doi:10.2533/chimia.2023.17 Chimia 77 (2023) 17–21

Tools for Synthesis Planning, Automation, and Analytical Data Analysis

Miruna Cretu[a], Marvin Alberts[a], Anubhab Chakraborty[a], Artem Leonov[a], Amol Thakkar*[ab], and Teodoro Laino[ab]

Abstract: Computer-aided synthesis design, automation, and analytics assisted by machine learning are promising resources in the researcher's toolkit. Each component may alleviate the chemist from routine tasks, provide valuable insights from data, and enable more informed experimental design. Herein, we highlight selected works in the field and discuss the different approaches and the problems to which they may apply. We emphasise that there are currently few tools with a low barrier of entry for non-experts, which may limit widespread integration into the researcher's workflow.

Keywords: Analytical Chemistry · Automation · Cheminformatics · Synthesis planning

1. Introduction

Artificial intelligence (AI) is becoming integral to chemical discovery pipelines as data and computational resources fuel a paradigm shift towards data-driven science.[1] AI research in chemistry commenced in the 1960s, intending to remove human bias and improve the efficiency of a chemist's workflow. Two pioneering projects emerged during the period. One was for the design of synthetic routes as proposed by E. J. Corey and led to the field of computer-aided synthesis planning (CASP).[2] The second, called Dendral, tackled the structure elucidation of unknown organic compounds.[3] The recent rise in the availability of publicly accessible data, computing power and AI algorithms has led to progress in developing models tailored to chemistry. As such, AI tools can now provide chemists with assistance in synthesis and automation tasks.

There exist numerous machine learning (ML) approaches to retrosynthesis, predictions of yields, and molecular properties.[4–7] With lab automation allowing the execution of such syntheses, the final link in the chain is analysis, where ML offers significant speed-up.[5,7–9] The availability of open-source tools and affordable hardware drives the synergy between digital and bench chemistry. Here we provide an overview into the process of this synergy's consolidation, by highlighting ways in which the strengths of AI are reflected into valuable tools that can be employed by bench chemists. We discuss the approaches AI takes to synthesis planning, automation, and, finally analytical chemistry (Fig. 1). In the process, we highlight freely-available or open-source software and discuss their limitations and ways of tackling them. Lastly, we stress the importance of chemists' contribution to crystallising progress by providing high-quality data and feedback on the existing tools.

2. Synthetic Route Design

Computer-aided synthesis planning (CASP) is a half-century-old area of research that recently culminated with the integration of ML into its development.[2] In this framework, synthesis planning is approached in two overarching steps: the prediction of individual transformations and the design of a search algorithm to complete a synthesis route. Although computational methods hold promise to augment synthetic route planning, they are not yet mainstream among chemists. This is partly due to their limitations, but equally due to their rapid emergence, with chemists often finding themselves overwhelmed by the number of CASP options available.[10,11]

2.1 *Retrosynthesis*

CASP methods for retrosynthesis are divided into 'expert' rule-based and ML-based techniques. The former include the Chematica/Synthia project and use hand-coded reaction rules in conjunction with guiding heuristics to navigate and select optimal synthetic pathways.[12] Although such methods excel in low-data regimes where few reactions apply to a particular transformation, they quickly become unfeasible when multiple reactions can match a transformation, especially with the number of reaction types discovered per year lying in the low thousands.[13]

ML-based CASP methods are gaining popularity in synthesis planning, and they can be largely categorized into data-driven methods, which entirely discard the use of rules, and hybrid methods. Similarly to expert rule-based approaches, hybrid methods rely on templates to identify relevant transformations for a target. However, they incorporate ML algorithms to assist the selection of optimal precursors and reactions in such a way as to avoid combinatorial explosion, as the number of precursors matching a target can easily reach a few thousands. Moreover, the possibility to extract template rules in an automated manner has emerged, releasing the burden on expert chemists to generate them manually.[14]

Entirely data-driven tools include the RXN platform, which uses an architecture dubbed the Molecular Transformer, that has been adapted from natural language processing (NLP) for synthesis planning tasks.[15] In this framework, chemical reactions are encoded as sentences using reaction SMILES, so that retrosynthesis prediction is regarded as a translation task from products to reactants, and reaction prediction as a translation task from reactants to products.[6,15] In retrosynthesis, after all possible transformations are generated by the Molecular Transformer, a beam search on the hypergraph of the proposed disconnections is used to find an optimal synthetic pathway.[6]

*Correspondence: Dr. A. Thakkar, Email: tha@zurich.ibm.com
[a]IBM Research Europe, Saumerstrasse 4, 8803 Rueschlikon, Switzerland; [b]National Center for Competence in Research-Catalysis (NCCR-Catalysis), Switzerland

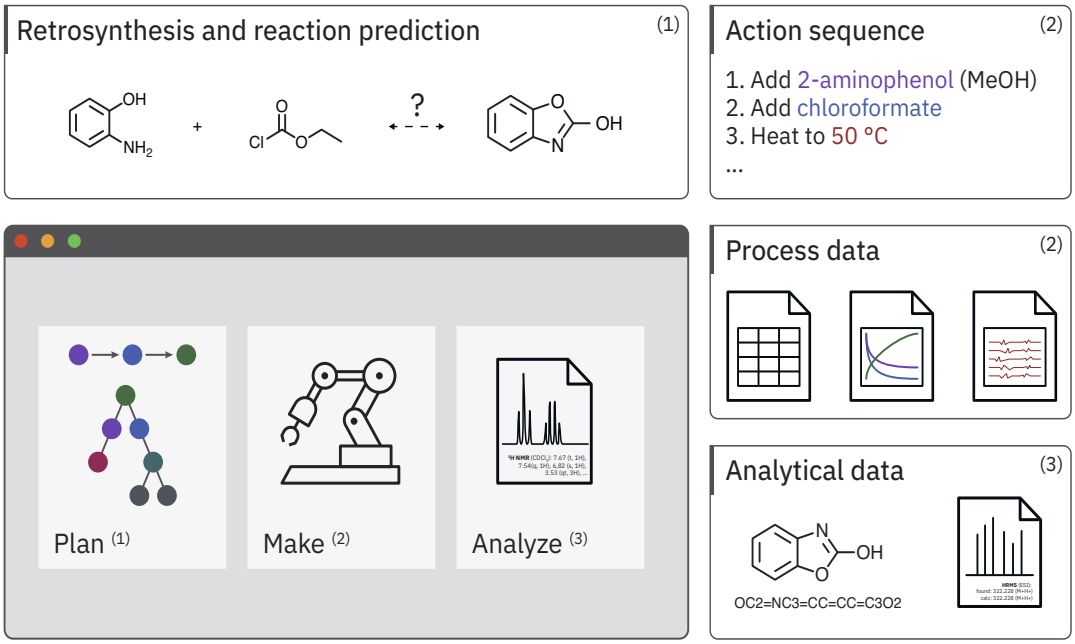

Fig. 1. Schematic showing the different types of AI-based tools that are currently accessible by the chemist.

2.2 *Synthetic Pathway Ranking*

Once multiple reaction steps are generated for a particular target, a ranking based on their likelihood is determined. This, however, often leaves users overwhelmed with the number of similar synthesis options, as the top-ranked options may differ only at a sub-portion level. Therefore, pathway-level guidance and output diversity have become active areas of research for predicting synthesis routes.[16] Heuristic strategies have been suggested to sort pathways by the number of steps involved (making sure few unproductive steps are present), and scoring functions have been defined to combine single-step reaction likelihood with degree of molecule simplification.[17] In addition, application-oriented metrics can be used to sort pathways, such as the price of the final precursors.[18] However, this requires a comprehensive database of buyable precursors, and consistency of such databases across different retrosynthesis predictions tools is not ensured.

2.3 *Tools for Computer-aided Synthesis Planning*

Out of the multiple technologies to employ hybrid template-based synthesis planning methods, ASKCOS and AiZynthFinder provide open-source software which are within accessible reach by chemists through graphical user interfaces (GUI).[19,20] AiZynthFinder provides the end-user with a command-line interface (CLI), as well as with a GUI through Jupyter Notebooks.[20] The GUI, however, lacks the distinctive features of the CLI, such as the abilities to process compounds in batch and to store detailed results on disc. ASKCOS path planning is available through a website interface that allows users to build retrosynthetic trees where each suggested reaction can be expanded and evaluated by the built-in reaction prediction tool.[21] Here, the user has the possibility to inspect recommended conditions, impurity, regio- and stereo-selectivity predictions generated by the same software. In fact, there have been major efforts to address the limited ability of data-driven models to handle regioselectivity. Ree *et al*. proposed a combination of QM/ML, which outperforms purely physics-based methods for predicting the regioselectivity of electrophilic aromatic substitution reactions.[22] Likewise, Guan *et al*. combine QM descriptors with ML to predict substitution reactions more generally.[23] Although selectivity models have limited user accessibility, they play a key role in assessing the quality of synthetic pathways.[6]

To facilitate user interaction, the RXN platform has a GUI that is freely available and includes additional features such as designating preferred disconnection sites in the product to guide the retrosynthesis.[24] Each step of the synthesis can be edited on the fly by adding missing reagents or solvents, and interactive features include further predicting synthesis routes for other compounds involved in the original synthesis.

Chemists' contribution to the development of AI tools for synthesis is crucial, both from a data perspective, where larger datasets and increased compound diversity can help models make more accurate predictions and improve model validation, and from a method development perspective, where chemists, as users, can offer invaluable feedback and motivation for improving the models. For this reason, freely-available retrosynthesis and reaction prediction platforms provide the end user the option to offer feedback on their experience and point towards limitations of the models.

3. Automation

The latest advances in ML within the chemistry domain demand high quality experimental data.[25] Although the vast majority of small molecule syntheses are still performed manually, the automation of chemical processes is becoming more accessible. With increased throughput, reproducibility, and precise records, automated systems facilitate chemical research, reaction optimisation, and discovery.[26,27]

3.1 *Digitisation of Reaction Procedures*

An important step towards laboratory automation is the adaptation of existing synthetic procedures to the available hardware platform. Various software tools facilitate the extraction of 'chemical keywords' (such as chemical names, actions, units, *etc*.)[28,29] to build a machine-readable procedure representation. Although such tools were not designed specifically for automation, they can be used to provide a list of instructions that can be adapted to automated execution. However, the adaptation process requires substantial human labour. To bridge the gap between procedure representation and executable code, Cronin *et al.* have proposed the chemical descriptive language, XDL, a markup language for chemistry.[30] A web-based application facilitates the translation of procedures to XDL, and outputs a hierarchy of chemical unit operations, which can be downloaded and executed on compatible hardware.[31] Despite the broad scope of chemical operations accounted for by XDL, its rule-based nature limits the possibilities in understanding the gamut of natural language.

An alternative approach using natural language processing (NLP) to extract chemical actions from procedures was suggested by Vaucher *et al*.[32] The resulting model, despite mistakes in the order of actions and quantities of materials, is capable of translating the procedure into a sequence of unit operations. The unit operations may be used to execute a synthesis either manually or on a robotic platform, as demonstrated by the RXN tool.[24] The concept was later extended to the prediction of an entire procedure based on the reaction SMILES which encodes the structure of the reactants and products.[33] In contrast to the XDL approach which translates a procedure into a set of unit operations, the one developed by Vaucher *et al*. only requires the encoded chemical species (including reagents, solvent, and catalyst) to output a set of unit operations. Despite its limited accuracy, to the best of our knowledge, it remains the only model capable of complete reaction procedure prediction including addition order and work up details, and is available for non-expert users through a web-browser.

3.2 *Synthesis Automation*

Automated systems are well suited to screening, optimisation problems, and execution of well-established unit operations.[34] Within chemical synthesis, this includes finding an initial or optimal set of reaction conditions, and expanding the substrate scope of a reaction.[35]

The basic requirements for an autonomous platform are hardware, analytical instruments, and orchestration software.[36] Most of the progress in automated systems was achieved in flow chemistry,[37,38] but the adoption of flow techniques requires specialised knowledge for establishing the reaction pathway and execution set-up. In contrast, batch systems leverage existing laboratory workflows, however the process of building such autonomous platforms is hampered by initial investments costs, personnel training, choice of vendors and running costs for maintenance and support.[39] Furthermore, the lack of standardised communication protocols between machines blocks autonomous decision making.[40,41] This means that data cannot be easily extracted, characterised, hypothesis generated, and tested in an autonomous manner, although each individual step is routinely conducted using manual labor. To address the communication issue, both proprietary and public toolkits have been developed, however they have not yet reached widespread adoption.[42] ChemOS is one such system, capable of managing hardware and orchestrating experimental planning, being backed with a diverse range of ML-based optimisation algorithms.[43] Furthermore, capabilities for database management support and the remote control of automatic hardware may drive the development of self-driving laboratories, as demonstrated for the optimisation of cross-coupling reactions.

Cronin *et al*. have proposed an open-source solution, namely The Chemputer (now ChemPU), that aims to automate traditional laboratory operations.[44] It is a modular robotic platform, capable of a diverse range of chemical operations – from synthesis to work-up and purification. The open-source approach to hardware enables the 3D printing of modules, which can be coupled with existing laboratory equipment, and using the XDL framework can translate high-level unit operations (*e.g*. add reagent) into low-level hardware primitives (*e.g*. syringe pump movements). This reduces the barrier of adoption for the community, and enables collaborative development. Recently, the platform has been coupled with a benchtop NMR for automated analysis and adjustment of reactions on the fly.[45]

In cases where the availability of capital may be a limiting factor, several groups have opted to repurpose and modify existing commercial hardware. Using in-house scripts coupled with the programming interface of the hardware, routine tasks such as controlling the addition of reagents may be automated.[46] Despite the restricted applicability of such an approach, they may help alleviate the researcher of routine operations and afford a greater degree of control. This is exemplified by the breadth of literature on cross-coupling reactions, discovery of new reactivities, and formulation development.[47–49]

Although execution of synthetic procedures has been demonstrated in the examples given, a system to automatically characterise and quantify the results is not yet implemented.

4. Analytical Tools

To close the synthesis prediction and automation loop, we need a system that can automatically characterise and quantify the components present in spectroscopic data.[9] ML shows promise as a core component of predictive analytics tools; however, the rate at which algorithms have been adopted in the analytical domain is comparatively lower than for other areas of chemistry.[50] Despite the plethora of analytical software on the market, only some provide functionality for the automated annotation and quantification of spectral components. Similarly, the integration of automated analysis pipelines without prior calibration or extensive planning into general laboratory automation is rare in synthetic chemistry.[51] As such, characterisation, annotation, and quantification still require a significant amount of manual labor and specialist knowledge, representing a bottleneck. Here, we highlight examples of ML in analytical chemistry, divided into three general areas: processing, prediction, and interpretation of spectra.

4.1 *Processing*

While spectra are routinely processed using both commercial and open-source programs, in many cases human efforts outperform these algorithms making spectral processing a tedious and time-consuming task.[52] In one recent example, Bruker introduced two deep learning algorithms for their TopSpin software. The first is based on a combination of convolutional and recurrent neural networks, allowing the automated correction of both phase information and the baseline of ^{1}H NMR spectra.[52] Using a similar approach, Bruker was also able to automate the peak picking of spectra.[53] Both algorithms outperform commonly used techniques and match human capabilities. In addition, there is a large body of research on processing techniques incorporating ML for other types of spectra.[54,55] However, so far all of these require expert knowledge to use.

4.2 *Prediction and Simulation*

Ab initio methods have traditionally been used for the prediction of spectra. However, recent studies have shown that ML approaches can reach comparable performance on a significantly faster time scale. This has been demonstrated for the prediction of IR, MS and NMR spectra.[56–58] Yet, ML methods are only starting to emerge in products usable by chemists.

At present, Modgraphs NMR Predict provides an approachable software integrating neural networks to predict ^{13}C spectra.[59] Alternatively, the MestreNova plugin NMR Predict provides an ensemble approach, based on Modgraph and Mestrelab for the prediction of ^{1}H and ^{13}C spectra.[60] Both methods are able to predict spectra in minutes, whereas *ab initio* methods often require hours. Methods incorporating ML for other spectra currently remain out of reach as they require expert knowledge to use.

4.3 *Interpretation*

In the field of automated spectra interpretation two schools of thought dominate: either the spectrum is assigned through comparison to other spectra, or by direct interpretation of the spectra, which can be further broken down into 'targeted' and 'untargeted' analysis. Targeted analysis is the most common, and used when there is prior knowledge of the components of the mixture. As such, methods based on comparison, *e.g*. database similarity searching, are commonly used to automatically interpret MS-MS data. A large number of algorithms exist for this purpose with

many offering GUIs or webtools.[61,62] Inroads are also being made in integrating ML into these methods.[63] However, these tools are limited by the database that is being searched and as such are often specialised to a certain domain (*e.g.* metabolomics).

A different comparison-based approach has been used to interpret NMR spectra. This method is based on a user predicted structure whose spectrum is calculated. Using the known experimental and the calculated spectrum, the peaks in the experimental spectrum are assigned. Concurrently, the probability of the experimental spectra matching the proposed structure is calculated. With this method the peaks of the spectra can be assigned but a major shortfall is the need of a user provided structure. MestreNova's Auto Assignment function follows this approach using the above-mentioned spectra prediction, with a similar approach being implemented by Goodman *et al*. in the DP4-AI framework.[64,65] Both tools implement a GUI, however DP4-AI requires a compute cluster and is based around DFT calculations making the use of MestreNova significantly easier for the non-expert user.

Research on the direct interpretation of spectra without requiring a database or the prediction of spectra is more limited. Numerous reports of ML algorithms used to directly predict structure from an NMR or MS-MS spectrum exist.[66–70] Similarly there are reports of the use of ML to predict functional groups from IR-spectra.[71] However, these techniques are still in their infancy with comparatively low accuracies and the accessibility being low.[72,73]

5. Conclusions

Overall the incorporation of ML methods into the synthetic chemistry workflow holds promise. Chemists are starting to use synthetic route design software that relies on ML approaches.[74] For the researcher, the biggest hurdle in the adoption of these methods is their ease of use. Currently, very few predictive models and automation platforms, whether for route design or analytics, provide a GUI. The few that do are typically integrated into commercial offerings at a more mature stage of their development, thus limiting the access researchers have to newly developed methods. While many reports on both automation and analysis can be found in the literature, the advancements so far do not translate into the day-to-day workings of chemists. This can largely be attributed to the amount of initial investment and expert knowledge required in the field of automation, and respectively the lack of accessible tools for analysis.

Nevertheless, efforts to collect and refine data persist and method development in ML is more active than ever. These trends, corroborated with a push from the community towards molecule discovery and synthesis that fit today's demanding timeframes, are setting the scene for a revolution in chemistry laboratories. The existent tools, although imperfect, are already equipping the chemist with valuable assistance, and the authors anticipate that more advancements will likely be transferred into the laboratory in the near future.

Acknowledgements

This publication was created as part of NCCR Catalysis (grant number 180544), a National Centre of Competence in Research funded by the Swiss National Science Foundation.

Received: December 1, 2022

[1] A. F. de Almeida, R. Moreira, T. Rodrigues, *Nat. Rev. Chem*. **2019**, *3*, 589, https://doi.org/10.1038/s41570-019-0124-0.
[2] E. J. Corey, W. T. Wipke, *Science* **1969**, *166*, 178, https://doi.org/10.1126/science.166.3902.178.
[3] R. K. Lindsay, B. G. Buchanan, E. A. Feigenbaum, J. Lederberg, *Artif. Intell*. **1993**, *61*, 209, https://doi.org/10.1016/0004-3702(93)90068-M.
[4] M. H. S. Segler, M. P. Waller, *Chem. Eur. J.* **2017**, *23*, 6118, https://doi.org/10.1002/chem.201604556.
[5] M. H. S. Segler, M. Preuss, M. P. Waller, *Nature* **2018**, *555*, 604, https://doi.org/10.1038/nature25978.
[6] 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, https://doi.org/10.1039/C9SC05704H.
[7] J. Shen, C. A. Nicolaou, *Drug Discov. Today: Technol*. **2019**, *32–33*, 29, https://doi.org/10.1016/j.ddtec.2020.05.001.
[8] P. Schwaller, A. C. Vaucher, T. Laino, J.-L. Reymond, *Mach. Learn.: Sci. Technol*. **2021**, *2*, 015016, https://doi.org/10.1088/2632-2153/abc81d.
[9] B. Debus, H. Parastar, P. Harrington, D. Kirsanov, *TrAC Trends Anal. Chem*. **2021**, *145*, 116459, https://doi.org/10.1016/j.trac.2021.116459.
[10] Z. Wang, W. Zhang, B. Liu, *Chin. J. Chem*. **2021**, *39*, 3127, https://doi.org/10.1002/cjoc.202100273.
[11] Y. Shen, J. E. Borowski, M. A. Hardy, R. Sarpong, A. G. Doyle, T. Cernak, *Nat. Rev. Meth. Primers* **2021**, *1*, 1, https://doi.org/10.1038/s43586-021-00022-5.
[12] S. Szymkuć, E. P. Gajewska, T. Klucznik, K. Molga, P. Dittwald, M. Startek, M. Bajczyk, B. A. Grzybowski, *Angew. Chem. Int. Ed*. **2016**, *55*, 5904, https://doi.org/10.1002/anie.201506101.
[13] S. Szymkuć, T. Badowski, B. A. Grzybowski, *Angew. Chem. Int. Ed*. **2021**, *60*, 26226, https://doi.org/10.1002/anie.202111540.
[14] C. W. Coley, W. H. Green, K. F. Jensen, *J. Chem. Inf. Model*. **2019**, *59*, 2529, https://doi.org/10.1021/acs.jcim.9b00286.
[15] P. Schwaller, T. Laino, T. Gaudin, P. Bolgar, C. A. Hunter, C. Bekas, A. A. Lee, *ACS Centr. Sci*. **2019**, *5*, 1572, https://doi.org/10.1021/acscentsci.9b00576.
[16] Y. Mo, Y. Guan, P. Verma, J. Guo, M. E. Fortunato, Z. Lu, C. W. Coley, K. F. Jensen, *Chem. Sci*. **2021**, *12*, 1469, https://doi.org/10.1039/D0SC05078D.
[17] K. Lin, Y. Xu, J. Pei, L. Lai, *Chem. Sci*. **2020**, *11*, 3355, https://doi.org/10.1039/C9SC03666K.
[18] T. Badowski, K. Molga, B. A. Grzybowski, *Chem. Sci.* **2019**, *10*, 4640, https://doi.org/10.1039/C8SC05611K.
[19] C. W. Coley, D. A. Thomas, J. A. M. Lummiss, J. N. Jaworski, C. P. Breen, V. Schultz, T. Hart, J. S. Fishman, L. Rogers, H. Gao, R. W. Hicklin, P. P. Plehiers, J. Byington, J. S. Piotti, W. H. Green, A. J. Hart, T. F. Jamison, K. F. Jensen, *Science* **2019**, *365*, eaax1566, https://doi.org/10.1126/science.aax1566.
[20] S. Genheden, A. Thakkar, V. Chadimová, J.-L. Reymond, O. Engkvist, E. Bjerrum, *J. Cheminf*. **2020**, *12*, 70, https://doi.org/10.1186/s13321-020-00472-1.
[21] ASKCOS Homepage, https://askcos.mit.edu/, accessed November 20, 2022.
[22] N. Ree, A. H. Göller, J. H. Jensen, *Digital Discov*. **2022**, *1*, 108, https://doi.org/10.1039/D1DD00032B.
[23] Y. Guan, C. W. Coley, H. Wu, D. Ranasinghe, E. Heid, T. J. Struble, L. Pattanaik, W. H. Green, K. F. Jensen, *Chem. Sci*. **2021**, *12*, 2198, https://doi.org/10.1039/D0SC04823B.
[24] IBM RXN, https://rxn.res.ibm.com/index.html, accessed November 20, 2022.
[25] Y. Shi, P. L. Prieto, T. Zepel, S. Grunert, J. E. Hein, *Acc. Chem. Res*. **2021**, *54*, 546, https://doi.org/10.1021/acs.accounts.0c00736.
[26] C. W. Coley, N. S. Eyke, K. F. Jensen, *Angew. Chem. Int. Ed*. **2020**, *59*, 22858, https://doi.org/10.1002/anie.201909987.
[27] C. W. Coley, N. S. Eyke, K. F. Jensen, *Angew. Chem. Int. Ed*. **2020**, *59*, 23414, https://doi.org/10.1002/anie.201909989.
[28] J. Mavračić, C. J. Court, T. Isazawa, S. R. Elliott, J. M. Cole, *J. Chem. Inf. Model.* **2021**, *61*, 4280, https://doi.org/10.1021/acs.jcim.1c00446.
[29] L. Hawizy, D. M. Jessop, N. Adams, P. Murray-Rust, *J. Cheminf*. **2011**, *3*, 17, https://doi.org/10.1186/1758-2946-3-17.
[30] S. Steiner, J. Wolf, S. Glatzel, A. Andreou, J. M. Granda, G. Keenan, T. Hinkley, G. Aragon-Camarasa, P. J. Kitson, D. Angelone, L. Cronin, *Science* **2019**, *363*, eaav2211, https://doi.org/10.1126/science.aav2211.
[31] ChemIDE, https://croningroup.gitlab.io/chemputer/xdlapp/, accessed November 20, 2022.
[32] A. C. Vaucher, F. Zipoli, J. Geluykens, V. H. Nair, P. Schwaller, T. Laino, *Nat. Commun*. **2020**, *11*, 3601, https://doi.org/10.1038/s41467-020-17266-6.
[33] A. C. Vaucher, P. Schwaller, J. Geluykens, V. H. Nair, A. Iuliano, T. Laino, *Nat. Commun.* **2021**, *12*, 2573, https://doi.org/10.1038/s41467-021-22951-1.
[34] S. V. Ley, D. E. Fitzpatrick, Richard. J. Ingham, R. M. Myers, *Angew. Chem. Int. Ed.* **2015**, *54*, 3449, https://doi.org/10.1002/anie.201410744.
[35] E. S. Isbrandt, R. J. Sullivan, S. G. Newman, *Angew. Chem. Int. Ed.* **2019**, *58*, 7180, https://doi.org/10.1002/anie.201812534.
[36] M. Seifrid, R. Pollice, A. Aguilar-Granda, Z. Morgan Chan, K. Hotta, C. T. Ser, J. Vestfrid, T. C. Wu, A. Aspuru-Guzik, *Acc. Chem. Res*. **2022**, *55*, 2454, https://doi.org/10.1021/acs.accounts.2c00220.
[37] C. P. Breen, A. M. K. Nambiar, T. F. Jamison, K. F. Jensen, *Trends Chem*. **2021**, *3*, 373, https://doi.org/10.1016/j.trechm.2021.02.005.
[38] C. Mateos, M. J. Nieves-Remacha, J. A. Rincón, *React. Chem. Eng*. **2019**, *4*, 1536, https://doi.org/10.1039/C9RE00116F.

[39] M. Christensen, L. P. E. Yunker, P. Shiri, T. Zepel, P. L. Prieto, S. Grunert, F. Bork, J. E. Hein, *Chem. Sci*. **2021**, *12*, 15473, https://doi.org/10.1039/D1SC04588A.
[40] A. J. S. Hammer, A. I. Leonov, N. L. Bell, L. Cronin, *JACS Au* **2021**, *1*, 1572, https://doi.org/10.1021/jacsau.1c00303.
[41] J. Bai, L. Cao, S. Mosbach, J. Akroyd, A. A. Lapkin, M. Kraft, *JACS Au* **2022**, *2*, 292, https://doi.org/10.1021/jacsau.1c00438.
[42] C. Elliott, V. Vijayakumar, W. Zink, R. Hansen, *JALA: J. Assoc. Lab. Autom*. **2007**, *12*, 17, https://doi.org/10.1016/j.jala.2006.07.012.
[43] L. M. Roch, F. Häse, C. Kreisbeck, T. Tamayo-Mendoza, L. P. E. Yunker, J. E. Hein, A. Aspuru-Guzik, *PLOS one* **2020**, *15*, e0229862, https://doi.org/10.1371/journal.pone.0229862.
[44] S. Rohrbach, M. Šiaučiulis, G. Chisholm, P.-A. Pirvan, M. Saleeb, S. H. M. Mehr, E. Trushina, A. I. Leonov, G. Keenan, A. Khan, A. Hammer, L. Cronin, *Science* **2022**, *377*, 172, https://doi.org/10.1126/science.abo0058.
[45] M. Bornemann-Pfeiffer, J. Wolf, K. Meyer, S. Kern, D. Angelone, A. Leonov, L. Cronin, F. Emmerling, *Angew. Chem. Int. Ed.* **2021**, *60*, 23202, https://doi.org/10.1002/anie.202106323.
[46] G. R. D. Prabhu, T.-H. Yang, C.-Y. Hsu, C.-P. Shih, C.-M. Chang, P.-H. Liao, H.-T. Ni, P. L. Urban, *Nat. Protoc*. **2020**, *15*, 925, https://doi.org/10.1038/s41596-019-0272-1.
[47] D. J. Blair, S. Chitti, M. Trobe, D. M. Kostyra, H. M. S. Haley, R. L. Hansen, S. G. Ballmer, T. J. Woods, W. Wang, V. Mubayi, M. J. Schmidt, R. W. Pipal, Greg. F. Morehouse, A. M. E. Palazzolo Ray, D. L. Gray, A. L. Gill, M. D. Burke, *Nature* **2022**, *604*, 92, https://doi.org/10.1038/s41586-022-04491-w.
[48] D. Caramelli, J. M. Granda, S. H. M. Mehr, D. Cambié, A. B. Henson, L. Cronin, *ACS Cent. Sci*. **2021**, *7*, 1821, https://doi.org/10.1021/acscentsci.1c00435.
[49] L. Cao, D. Russo, K. Felton, D. Salley, A. Sharma, G. Keenan, W. Mauer, H. Gao, L. Cronin, A. A. Lapkin, *Cell Rep. Phys. Sci*. **2021**, *2*, 100295, https://doi.org/10.1016/j.xcrp.2020.100295.
[50] L. B. Ayres, F. J. V. Gomez, J. R. Linton, M. F. Silva, C. D. Garcia, *Anal. Chim. Acta* **2021**, *1161*, 338403, https://doi.org/10.1016/j.aca.2021.338403.
[51] C. Houben, A. A. Lapkin, *Curr. Opin. Chem. Eng*. **2015**, *9*, 1, https://doi.org/10.1016/j.coche.2015.07.001.
[52] S. Bruderer, F. Paruzzo, C. Bolliger, 'Deep learning-based phase and baseline correction of 1D ^{1}H NMR Spectra', Bruker.
[53] F. Paruzzo, S. Bruderer, Y. Janjar, B. Heitmann, C. Bolliger, 'Automatic Signal Region Detection in ^{1}H NMR Spectra Using Deep Learning', Bruker, **2020**.
[54] Z. A. Al-Mualem, C. R. Baiz, *J. Phys. Chem. A* **2022**, *126*, 3816, https://doi.org/10.1021/acs.jpca.2c02605.
[55] U. W. Liebal, A. N. T. Phan, M. Sudhakar, K. Raman, L. M. Blank, *Metabolites* **2020**, *10*, 243, https://doi.org/10.3390/metabo10060243.
[56] C. McGill, M. Forsuelo, Y. Guan, W. H. Green, *J. Chem. Inf. Model*. **2021**, *61*, 2594, https://doi.org/10.1021/acs.jcim.1c00055.
[57] F. Allen, R. Greiner, D. Wishart, *Metabolomics* **2015**, *11*, 98, https://doi.org/10.1007/s11306-014-0676-4.
[58] E. Jonas, S. Kuhn, *J. Cheminf*. **2019**, *11*, 50, https://doi.org/10.1186/s13321-019-0374-3.
[59] Modgraph C13 NMR Prediction: CSEARCH, http://www.modgraph.co.uk/product_nmr_network.htm, accessed November 21, 2022.
[60] Ensemble NMR Prediction, https://resources.mestrelab.com/ensemble-nmr-prediction/, accessed November 21, 2022.
[61] K. Dührkop, M. Fleischauer, M. Ludwig, A. A. Aksenov, A. V. Melnik, M. Meusel, P. C. Dorrestein, J. Rousu, S. Böcker, *Nat. Meth*. **2019**, *16*, 299, https://doi.org/10.1038/s41592-019-0344-8.
[62] K. Dührkop, H. Shen, M. Meusel, J. Rousu, S. Böcker, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 12580, https://doi.org/10.1073/pnas.1509788112.
[63] K. Scheubert, F. Hufsky, S. Böcker, *J. Cheminf*. **2013**, *5*, 12, https://doi.org/10.1186/1758-2946-5-12.
[64] ^{1}H NMR Auto-assignments, https://resources.mestrelab.com/auto-assignments/, accessed November 21, 2022.
[65] A. Howarth, K. Ermanis, J. M. Goodman, *Chem. Sci*. **2020**, *11*, 4351, https://doi.org/10.1039/D0SC00442A.
[66] C. Zhang, Y. Idelbayev, N. Roberts, Y. Tao, Y. Nannapaneni, B. M. Duggan, J. Min, E. C. Lin, E. C. Gerwick, G. W. Cottrell, W. H. Gerwick, *Sci. Rep*. **2017**, *7*, 14243, https://doi.org/10.1038/s41598-017-13923-x.
[67] J. Zhang, K. Terayama, M. Sumita, K. Yoshizoe, K. Ito, J. Kikuchi, K. Tsuda, *Sci. Technol. Adv. Mater*. **2020**, *21*, 552, https://doi.org/10.1080/14686996.2020.1793382.
[68] Z. Huang, M. S. Chen, C. P. Woroch, T. E. Markland, M. W. Kanan, *Chem. Sci*. **2021**, *12*, 15329, https://doi.org/10.1039/D1SC04105C.
[69] M. A. Stravs, K. Dührkop, S. Böcker, N. Zamboni, *Nat. Meth*. **2022**, *19*, 865, https://doi.org/10.1038/s41592-022-01486-3.
[70] A. A. Bondarenko, Y. S. Vlasova, M. V. Polynski, V. V. Ilyushenkova, V. P. Ananikov, *Inorg. Chem. Front*. **2022**, *9*, 5279, https://doi.org/10.1039/D2QI01098D.
[71] A. A. Enders, N. M. North, C. M. Fensore, J. Velez-Alvarez, H. C. Allen, *Anal. Chem*. **2021**, *93*, 9711, https://doi.org/10.1021/acs.analchem.1c00867.
[72] E. Szymańska, *Anal. Chim. Acta* **2018**, *1028*, 1, https://doi.org/10.1016/j.aca.2018.05.038.
[73] 'Theory and Practical Applications of Automation Solutions in Analytical Measurements', http://www.americanlaboratory.com/914-Application-Notes/360982-Theory-and-Practical-Applications-of-Automation-Solutions-in-Analytical-Measurements/, accessed November 23, 2022.
[74] J. P. Janet, A. Tomberg, J. Boström, *Nat. Mach. Intell*. **2021**, *3*, 572, https://doi.org/10.1038/s42256-021-00367-2.

License and Terms

The definitive version of this article is the electronic one that can be found at https://doi.org/10.2533/chimia.2023.17