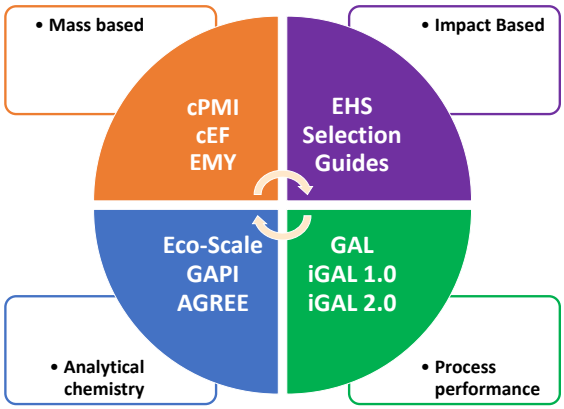
Graphical Abstract

Software tools for green and sustainable chemistry

Ivan N. Derbenev, James Dowden, Jamie Twycross, Jonathan D. Hirst



Highlights

Software tools for green and sustainable chemistry

Ivan N. Derbenev, James Dowden, Jamie Twycross, Jonathan D. Hirst

- The need for universal green chemistry metrics and tools
- Fusion of computational and laboratory chemistry software
- Importance of solvent selection tools
- Greener retrosynthesis

Software tools for green and sustainable chemistry

Ivan N. Derbenev[a,b], James Dowden[b], Jamie Twycross[c], Jonathan D. Hirst[b]

[a]*Digital Research Service, University of Nottingham, Jubilee Campus, Nottingham, NG8 1BB, UK*
[b]*School of Chemistry, University of Nottingham, University Park, Nottingham, NG7 2RD, UK*
[c]*School of Computer Science, University of Nottingham, Jubilee Campus, Nottingham, NG8 1BB, UK*

Abstract

In this review, we consider green chemistry metrics, related software tools, and the opportunities and challenges for their use in research laboratories. We provide an overview of state-of-the-art software designed both to aid researchers in planning and conducting chemical experiments and to assess sustainability of individual reactions and synthetic routes. The increasing digitalisation of research means that there is great opportunity for more extensive use of computational tools by synthetic chemists and for closer integration of green chemistry principles into the routine work of chemical laboratories. We discuss the scope for using software tools in the laboratory and assisting synthetic chemists in the adoption of green and sustainable chemistry approaches that are suitable for their specific purposes.

Keywords: sustainable chemistry, green chemistry metrics, research software, CHEM21

PACS: 0000, 1111

2000 MSC: 0000, 1111

1. Introduction

Today, green chemistry is an increasingly innovative and interdisciplinary research area [1]. One of the main challenges is the evaluation of the greenness of chemical processes [2] and setting this in the broader context of life cycle analysis [3], for which there are a variety of assessment tools. Various metrics of different types (see Figure 1) are used to quantify the environmental, economic, and health and safety ramifications of chemical processes. The choice of metrics mostly depends on personal preference or application context, which means that a universal set of metrics suitable for pharmaceutical companies, manufacturing organisations, or university laboratories does not currently exist [4]. However, several methods and toolkits are available for specific user needs [5]. These tools rely on well-known metrics, such as process mass intensity (PMI), environmental factor (E-factor), yield, etc., to assess whether a chemical process is efficient and the chemicals involved are safe or usage of hazardous chemicals is minimised. Furthermore, new metrics have been introduced, for example, in analytical chemistry: Eco-Scale, GAPI (green analytical procedure index), AGREE (analytical greenness calculator metric) [2], or in pharmaceutical chemistry: iGAL (innovation green aspiration level) [6], AMGS (Analytical Method Greenness Score) [7], eco-label [8].

Another challenge is the development of software tools to help synthetic chemists assess the sustainability and greenness of chemical synthesis. Some of this software has a potential to be adopted by chemists in academia and industry if it can provide real advantage [9]. For example, some electronic laboratory notebooks (ELNs) do contain a section for green chemistry met-

rics. However, their visibility may be limited amongst numerous other features. Moreover, only a small minority of researchers use ELNs, whereas most chemists still prefer paper laboratory notebooks [10, 11]. The reasons for not adopting ELNs include cost, usability, and accessibility issues across different devices and operating systems [12]. Even if these barriers are overcome and chemists are able to calculate various green chemistry metrics with the help of ELNs, this would not guarantee the choice of more sustainable synthetic routes. Knowledge of those metrics is only the starting point for further analysis of individual synthetic steps and, more importantly, for decision making on possible solutions to improve the reaction route or (ideally) to choose the greenest one. For this purpose, chemists can use retrosynthetic software that exploits artificial intelligence (AI) along with extensive databases of known chemical reactions to navigate alternative reaction sequences (Y Lin et al., chemrxiv doi: 10.26434/chemrxiv.12765410.v1). Despite the improvement of machine learning methods in chemistry, these software tools are still largely in the domain of computational chemists rather than researchers in laboratories [11, 13]. In the following sections, we survey some of the existing software that could be useful for implementing green chemistry approaches in a research laboratory.

2. Predictors of green metrics: standalone and within ELNs

In this review, we have separated software for calculating green chemistry metrics into two types. The first type includes tools that use statistical analysis or machine learning techniques to estimate green metrics when reaction data are incomplete or when we need to assess a reaction route. These tools

require an accessible database of known reaction data [14]. Despite some initiatives to create an open-access database of organic reactions from the global research community [15, 16], they are still more limited than commercially available solutions [17]. Aside from the issue of data availability, the existing predictors of green metrics often focus only on one specific metric. The American Chemical Society Green Chemistry Institute Pharmaceutical Roundtable (ACS GCIPR) has developed a web-based toolbox [18] for predicting a set of the following green metrics for a given synthetic route: cumulative PMI (cPMI) [19], iGAL [6], AMGS [7], and selection tools discussed in Section 4. This software calculates a distribution of cPMI for a given synthetic route using a Monte Carlo procedure to draw samples from a uniform distribution for stoichiometry ranges, and from a bivariate normal distribution for PMI and yield. Here, the PMI metric is upgraded to allow route comparison at the early stage of drug design. The ACS GCIPR toolbox also includes a reagent selection guide [20], where one can visualise reagents in a Venn diagram constructed based on the criteria of wide utility, scalability and greenness. The ACS GCIPR Biocatalysis Guide [21] shows the most used biocatalysis transformations, which can inform their consideration in retrosynthetic analyses.

The second type of software represents tools that calculate green metrics directly from input data. This approach is implemented in some ELNs that include a green chemistry section. Chemotion is an open source ELN supporting the acquisition, storage and management of chemical data [22]. In addition to standard ELN features of processing molecules and reactions, there is also an option to calculate mass-based metrics such as E-factor,

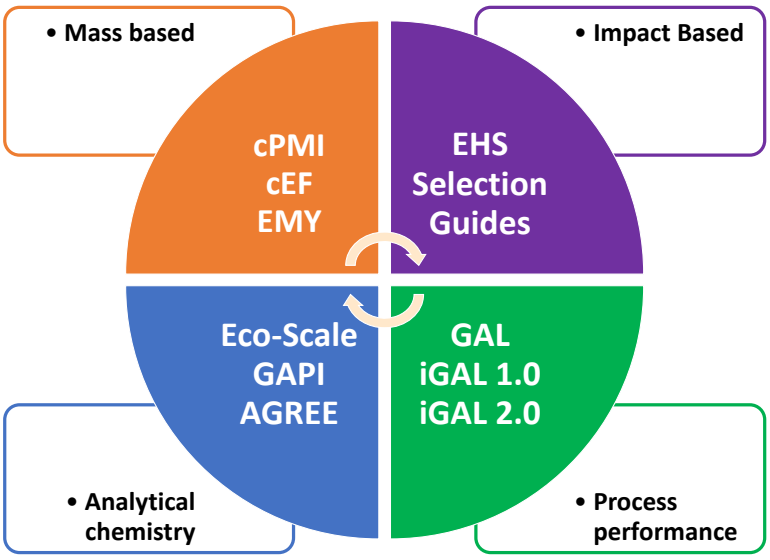


Figure 1: Types of green chemistry metrics with examples: cPMI (cumulative process mass intensity [19]), cEF (complete environmental factor [6]), EMY (effective mass yield [19]), EHS (environment, health, and safety [19]), GAPI (green analytical procedure index [2]), AGREE (analytical greenness calculator metric [2]), iGAL (innovation green aspiration level [6])

PMI, and atom economy. Other tools are specifically focused on green metric calculation, e.g., CHEM21 Toolkit [23]. These tools allow one to calculate metrics for a single reaction; some of them also propose impact base metrics (e.g., solvent colour codes in CHEM21). CHEM21, for example, is being used by chemists, with benefit, in the "Accelerated Discovery and Development of New Medicines: Prosperity Partnership for a Healthier Nation" project [24]. However, they are difficult to apply to synthetic route assessment or to reactions with incomplete data.

In summary, current research and development activities are mostly focused on the the standalone predictors and direct calculators of green metrics are underrepresented within existing ELNs. This may be for several reasons. First, a lot of reaction data has been accumulated over the last several decades, requiring new computational techniques to process the data [14]. Second, exploiting these chemical data requires the development of new machine learning approaches [25]. Finally, the lack of green chemistry sections in ELNs is explained by low popularity of ELNs in general, as discussed in the introduction [11]. Future perspectives of this software are discussed in Section 6. The metrics predictors described in this section are presented in Table 1.

3. Yield predictors

Reaction yield is a measure that can only be calculated after the completion of a reaction. In patent data from 1976-2016 ($\sim$1 million unique reactions), only 53% of all reactions report a yield value either directly in the patent text or calculated from the reported data [26]. However, the number of

Tool	Description
PMI Predictor [19]	Calculates Process Mass Intensity (PMI) for a chemical synthesis route using the sequence of synthetic steps and step information https://acsgcipr-predictpmi.shinyapps.io/pmi_calculator/
iGAL 2.0 Scorecard Calculator [6]	Produces the scorecard output for an active pharmaceutical ingredient (API) using the model of innovative green aspiration level - iGAL 2.0 https://www.acsgcipr.org/tools-for-innovation-in-chemistry/green-chemistry-innovation-scorecard-calculator-igal/
AMGS Calculator [7]	Calculates an analytical method greenness score (AMGS) to enable the comparison of separation methods used in drug development https://www.acsgcipr.org/amgs/
Chemotion [22]	Open Source electronic lab notebook for researchers with a green chemistry tab https://github.com/ComPlat/chemotion_ELN
CHEM21 toolkit [23]	Unified metrics toolkit evaluating sustainability of reactions, encompassing a comprehensive and holistic range of criteria for measuring how green a reaction is, covering quantitative and qualitative criteria both upstream and downstream of the reaction itself

Table 1: Green metric predictors and electronic lab notebooks.

reactions with reliable yield data is even less, due to a discrepancy of $> 10\%$ between text-mined and calculated yield values [26]. Thus, yield-predicting

software can help overcome these challenges and open up new perspectives in chemical synthesis planning.

In addition to traditional regression-based tools for reaction analysis, new methods based on a range of machine learning approaches have been introduced [27]. Random forest algorithms using quantum chemistry descriptors as inputs and reaction yield as output demonstrated better predictive performance than linear regression methods. However, the choice of molecular descriptors can significantly improve the performance of support vector regression (SVR) models [28]. In particular, the SVR models built on molecular fingerprints outperformed the models based on quantum chemical descriptors and on molecular graphs. Deep learning methods based on SMILES input can predict the yield for various types of reactions with up to 99% accuracy without the use of molecular graphs, molecular fingerprints or quantum chemistry descriptors [29].

Despite these promising results, machine learning tools can fail to predict patent reaction yields accurately if input data are of low quality. The challenge is that the reported yields are heterogeneous and inconsistent due to the large variability of methods of purification and reporting yields and the different optimisation in each reported reaction [30]. The yield predictors described in this section are collected in Table 2.

4. Solvent selection

In addition to software that predicts mass-based metrics, there are also some tools assessing the environmental impact of chemical reactions and compounds. These tools can help in selecting more sustainable reaction

Tool	Description
rxnpredict [27]	Predicts yield of a palladium-catalyzed Buchwald-Hartwig cross-coupling of aryl halides with 4-methylaniline in the presence of various potentially inhibitory additives using random forest algorithms with quantum chemical descriptors https://github.com/doylelab/rxnpredict
yield_prediction [28]	Predicts yield of Buchwald-Hartwig amination reactions using support vector regression with structure-based descriptors https://github.com/alexehaywood/yield_prediction
rxn_yield [29]	Predict yield applying natural language processing architectures to reaction SMILES representation, using an encoder transformer model combined with a regression layer https://rxn4chemistry.github.io/rxn_yields/

Table 2: Reaction yield prediction software.

conditions or greener reagents [21]. Solvents represent at least half of the material mass used in chemical industry. Hence, solvent selection tools can assist in reducing the overall environmental impact of active pharmaceutical ingredients, for example [31]. Quite a few solvent selection methods have been developed and some of them are integrated in software tools [31, 32, 33].

The common principle of the existing software is based on classification of and search for greener alternatives in solvent space [34]. For example, the Sustainable Solvents Selection and Substitution Software (SUSSOL) uses a neural network (Self-organizing Map of Kohonen) to cluster a solvent

database based on the physical properties of the solvents [35]. A green-solvent selection tool developed for printed electronics organizes a large set of solvents according to their Hansen solubility parameters (the dispersion, polar, and hydrogen-bonding energies), ink properties, and sustainability descriptors [36]. This software also introduces a composite score value G - a new metric involving scores of health, safety, environment, and waste disposal categories from the GSK solvent selection guide [31]. The existing solvent selectors rely on the experimentally established greenness of solvents (as implemented in CHEM21 [31] or ACS GCIPR Solvent Tool [37]). If a solvent is novel, then we have no tool to categorize it as green or otherwise. The same is related to reaction conditions in general. Even though existing software can predict reaction conditions [5], they do not provide an assessment of greenness. The solvent selection tools from this section are summarized in Table 3.

5. Retrosynthetic analysis

Retrosynthesis is challenging due to the enormous search space. At least 10^7 reactions are contained in commercially available data sets (C Yan et al., arXiv doi: 2011.02893). Therefore, computer-aided synthesis planning (CASP) can suggest synthetic routes to a desired product from available chemicals through retrosynthesis [38, 39]. The development of software implementing retrosynthetic analysis is of significant interest, but there is a gap between the computationally generated fragments and available reagents. To bridge this gap, one can use software like RetroXpert (C Yan et al., arXiv doi: 2011.02893) or SynthI [40] that generates synthons associated with actual

Tool	Description
SUSSOL [35]	Processes solvents by a neural network, the Self-organizing Map of Kohonen, which results in a 2D map of clusters https://github.com/SUSSOLKDG/Sussol
Green Solvent Selection Tool [36]	Organizes a large set of solvents according to their Hansen solubility parameters, ink properties, and sustainability descriptors, and suggests green alternative solvents with similar dissolution capacity as the current non-sustainable solvent http://www.opeg-umu.se/green-solvent-tool
CHEM21 solvent selection guide [31]	Ranks solvents using a set of Safety, Health and Environment criteria is proposed, aligned with the Global Harmonized System (GHS) and European regulations
ACS GCIPR Solvent Tool [37]	An interactive tool based on consideration of chemical functionality, physical properties, regulatory concerns, and safety/health/environmental (SHE) impact https://www.acsgcipr.org/tools-for-innovation-in-chemistry/solvent-tool/

Table 3: Solvent selectors.

fragments or equivalent reactants. Another example is the approach of Lin et al. [41], where an attention-based machine translation model Molecular Transformer [42] is used to build retrosynthetic routes for 12 investigational COVID-19 therapeutics. A different algorithm underpins the AiZynthFinder software which is based on Monte Carlo tree search (MCTS) guided by an artificial neural network to suggest purchasable precursors for a target molecule

[43]. This software and other similar tools (ASKCOS [44], LillyMol [44] etc.) do not provide holistic green chemistry analysis [45]. Only in the work of Christensen et al. [45] is each reaction in the routes proposed by MCTS evaluated by the "greenness" of neural network predicted solvents [46]. Table 4 contains the retrosynthetic software discussed here.

6. Future perspectives

As emphasized in Section 2 above, one of the key challenges in software development for green chemistry is the availability of reaction data. More broadly, a significant barrier to the development and application of tools is that much chemical information is not machine readable or curated in a usable form. Initiatives to create a global open-access reaction database can enable the development of new green chemistry tools and liberate researchers from commercially curated data. Open software, in turn, will enable chemists in research laboratories to incorporate green chemistry methods for planning and analysing their experiments. There may no single metric or single tool that is enough for a complete assessment of greenness or sustainability. Nevertheless, integrating green metric predictors and solvent selectors into ELNs can also make digital tools more attractive for experimental chemists. However, the limitations of integrated metrics in this software should be taken into account to avoid biased decision making. For example, a synthetic route may seem satisfactory according to the metrics from a certain software, but this route may have toxic intermediate products not detected by the software. Therefore, it is chemists who should make an informed interpretation of metrics to plan chemical synthesis. There are many tools to help with tox-

Tool	Description
RetroXpert (C Yan et al., arXiv doi: 2011.02893)	Identifies the potential reaction centre of the target molecule through a graph neural network, generates intermediate synthons and the reactants associated with synthons https://github.com/uta-smile/RetroXpert
SynthI [40]	Merges the space of the retrosynthetically generated fragments and the pool of available reagents into a single synthons space https://github.com/Laboratoire-de-Chemoinformatique/SynthI
covidroutes [41]	Presents predicted retrosynthetic routes to 12 diverse COVID-19 therapeutic candidates http://covidroutes.cernaklab.com/
AiZynthFinder [43]	Uses a Monte Carlo tree search guided by an artificial neural network to recursively break down a molecule to purchasable precursors http://www.github.com/MolecularAI/aizynthfinder
ASKCOS [44]	Predicts feasible synthetic routes towards a desired compound and associated tasks related to synthesis planning https://github.com/ASKCOS/ASKCOS
LillyMol [44]	Uses reaction transformation rules learned from a large patent reaction dataset https://github.com/EliLillyCo/LillyMol

Table 4: Retrosynthetic software.

icity and environmental fate prediction, some of which have been reviewed recently [47, 48]. In general, automation of chemical experiments provides new perspectives on integration of computer science, machine vision, analytical software, and robotics [45]. Proprietary data present a particular challenge. Recent work [49] has considered federated privacy-preserving machine learning methods in drug design with an appropriate split of chemical data between training, validation and test set as a means of circumventing difficulties in working with commercially sensitive data. On the other hand, usage of social media for the exchange of open chemical data and integration with chemistry software tools [50] may be an enabling step towards new quality of research across the globe, especially if online communication continues to prevail over face-to-face meetings.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) [grant number EP/S035990/1]. J.D.H. is supported by the Royal Academy of Engineering under the Chairs in Emerging Technologies scheme. We thank Professor Pete Licence, Director of the GlaxoSmithKline Carbon Neutral Laboratory at the University of Nottingham, and our research co-workers Chris Handley, Samuel Boobier, Joe Davies, Zakariyyaa

Siddiq, and Aimie Garces for many stimulating discussions.

References

[1] D. J. Constable, Green and sustainable chemistry – the case for a systems-based, interdisciplinary approach, iScience 24 (2021) 103489. doi:10.1016/j.isci.2021.103489.

[2] M. Sajid, J. Płotka-Wasylka, Green analytical chemistry metrics: A review, Talanta 238 (2022) 123046. doi:10.1016/j.talanta.2021.123046.

[3] C. Jiménez-González, D. J. C. Constable, C. S. Ponder, Evaluating the "greenness" of chemical processes and products in the pharmaceutical industry—a green metrics primer, Chem. Soc. Rev. 41 (2012) 1485–1498. doi:10.1039/C1CS15215G.

[4] L. Rogers, K. F. Jensen, Continuous manufacturing – the green chemistry promise, Green Chem. 21 (2019) 3481–3498. doi:10.1039/C9GC00773C.

[5] R. A. Sheldon, Metrics of green chemistry and sustainability: Past, present, and future, ACS Sustain. Chem. Eng. 6 (2018) 32–48. doi:10.1021/acssuschemeng.7b03505.

[6] F. Roschangar, J. Li, Y. Zhou, W. Aelterman, A. Borovika, J. Colberg, D. P. Dickson, F. Gallou, J. D. Hayler, S. G. Koenig, M. E. Kopach, B. Kosjek, D. K. Leahy, E. O'Brien, A. G. Smith, M. Henry, J. Cook, R. A. Sheldon, Improved iGAL 2.0 metric empowers pharmaceutical

scientists to make meaningful contributions to United Nations sustainable development goal 12, ACS Sustain. Chem. Eng. (2022 (in press)). doi:10.1021/acssuschemeng.1c01940.

[7] M. B. Hicks, W. Farrell, C. Aurigemma, L. Lehmann, L. Weisel, K. Nadeau, H. Lee, C. Moraff, M. Wong, Y. Huang, P. Ferguson, Making the move towards modernized greener separations: introduction of the analytical method greenness score (AMGS) calculator, Green Chem. 21 (2019) 1816–1826. doi:10.1039/C8GC03875A.

[8] U. Onken, A. Koettgen, H. Scheidat, P. Schueepp, F. Gallou, Environmental metrics to drive a cultural change: Our green eco-label, Chimia 73 (2019) 730–736. doi:doi:10.2533/chimia.2019.730.

[9] M. Poliakoff, J. M. Fitzpatrick, T. R. Farren, P. T. Anastas, Green chemistry: Science and politics of change, Science 297 (2002) 807–810. doi:10.1126/science.297.5582.807.

[10] T. Kranjc, Introduction to Laboratory Software Solutions and Differences Between Them, John Wiley & Sons, Ltd, 2021, Ch. 3, pp. 75–84. doi:10.1002/9783527825042.ch3.

[11] S. Kanza, C. Willoughby, N. Gibbins, R. Whitby, J. G. Frey, J. Erjavec, K. Zupančič, M. Hren, K. Kovač, Electronic lab notebooks: can they replace paper, J. Cheminformatics 9 (2017) 1–15.

[12] B. Gerlach, C. Untucht, A. Stefan, Good Research Practice in Non-Clinical Pharmacology and Biomedicine, Springer Nature, 2020, Ch.

Electronic Lab Notebooks and Experimental Design Assistants, pp. 257–275. doi:10.1007/10.1007/978-3-030-33656-1.

[13] S. Kanza, C. L. Bird, M. Niranjan, W. McNeill, J. G. Frey, The AI for scientific discovery network+, Patterns 2 (2021) 100162. doi:10.1016/j.patter.2020.100162.

[14] P. Baldi, Call for a public open database of all chemical reactions, J Chem Inf Model (2021). doi:10.1021/acs.jcim.1c01140.

[15] D. Lowe, Chemical reactions from US patents (1976-sep2016) (Jun 2017). doi:10.6084/m9.figshare.5104873.v1.

[16] S. M. Kearnes, M. R. Maser, M. Wleklinski, A. Kast, A. G. Doyle, S. D. Dreher, J. M. Hawkins, K. F. Jensen, C. W. Coley, The open reaction database, J. Am. Chem. Soc. 143 (2021) 18820–18826. doi:10.1021/jacs.1c09820.

An open reaction database is proposed along with data, schema, supporting code, and a web user interface. This database has a consistent data representation and infrastructure to ease data sharing and training of machine-learning models.

[17] A. Thakkar, T. Kogej, J.-L. Reymond, O. Engkvist, E. J. Bjerrum, Datasets and their influence on the development of computer assisted synthesis planning tools in the pharmaceutical domain, Chem. Sci. 11 (2020) 154–168. doi:10.1039/C9SC04944D.

[18] L. J. Diorazio, P. Richardson, H. F. Sneddon, A. Moores, C. Briddell, I. Martinez, Making sustainability assessment accessible: Tools developed by the ACS green chemistry institute pharmaceutical roundtable, ACS Sustain. Chem. Eng. 9 (2021) 16862–16864. doi:10.1021/acssuschemeng.1c07651.

[19] A. Borovika, J. Albrecht, J. Li, A. S. Wells, C. Briddell, B. R. Dillon, L. J. Diorazio, J. R. Gage, F. Gallou, S. G. Koenig, et al., The PMI predictor app to enable green-by-design chemical synthesis, Nat. Sustain. 2 (2019) 1034–1040. doi:10.1038/s41893-019-0400-5.

[20] S. G. Koenig, C. Bee, A. Borovika, C. Briddell, J. Colberg, G. R. Humphrey, M. E. Kopach, I. Martinez, S. Nambiar, S. V. Plummer, S. D. Ribe, F. Roschangar, J. P. Scott, H. F. Sneddon, A green chemistry continuum for a robust and sustainable active pharmaceutical ingredient supply chain, ACS Sustain. Chem. Eng. 7 (2019) 16937–16951. doi:10.1021/acssuschemeng.9b02842.

[21] C. Jimenez-Gonzalez, C. Lund, Green metrics in pharmaceutical development, Curr. Opin. Green Sustain. Chem. 33 (2022) 100564. doi:10.1016/j.cogsc.2021.100564.

A review article highlighting key green chemistry metrics used in pharmaceutical processes and future perspectives on sustainability assessment.

[22] P. Tremouilhac, A. Nguyen, Y.-C. Huang, S. Kotov, D. S. Lütjohann, F. Hübsch, N. Jung, S. Bräse, Chemotion ELN: an open source electronic

lab notebook for chemists in academia, J. Cheminformatics 9 (2017) 1–13. doi:10.1186/s13321-017-0240-0.

[23] C. R. McElroy, A. Constantinou, L. C. Jones, L. Summerton, J. H. Clark, Towards a holistic approach to metrics for the 21st century pharmaceutical industry, Green Chem. 17 (2015) 3111–3121. doi:10.1039/C5GC00340G.

[24] B. Martín-Matute, M. A. R. Meier, T.-X. Métro, S. G. Koenig, H. F. Sneddon, P. Sudarsanam, P. Watts, Sustainable chemistry and engineering in pharma, ACS Sustain. Chem. Eng. 9 (2021) 13395–13398. doi:10.1021/acssuschemeng.1c06526.

[25] P. Baldi, Deep Learning in Science, Cambridge University Press, 2021.

[26] A. F. de Almeida, R. Moreira, T. Rodrigues, Synthetic organic chemistry driven by artificial intelligence, Nat. Rev. Chem. 3 (2019) 589–604. doi:10.1038/s41570-019-0124-0.

This review discusses the quality of publicly available chemical data and the recent impact of AI on different tasks of synthetic chemistry. It also dissects selected examples from the literature to make it clear for organic chemists.

[27] D. T. Ahneman, J. G. Estrada, S. Lin, S. D. Dreher, A. G. Doyle, Predicting reaction performance in C-N cross-coupling using machine learning, Science 360 (2018) 186–190. doi:10.1126/science.aar5169.

[28] A. L. Haywood, J. Redshaw, M. W. D. Hanson-Heine, A. Taylor, A. Brown, A. M. Mason, T. Gärtner, J. D. Hirst, Kernel methods for

predicting yields of chemical reactions, J Chem Inf Model (2022 (in press)). doi:10.1021/acs.jcim.1c00699.

[29] S. Jiang, Z. Zhang, H. Zhao, J. Li, Y. Yang, B.-L. Lu, N. Xia, When SMILES smiles, practicality judgment and yield prediction of chemical reaction via deep chemical language processing, IEEE Access 9 (2021) 85071–85083. doi:10.1109/ACCESS.2021.3083838.

[30] P. Schwaller, A. C. Vaucher, T. Laino, J.-L. Reymond, Prediction of chemical reaction yields using deep learning, Mach. Learn.: Sci. Technol. 2 (2021) 015016. doi:10.1088/2632-2153/abc81d.

[31] D. Prat, A. Wells, J. Hayler, H. Sneddon, C. R. McElroy, S. Abou-Shehada, P. J. Dunn, Chem21 selection guide of classical- and less classical-solvents, Green Chem. 18 (2016) 288–296. doi:10.1039/C5GC01008J.

[32] M. González-Miquel, I. Díaz, Green solvent screening using modeling and simulation, Curr. Opin. Green Sustain. Chem. 29 (2021) 100469. doi:10.1016/j.cogsc.2021.100469.

[33] S. Boobier, D. R. Hose, A. J. Blacker, B. N. Nguyen, Machine learning with physicochemical relationships: solubility prediction in organic solvents and water, Nat. Commun. 11 (2020) 1–10. doi:10.1038/s41467-020-19594-z.

[34] M. Bystrzanowska, M. Tobiszewski, Chemometrics for selection, prediction, and classification of sustainable solutions for green chemistry—a review, Symmetry 12 (2020) 2055. doi:10.3390/sym12122055.

[35] H. Sels, H. De Smet, J. Geuens, Sussol—using artificial intelligence for greener solvent selection and substitution, Molecules 25 (2020) 3037. doi:10.3390/molecules25133037.

[36] C. Larsen, P. Lundberg, S. Tang, J. Ràfols-Ribé, A. Sandström, E. M. Lindh, J. Wang, L. Edman, A tool for identifying green solvents for printed electronics, Nat. Commun. 12 (2021) 1–7. doi:10.1038/s41467-021-24761-x.

[37] L. J. Diorazio, D. R. J. Hose, N. K. Adlington, Toward a more holistic framework for solvent selection, Org. Process Res. Dev. 20 (2016) 760–773. doi:10.1021/acs.oprd.6b00015.

[38] J. L. Baylon, N. A. Cilfone, J. R. Gulcher, T. W. Chittenden, Enhancing retrosynthetic reaction prediction with deep learning using multiscale reaction classification, J. Chem. Inf. Model. 59 (2019) 673–688. doi:10.1021/acs.jcim.8b00801.

[39] J. S. Schreck, C. W. Coley, K. J. M. Bishop, Learning retrosynthetic planning through simulated experience, ACS Cent. Sci. 5 (2019) 970–981. doi:10.1021/acscentsci.9b00055.

[40] Y. Zabolotna, D. M. Volochnyuk, S. V. Ryabukhin, K. Gavrylenko, D. Horvath, O. Klimchuk, O. Oksiuta, G. Marcou, A. Varnek, Synthi: A new open-source tool for synthon-based library design, J Chem Inf Model (2021 (in press)). doi:10.1021/acs.jcim.1c00754.

[41] Y. Lin, Z. Zhang, B. Mahjour, D. Wang, R. Zhang, E. Shim, A. McGrath, Y. Shen, N. Brugger, R. Turnbull, et al., Reinforcing the supply

chain of umifenovir and other antiviral drugs with retrosynthetic software, Nat. Commun. 12 (2021) 1–8. doi:10.21203/rs.3.rs-820369/v1.

[42] A. A. Lee, Q. Yang, V. Sresht, P. Bolgar, X. Hou, J. L. Klug-McLeod, C. R. Butler, Molecular transformer unifies reaction prediction and retrosynthesis across pharma chemical space, Chem. Commun. 55 (2019) 12152–12155. doi:10.1039/C9CC05122H.

[43] S. Genheden, A. Thakkar, V. Chadimová, J.-L. Reymond, O. Engkvist, E. Bjerrum, Aizynthfinder: a fast, robust and flexible open-source software for retrosynthetic planning, J. Cheminformatics 12 (2020) 1–9. doi:10.1186/s13321-020-00472-1.

[44] I. A. Watson, J. Wang, C. A. Nicolaou, A retrosynthetic analysis algorithm implementation, J. Cheminformatics 11 (2019) 1–12. doi:10.1186/s13321-018-0323-6.

[45] M. Christensen, L. P. E. Yunker, P. Shiri, T. Zepel, P. L. Prieto, S. Grunert, F. Bork, J. E. Hein, Automation isn't automatic, Chem. Sci. (2021) 15473–15490doi:10.1039/D1SC04588A.

An overview of automated systems for synthetic chemistry that integrates robotic and computational components into an iterative workflow.

[46] H. Gao, T. J. Struble, C. W. Coley, Y. Wang, W. H. Green, K. F. Jensen, Using machine learning to predict suitable conditions for organic

reactions, ACS Cent. Sci. 4 (2018) 1465–1476. doi:10.1021/acscentsci.8b00357.

[47] S. Slavov, R. D. Beger, Quantitative structure–toxicity relationships in translational toxicology, Curr. Opin. Toxicol. 23-24 (2020) 46–49. doi:https://doi.org/10.1016/j.cotox.2020.04.002.

[48] S. Lorenz, A.-K. Amsel, N. Puhlmann, M. Reich, O. Olsson, K. Kümmerer, Toward application and implementation of in silico tools and workflows within benign by design approaches, ACS Sustain. Chem. Eng. 9 (2021) 12461–12475. doi:10.1021/acssuschemeng.1c03070.

[49] J. Simm, L. Humbeck, A. Zalewski, N. Sturm, W. Heyndrickx, Y. Moreau, B. Beck, A. Schuffenhauer, Splitting chemical structure data sets for federated privacy-preserving machine learning, J. Cheminformatics 13 (2021) 96. doi:10.1186/s13321-021-00576-2.

[50] Y. Naruki, K. Ryuichi, K. Z. Yamamoto, Twitter integration of chemistry software tools, J. Cheminformatics 13 (2021). doi:10.1186/s13321-021-00527-x.

A new concept of using social media as an integration platform for cheminformatics software is presented. Three case studies of using Twitter bots (retrosynthesis bot, molecule viewer, and chemical structural formula editor) are reported.