took place both in high yield and stereoselectively. Introducing TMSCI was critical to facilitating the reaction and providing the desired material (194). Transmetallation and quenching with allyl bromide gave the allyl ether (195), and subsequent Claisen rearrangement afforded a mixture of syn- (196) and anti-isomers (197). The syn-isomer (196) was taken forward through a sequence of deprotection, alcohol oxidation, and Wacker oxidation to unmask the penultimate tricarbonyl 198. The last, and perhaps hopeful, step was heavily reliant on in-house-optimised Anders–Wittig chemistry (Anders–Gaßner reaction).\(^{[119]}\) which delivered the target (i.e. 180 as the corresponding enantiomer 200), albeit in a low yield of 26% using the Wittig reagent 199 (Scheme 19).\(^{[114]}\)

Although the Claisen rearrangement gave a mixture of syn- and anti-isomers, it was possible to convert the anti-isomer into vibsanin E (179) (Fig. 3), although this was only undertaken in the racemic series.\(^{[113]}\) However, although a well-devised synthetic route was in hand for these systems, it was not guaranteed that other members of the vibsan family could be conquered, and a spectacular failure was endured with attempts to synthesise 3-hydroxyvibsanin E (201). The synthetic route above that was arduously devised for vibsanin E and 5-epi-vibsanin E came completely unstuck in the penultimate step in the campaign to 3-hydroxyvibsanin E, when t-butyldimethylsilyl deprotection could not be achieved, giving 3-OTBS-vibsanin E (202) (Fig. 3).\(^{[120]}\)

**Collective Total Synthesis**

The term collective total synthesis is a fairly new concept that was first put forward by MacMillan in 2011, when his group achieved the synthesis of six alkaloids from the same advanced intermediate.\(^{[121]}\) MacMillan argued that although syntheses targeting an advanced core structure applicable to the synthesis of closely related natural products within a family had been previously reported, it was much less common to find examples where the preparation of an advanced intermediate with appropriate functionality was amenable to the construction of structurally diverse natural products in different families, the latter constituting the definition of the term collective total synthesis, which has analogies with a biosynthetic pathway providing collections of metabolites.\(^{[121]}\) Since that time, further examples using different natural product classes have been added,\(^{[122]}\) but the term is somewhat contentious, because there

![Scheme 18. Retrosynthetic analysis of vibsanin E (179) and certain isomers.](image-url)
is an element of false pretences (i.e. the deck can be stacked by making a judicious choice of targets to match the key advanced intermediate). Alternatively, a collective total synthesis can only be achieved if enough natural product elucidation has been performed such that there is sufficient choice to undertake such an endeavour. Overall, however, the concept has substantial merit in that it aims to make one advanced intermediate multi-task for multiple target syntheses, which has several important
advantages. For example, providing a wide array of targets for biological testing saves time and effort designing individual synthetic routes, and is green chemistry-orientated (i.e. limiting waste streams).

The tetraterpene sesquiterpenes (limonoids) are such a natural product class that fit the collective total synthesis criteria, in that a huge amount of structure elucidation has been undertaken on many different species within the meliaceae and rutaceae families.\[^{[123]}\] A case of serendipitous collective total synthesis is that of the bicyclogenolides (e.g. mexicanolide (203)), whereby the retrosynthetic analysis aimed to access azedaralide (204) as the key advanced intermediate to achieve a total synthesis of mexicanolide (203), via a pre-established proposed biosynthetic route (i.e. via 205). Entry to azedaralide (204) was envisioned to be directly achievable from pyroangolensolide (206), a known degradation product that had been synthesised several times previously, and could be obtained starting from 2,6-dimethyl-cyclohexenone (207) (Scheme 20). In this case not only were two limonoid natural products synthesised (i.e. 203 and 204), but an additional three were also achieved (i.e. 208–210).

In the forward direction, the asymmetric synthesis of (+)-azedaralide (216) followed the racemic synthetic route (i.e. 204).\[^{[38]}\] First, cyclohexenone 211 was subjected to in-house optimised Morita–Baylis–Hillman conditions,\[^{[29]}\] which afforded the extended ketone (212) after TBS protection. Methylation (i.e. 213), followed by a (−)-disopinocampheylcloborane- (see 215 for (+)-DIP-Cl isomer; DIP = disopinocamphelyborane) mediated asymmetric diastereoselective aldol, provided the furalcohol (214). The alcohol function was then acylated to facilitate cyclisation, followed by TBS deprotection, to give the first natural product as the advanced intermediate (i.e. (+)-azedaralide (216)). A rarely utilised ketal-Claisen rearrangement was used to attach the A and B ring components of the targets to azedaralide. For this retrosynthetic manoeuvre to even be considered, a suitably functionalised cyclohexenene ring, matching the features seen in 217, was required. This goal was achieved in both racemic and asymmetric form using the following sequence, the only difference being that in the racemic procedure, a Mukiyama aldol was deployed, whereas the enantioselective route employed a boron-mediated aldol using (+)-DIP-Cl (215). In brief, treating aldehyde 218 with the boron enolate of 2-butanone, followed by potassium hydride-induced cyclisation gave the desired cyclohexanone (219), which was easily converted into the required enol ether (217) poised for condensation and rearrangement. Heating a mixture of 216 and 217 in the presence of anhydrous toluene acid afforded, via the extended enol ether 220, the second natural product, cipadonoi B (208), as the major product.\[^{[124]}\] Epoxidation of cipadonoi B to introduce the additional ring oxygenation proceeded smoothly. Interestingly, it was not possible to reduce the epoxide to access the proposed intermediate 205 (Scheme 20) using traditional reagents. However, treatment of the epoxide (221) with aluminium amalgam not only regioselectively reduced the epoxide ring, but the resulting enolate subsequently underwent a 1,6-conjugate addition to afford the third natural product, proceranolide (209). This fortuitous result opened the opportunity to synthesise the fourth natural product, khayasin (210), using a peptide coupling protocol to install the hindered isopropyl ester. In addition, the original target and the fifth natural product, mexicanolide (203), could also be synthesised from proceranolide (209) using the Jones oxidation protocol (Scheme 21).\[^{[125]}\]

Chemical Database-Aided Synthesis

Several searchable chemical databases are available to the practitioner. Reaxys and SciFinder are the main staple in this regard. Key features include the capability to find specific compounds and substructures, general and structure specific reactions, physical properties (experimental and calculated), and spectroscopic data (experimentally determined and calculated).

In the context of database searching as applied to retrosynthesis, D’Angelo and Smith put forward the notion that retrosynthesis should be used to initially find a product from
undertaking a disconnection, which then enables a computer search to generate the remainder of the retrosynthesis. They termed this process as ‘a hybrid retrosynthesis approach’, because the concept utilises the basics of retrosynthetic analysis to identify a working intermediate that is subsequently searched for in a database using a computer.[126]

Computer-Aided Synthesis and Artificial Intelligence

The use of computers in chemistry has been around nearly as long as personal computers themselves,[127] and their possible application to planning synthetic routes has been realised for nearly as long.[128] There have been several comprehensive reviews on the topic,[13c,129] and so here we will only cover the topic briefly and highlight a few key programs and ideas. A selection of past and present programs is listed in Table 1, which by no means covers every program in the literature.

One of the first challenges to overcome in the realm of computers in chemistry was to develop a method to code chemical structures and atom connectivity in a way that was machine-readable. The most commonly used method is molecular graph theory, in which the molecule is represented by nodes and edges. Feature matrices are then used to encode various properties of the atoms (i.e. nodes) and relationships or bonds between them (i.e. edges), meaning that a 2D graph can encode 3D information such as stereochemistry.[130] Graph theory has been particularly useful in describing, and exploring, the small-molecule chemical universe.[131]

At its heart, retrosynthetic analysis can be seen as a pattern recognition exercise, an activity that computers are exceptionally good at. A set (albeit an ever-expanding set) of fairly simple rules can be used to solve even incredibly complex problems. In this regard, the parallels between retrosynthetic analysis and chess have been discussed before.[129a] In 1997, a computer called Deep Blue famously beat the world chess champion in a six-game match after making a move that has variously been attributed to a bug in the software and to superior...
that even in their initial publication on SYNCHEM, the authors note they have developed a ‘disaffection’ for its successor, SYNCHEM2. For instance, early publications by the same authors, and the sheer amount of time that developing these programs takes, that they were under development for some time before.

| Name                  | Type                  | Year | Notes                                                                 | External database                                                                 | Readily available |
|-----------------------|-----------------------|------|----------------------------------------------------------------------|-----------------------------------------------------------------------------------|-------------------|
| OCSS/LHASA[124]        | Retrosynthetic        | 1969 | Human-coded rules                                                    | None                                                                               | No                |
| SYNCHEM[130]           | Retrosynthetic        | 1977 | Knowledge based, heuristic search, non-interactive                   | Aldrich Chemical Co. and common compounds                                         | No                |
| SECS[136]              | Forward prediction    | 1978 | Logic-centred, knowledge based, heuristic search                     | None                                                                               | No                |
| EROS[138]              | Forward prediction    | 1978 | Formal reaction generators that regard reactions as bond- and electron-shifting processes | None                                                                               | No                |
| CAMEO[139]             | Forward prediction    | 1980 | Human-programmed mechanistic rules                                   | None                                                                               | No                |
| CHIRON[140]            | Starting material     | 1984 | Analyses input for stereochemical information and searches a database of starting materials for matching compounds | Manually compiled compound catalogue with well-defined stereochemical features     | No                |
| SSI[141]               | Starting material     | 1984 | Hierarchical search within a starting material library              | A portion of the Aldrich catalogue                                               | No                |
| SYNGEN[142]            | Retrosynthetic        | 1989 | Programmed rules to mechanically test all possible reactions        | REACCS, SYNLIB                                                                    | No                |
| RETROSYN[143]          | Retrosynthetic        | 1990 | Reaction classes from a database to suggest reactions                | A subset of the ORGYSYN Database of Molecular Design                             | No                |
| WODCA[144]             | Retrosynthetic        | 1990 | Logic-oriented heuristics, interfaced with EROS for forward prediction | Built in starting material database                                               | No                |
| Beppio[145]            | Forward prediction    | 1992 | Classification by similarity, based on modelling                    | None                                                                               | No                |
| SOPHIA[146]            | Forward prediction    | 1995 | Heuristic approach                                                  | SYNLIB plus an original database constructed from textbook reactions             | No                |
| SEXAM[147]             | Starting material     | 1998 | Input targets and starting materials are described in connectivity tables, considers only the carbon skeleton | None                                                                               | No                |
| KOSP[148]              | Retrosynthetic        | 1999 | Reaction knowledge base derived automatically from reaction database | AIPHOS, Available Chemicals Directory (ACD)                                       | No                |
| ROBIA[149]             | Forward prediction    | 2005 | Programmed rules and molecular modelling                             | None                                                                               | No                |
| ICSynth[150]           | Retrosynthetic        | 2005–2013, Java applet; 2014 onwards, HTML5 | Machine learning to generate sets of chemical rules from reaction databases | SPRESI, option to incorporate others                                               | Yes – commercial |
| ReactionPredictor[151] | Forward prediction    | 2012 | Machine learning approach, mechanistic                              | Hand-curated from literature and graduate level textbooks                         | Yes               |
| Chematica/ Synthia[152] | Retro and forward    | 2013 (as a poster presentation at the 245th ACS national meeting)[153] | Algorithms that draw from a database of hand-coded rules | Network of Organic Chemistry, MilliporeSigma reagent database                     | Yes – commercial |
| Reaction prediction seq_2_seq[154] | Retrosynthetic | 2017 | Machine learning, Seq2Seq                                           | USPTO                                                                              | Yes               |
| IBM RXN[155]           | Forward prediction    | 2018 | Natural language approach, Seq2Seq, trained on automatically extracted reaction data | USPTO[156]                                                                         | Yes               |
| LillyMol[157]          | Retrosynthetic        | 2018 | Machine learning, trains reaction transformation rules, requires atom mapping | USPTO, option to use in-house databases                                           | Yes               |
| Spaya AI[158]          | Retrosynthetic        | 2020 | Deep machine learning                                               | Pistachio database Mcule, EMolecules, Chemspace libraries                        | Yes – exited beta in 2021   |
| AutoSynRoute[134]      | Retrosynthetic        | 2020 | Transformer-based Seq2Seq model and Monte Carlo tree searching       | USPTO (USPTO_MIT)                                                                 | Yes               |
| AiZynthFinder[159]     | Retrosynthetic        | 2020 | Monte Carlo tree search guided by an artificial neural network policy | USPTO                                                                              | Yes               |

*Indicates the year that a publication, conference proceeding, or press release explicitly discussing the program appeared. It is clear both from previous publications by the same authors, and the sheer amount of time that developing these programs takes, that they were under development for some time before this. For instance, SYNCHEM, first disclosed in 1977, was already under development at the time that LHASA was announced in 1969. Similarly, the origins of Chematica/Synthia, originally published in 2013, are traced back to an MIT bridge club in 2001. It is perhaps telling of the speed at which the field moves that even in their initial publication on SYNCHEM the authors note they have developed a ‘disaffection’ for SYNCHEM and have abandoned its development in favour of its successor, SYNCHEM2.
intelligence. Whether this move was a software bug or deliberately executed (and indeed whether a computer can be said to ‘deliberately’ do anything) remains up for debate. Deep Blue relied on human-programmed rules and strategies, and could not learn new tactics simply by observing other players during the game. In this way, Deep Blue was analogous to the early retrosynthesis programs such as LHASA and can be described as a ‘brute-force’ player, defeating opponents through sheer knowledge and ability to quickly evaluate every known (to the computer) option. More recently, board game-playing computers have used deep neural networks and are capable of not only playing the games, but also of independently developing new strategies. These neural network computers display a game style described as more ‘human’ and ‘elegant’ than the earlier computers such as Deep Blue, and are regularly used as training tools by today’s competitors. There is also a great deal of research into using neural networks for chemical synthesis and retrosynthesis (see below). While neural networks show great promise for this type of problem, it is important to note that they are not without their issues, often being described as ‘brittle’, i.e. incredibly strong until they are faced with something unknown, carelessly entered, or deliberately edited, when they can break in spectacular fashion.

Historically, two general methods of building computer-aided retrosynthesis programs have emerged: (1) human-coded rules combined with a heuristic search; and (2) machine learning techniques that automatically generate rules from a provided reaction database. Human-coded rule programs dominated up until the mid-2000s, starting with Corey’s pioneering work in the late 1960s. OCSS (Organic Chemical Simulation of Synthesis),[12a] which evolved to become LHASA (Logic and Heuristics Applied to Synthetic Analysis),[135] is considered the first true computer-assisted retrosynthesis planning tool. It was designed to be interactive to allow the chemist to choose which branches of the synthesis tree were worth exploring based on their own knowledge, favoured reactions, and experience in the laboratory. This concept of being able to explore a synthesis tree has been common to nearly all programs developed since, although the tree has sprouted considerably more branches (Fig. 4).

Even though the methods employed by the programs can be described by umbrella terms (e.g. hand-coded rule set, machine learning, a combination of both), often the authors of a program would select an overarching approach that differentiated it from other contemporary programs. For example, SYNCHEM was unusual in that it first considered only the carbon skeleton of the target molecule and systematically broke C–C bonds to find the most efficient (convergent) route from a database of starting materials; once starting materials were found, SYNCHEM generated intermediates by adding all possible functionalities to the carbon skeletons and mechanistically testing the reactions and deleting non-viable ones, whereas other programs would start from considering the target molecule as a whole.[137] For forward prediction programs, EROS considered reactions as bond- and electron-shifting processes, and then implemented evaluation processes to select only chemically viable reactions, whereas many other prediction programs focus on functional groups and their transformations as the basis of chemical reactions.[138]

Although it has been proposed that the synthesis of longifolene (222) was Corey’s inspiration for LHASA, and likely the birthplace of the term retrosynthesis, it was not subjected to LHASA.[67] That said, it did form its ground rules.[12c] The synthetic approach also utilised as the key step the santonic acid degradation performed by Woodward (see Scheme 11 above).[67] Therefore, based on this premise, it is worthwhile...
discussing the longifolene synthesis in detail. The initial retro-synthetic manoeuvre consisted of introducing carbonyls adjacent to the gem-dimethyl group, and in place of the exocyclic double bond (i.e. 223). These key carbonyl insertions provided retro-conjugate addition synthons to be formulated, which resulted in the enone (224) being proposed. Rationalising the three-dimensional structure in two dimensions reveals 225, which could be derived from the Wieland–Miescher ketone (226), accessed via a Robinson annulation of methylcyclohexandione (227) with methylvinylketone (228) (Scheme 22).[2b]

Starting from the Wieland–Miescher ketone (229), the A-ring carbonyl was protected as a ketal (230), which enabled the enone to be chain-extended (i.e. 231), thus setting the stage for a pinacol rearrangement via diol 232. The six- to seven-membered ring enlargement was the first key hurdle in the synthetic approach, and although proceeding smoothly in the end, it did require activating conditions to promote the pinacol rearrangement and afford the desired ring system (i.e. 234), albeit a different isomer to that proposed in the retrosynthesis (i.e. 225). The second key hurdle followed directly after in that conditions were required to mediate cage bicyclic formation. This was eventually achieved using triethylamine at high temperature, and only produced material (i.e. 223) in the order of 12 % yield. Having surmounted these hurdles, the remaining part of the synthetic campaign amounted to a single skeleton embellishment, and several FGIs, for example, completion of the gem-dimethyl function (i.e. 235), followed by removal of the adjacent ketone to give 237 via reduction of a dithiane (i.e. 236). Final oxidation of alcohol 237 and subsequent methylation over two steps gave the long-awaited and seminal target longifolene (222)[160] (Scheme 23).

Interestingly, the Corey group also spawned many of the next generation of programs when those that worked on OCSS and LHASA moved on to their own independent careers. Jorgensen, who worked on the LHASA project as a graduate student in Corey’s laboratory, went on to develop CAMEO (Computer-Assisted
Mechanistic Evaluation of Organic Reactions.\textsuperscript{139,161} Instead of retrosynthetic analysis, this was a forward prediction program, which relied on human-programmed mechanistic rules to predict the outcome of a reaction. In total, 26 papers detailing the development and capabilities of CAMEO were published; the last appeared in 1995.\textsuperscript{162} It seems that ceasing work on CAMEO was a conscious though difficult decision for Jorgensen, as he discussed in his autobiography in 2015.\textsuperscript{163} Wipke also worked on OCSS as a postdoctoral researcher and went on to develop both SSY\textsuperscript{41}\textsuperscript{1} and SECS.\textsuperscript{1,13b} It is likely that many early programs were abandoned because upkeep and expansion of the rule set became too time-consuming, and the results were not promising enough to warrant the commitment.

Since the 2000s, advances in machine learning and computing power have opened the door to more complex machine learning approaches. In terms of synthetic chemistry, three learning modes are used: (1) supervised learning, where both the input (reactants, reagents, conditions) and output (products, yield) are provided to the computer in the form of a training set; (2) unsupervised learning, in which only the input data are provided to the computer; and (3) reinforcement learning, where the computer aims to find the optimal path towards the goal and is given a reward for each step.\textsuperscript{1,164} The method or mode of learning is independent of the end-goal of the program, i.e. a supervised learning mode can be utilised in a retrosynthesis program, a forward prediction program, or any other end-goal.

In machine learning, the programs are trained using databases of chemical reactions, and the quality of data (e.g. experimental details, stereochemistry, regioselective information) varies greatly between different databases. Often the information is extracted automatically and this can result in errors or missed information, which would have otherwise been identified by a human, especially when the information is being extracted from prose (written experimental procedures, patents) rather than drawn reaction schemes (although drawn reaction schemes often omit crucial information such as temperature and time).\textsuperscript{1,15b,15b}\textsuperscript{1} In a recent interview about a new collaboration, MIT researchers stated that they hoped the quality of data in the CAS (Chemical Abstracts Service) database would improve the predictive power of their machine learning algorithm that has previously only been trained on freely available datasets.\textsuperscript{1,165} While it seems intuitive that an algorithm trained on data rich in, for example, stereochemical information would be better at predicting stereochemical outcomes of reactions, there is some evidence that the dataset an algorithm is trained on may have little effect on the quality of its predictions.\textsuperscript{1,166}

A final consideration for both industry and academic chemists is that confidentiality and intellectual property is of utmost importance. Many of the browser-based free programs (e.g. IBM RXN) explicitly state that any information submitted through their website (i.e. into the retrosynthesis program) is visible to the developers. IBM RXN go on to state that by providing information, the user is granting IBM unrestricted use rights.\textsuperscript{1} Some commercial services (e.g. SpyralAI) also make the search input data visible to the developers.\textsuperscript{1,15b} Others (e.g. ICsynth) can be installed locally and therefore the data are only visible to the user.\textsuperscript{1,15b} For these subscription services, there is also the issue of what happens to the data on termination of the licence or agreement. As the capabilities and availability of computer-assisted synthesis programs expand, the status of the ownership of the input and output data will continue to evolve.

Chiral Pool

The chiral pool, which has already been mentioned above (see for example Scheme 17), is the collection of optically active starting materials that are available from natural sources, such as amino acids, carbohydrates, and terpenes. They are highly utilised in synthesis, because of their exceptional enantiopurity, and thus can be incorporated into the final structure of the target or as chiral auxiliaries. Furthermore, the chiral pool presents an interesting opportunity for the use of computers to recognise structural relationships between targets and starting materials.

As a prime example, the chiral pool approach was used by Corey for the synthesis of helminthosporal (239).\textsuperscript{1,167} Starting with S-(+)-carvone (240), a natural product readily available from caraway seeds, provided both a chiral centre and the isopropyl side chain, which inserted the correct configuration of the isopropyl group in the target. In brief, hydrogenation of 240 gave the saturated ketone (241), which was converted into a sacrificial 1,3-dicarbonyl (i.e. 242) to induce smooth albeit slow conjugate addition to methylvinylketone (228). Tricarbonyl 243 was then decarbonylated (i.e. 244), and subsequently cyclised to the bicyclo[3.3.1] system (i.e. 245) through treatment with the Lewis acid boron trifluoride. Wittig-mediated chain extension installed the masked aldehyde (246), which underwent a trans-protection (i.e. direct interconversion of protecting groups\textsuperscript{1,168}) with ethylene glycol to give the ketal (247) (note no change in oxidation state converting an enol ether to a ketal). Dihydroxylation gave the diol (248), which was oxidatively cleaved to give the dicarbonyl (249) as the key step in setting up the ring-contraction strategy towards the bicyclo[3.2.1] system, the latter being achieved via a Claisen condensation affording the cage bicyclic enal (250). The final step entailed a simple acid-mediated deprotection to unmask the remaining aldehyde in helminthosporal (239) (Scheme 24).

Such a synthesis is described as taking a chiron approach to retrosynthetic planning. Often, however, the relationship between the structure of the target and the chiral pool is not immediately obvious, and in the case of sugars can result in much of the initial functionality being removed (e.g. to access the carbon framework) to achieve synthesis of the target (e.g. leukotriene B\textsubscript{5} (251) from tri-O-acetylt-glucal (252)).\textsuperscript{1,169} In the helminthosporal example, a substantial clue to considering a chiron approach, beyond the importation of stereochemistry, was the isopropyl group and the cyclohexane moiety that S-(+)-carvone provides. Furthermore, when the hydrogenated material (i.e. 241) is redrawn in the chair configuration, the relationship between the starting material and target is more apparent, and now lends more weight to the adopted retrosynthetic design (Scheme 25).

Actually, the term chiron (chiral synthon) was coined in the early 1980s by Hanessian,\textsuperscript{1,170} who also developed a computer program specifically to search for chiral starting materials given a target molecule.\textsuperscript{1,14a} CHIRON was an interactive program that projected different perspectives of a drawn molecule, allowing the chemist to ‘see’ the compound in new ways. The program also mapped stereochemical information onto the target and searched its hand-assembled starting material database for materials with maximal overlap in functional and stereochemical features. The aim was to find starting materials that would require the least transformation to reach the product, specifically focussing on the carbon framework and rings.
Currently Available Programs

Reaction Predictor, Seq2Seq, IBM RXN for Chemistry, LillyMol, SpayaAI, AutoSynRoute, and AiZynthFinder are all currently freely available, either as browser-based applications or as code downloadable from GitHub. They all rely on machine learning methods for training, and do not include any hand-coded reaction rules. ICSynth, a commercial machine learning-based program, has seen use in the pharmaceutical industry for route design.

Arguably the most advanced computer-aided synthesis tool today is Synthia (formerly Chematica) developed by Grzybowska et al. and recently licenced by Merck. From its conceptual beginnings in 2001 at the MIT Bridge Club to being validated by in-laboratory results in 2018, the road to this autonomous synthesis planning tool was certainly a long one. The technical details of how Synthia works are covered in detail elsewhere, and so here only a few differences and similarities between this program and others that have come before will be mentioned.

First, despite the rapid advances and interest in machine learning of late, at the core of Synthia’s workings is a massive database of over 50000 human-coded reaction rules. These rules were compiled by chemists over many years and, much like the early now-abandoned programs that relied on this type of database, will need to be kept up to date for the program to continue to be of use as new chemistry is discovered. Second, Synthia has a library of algorithms that it can draw on. The algorithms can, for example, augment the reaction rules to improve regio- and stereo-selectivity predictions, allow the program to plan over several steps, or combine reactions that proceed under the same conditions (i.e. tandem reactions) into one step. These algorithms set it apart from other programs, which can only plan one step at a time.

This hybrid hand-coded-rules-and-algorithm approach means that Synthia approaches retrosynthetic problems in a seemingly more ‘human’ way than other programs, which can only apply hand-coded rules strictly in the way they were written, or may miss an important reaction that is not covered in the database it was trained on. Synthia has probably the largest chemical reaction rule knowledge set of any program thus far (both as a consequence of the effort put in to coding the rules and by virtue of there simply having been more chemistry known now than in the past).

In 2018, verification of Chematica’s capabilities was published, with the in-laboratory confirmation of several routes devised by the program (Fig. 5). The Chematica-devised
syntheses all provided improvements on the literature procedures in terms of yield and/or number of steps (253–257). In two examples (i.e. 259 and 260), Chematica was used to specifically avoid patented routes and provided comparable or significantly improved yield. One was a natural product (i.e. 258) that had not previously been synthesised. This verification clearly showed that there has been great improvement in the synthesis planning capabilities of computer programs over the last 50 years; however, the commercial aspect of the more capable programs (Synthia and ICSynth) means that their impact on the wider research community will be limited. Synthia also risks encountering the same fate as earlier hand-coded programs, abandoned owing to the time required and difficulty in keeping them up to date. Additionally, there is a great amount of research being carried out in the area of automatic data extraction, neural networks, natural language sequence to sequence processing, and machine learning, which negates the need for manual extraction of data and updating reaction rules.

**Discussion**

Having canvassed both humanistic and computer-assisted retrosynthetic methods and practices in detail, it is important that with this knowledge, questions be posed and considered to better understand a probable direction this art form will take in the future, i.e. in the context of a computation era attempting to revolutionise the discipline of chemical synthesis through both retrosynthetic design and forward synthetic planning.

**Is the Target Real?**

Both natural and synthetic products are regularly assigned incorrect structures, and it is often not until practitioners attempt their synthesis for the first time (e.g. trunkamide A (261)) or repeat a synthesis (e.g. tetrahydrooxazepine (262)) that the error is detected and corrected. Compounding this problem, in some cases even achieving a synthesis of the proposed target may not offer any additional information on the proposed structure (e.g. afzeliindanone (263)). Another dimension to this problem is that some structures are extremely difficult to solve (e.g. EBC-232 (264)), or an outright solution is not possible (e.g. a flat structure lacking relative and/or absolute stereochemistry). Therefore, the question in this context is whether AI development will in the future attempt to verify the input target structure has been correctly elucidated, which is especially important for those suspicious structure proposals that defy physical organic concepts such as aromaticity (e.g. 265) and Bredt’s rule (e.g. 266) (Fig. 6). Indeed, when describing an earlier incarnation of Chematica (pre 2016), Grzybowski et al. stated that ‘In full disregard of Bredt’s rules, it installed double bonds at bridgehead atoms’, and described this as a ‘creative’ and ‘non-existent or improbable’ structural feature. The structures synthesised via Chematica suggested routes in 2018 (Fig. 5) do not contain rare or unusually strained motifs, so it remains to be seen how Chematica would deal with such a structure now.

It is certainly conceivable that AI will indeed adopt an elucidation aspect, because it is currently possible through...
modern Computer-Assisted Structural Elucidation (CASE)\(^{[177,181]}\) to screen target structures for validity. More importantly, this feature can be undertaken rapidly with current density functional theory (DFT) and associated methods\(^{[182]}\) to give surprisingly accurate re-assignments, for example, the proposed structure for botryosphaerihydrofuran (267) and its revised structure 268 (Fig. 7).\(^{[183]}\) Assigning absolute stereochemistry is a developing possibility with such methods.\(^{[184]}\)

With that said, there are risks associated with automated NMR-based structural elucidation in terms of DFT energy miscalculations and subsequent erroneous consequences.\(^{[185]}\)

**Will AI Consider Model Systems?**

Model systems (or model studies) in organic synthesis are loosely defined as testing a ‘key step’ in a proposed synthesis without probing the real system. Such studies have been widely used, and in some cases with great success, for example, establishing whether ring-closing metathesis was a viable option for the construction of the D ring seen in manzamine A (269) (Fig. 8).\(^{[186]}\) However, the general view of the community is that a ‘model is a model, and real is real’,\(^{[7]}\) meaning that the model system may or may not work, and that outcome is not a true reflection of whether the proposed transformation would work (or not) in the real system. Therefore, it could be justified that model systems be deemed obsolete if the practitioner has embraced full AI assistance.

**Will AI Discount Fundamental Molecules?**

Fundamental molecules, such as strained hydrocarbons (e.g. cubane (54)\(^{[48]}\) and [1.1.1]propellane (270)\(^{[187]}\)), have fascinated theoretical and synthetic chemists since the discipline of organic chemistry began. However, the issue with many of these systems is undesirable physical properties, for example, explosivity and impact sensitivity. Therefore, will future AI retrosynthesis and synthetic planning programs take into account these details, which are a potential safety concern if not handled with practical knowledge and expertise? Cubanes and bicyclo[1.1.1]pentanes are such examples, but their impact sensitivity depends solely on the functional group appended to the parent skeleton (Fig. 9). For example, iodinated systems (e.g. 271 and 272) are quite sensitive, but not methyl esters (e.g. 273),\(^{[50b]}\) although some carboxylic acids are strong candidates (e.g. 274)\(^{[50a]}\). Hence, will AI search the literature for this know-how or calculate strain energy before making such synthetic design proposals? With available literature, and the ability to calculate strain energy and impact sensitivity,\(^{[50a]}\) it should be feasible to introduce these features. Information such as impact sensitivity, however, is often only reported in the written text and has no standardised reporting protocol, meaning that this information is very difficult to automatically extract from the literature.
Overall, however, these intriguing molecule types are playing an increasingly important role in bioactive molecule discovery, and thus cannot be eliminated from consideration.

Will AI Learn from Practical Synthetic Problems?

It is often the case with synthetic practice that retrosynthesis design is very much considered a rough plan for the forward synthesis, which in turn becomes a trial-and-error process to achieve the target. This process has been elegantly described as ‘dead ends and detours’. Given this aspect is an inherent process of chemical synthesis, it is likely that AI will by default follow a similar path, but in collaboration with the practitioner to determine what works in reality. This essential feedback loop provides additional information on unexpected failures arising from the initial AI retrosynthetic plan and projected synthetic approach. This process will ultimately achieve the stated goal, but it needs to be determined whether this approach is any different for what is currently undertaken based on solely human design. Efforts to evaluate such AI performance are ongoing; see for example Fig. 5.

Another aspect for consideration under this section is that of reaction yields. The main issue is that reported yields vary in accuracy owing to a range of factors (e.g. scale of reaction). To complicate matters further, academic laboratories do not determine yield values to the same level of accuracy being undertaken by industry. Hence, this can be considerably misleading when coding software for predictive forward syntheses. This complication is unlikely to change anytime soon.

How will Protecting Groups be Interpreted?

The famous opening chapter to ‘Protecting Groups’ by Kocienski, states up front ‘Death, Taxes and Protecting Groups’. This very statement sets the stage for the necessary evil that Kocienski follows by introducing exacerbated by an increasingly complex target that demands an protecting group. There are numerous syntheses that have been completed without the use of a single protection or deprotection, and this is a preferred situation as it limits synthetic steps, is economical, and limits waste. However, there is no denying that syntheses without protecting group deployment are difficult, and if the synthetic route to the target (or the target itself) is not conducive to it, then the effort involved in attempting to achieve this goal may end up using many more resources (e.g. labour) than adopting a protecting group or two. That is, protecting group-free synthesis is an admirable goal, but must be viewed in realistic terms.

Can Domino and Cascade Reactions be Readily Predicted?

Domino and cascade reactions are always a welcome opportunity for the synthetic practitioner to implement because of numerous benefits, for example increased efficiency and waste minimisation of solvents, reagents, adsorbents, energy, and labour. Therefore, such reactions facilitate ecologically and economically favourable production and manufacture, and would greatly assist meeting Mulzer’s 25-step limit for any complex synthesis. The terms domino, cascade, tandem, and one-pot are used apparently interchangeably in much of the literature, but they are used to describe very different reaction procedures. A domino reaction, as defined by Tietze, forms several bonds in one sequence without isolating the intermediates, changing the reaction conditions, or adding reagents. Importantly, in a domino reaction, each bond-forming reaction depends on the reaction conditions, whereas a tandem reaction has been defined by Nicolaou et al. as two or more reactions occurring on the same substrate and under the same conditions, but essentially in isolation of one another. Tandem is commonly used to describe reactions that would be more accurately defined as one-pot or sequential; however, tandem reactions are not time-resolved and so this usage is not preferred as justified by Tietze. According to Nicolaou et al., cascade can be used as an umbrella term for all domino, one-pot, and tandem reactions. Tietze argues that while cascade is often used to describe domino reactions, it is ambiguous and the latter term is preferred.

As much as these reactions are highly desirable, their design is by no means straightforward, and is generally associated with natural product syntheses due to inspiration provided by Mother Nature (see the Biomimetic Synthesis section above). In terms of AI, this places a significant demand on programming. Synthia will identify reactions that can take place under the same reaction conditions, which addresses simultaneous or tandem reactions, using the definition of Nicolaou et al. Tactical combinations, as coined by Corey, can be defined as strategic retrosynthetic transformation sequences, the first of which does not immediately or obviously produce any benefit and often increases molecular complexity, but sets the scene for the
second transformation, which results in molecular simplification. The term can in theory be applied to any combination of two transformations that fit the criteria; however, previously only ~500 such combinations had been explicitly identified in the literature and catalogued. Grzybowski et al. used Chematica to discover ~4.85 million tactical combinations and subsequently explicitly coded them into the program (along with the ~500 ‘known’ tactical combinations). This is part of the reason why Synthia can plan over multiple steps, but currently only two-step combinations have been included. These tactical combinations are not necessarily suitable for domino reactions, however, because there is no requirement for them to preclude the isolation of intermediates.

Overall, it is possible that Synthia would identify a route that would, in practice, result in a domino process. However, it seems that the user would have to manually review the retrosynthetic tree and identify that each reaction step would take place sequentially and under the exact same reaction conditions, unless the domino sequence was already known and was coded in as a reaction rule. Therefore, this implementation could be some time away and is far from being perfected.

Assessing Molecular Complexity

In light of what we know about computer-assisted and AI methods for devising retrosyntheses and route planning, there is a requirement for the computer to understand complexity. Complexity is two-fold in this context, although related: (1) molecular or target complexity, and (2) synthetic design or route complexity. Note that the molecular complexity of the target may or may not be directly proportional to the number of steps required to synthesise that target. Therefore, should there be some consideration as to whether this impacts AI performance in undertaking retrosynthetic design and predicting forward syntheses more efficiently? Obviously, AI will inherently attempt to provide the ‘best’ synthetic approach following consideration of the proposed retrosynthetic plan.

How is complexity perceived? Although molecular complexity is difficult for the practitioner to define, it is generally intuitively understood. For example, taking into account the target’s carbon skeleton, number of appendages, and functional groups, stereochemistry and conformation, one can quickly ascertain a level of molecular complexity, and subsequently synthetic difficulty. As mentioned above, the latter is not necessarily determined by molecular complexity, i.e. complex molecules can be readily synthesised, but it is all dependent on retrosynthetic planning. When coding tactical combinations, molecular complexity had to be considered by the Chematica/Synthia team, and was quantified by the number of rings, the number of stereocentres, and the length of the SMILES string code of the molecule. Some targets, such as Taxol® are currently ‘too complex’ for Synthia, and the route search times out (i.e. exceeds the available RAM).

In an effort to conceptualise retrosynthetic planning based on complexity, there are three general parameters that have been used by the lead author over the years to categorise synthetic approaches to targets. These three categories, which are essentially a ranking system of synthetic difficulty, are best imagined as analogies to proton NMR splitting patterns, which include first-order, second-order, and non-first-order interpretations.

A good first-order example is the total synthesis of EBC-329 as it is mostly first-order $(n + 1)$ retrosynthetic bond breaking. When further simplified by adopting a chiral pool approach, as suggested by the cyclopropyl-gem-dimethyl residue, the conjugated enlactone is disconnected in the first instance to give an enyne as conjured from metal coupling. Further disconnection of the enyne to a methyl ketone (i.e. scheme 276) is inspired by Wittig chemistry, as is the inspiration for the subsequent disconnect to the dicarbonyl (278). Two carbonyls distanced by six carbon units imply an oxidative ring cleavage strategy fitting carene (279) (Scheme 26).

Based on this first-order retrosynthesis, the forward synthesis almost became pedestrian, albeit with one minor hurdle. Oxidative cleavage performed best over two steps to give 278, which reacted smoothly with the phosphorane to afford 277. Some experimentation was required to procure the second olefination, but this was eventually achieved using Horner–Wadsworth–Emmons methodology, although a mixture of and -isomers was obtained (i.e. 282). The final step consisted of installing the alkylidenebutenolide function, which was accomplished in one step via a copper-mediated reaction with iodoacid 283 (Scheme 26). HPLC purification then afforded the pure natural product, establishing the absolute stereochemistry in a total of seven steps (Scheme 27).

Second-order considerations include multiple bond disconnections at once, which would equate to strategies consisting of skeletal rearrangements (e.g. Scheme 21), cascade reactions (e.g. Scheme 5), and cycloadditions (e.g. Scheme 13) as prime examples in this category.

Non-first-order retrosynthesis (akin to aromatic ring-splittting patterns) encompasses irrational bond breaking and forming retrosynthetic manoeuvres that generally only become clear once viewed. A seminal contribution to this category was...
the total synthesis of nominine (284) by Peese and Gin.[200] The retrosynthesis boldly aimed to build four bonds using a convergent dual-cycloaddition approach via key intermediate 285 (Scheme 28).

The elegant forward synthesis is immediately daring by not only starting with an unstable dimethyl acetal (i.e. 286), but performing an ortho-lithiation on a substrate where multiple positions could have in principle been lithiated. The product (288) was converted into a functionally dense azide (289), which was converged with a cyclohexylaldehyde (290) as obtained from enone 20 via triflate 291. The audacious Staudinger reaction afforded the adduct (292) in high yield, which on treatment with acid gave the desired zwitterion (293), setting the stage for the first key 1,3-dipolar cycloaddition. Astoundingly, heating 293 at 180°C gave in one step the majority of the target skeleton (i.e. 294). FGI gave rise to the second cycloaddition precursor, alkene 297, via intermediates 295 and 296. Enamine formation derived from 297 facilitated the second cycloaddition to give 298 in 78% yield. Installation of the exocyclic double bond followed by allylic oxidation gave nominine (284) (Scheme 29). This bold and exemplary case reduced a previously 40-step synthesis to just 15 steps.

Given practitioners can intuitively categorise both forms of complexity (molecular and synthetic), an in silico complexity index has been developed to rank both molecule and synthetic complexity,[201] which opens the door for AI adoption if a requirement is justified. The overall message being conveyed, however, is that complexity is much better understood by the practitioner than the computer at this point in time.

**Mechanism**

Understanding the mechanism of a chemical process is a fundamental part of the practitioner’s toolbox, as it enables further discovery and innovation. However, is this process important for retrosynthesis and forward synthesis if it merely comes down to bond breaking and bond forming from a graph theory or AI perspective? This aspect is difficult to judge because some chemical processes take decades to elucidate, although not hindering their use and adoption by the community. The Ley–Griffith oxidant, tetra-n-propylammonium perruthenate (TPAP), is a good example that was routinely used for mild and controlled oxidation of alcohols and many other substrates, but whose mechanism of oxidation was only recently elucidated, having stood unknown for almost four decades.[202] In the eyes of the computer, it is likely going to be a case that if it has literature precedent as working, then it can be adopted irrespective if the mechanism is understood or not.

**Synthetic Methodology**

From the perspective of AI coding, and current day AI performance (e.g. Synthia), has sufficient methodology already been discovered? If the paradigm shifts introduced by ring-closing metathesis[203] and C–H activation[10] are any indication, then methodological development must continue, whether by rational design or accidental discovery. However, what must not happen is the removal of old methodology from consideration as new methodology is developed, i.e. this is not a case of replacement, but bolstering the repertoire.

**Will AI Develop a Chemical Mind or Brain Equivalent?**

Perhaps the most important, albeit emotionally confronting, question for practitioners of synthetic chemistry is whether they...
will become obsolete owing to AI thinking and robotics (e.g. performing chemical reactions). In fact, this is an area of debate for humankind in general and across all facets of life.

As early as 1950, the question ‘Can machines think?’ was proposed by Turing, who also embarked on defining the term ‘think’. Some 30 years later, Searle proposed essentially the same question, ‘Could a machine think?’ but attempted to address the mind, brain, and program interface using the argument of intentionality in human beings. He then went on to differentiate the brain–mind relationship (dualism), often called the mind–body problem, by considering the brain as a digital computer, and the mind as a computer program. More recently, Signorelli considered the idea of machines overcoming humans and, if so, that they must be intrinsically related to conscious machines.

The Turing test, originally called the imitation game by Turing, essentially asks whether an interrogator would be able to tell the difference between two participants, one of which is human and one a machine. Importantly, the Turing test examines behaviour (output) rather than mechanism or working (process). There have been many variations of the Turing test, and the subject matter expert or Feigenbaum test is particularly applicable to chemical synthesis programs. Chematica was recently subjected to the Feigenbaum or Turing test by having 18 experts grade the likelihood of 40 (20 from the literature and 20 from Chematica) synthetic routes to be human- or computer-designed. Overall, the experts were unable to differentiate between the human- and computer-designed routes, and even scored the computer-designed routes as being slightly more ‘elegant’ than the human-designed routes.

With that said, all these considerations are philosophical debate, and are in the process of being translated, tested, and determined in silico. Therefore, it seems a long way off before synthetic chemists undergo complete morphological change into scientific generalists, i.e. when a ‘scientist’ of the future acquires molecules (based on society drivers) using the push of a button.

**Conclusion**

There is no doubt that synthetic chemistry has been and will continue to be a substantial enabling foundation discipline for humankind. The past and present output of pharmaceuticals, agrochemicals, and materials (e.g. plastics, microelectronics) is
a reflection of the magnitude of this discipline being deployed by industry.\textsuperscript{210}

However, even though this branch of science is highly sophisticated, it is by no means trouble-free, as Cornforth flagged back in 1993,\textsuperscript{211} likely with Seebach’s 1990 proposal of ‘Organic synthesis – where now?’\textsuperscript{212} front of mind. If we accept the answer to Seebach’s question is ‘everywhere’ and that Cornforth’s conjecture remains correct, then evolution of the discipline is inevitable, and in modern times that requires integration of algorithmic methods to assist in prompting human decision-making. Therefore, with the emergence of programs such as Synthia, the real quandary is whether the art of organic synthesis will be maintained, i.e. will the syntheses be elegant, and meet all social related desires of efficiency and environmental impact? Beyond that point, will computer-assisted methods and AI truly be revolutionary for the discipline, and importantly should they be relied on completely as Judson intimates?\textsuperscript{213}

‘The GPS navigation device may render paper maps redundant – but not the driver of the car.’\textsuperscript{214}

Lastly, the present article expresses certain thoughts and philosophies on retrosynthesis and forward synthetic planning that are designed to generate discussion among students and colleagues alike. It is fully realised that this important and ever-evolving area comes with many individual views and experiences that constantly drive synthetic organic chemistry thinking, and Sir John Warcup Cornforth\textsuperscript{215} is one of several iconic individuals who have made significant contributions to the evolution of chemical synthesis.

Conflicts of Interest
The authors declare no conflicts of interest.

Declaration of Funding
This research did not receive any specific funding. The University of Queensland provided all the resources to enable the preparation of this manuscript.

Acknowledgements
C.M.W. is eternally grateful to the many bright and talented students and post-doctoral researchers who have passed through his laboratories, bringing culture, philosophy, and humour to the Williams research group enterprise.

C.M.W. is profoundly grateful for financial and collaborative support over the last 20 years from the Australian Research Council (ARC), National Health and Medical Research Council (NHMRC), State and Federal Government departments (e.g. DPI, DAWE), Defence Science Technology Group (DSTG), Commonwealth Scientific and Industrial Research Organisation (CSIRO), and local and international industry (e.g. EcoBiotics Ltd, Bayer, Dow, DuPont). C.M.W. is eternally grateful for the many wonderful scientific collaborations that have developed over time, and in turn allowed his research to blossom in a multitude of unexpected directions. Such relationships with both academic and industrial collaborators are dearly cherished. M.A.D gratefully acknowledges the Australian Government Research Training Program Scholarship.

C.M.W. is especially grateful to Emeritus Professor Curt Wentrup (University of Queensland) for the invitation to write an Australian Journal of Chemistry Sir John Cornforth review article. Beyond the invitation itself, C.M.W. is especially honoured to contribute to this series having had the fortunate opportunity to meet Sir John W. Cornforth Jr. during the course of being interviewed for a Lectureship in 1999 at the University of Sussex, Brighton, UK. Anecdotally, Sir John Cornforth sat in the front row of the audience about 1 m away from the presenting interviewees, and at the end of C.M.W.’s lecture, asked (shouted owing to deafness) a stereochemistry question regarding the proposed research program, which was not likely answered satisfactorily. At the end of the lecture C.M.W. asked his host Professor Phil Parsons (now at Imperial College London), ‘Who was that?’, to which he laughed and replied, ‘That was John Cornforth!’ Perhaps it was no surprise that Associate Professor Gareth Rowlands (now at Massey University, NZ) ended up getting the position.

Lastly, the authors are most grateful to those that have provided feedback on this topic over the years and the ensuing manuscript.

References
[1] (a) E. J. Corey, X.-M. Cheng, The Logic of Chemical Synthesis 1995 (Wiley: New York, NY).
(b) E. J. Corey, Chem. Soc. Rev. 1988, 17, 111. doi:10.1039/CS9881700111
(c) E. J. Corey, Pure Appl. Chem. 1967, 14, 19. doi:10.1351/PAC196714010019
[2] (a) The Nobel Prize, The Nobel Prize in Chemistry 1990. Nobel Media AB 2020. Available at: https://www.nobelprize.org/prizes/chemistry/1990/summary (accessed 11 Jun 2020)
(b) E. J. Corey, Angew. Chem. Int. Ed. Engl. 1991, 30, 455. doi:10.1002/ANIE.199104553
[3] (a) S. Warren, Designing Organic Syntheses: A Programmed Introduction to the Syntaxon Approach 1982 (Wiley: New York, NY).
(b) S. Warren, P. Wyatt, Organic Synthesis: The Disconnection Approach, 2nd Edn 2008 (John Wiley & Sons: New York, NY).
[4] M. B. Smith, J. Chem. Educ. 1990, 67, 948. doi:10.1021/ED067p948
[5] (a) J. Saunders, Top Drugs: Top Synthetic Routes 2000 (Oxford University Press: Oxford).
(b) T. P. Stockdale, C. M. Williams, Chem. Soc. Rev. 2015, 44, 7737. doi:10.1039/C4CS04477A
[6] K. C. Nicolaou, E. J. Sorensen, Classics in Total Synthesis: Targets, Strategies, Methods 1996 (Wiley-VCH: Weinheim).
[7] K. C. Nicolaou, S. A. Snyder, Classics in Total Synthesis II: More Targets, Strategies, Methods 2003 (Wiley-VCH: Weinheim).
[8] K. C. Nicolaou, J. S. Chen, Classics in Total Synthesis III: Further Targets, Strategies, Methods 2011 (Wiley-VCH: Weinheim).
[9] (a) D. G. Brown, J. Boström, J. Med. Chem. 2016, 59, 4443. doi:10.1021/ACS.JMEDCHEM.5B01409
(b) S. D. Roughley, A. M. Jordan, J. Med. Chem. 2011, 54, 3451. doi:10.1021/JM200187Y
[10] (a) H. M. L. Davies, K. Liao, Nat. Rev. Chem. 2019, 3, 347. doi:10.1038/S41570-019-0099-X
(b) Y. Wei, P. Hu, M. Zhang, W. Su, Chem. Rev. 2017, 117, 8864. doi:10.1021/ACS.CHERMREV6B00516
(c) J. He, M. Wasa, K. S. L. Chan, Q. Shao, J.-Q. Yu, Chem. Rev. 2017, 117, 8754. doi:10.1021/ACS.CHERMREV6B00622
[11] D. Petzold, M. Giedyk, A. Chatterjee, B. König, Eur. J. Org. Chem. 2020, 1193. doi:10.1002/EJOC.201901421
[12] (a) E. J. Corey, W. T. Wipke, Science 1969, 166, 178. doi:10.1126/SCIENCE.166.3902.178
(b) E. J. Corey, W. J. Howe, H. W. Orf, D. A. Pensak, G. Peterson, J. Am. Chem. Soc. 1975, 97, 6116. doi:10.1021/JA00854A026
(c) E. J. Corey, W. L. Jorgensen, J. Am. Chem. Soc. 1976, 98, 189. doi:10.1021/JA00417A030
(d) E. J. Corey, A. K. Long, S. D. Rubenstein, Science 1985, 228, 408. doi:10.1126/SCIENCE.3838594
(e) K. Lin, Y. Xu, J. Pei, L. Lai, Chem. Sci. 2020, 11, 3355. doi:10.1039/C9SC03666K
(f) P. Schwaller, R. Petraglia, V. Zullo, V. H. Nair, R. A. Haeuselmann, R. Pisoni, C. Bekas, A. Iuliano, T. Laino, Chem. Sci. 2020, 11, 3316. doi:10.1039/C9SC05704H
[13] (a) H. L. Gelernter, A. F. Sanders, D. L. Larsen, K. K. Agarwal, R. H. Boivie, G. A. Spritzer, J. E. Searleman, Science 1977, 197, 1041. doi:10.1126/SCIENCE.197.4308.1041
(b) W. T. Wipke, G. I. Ouchi, S Krishnan, Artif. Intell. 1978, 17, 133. doi:10.1016/0004-3702(78)9006-4
(c) S. Szymkuc’, E. P. Gajewska, T. Klucznik, K. Molga, P. Dittwald, M. Startek, M Bajczyk, B. A. Grzybowsk, Angew. Chem. Int. Ed. 2016, 55, 5904. doi:10.1002/ANIE.201506101
See for example: A. Greule, J. E. Stok, J. J. De Voss, M. J. Cryle, N. X. Zhang, E. King-Smith, L.-Bi. Dong, L.-C. Yang, J. D. Rudolf, B. D. M. Pinkerton, T. J. Vanden Berg, P. V. Bernhardt, C. M. Williams, Chem. – Eur. J. 2017, 23, 2282. doi:10.1002/chem.201605751

(b) D. M. Pinkerton, P. V. Bernhardt, G. P. Savage, C. M. Williams, Asian J. Org. Chem. 2017, 6, 583. doi:10.1016/j.ajoc.201700013

c) D. M. Pinkerton, S. Chow, N. H. Eisa, K. Kaith, T. J. Vanden Berg, J. M. Burns, L. W. Gudat, G. P. Savage, A. Chadli, C. M. Williams, Chem. – Eur. J. 2019, 25, 1451. doi:10.1002/chem.201805420

IBX has sufficient acidity to promote side reactions and/or acid catalysed deprotections, see ref. [28] and T. Waters, J. Boulton, T. Clark, M. J. Gallen, C. M. Williams, R. A. J. O’Hair, Org. Biomol. Chem. 2008, 6, 2530. doi:10.1039/B803452D

(a) M. J. Schnermann, R. A. Shenvi, Nat. Prod. Rep. 2013, 30, 543. doi:10.1039/C3NP00109E

(b) M. J. Garson, J. S. Simpson, Nat. Prod. Rep. 2004, 21, 164. doi:10.1039/B302359C

E. J. Corey, P. A. Magriotis, J. Am. Chem. Soc. 1987, 109, 287. doi:10.1021/JA00235A052

(a) K. A. Fairweather, L. N. Mander, Org. Lett. 2006, 8, 3395. doi:10.1021/OL061228F

(b) K. A. Fairweather, S. R. Crabtree, L. N. Mander, in Strategies and Tactics in Organic Synthesis (Ed. M. Harmata) 2012, Vol. 8, Ch. 2, pp. 35–58 (Academic Press: New York, NY). doi:10.1016/B978-0-12-386540-3.00017-4

(b) D. Schwartz, C. M. Williams, E. Anders, P. V. Bernhardt, Tetrahedron 2008, 64, 6482. doi:10.1016/J.TET.2008.04.068

(b) D. Schwartz, J. R. Denton, P. V. Bernhardt, H. M. L. Davies, C. M. Williams, Synthesis 2009, 2840. doi:10.1055-S-0029-1216912

(b) S. Jones, B. Simmons, A. Mastracchio, D. W. C. MacMillan, Nature 2011, 473, 183. doi:10.1038/NATURE10232

See for example: (a) J. Wang, S.-G. Chen, B.-F. Sun, G.-Q. Lin, Y.-J. Shang, Chem. – Eur. J. 2013, 19, 2539. doi:10.1002/chem.201203467

(b) W. Xu, J. Zhao, C. Tao, H. Wang, Y. Li, B. Cheng, H. Hai, Org. Lett. 2018, 20, 1509. doi:10.1021/OL052005C

(c) C. Qiao, W. Zhang, J.-C. Han, W.-M. Dai, C.-C. Li, Tetrahedron 2019, 75, 1739. doi:10.1016/J.TET.2018.11.019

(d) D.-W. Zhang, H.-L. Fan, W. Zhang, C.-J. Li, S. Luo, H.-B. Qin, Chem. Commun. 2020, 56(33), 4751. doi:10.1039/D0CC04064A

(e) S.-H. Wang, R.-Q. Si, Q.-B. Zhang, X. Guo, T. Ke, X.-M. Zhang, F.-M. Zhang, Y.-Q. Tu, Angew. Chem. Int. Ed. 2020, 59, 21954. doi:10.1002/ANO202009238

(b) A. V. Lakshmi, P. Gupta, Nat. Prod. Res. 2008, 22, 1197. doi:10.1080/147864107017654909

(b) D. A. Mulholland, B. Parele, P. H. Coombes, Curr. Org. Chem. 2000, 4, 1011. doi:10.1080/1478641001002137941

(c) J. Wu, Q. Xioo, J. Xu, M.-Y. Li, J.-Y. Pan, M.-h. Yang, Nat. Prod. Rep. 2008, 25, 955. doi:10.1039/B003745A

J. M. Faber, C. M. Williams, Chem. Commun. 2011, 2258. doi:10.1039/C0CC04698A

J. M. Faber, W. A. Eger, C. M. Williams, J. Org. Chem. 2012, 77, 8913. doi:10.1021/JOC11812F

J. D’Angelo, M. B. Smith, Hybrid Retrosynthesis Organic Synthesis Using Reaxys and SciFinder 2015 (Elsevier: Amsterdam).

G. E. Vleéduits, V. K. Finn, Inf. Storage Retr. 1963, 1, 101. doi:10.1007/BF01075012-3

M. Bersohn, A. Esack, Chem. Rev. 1976, 76, 269. doi:10.1021/FCHEM.18009199

(a) M. H. Todd, Chem. Soc. Rev. 2005, 34, 247. doi:10.1039/B010460A

(b) A. Cook, A. P. Johnson, L. Law, M. Mirzazadeh, O. Ravitz, A. Simon, WIREs Comput. Mol. Sci. 2012, 2, 79. doi:10.1002/WCMS.61

(c) O. Engkvist, P. O. Norby, N. Selmi, Y. Lam, Z. Peng, E. C. Sherer, W. Amberger, T. Erdahl, L. A. Smyth, Drag Digest. Today 2018, 23, 1203. doi:10.1001/DRUDIS.2018.02.014

(d) F. Feng, L. Lai, J. Pei, Front. Chem. 2018, 6, 199. doi:10.3389/FCHEM.2018.00199

(e) F. Streith-Kalthoff, S. Sandfort, M. H. S. Segler, F. Glorius, Chem. Soc. Rev. 2020, 49, 6154. doi:10.1039/C9CS00786E

(f) P. M. Pflüger, F. Glorius, Angew. Chem. Int. Ed. 2020, 59, 18860. doi:10.1002/ANIE.202008366

(g) W.-D. Ihlenfeldt, J. Gasteiger, Angew. Chem. Int. Ed. 1996, 34, 2613. doi:10.1002/ANIE.199526131

(h) G. Schneider, D. E. Clark, React. Chem. Eng. 2019, 58, 10792. doi:10.1002/ACIE.201814681

(i) A. Thakkar, S. Johansson, K. Jorner, D. Buttar, J.-L. Reymond, O. Engkvist, React. Chem. Eng. 2021, 6, 27. doi:10.1039/D0RE00340A

(a) A. Bondy, U. S. R. Murty, Graph Theory 2008 (Springer-Verlag: London).

(b) L. David, A. Thakkar, R. Mercado, O. Engkvist, J. Cheminform. 2020, 12, 56. doi:10.1002/1332-020-00460-5

T. Fink, H. Bruggesser, J.-L. Reymond, Angew. Chem. Int. Ed. 2005, 44, 1504. doi:10.1002/ANIE.200402457
[206] J. R. Searle, *Behav. Brain Sci.* 1980, 3, 417. doi:10.1017/S0140525X00005756

[207] J. R. Searle, *Proc. Addresses Am. Philos. Assoc.* 1990, 64, 21. doi:10.2307/3130074

[208] C. M. Signorelli, *Front. Robot. AI* 2018, 5, 121. doi:10.3389/FROBT.2018.00121

[209] E. A. Feigenbaum, *J. ACM* 2003, 50, 32. doi:10.1145/602382.602400

[210] See for example: P. A. Wender, B. L. Miller, *Nature* 2009, 460, 197. doi:10.1038/460197A

[211] J. W. Cornforth, *Aust. J. Chem.* 1993, 46, 157. doi:10.1071/CH9930157

[212] D. Seebach, *Angew. Chem. Int. Ed. Engl.* 1990, 29, 1320. doi:10.1002/ANIE.199013201

[213] P. Judson, *Knowledge-based Expert Systems in Chemistry: Not Counting on Computers* 2009 (RSC: Cambridge). doi:10.1039/9781847559807

[214] H. Else, ‘Need to make a molecule? Ask this AI for instructions’, *Nature News*, 28 March 2018. doi:10.1038/D41586-018-03977-W

[215] A. R. Battersby, D. W. Young, *Biogr. Mem. Fellows R. Soc.* 2016, 62, 19. doi:10.1098/RSBM.2015.0016

Handling Editor: Curt Wentrup