

# ChemPager: Now Expanded for even Greener Chemistry

Hugo Loureiro<sup>a§</sup>, Michael Prem<sup>b§</sup>, and Georg Wuitschik<sup>a\*</sup>

**Abstract:** ChemPager is a freely available data analysis tool for analyzing, comparing and improving synthetic routes. Here, we present an expansion of this application that makes use of the functionality of the *PMI Predictor*, which the ACS Green Chemistry Institute Pharmaceutical Roundtable has recently published as a web application. This addition enables ChemPager to predict the cumulative process mass intensity of chemical routes, irrespective of their development status, by comparison with a set of reactions executed on large scale. The prediction of this core green chemistry metric aims to improve existing routes and help the decision-making process among route alternatives without the need for experimental data.

**Keywords:** Data analysis · Green chemistry · PMI · Process Development · Route scouting



**Georg Wuitschik** leads a laboratory in Process Chemistry at Roche. He has worked on a number of small molecule projects after joining Roche in 2011. His interests cover API process R&D, drug substance manufacturing, and data analytics. He received his diploma from Technical University of Munich with a master thesis under the supervision of Prof. Barry M. Trost and got his PhD from ETH Zurich under the supervision of Prof. Erick M. Carreira with the title *Oxetanes in Medicinal Chemistry*, followed by a postdoc in the group of Prof. Steve V. Ley at Cambridge University.



**Hugo Loureiro** is an intern at the Process Chemistry division of Roche. His interests lie in the intersection of life sciences and technology, more specifically in bioinformatics, biostatistics and data science. He received his MSc degree in electrical engineering from Instituto Superior Técnico in Lisbon. There he conducted his master thesis in survival analysis applied to bone metastatic cancer under the supervision of Prof. Susana Vinga.



**Michael Prem** works as an intern in the Drug Substance Scale-up & Supply department of Roche. Currently he is most interested in full stack development and data engineering. He has a MSc in biochemistry from the University of Würzburg and did his master thesis in the area of geometry optimization under the supervision of Prof. Bernd Engels.

## 1. Introduction

The desire to build sustainable businesses prompts companies to minimize the negative impact they have on society and the environment. For chemists, the main levers to achieve this goal are reduction of cost and ecological footprint, while maximizing process robustness and safety of chemical processes. These goals are mostly not at odds with each other. Many factors will influence the individual conclusion as to what is the best route to prepare a given molecule. We recently published ChemPager,<sup>[1]</sup> a data analysis tool to guide the development of chemical processes and compare them.

ChemPager provides an intuitive user interface with visualizations and tabular representations of various key performance indicators (KPIs) at the level of individual batches, steps and campaigns, or over the whole project portfolio. This includes, for example, process mass intensity (PMI),<sup>[2]</sup> production cost, solvent usage and volume time yield. ChemPager enables the chemist to make informed decisions and serves as a catalyst that drives the optimization of chemical syntheses.<sup>[1]</sup>

While ChemPager already offers scenario planning, it lacked a component to compare routes of significantly different development levels, for example commercial production data, lab results and even paper chemistry generated in brainstorming. We aim to achieve this goal partly by integrating some of the functionality of the free *PMI Predictor* web application described by Borovika and coworkers.<sup>[3]</sup> This latest iteration of ChemPager is available free of charge as part of the article's supplementary information.

The *PMI Predictor* web app provides a range of cumulative PMI (cPMI) values for a given synthesis route based on the sequence of steps and each step's reaction type.<sup>[3a]</sup> By estimating PMI values, the application is able to evaluate the environmental impact of current and proposed syntheses. This functionality in combination with its ease of use makes the *PMI Predictor* a great instrument for green chemistry.

We built on this foundation to extend the functionality of ChemPager. Apart from predicting the cPMI probability distribution, ChemPager has now been extended to highlight steps in the current route where the highest potential for PMI reduction is

\*Correspondence: Dr. G. Wuitschik<sup>a</sup>, E-mail: georg.wuitschik@roche.com

<sup>a</sup>Small Molecules Technical Development, Process Chemistry & Catalysis, F. Hoffmann-La Roche Ltd., Grenzachstrasse, CH-4070 Basel; <sup>b</sup>Small Molecules Technical Development, Drug Substance Scale-up & Supply, F. Hoffmann-La Roche Ltd., Grenzachstrasse, CH-4070 Basel

<sup>§</sup>These authors contributed equally to this work.

seen. It also provides context as to which solvents and reagents are used in the particular step and how they are being disposed of. We hope these additions make ChemPager a valuable decision-making tool when developing greener processes in the pharmaceutical industry.

## 2. ChemPager

ChemPager is built on top of the TIBCO Spotfire platform<sup>[4]</sup> with source data being stored in conventional spreadsheets, *i.e.* Microsoft Excel or Google Sheets.<sup>[5]</sup> The application offers a concise summary of the current project status as well as structured access to chemical information (Fig. 1). This range spans from individual reaction components to the overall project portfolio. ChemPager calculates commonly used KPIs like PMI, cost of inputs, labor and equipment, or volume-time-output.<sup>[6]</sup> In addition, project data is distilled into scores for robustness, safety, economy, greenness, and project difficulty.<sup>[1]</sup>

All data is presented in interactive visualizations that offer an immediate comparison between campaigns or synthesis routes. This provides the user with a way to drill down into individual steps, batches and process inputs, and how they contribute to the overall result. Based on that data, ChemPager highlights the steps and reagents for which an improvement in price or yield would result in the largest cost savings. The tool also allows complex scenarios to be created. These help, for example, to decide whether a known increase in yield would compensate for the cost of additional reagent. They also allow to elucidate which boundary conditions have to be met to reach a certain cost/PMI target.<sup>[1]</sup>

## 3. Improving ChemPager

We felt an expansion was needed to enable the synthetic chemist to compare the greenness of developed syntheses with industry and literature data. Various metrics have been proposed with the aim to quantify the environmental impact of chemical

syntheses.<sup>[7]</sup> The PMI and the closely related E factor<sup>[8]</sup> have been used the most<sup>[9]</sup> in the pharmaceutical industry. PMI is the guiding metric of the American Chemical Society Green Chemistry Institute Pharmaceutical Roundtable (ACS GCIPR), when it comes to process greenness.<sup>[2]</sup> However, great care has to be taken when comparing different processes, since the choice of synthesis starting point can greatly change PMI.<sup>[10]</sup> Nevertheless, process improvements are often directly reflected in its reduction.<sup>[12]</sup> However, PMI or any other quantitative metric cannot represent all aspects of sustainability. Thus, a range of approaches has been proposed: qualitative metrics,<sup>[7b,11]</sup> improved metrics,<sup>[12]</sup> greenness scores,<sup>[1,12a]</sup> and Life Cycle Assessments.<sup>[2,7b,13]</sup>

## 4. The PMI Predictor

The *PMI Predictor* makes previous work by Eastgate and co-workers<sup>[10,14]</sup> accessible in a clean and intuitive web interface. In order to enable the estimation of PMI values, member companies of the ACS GCIPR gathered, in a joint effort, PMI and yield data on large-scale reactions. After expanding this dataset with scale-up information from the published literature, Borovika *et al.* categorized the reactions according to the type of chemistry used.<sup>[3a,15]</sup>

Using this information, the *PMI Predictor* applies a Monte Carlo-based algorithm to calculate a probability distribution of the cPMI and displays it as a histogram. This allows to compare the cPMI of established processes with what could be expected upon scale-up based solely on the chemical transformations used.<sup>[3a]</sup> While the *PMI Predictor* offers an easy way to enter synthesis information for *one* route, it currently lacks the capability to visualize several campaigns side-by-side. ChemPager includes this functionality.

Secondly, we use the data already available in ChemPager to provide a new tabular visualization highlighting the solvent types employed and how the reaction components are disposed of. This complements the quantitative approach of PMI prediction

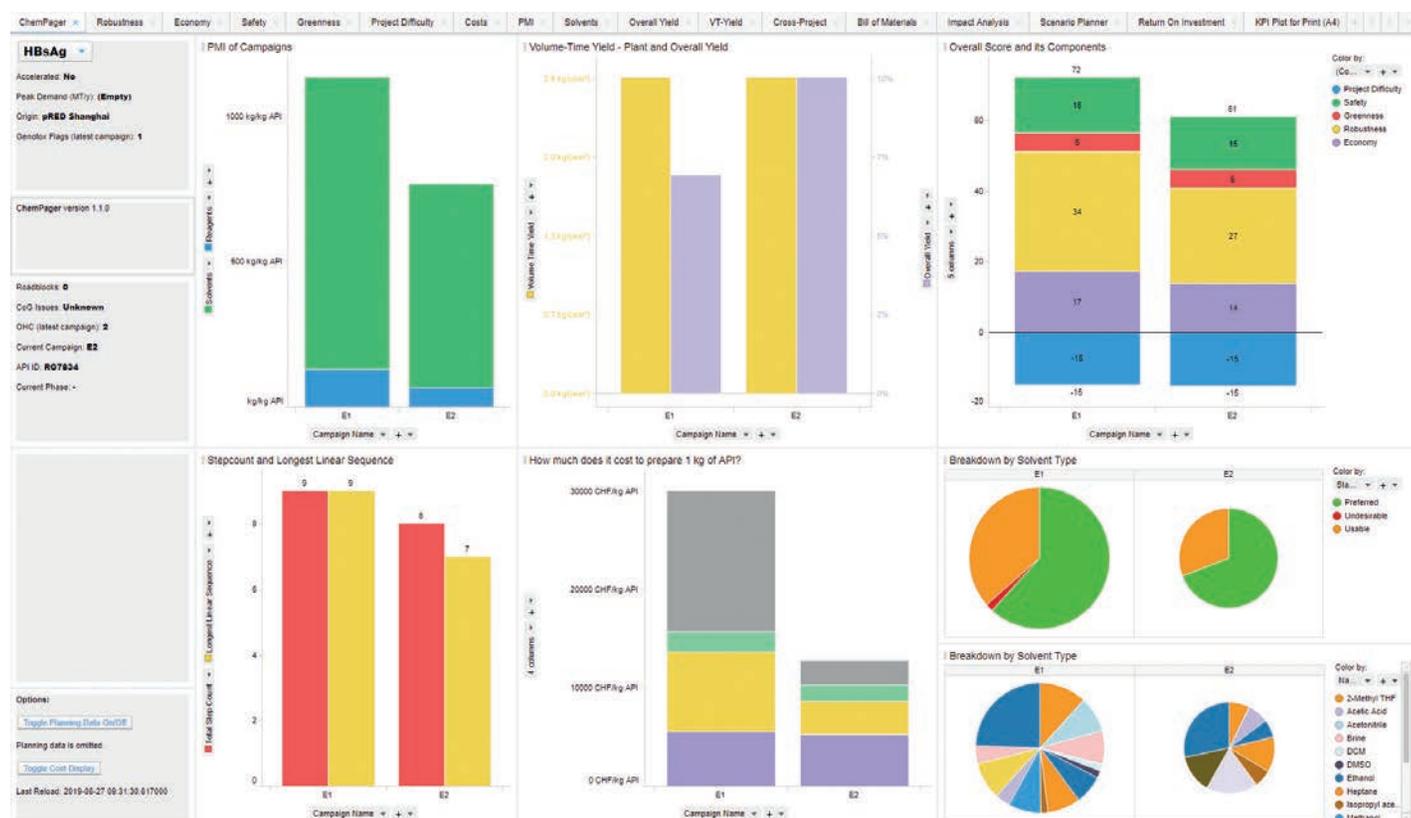


Fig. 1. ChemPager entrance page displaying key information on the project in question (bar on the left), PMI (top left), yield (top middle), project scores (top right), step count (bottom left), production cost (bottom middle) and solvent usage (bottom right). The bar on top also shows some of the detail visualization pages for further analysis. For further information on how the scores are calculated, please refer to our previous publication.<sup>[1]</sup>

with a view on what the actual inputs are, and what happens to them. Solvents account for a major component of the mass used in each step and consequently for the amount of waste that each step generates.<sup>[2]</sup> Solvent waste has three main ends: recycling, waste water and incineration. Depending on the solvent and the circumstances, any of these disposal paths can be the greenest.

## 5. Use Case

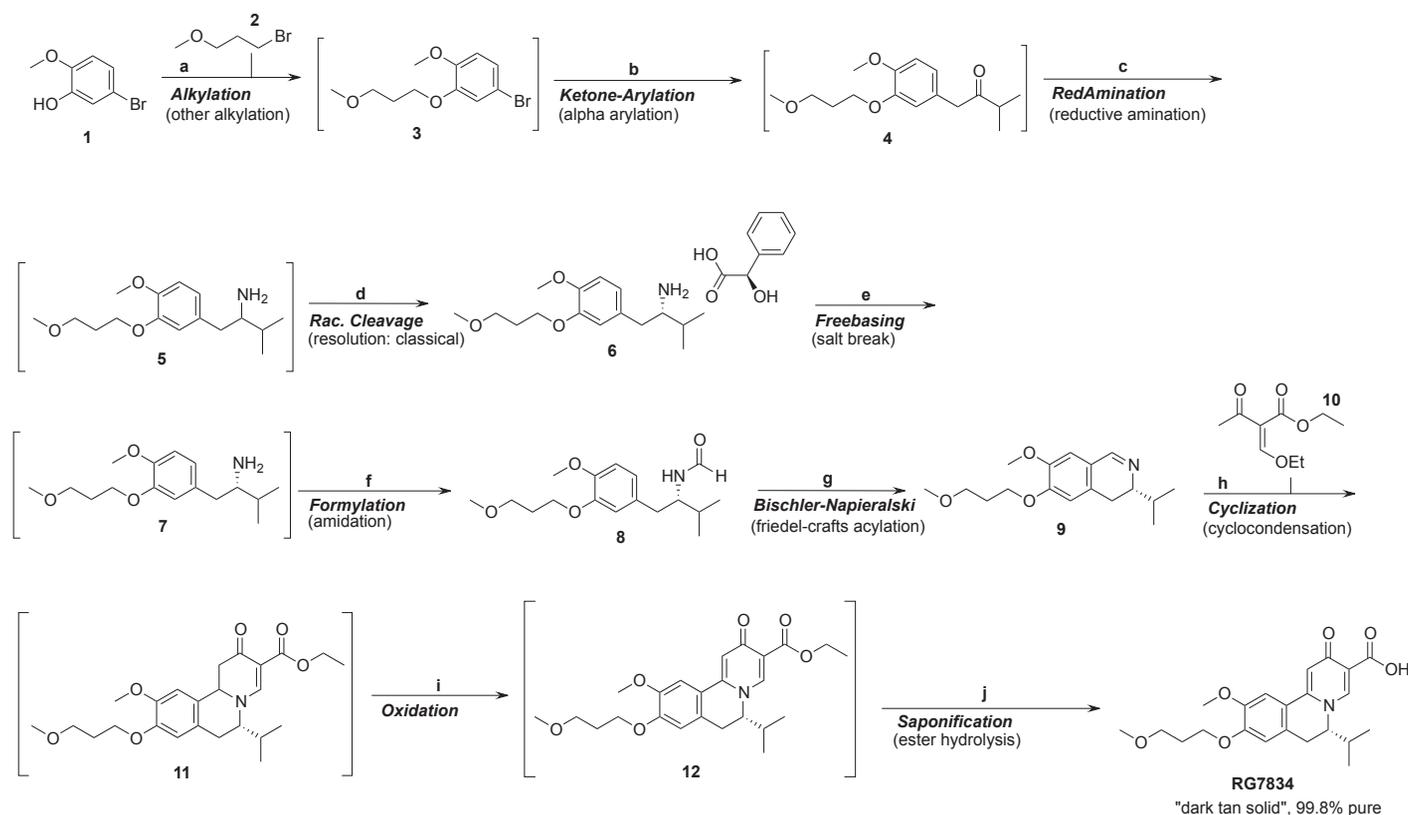
RG7834 was a drug candidate under development for the treatment of hepatitis B.<sup>[16]</sup> This project included several campaigns with two distinct synthesis routes. The two campaigns discussed here supplied clinical studies and were conducted at Asymchem Inc. (Tianjin). They delivered 10.75 kg (E1, phase 1 supply) and 37 kg (E2, phase 2 supply) of RG7834, respectively.

While the E1 route (Scheme 1) already incorporated a number of changes rendering it superior to preceding campaigns, there was room for further improvement. In terms of safety, the objective was to find alternative oxidation conditions for the preparation of ketoester **12**. These should avoid the combination of the potentially incompatible iodine and DMSO/MeCN (Scheme 1, step i). Also, hydrogen iodide produced during the oxidation as byproduct mediates the cleavage of the ester functionality. This will result in significant levels of known mutagen ethyl iodide.<sup>[17]</sup> Quality-wise, we were concerned that the telescoping of the last three steps, leading to the presence of colored impurities in RG7834, would not be robust upon further scale-up. We envisioned that isolation and purification of ester **12** could markedly improve quality consistency of RG7834. Finally, the diastereomeric salt formation of mandelate **6** presented a cost challenge in that at least half of racemic amine **5** was lost. Amine **5** itself was prepared in three steps using significant quantities of expensive palladium. In order to avoid both the palladium-catalyzed step and the racemate cleavage, we envisioned a chiral pool-approach using valine.

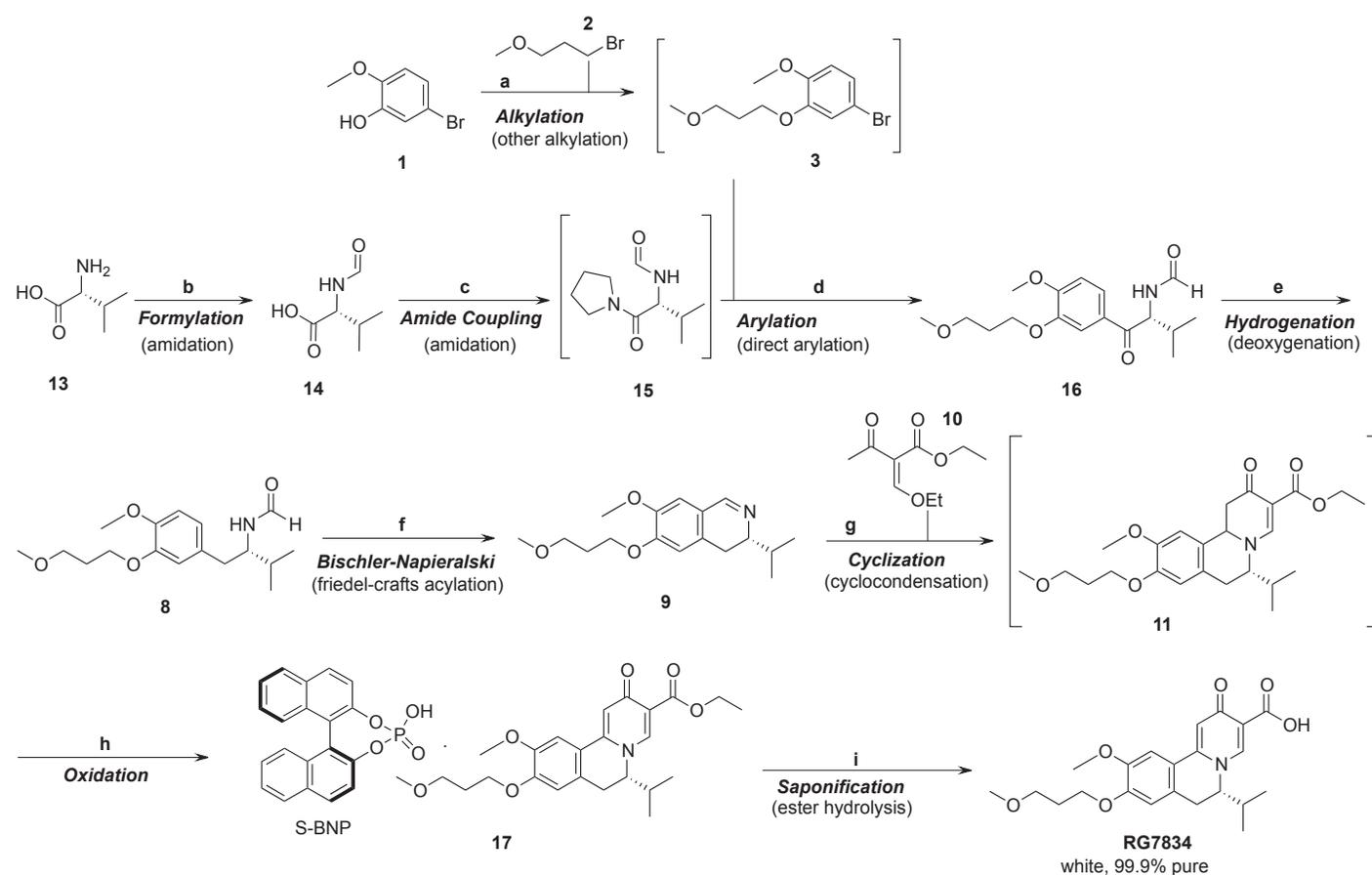
Beside achieving the goals outlined above, the developed E2 route in Scheme 2<sup>[18]</sup> also included detail improvements, for example to the Bischler-Napieralski reaction leading to imine **9**. For an in-depth discussion of the chemistry and the concomitant improvements in PMI, cost and overall yield, refer to the original ChemPager publication.<sup>[1]</sup>

ChemPager is useful to compare two existing routes or to guide their further improvement, but is not of much help to assess their potential, especially if the state of development is vastly different. In our case, the *PMI Predictor* module can help and would have uncovered the potential PMI advantage of the chiral pool route. Its predicted PMI distribution, shown in Fig. 2 in blue, is markedly shifted to lower PMIs. The real PMI value, shown as a vertical line, lies within this prediction.<sup>[19]</sup> More detailed step-level data is provided in tabular form at the bottom of the page and in two visualizations on the right. There, for each step the actually achieved PMI and yield is compared with the median PMI or yield. Both are extracted from the underlying set of reactions published by Borovika *et al.*<sup>[3a,c]</sup> This highlights steps which deviate either in yield or PMI from the precedent of the presets file. As only the types of chemistry are used for categorization, many differences thus uncovered will be the result of conditions specific to the chemistry at hand.<sup>[20]</sup> In a lot of instances, however, these visualizations will guide the development chemist towards opportunities for improvement.<sup>[21]</sup>

The second page depicted in Fig. 3 provides a more granular view on what is happening at the step level and tries to supplement the quantitative nature of PMI with qualitative data relating to solvent category and disposal. The second page is composed of four visualizations. The one on the top left is focused on the classification of solvents employed in the synthesis at different stages.<sup>[22,23]</sup> The one in the top middle shows how the reaction components are disposed of afterwards, an important consideration especially for commercial production.<sup>[24]</sup> The visualization on the right shows how the PMI of each step compares with similar chemistry cap-



Scheme 1. Synthesis route of the E1 campaign (10.75 kg RG7834). Displayed in bold are the step names used in ChemPager and in brackets, the reaction classification in the PMI Predictor. a)  $K_2CO_3$ , EtOH, 80 °C, 99%; b) 2mol%  $Pd_2(dba)_3$ , 3mol% xantphos, NaOtBu, MeTHF, 50 °C, 86%; c) 10 eq  $NH_4OAc$ ,  $NaBH_3CN$ , MeOH; d) R-Mandelic acid, TBME, 33% (2 steps); e) aq. NaOH, TBME, 85%; f) formic acid, MeTHF, 80 °C, 85%; g)  $POCl_3$ , MeCN, 75%; h) water, reflux; i)  $I_2$ , DMSO, MeCN, 65 °C, 87%; j) aq. LiOH, MeTHF, then  $H_3PO_4$ , 44% (3 steps).



Scheme 2. Synthesis route of the E2 campaign (37 kg RG7834). Displayed in bold are the step names used in ChemPager and in brackets the reaction classification in the PMI Predictor. a)  $\text{K}_2\text{CO}_3$ , EtOH, 80 °C, 98%; b) formic acid, PhMe, reflux, 70%; c) IBCF, NMM, THF, -20 °C, then pyrrolidine, 62%; d) MeLi, THF, -20 °C, then cat.  $\text{iPrMgCl}\cdot\text{LiCl}$ , then BuLi, 75%; e)  $\text{H}_2$  (20 bar), AcOH, Pd/C, 80 °C, 79%; f)  $\text{POCl}_3$ , PhMe, 25-45 °C, 85%; g) Brine/MeTHF, 75 °C; h)  $\text{I}_2$ , pyridine, MeTHF, then S-BNP, 60 °C, 87% (2 steps); i)  $\text{NEt}_3$ , MeTHF, then aq. NaOH, then  $\text{H}_3\text{PO}_4$ , 58%. IBCF = isobutyl chloroformate, NMM = N-Methylmorpholine

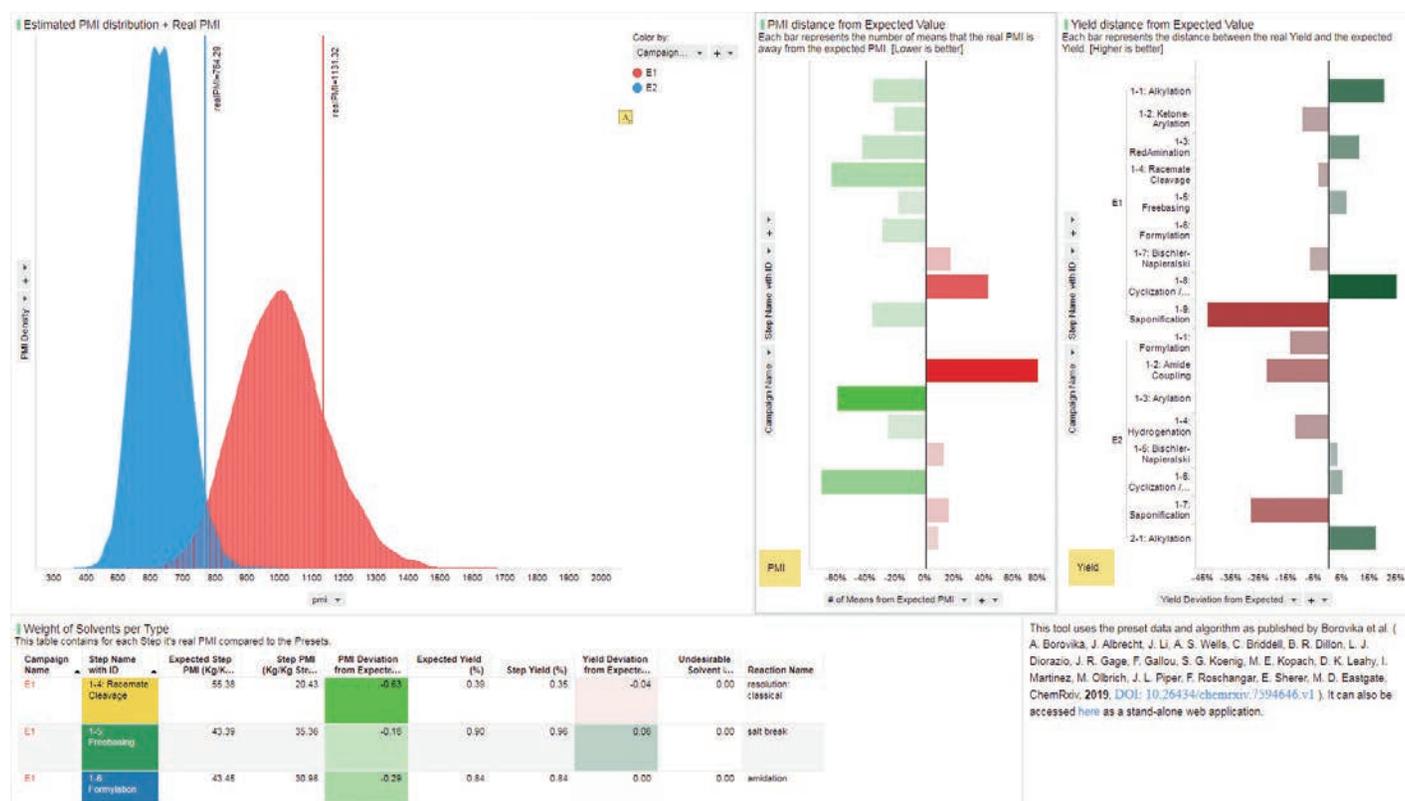


Fig. 2. Visual comparison of the estimated cPMI distribution with the actual cPMI. The panel on the top left shows for each campaign the estimated cPMI distribution and the corresponding real cPMI value (vertical line). The visualizations on the top right display for each step the deviation of yield and PMI from the respective means of the presets file.

tured in the presets file. The table at the bottom provides detailed information on the reaction components of each step. This way the user can discern immediately which solvents or reagents are behind any of the bars in the visualizations above.<sup>[25]</sup>

In the specific example of RG7834, the undesirable solvent marked red in the top left visualization of Fig. 3 becomes apparent and a click on it reveals that it is DCM and that it is incinerated. Shown are also the improvements in overall PMI and where the improvements come from.

## 6. Conclusion and Outlook

Chemists need intuition and experience to choose the right synthetic route, but tools can help greatly. We have shown that a unified tool for data analysis can aid process understanding and improve decision-making during the initial phases of synthesis route design, reducing the environmental impact of chemical syntheses. In conclusion, we have expanded ChemPager with the capability to estimate the PMI of syntheses and compare the estimate with actual data on a step as well as campaigns. This not only facilitates decision-making among routes, but may also guide their further improvement.

Future development at Roche will likely lead to a replacement of the currently used spreadsheet data stores and manual data entry. In addition, the connection to other company-internal databases is currently being established, turning ChemPager into an all-in-one platform for process chemistry.

## Supplementary Information

Supplementary information on ChemPager is available on <https://www.ingentaconnect.com/content/scs/chimia>

## Acknowledgements

The authors would like to thank Dr. Alina Borovika (Bristol-Myers Squibb) for her support. The authors would also like to thank Dr. Frank Roschangar (Boehringer-Ingelheim) and Dr. Edouard Godineau (Syngenta) for input and useful discussions.

Received: July 4, 2019

- [1] D. Kaiser, J. Yang, G. Wuitschik, *Org. Process Res. Dev.* **2018**, *22*, 1222, DOI: 10.1021/acs.oprd.8b00199.
- [2] C. Jiménez-González, C. S. Ponder, Q. B. Broxterman, J. B. Manley, *Org. Process Res. Dev.* **2011**, *15*, 912, DOI: 10.1021/op200097d.
- [3] a) 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, M. E. Kopach, D. K. Leahy, I. Martinez, M. Olbrich, J. L. Piper, F. Roschangar, E. Sherer, M. D. Eastgate, *ChemRxiv* **2019**, DOI: 10.26434/chemrxiv.7594646.v1; b) J. Albrecht, J. Li, A. Borovika, M. D. Eastgate, PMI Predictor, <https://www.acsgcipr.org/tools-for-innovation-in-chemistry/process-mass-intensity-prediction-calculator/>; [https://acsgcipr-predictpmi.shinyapps.io/pmi\\_calculator/](https://acsgcipr-predictpmi.shinyapps.io/pmi_calculator/), both accessed June, 21, 2019; c) J. Albrecht, J. Li, A. Borovika, M. D. Eastgate, PMI Predictor (source code repository), [https://github.com/chepyle/PMI\\_Calculator](https://github.com/chepyle/PMI_Calculator), accessed June, 21, 2019.
- [4] Spotfire Analyst 7.11 LTS, TIBCO Software Inc., Palo Alto (CA), **2017**.
- [5] At Roche, we are currently working on a system that is foreseen to replace these spreadsheets and most of the associated manual data entry.
- [6] R. Dach, J. J. Song, F. Roschangar, W. Samstag, C. H. Senanayake, *Org. Process Res. Dev.* **2012**, *16*, 1697, DOI: 10.1021/op300144g.
- [7] a) D. J. C. Constable, in 'Handbook of Green Chemistry Volume 11: Green Metrics', Eds. D. J. C. Constable, C. Jiménez-González, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, **2018**, p. 1; b) R. A. Sheldon, *ACS Sustainable Chem. Eng.* **2018**, *6*, 32, DOI: 10.1021/acssuschemeng.7b03505; c) C. Jiménez-González, D. J. C. Constable, C. S. Ponder, *Chem. Soc. Rev.* **2012**, *41*, 1485, DOI: 10.1039/C1CS15215G.
- [8] R. A. Sheldon, *Green Chem.* **2017**, *19*, 18, DOI: 10.1039/C6GC02157C.
- [9] W. J. W. Watson, *Green Chem.* **2012**, *14*, 251, DOI: 10.1039/C1GC15904F.
- [10] J. Li, E. M. Simmons, M. D. Eastgate, *Green Chem.* **2017**, *19*, 127, DOI: 10.1039/c6gc02359b.
- [11] A. P. Dicks, A. Hent, in 'Green Chemistry Metrics: A Guide to Determining and Evaluating Process Greenness', Springer International Publishing, Cham, Switzerland, **2015**, p. 69.
- [12] a) F. Roschangar, Y. Zhou, D. J. C. Constable, J. Colberg, D. P. Dickson, P. J. Dunn, M. D. Eastgate, F. Gallou, J. D. Hayler, S. G. Koenig, M. E. Kopach, D. K. Leahy, I. Mergelsberg, U. Scholz, A. G. Smith, M. Henry, J. Mulder, J. Brandenburg, J. R. Dehli, D. R. Fandrick, K. R. Fandrick, F. Gnad-Badouin, G. Zerban, K. Groll, P. T. Anastas, R. A. Sheldon, C. H. Senanayake, *Green Chem.* **2018**, *20*, 2206, DOI: 10.1039/C8GC00616D; b) F. Roschangar, R. A. Sheldon, C. H. Senanayake, *Green Chem.* **2015**, *17*, 752, DOI: 10.1039/C4GC01563K.
- [13] C. Jiménez-González, A. D. Curzons, D. J. C. Constable, V. L. Cunningham, *Int. J. LCA* **2004**, *9*, 114, DOI: 10.1007/BF02978570.
- [14] J. Li, J. Albrecht, A. Borovika, M. D. Eastgate, *ACS Sustainable Chem. Eng.* **2018**, *6*, 1121, DOI: 10.1021/acssuschemeng.7b03407.

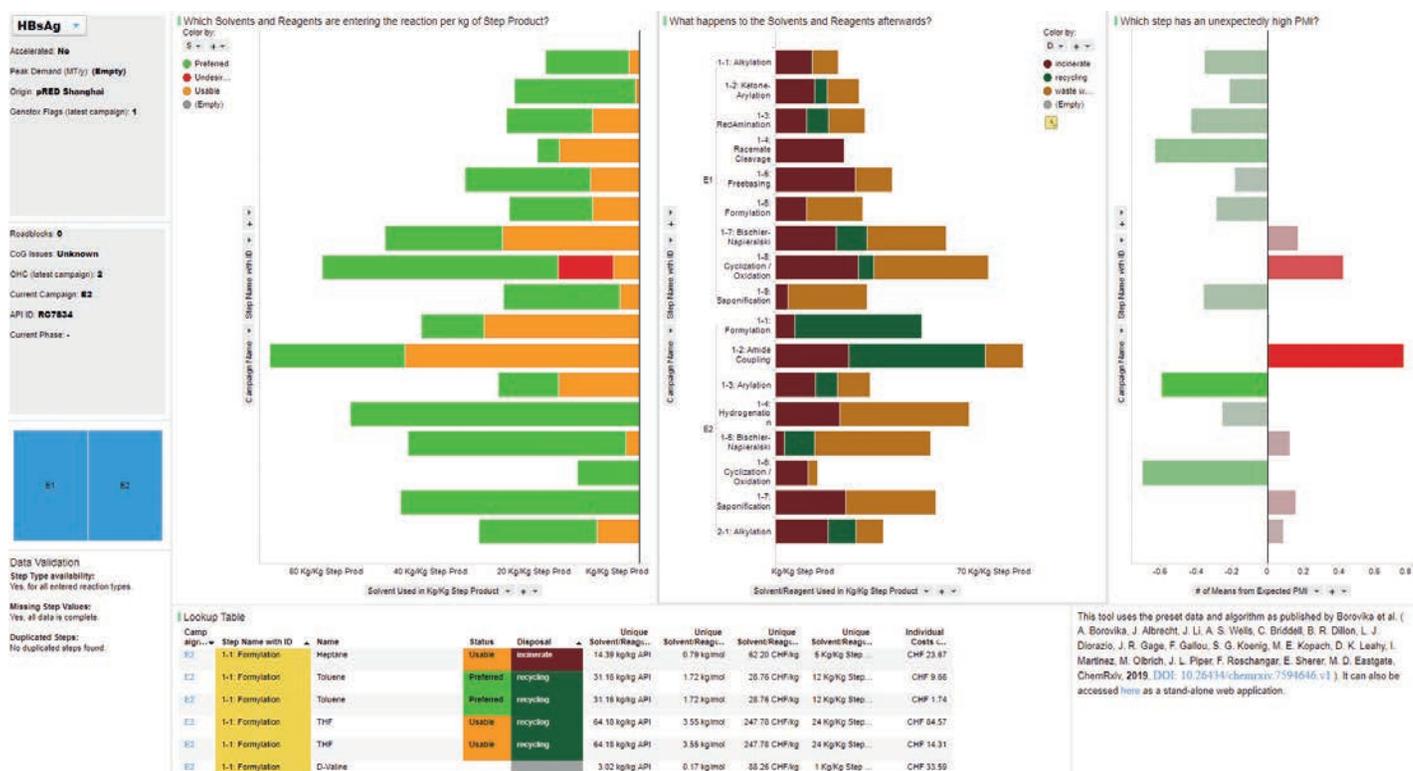


Fig. 3. Step-level visualization of solvents employed and how they are categorized (top left) as well as disposal of the reaction components (top middle) and how the PMI achieved in each step compares to the median PMI of comparable chemistry recorded in the presets file (top right).

- [15] For details on the data used by ChemPager, see the supplementary information of this article.
- [16] a) H. Mueller, S. Wildum, S. Luangsay, J. Walther, A. Lopez, P. Tropberger, G. Ottaviani, W. Lu, N. J. Parrott, J. D. Zhang, R. Schmucki, T. Racek, J.-C. Hoflack, E. Kueng, F. Point, X. Zhou, G. Steiner, M. Lütgehetmann, G. Rapp, T. Volz, M. Dandri, S. Yang, J. A. T. Young, H. Javanbakht, *J. Hepatol.* **2018**, *68*, 412, DOI: 10.1016/j.jhep.2017.10.014; b) R. F. Schinazi, M. Ehteshami, L. Bassit, T. Asselah, *Liver Int.* **2018**, *38*, 102, DOI: 10.1111/liv.13656; c) J. E. Arends, F. I. Lieveld, S. Ahmad, A. Ustianowski, *Infect. Dis. Ther.* **2017**, *6*, 461, DOI: 10.1007/s40121-017-0173-y; d) S. Wai-Kay, Y. Man-Fung, *Clinical Liver Disease* **2016**, *8*, 83, DOI: 10.1002/cld.577; e) X. Han, H. Javanbakht, M. Jiang, C. Liang, J. Wang, Y. Wang, Z. Wang, R. J. Weikert, S. Yang, C. Zhou, WO2015113990, **2015**.
- [17] K. Hemminki, K. Falck, H. Vainio, *Arch. Toxicol.* **1980**, *46*, 277, DOI: 10.1007/BF00310445.
- [18] G. Wuitschik, M. Mosch, WO2018041873, **2018**.
- [19] This is confirmed in a number of other internal projects.
- [20] Examples for this include the unusually low PMI for the oxidation in E1 which is the result of telescoping and the concomitant omission of solvent used in this step, or the decreased yield for the saponification step in E1, since the previous step yields were set to 100% and the low yield of the saponification represents the yield of the telescoped sequence.
- [21] Examples for this are the unusually high PMI and low yield for the racemate cleavage in E1 and the amide coupling in E2.
- [22] The solvent categorization used for ChemPager is Roche-internal, but can easily be modified to a scheme such as the one presented in the CHEM21 selection guide, see ref. [23].
- [23] a) D. Prat, A. Wells, J. Hayler, H. Sneddon, C. R. McElroy, S. Abou-Shehata, P. J. Dunn, *Green Chem.* **2016**, *18*, 288, DOI: 10.1039/C5GC01008J; b) CHEM21, Online Learning Platform, <http://learning.chem21.eu/methods-of-facilitating-change/tools-and-guides/solvent-selection-guides/guide-tables/>, accessed August, 13, **2019**.
- [24] The disposal attributes shown are not the real ones and have been reassigned.
- [25] The drill-down into the data is performed simply by marking individual bar segments, items in the legend or whole steps/campaigns.