

Driving Aspirational Process Mass Intensity Using SMART-PMI and Innovative Chemistry

Edward C. Sherer^{1,#}, Ansuman Bagchi², Birgit Kosjek¹, Kevin M. Maloney¹, Zhengwei Peng³, Sandra A. Robaire⁴, Robert P. Sheridan³, Essam Metwally⁵, Louis-Charles Campeau¹

¹ Department of Process Research & Development, MRL, Merck & Co., Inc., Rahway, NJ, 07065, USA.

² Department of Applied Mathematics & Modeling, Global Services, Merck & Co., Inc., Rahway, NJ, 07065, USA.

³ Department of Computational and Structural Chemistry, MRL, Merck & Co., Inc., Kenilworth, NJ, 07065, USA.

⁴ Department of Chemical Commercialization Technology, MMD, Merck & Co., Inc., Rahway, NJ, 07065, USA.

⁵ Department of Computational and Structural Chemistry, MRL, Merck & Co., Inc., S. San Francisco, CA, 94080, USA.

corresponding author; email: edward_sherer@merck.com

KEYWORDS process mass intensity, PMI, machine learning, green chemistry, SMART-PMI

ABSTRACT: An important metric for gauging the impact a synthetic route has on chemical resources, cost, and sustainability is process mass intensity (PMI). Calculating the overall PMI or step PMI for a given synthesis from a process description is more and more common across the industry. Our company has established a strong track record of delivering on our Corporate Sustainability goals, being recognized with seven EPA Green Chemistry Challenge Awards in the last 15 years. While green chemistry principles help in optimizing PMI and developing more sustainable processes, a key challenge for the field is defining what ‘good’ looks like for any given molecule. Predicting aspirational PMI for a synthetic target is not yet possible from chemical structure alone. The only tool chemists have at their disposal to predict PMI requires the synthetic route to be available, which is inherently retrospective. We have developed SMART-PMI (in-Silico MSD Aspirational Research Tool) to fill this glaring gap. Using only a 2D chemical structure, which enables a measure of molecular complexity, we can generate a predicted SMART-PMI using historical PMI data from our company’s clinical and commercial portfolio of processes. From this SMART-PMI prediction, we have established target ranges for Successful, World Class, and Aspirational PMI. Using this model, chemists can develop powerful synthetic strategies that make the biggest impact on PMI and, in turn, drive improvements to the model. The potential of SMART-PMI to set industry-wide aspirational PMI targets is discussed.

The world of pharmaceutical manufacturing is facing increasing challenges of cost and sustainability. Today, there is heightened awareness of the environmental impacts of manufacturing activities in society. Scientists tasked with designing these manufacturing processes have a responsibility to ensure that these practices are sustainable, limiting impact on climate change and waste generation. In order to achieve these critical goals, pharmaceutical scientists drive chemistry innovation to design and optimize synthetic routes for drug manufacture. Given multiple possible routes to the same synthetic target, chemists are faced with the challenge of maximizing yield, reducing cost of goods, eliminating hazardous reagents/solvents, and reducing waste. Despite these well-established parameters, it remains incredibly difficult to assess if a given process is “good enough” on any of these dimensions. Ultimately, metrics are used to evaluate progress and determine if goals are reached. We believe these aspirational targets are needed to drive continuous innovation into new synthetic and cheminformatics methods that help scientists find the most direct and sustainable route from commodity chemicals to the active pharmaceutical ingredient (API). If the field relies purely on established synthetic precedent, our drive for innovations in green chemistry will invariably stagnate. To avoid this torpor, we need to establish predictive models that can learn from new innovations and continue to set evermore aspirational targets for the field.

The use of process mass intensity (PMI) to gauge the greenness of a chemical route is well-established.^{1, 2} This metric accounts for the total mass of all components of a

synthesis normalized per 1 kg of drug substance synthesized. With PMI information on a route, chemists can make decisions about which steps should be areas of focus or which synthetic routes should be pursued. Useful statistics for a route include the total PMI and the step-PMI for each individual chemical step. Additional metrics for determining the environmental impact of chemical syntheses include the Green Aspirational Level (GAL) amongst others.¹⁻¹⁰ It is common in the life cycle of any given synthetic route for the PMI to drastically improve. This outcome is expected since mitigation of waste generated from any single synthesis event becomes more important as molecules are synthesized on larger scale. For example, it is common for inefficient workups and preparative chromatography to be used in early small-scale laboratory syntheses. As a molecule progresses into pre-clinical and clinical development, these operations are usually optimized or completely removed in favor of greener isolation and purification methods. For this reason, PMI is usually only evaluated in process chemistry settings. Given a synthetic route and process description, it is straightforward to calculate the PMI for a drug molecule.

Recently, it was demonstrated that a range of PMIs can be predicted for a given synthetic route.^{11, 12} “The PMI Predictor” can be used to calculate a range of PMI values for a detailed synthetic scheme. Understanding where a given PMI falls across this calculated range provides insight into how well a route has been optimized. However, while route optimization typically leads to reduction in PMI over time, this impact is often modest because it is unable to overcome the selection of a suboptimal route, nor does it challenge a

chemist to envision an aspirational route. Upon entering different routes, "The PMI Predictor" does identify which route is likely to lead to lower PMI, thus enabling scientists to make more informed choices when comparing known routes. However, subtle differences between routes may be hard to discern given the wide error ranges. Chemists are still unable to set an aspirational PMI bar for a molecule without considering the synthetic route. Herein, we describe a tool for setting aspirational PMI targets predicted from chemical structure alone. In doing so, it enables a continuous improvement mindset and can be used to drive synthetic strategy and significant investments toward aspiration manufacturing routes which improve sustainability goals.

At our company, it is typical for the chemistry routes used to manufacture clinical supplies early in development to be changed and optimized for the appropriate measures of success for a long-term commercial process. This route optimization effort includes a focus on reducing PMI. We previously identified a relationship between molecular complexity (as determined by a machine learning (ML) model derived from chemists' intuition) and PMI.¹³ With these initial observations, we were inspired to further explore the underlying molecular descriptors that could be used to drive aspirational synthetic targets. A data set was constructed (Supporting Information Table S1) for all compounds for which PMI values were available; in total, this composite includes 30 compounds and 36 PMIs (six molecules have two routes available). The stage of development varied for this set from early clinical supply (31) up to commercial manufacturing (5). This set is, therefore, biased toward less efficient routes.

PMI was experimentally calculated from the synthetic route definition according to guidelines developed by the American Chemical Society Green Chemistry Institute (PMI-LCA – life cycle assessment tool). Using this dataset, we built a linear combination of molecular complexity values generated using our ML model and molecular weight (MW). This model, which we call SMART-PMI (in-Silico MSD Aspirational Research Tool), was used to predict PMI. We investigated several other molecular descriptors for underlying correlations to PMI, but MW was the only metric that helped. Owing to the larger sampled chemical space of the complexity training set, ML models built directly to PMI from a limited number of data points will be less robust and were not pursued.

The MW of more complex molecules tends to be greater than less complex molecules. However, at any given MW range, the complexity values can vary considerably. For this reason, the two descriptors were combined into a linear model to predict SMART-PMI. The resultant best fit to the synthetically determined PMI is provided by Equation 1.

Equation 1

$$\text{SMART-PMI} = (0.13 \times \text{MW}) + (177 \times \text{Complexity}) - 252$$

Shown in Figure 1 is the experimental PMI for each API plotted against the predicted SMART-PMI. As expected, as the MW and complexity of a molecule increase, the SMART-PMI increases. The best fit line to all data points represents a target SMART-PMI averaged across all stages of route development with an R^2 of only 0.33 (MAE and MUE is 0.2 and 144, respectively). For this reason, it is a rough guide to how good a given synthetic route is at any point in time. Values that lie significantly above the best fit line indicate that the synthetically determined PMI is less optimal, even for early clinical supply chemistry. This outcome may be due to either sub-optimal chemistry in the route or a less efficient chemical route to the target API.

Our ultimate goal was to use this tool to set aspirational targets for commercial manufacturing routes, therefore helping in investment decisions toward new routes that meet or exceed our company's long-term sustainability goals.^{14, 15} A comparison of APIs, for which we have both an early clinical supply route PMI and commercial manufacturing route PMI, is presented in Figure 2. Several of the early routes (3/6) have synthetic PMIs much larger than that predicted from the model with the other three in-line or slightly better. For these three, the PMIs were decreased by an average of 66%. Despite a significantly better starting point, for the other three API routes, PMI was still improved 47% on average.

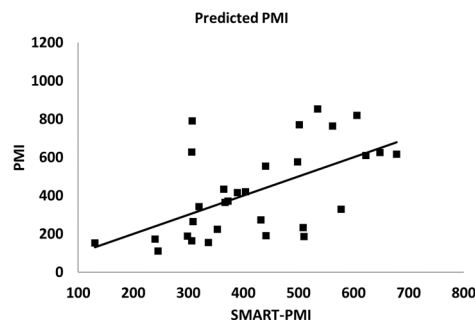


Figure 1: SMART-PMI in relation to the minimum synthetic PMI for each molecule

Considering this analysis and focusing on the data points from commercial manufacturing routes, we developed guidelines with a continuous improvement mindset that set an aspirational bar for the PMI of our future products. With the best fit line roughly equivalent to $\text{PMI} = \text{SMART-PMI}$, three categories were set in relation to SMART-PMI values: Successful = 0.9 to $1.1 \times \text{SMART-PMI}$, World Class = 0.5 to $0.9 \times \text{SMART-PMI}$, and Aspirational = $< 0.5 \times \text{SMART-PMI}$. A graphical representation of SMART-PMI ranges for Successful, World Class, and Aspirational are presented in Figure 3.

When comparing the values of SMART-PMI to those from "The PMI Predictor" for any given route, we typically see lower targets for SMART-PMI, as it includes a

disproportionate number of PMIs from commercial manufacturing routes in the training set, several of which have not been previously disclosed. In order to pressure-test these aggressive targets, we mapped processes that have received external recognition for Green Chemistry onto the model. Over the last 15 years, our company has been recognized with the Environmental Protection Agency's (EPA) Green Chemistry Challenge Award on six occasions for commercial products (Emend®, Prevymis™, Zerbaxa®, Delstrigo™, Januvia® (x2)). The location of these six APIs in Figure 3 is quite striking when plotted onto the SMART-PMI model. The selection of the Aspirational target having a value of $< 0.5 \times \text{SMART-PMI}$, when indicated on the plot, sends a green line directly through the path of the six award-winning synthetic routes. We believe this threshold strengthens the case that Aspirational SMART-PMI will drive innovative chemistry

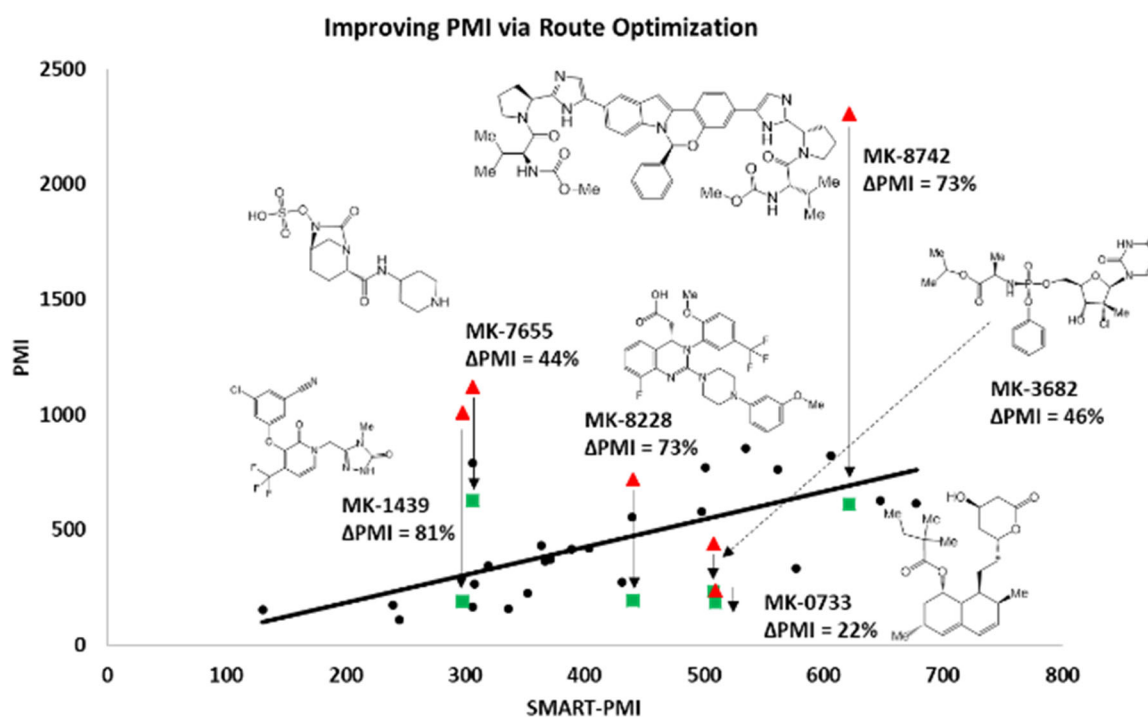
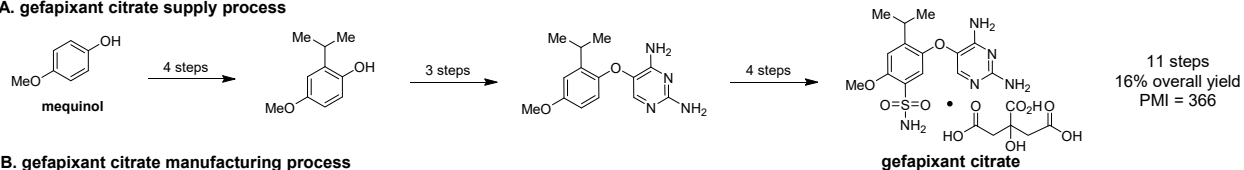


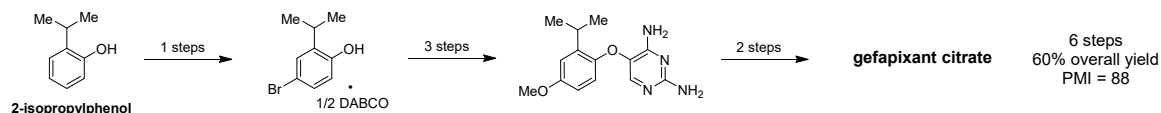
Figure 2: Improvement in PMI for six APIs through route optimization efforts.

Scheme 1

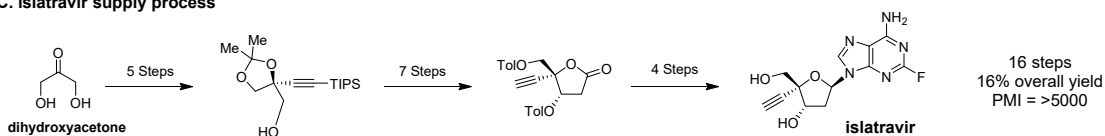
A. gefapixant citrate supply process



B. gefapixant citrate manufacturing process



C. islatravir supply process



D. islatravir manufacturing process

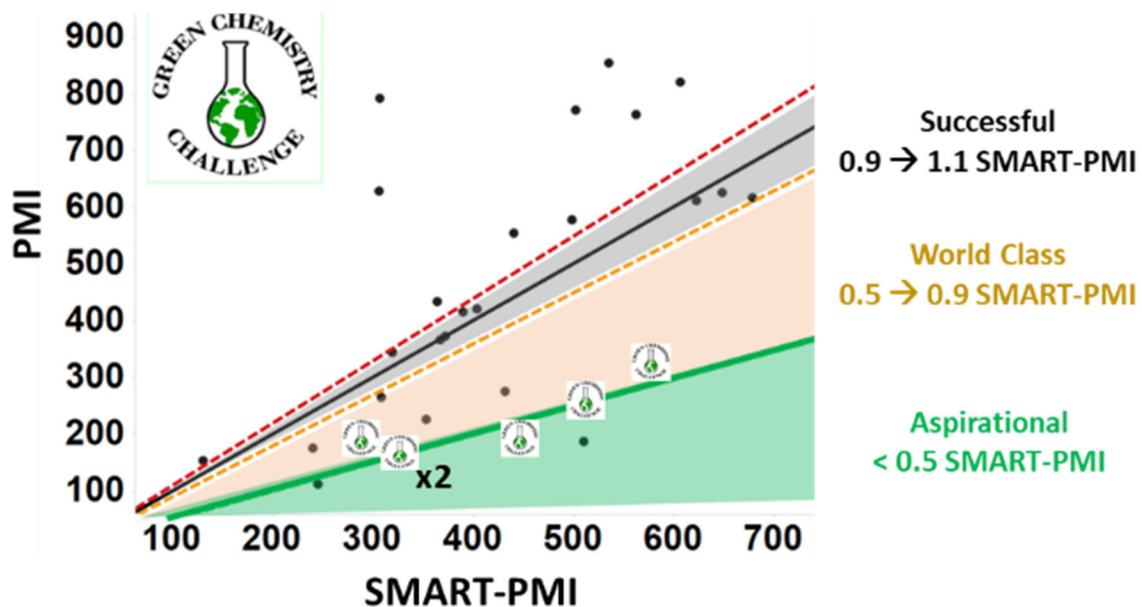
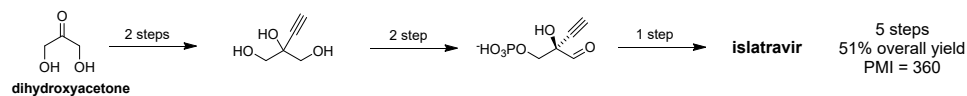
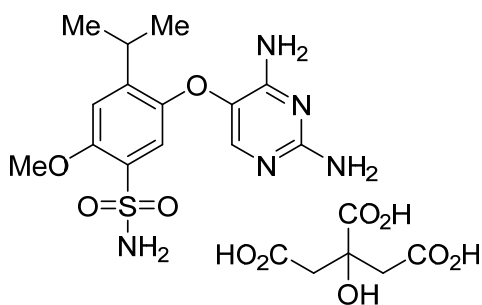
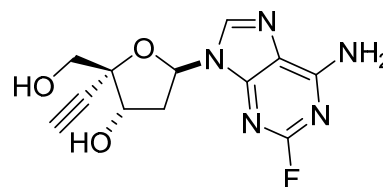


Figure 3: The SMART-PMI model with Successful, World Class, and Aspirational categories overlaid. MRL has received six EPA Green Chemistry Challenge Awards which are indicated on the plot, all of which lie on the Aspirational line.



Gefapixant citrate

Calculated complexity = 2.4 (w/o citrate)
 MW = 353 (w/o citrate)
 Predicted PMI = 189+20 = 209 (w/ citrate)
 PMI Successful Range = 216-259
 PMI World Class Range = 129-215
PMI Aspiration Range <129



Islatravir

Calculated complexity = 3.5
 MW = 293
 Predicted PMI = 404
 PMI Successful Range = 364-445
 PMI World Class Range = 202-363
PMI Aspiration Range <202

Figure 4: SMART-PMI predictions for gefapixant citrate and Islatravir. Since the model predicts based on API 2D structure alone, the step-PMI for citrate of 20 was added to the initial predicted PMI of gefapixant.

over time and ultimately reduce the waste produced in pharmaceutical manufacturing to the greatest extent over time. In the spirit of continuous improvement, the aspirational target line needs to continue to move downward.

We recently published the commercial manufacturing route to gefapixant (MK-7264).¹⁶ When our work on this API was initiated, an existing route had been proposed as the commercial manufacturing route (Scheme 1). We assessed the PMI of this route relative to both the “PMI Predictor” and our own SMART-PMI. Using the “PMI Predictor” we obtained a PMI range of 876-1482 for the API Supply Process in Scheme 1. The actual route PMI was only 366. By this analysis, one might conclude that an aspirational PMI had been achieved with this route, rendering it suitable for commercial manufacturing. However, SMART-PMI (of API with citrate) predicts a Successful SMART-PMI of 216-259 and an Aspirational target of 129 or less (Figure 4). This prediction clearly demonstrated that the current route fell short of our sustainability goals, in addition to key concerns around the use of undesirable reagents leading to higher cost of goods. This inspired us to invest in a campaign to redesign the manufacturing process resulting in a new route with a PMI of 88 (Scheme 1) and a reduction in PMI of 89%. The “PMI Predictor” indicates a PMI range of 159-248 is realistic for the new Commercial Route in Scheme 1 demonstrating the effect that this tool plays in comparing routes.

Likewise, another recent example of this impact can be gleaned from our work on islatravir (MK-8591). A highly efficient chemical process for the manufacture of clinical supplies was in place (Scheme 1). Using the “PMI Predictor”, we obtained a PMI range of 2458-3187 for the islatravir supply route in Scheme 1. The PMI of this route was >5000. However, the SMART-PMI (Figure 4) for “World-Class” range was 203-363 (Aspirational < 202). Optimizing the current route could allow us to meet the “PMI Predictor” ranges. However, it would be impossible to overcome selection of this suboptimal route to realize the improvement by an order of magnitude predicted for the “World-Class” range with SMART-PMI. As such a completely revamped and streamlined approach to this nucleoside analogue was developed leveraging a biocatalytic cascade that significantly reduces the number of steps and isolation for its preparation.¹⁷ Using an improved process based on this route, PMI was lowered by 93% to 360, in line with the “World-Class” range.

The SMART-PMI model demonstrates the ability to predict PMI from 2D chemical structure alone and can be used to gauge the success of both chemical and biocatalytic routes. No information is needed about the chemical steps or synthetic strategy. Chemists can draw a structure into a chemical sketcher and then execute the prediction of SMART-PMI in real-time, providing an instantaneous estimate with Aspirational targets that then inspire innovative chemistry needed to achieve them. As our teams continue to meet and

exceed this bar, improved PMI data is incorporated into the analysis, the model will trend toward lower SMART-PMI predictions which will lead to continued improvements in chemistry and decreased environmental impact.

ASSOCIATED CONTENT

Supporting Information. Table of PMI values. Table of descriptors. Structures of compounds used to build the model. This material is available free of charge via the Internet at <http://pubs.acs.org>. The SMART-PMI model can be accessed at <https://github.com/Merck/compoundcomplexity>.

AUTHOR INFORMATION

Corresponding Author

* email: edward_sherer@merck.com

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding Sources

Not applicable

ACKNOWLEDGMENT

We thank Chris Culberson, Rebecca Ruck, Guy Humphrey and Benjamin Sherry for helpful discussions about this work. We thank Hong Ren for naming the SMART-PMI model.

ABBREVIATIONS

MRL, the research laboratories of Merck & Co., Inc., Kenilworth, NJ, USA; MMD, the manufacturing division of Merck & Co., Inc., Kenilworth, NJ, USA; MSD, Merck & Co., Inc., Kenilworth, N.J., U.S.A., which is known as MSD outside the U.S. and Canada; others defined within text

REFERENCES

- Jimenez-Gonzalez, C.; Ponder, C. S.; Broxterman, Q. B.; Manley, J. B., Using the Right Green Yardstick: Why Process Mass intensity Is Used in the Pharmaceutical Industry To Drive More Sustainable Processes. *Org. Process. Res. Dev.* **2011**, *15* (4), 912-917.
- Cespi, D.; Beach, E. S.; Swarr, T. E.; Passarini, F.; Vassura, I.; Dunn, P. J.; Anastas, P. T., Life cycle inventory improvement in the pharmaceutical sector: assessment of the sustainability combining PMI and LCA tools. *Green Chem.* **2015**, *17* (6), 3390-3400.
- Bryan, M. C.; Dunn, P. J.; Entwistle, D.; Gallou, F.; Koenig, S. G.; Hayler, J. D.; Hickey, M. R.; Hughes, S.; Kopach, M. E.; Moine, G.; Richardson, P.; Roschangar, F.; Steven, A.; Weiberth, F. J., Key Green Chemistry research areas from a pharmaceutical manufacturers' perspective revisited. *Green Chem.* **2018**, *20* (22), 5082-5103.
- Koenig, S. G.; Bee, C.; Boroyika, A.; Briddell, C.; Colberg, J.; Humphrey, G. R.; Kopach, M. E.; Martinez, I.; Nambiar, S.; Plummer, S. V.; Ribe, S. D.; Roschangar, F.; Scott, J. P.; Sneddon, H. F., A Green Chemistry Continuum for a Robust and Sustainable Active Pharmaceutical Ingredient Supply Chain. *ACS Sustain. Chem. Eng.* **2019**, *7* (20), 16937-16951.
- Roschangar, F.; Colberg, J.; Dunn, P. J.; Gallou, F.; Hayler, J. D.; Koenig, S. G.; Kopach, M. E.; Leahy, D. K.; Mergelsberg, I.; Tucker, J. L.; Sheldon, R. A.; Senanayake, C. H., A deeper shade of green: inspiring sustainable drug manufacturing. *Green Chem.* **2017**, *19* (1), 281-285.
- Roschangar, F.; Sheldon, R. A.; Senanayake, C. H., Overcoming barriers to green chemistry in the pharmaceutical industry - the Green Aspiration Level (TM) concept. *Green Chem.* **2015**, *17* (2), 752-768.
- Roschangar, F.; Zhou, Y. Y.; Constable, D. J. C.; Colberg, J.; Dickson, D. P.; Dunn, P. J.; Eastgate, M. D.; Gallou, F.; Hayler, J. D.; Koenig, S. G.; Kopach, M. E.; Leahy, D. K.; Mergelsberg, I.; Scholz, U.; Smith, A. G.; Henry, M.; Mulder, J.; Brandenburg, J.; Dehli, J. R.; Fandrick, D. R.; Fandrick, K. R.; Gnad-Badouin, F.; Zerban, G.; Groll, K.; Anastas, P. T.; Sheldon, R. A.; Senanayake, C. H., Inspiring process innovation via an improved green manufacturing metric: iGAL. *Green Chem.* **2018**, *20* (10), 2206-2211.
- Leahy, D. K.; Tucker, J. L.; Mergelsberg, I.; Dunn, P. J.; Kopach, M. E.; Purohit, V. C., Seven Important Elements for an Effective Green Chemistry Program: An IQ Consortium Perspective. *Org. Process. Res. Dev.* **2013**, *17* (9), 1099-1109.
- Li, T.; Li, X., Comprehensive mass analysis for chemical processes, a case study on L-Dopa manufacture. *Green Chem.* **2014**, *16* (9), 4241-4256.
- Hicks, M. B.; Farrell, W.; Aurigemma, C.; Lehmann, L.; Weisel, L.; Nadeau, K.; Lee, H.; Moraff, C.; Wong, M. L.; Huang, Y.; Ferguson, P., Making the move towards modernized greener separations: introduction of the analytical method greenness score (AMGS) calculator. *Green Chem.* **2019**, *21* (7), 1816-1826.
- Borovika, A.; Albrecht, J.; Li, J.; Wells, A. S.; Briddell, C.; Dillon, B. R.; Diorazio, L. J.; Gage, J. R.; Gallou, F.; Koenig, S. G.; Kopach, M. E.; Leahy, D. K.; Martinez, I.; Olbrich, M.; Piper, J.

- L.; Roschangar, F.; Sherer, E. C.; Eastgate, M. D., The PMI Predictor app to enable green-by-design chemical synthesis. *Nat. Sustain.* **2019**, 2 (11), 1034-1040.
12. Li, J.; Simmons, E. M.; Eastgate, M. D., A data-driven strategy for predicting greenness scores, rationally comparing synthetic routes and benchmarking PMI outcomes for the synthesis of molecules in the pharmaceutical industry. *Green Chem.* **2017**, 19 (1), 127-139.
 13. Sheridan, R. P.; Zorn, N.; Sherer, E. C.; Campeau, L. C.; Chang, C. Z.; Cumming, J.; Maddess, M. L.; Nantermet, P. G.; Sinz, C. J.; O'Shea, P. D., Modeling a crowdsourced definition of molecular complexity. *J. Chem. Inf. Model.* **2014**, 54 (6), 1604-1616.
 14. MSD Corporate Responsibility Report. <https://www.msdrresponsibility.com/environmental-sustainability/>.
 15. MSD Corporate Responsibility Report: API. <https://www.msdrresponsibility.com/environmental-sustainability/product-stewardship/green-sustainable-science/>.
 16. Ren, H.; Maloney, K. M.; Basu, K.; Di Maso, M. J.; Humphrey, G. R.; Peng, F.; Desmond, R.; Otte, D. A. L.; Alwedi, E.; Liu, W. J.; Zhang, S. W.; Song, S. Q.; Arvary, R. A.; Zompa, M. A.; Lehnher, D.; Martin, G. E.; Chang, H. Y. D.; Mohan, A. E.; Guzman, F. J.; Jellett, L.; Lee, A. Y.; Spencer, G.; Fisher, E. S.; Naber, J. R.; Gao, H.; Lohani, S.; Ruck, R. T.; Campeau, L. C., Development of a Green and Sustainable Manufacturing Process for Gefapixant Citrate (MK-7264) Part 1: Introduction and Process Overview. *Org. Process. Res. Dev.* **2020**, 24 (11), 2445-2452.
 17. Huffman, M. A.; Fryszkowska, A.; Alvizo, O.; Borra-Garske, M.; Campos, K. R.; Canada, K. A.; Devine, P. N.; Duan, D.; Forstater, J. H.; Grosser, S. T.; Halsey, H. M.; Hughes, G. J.; Jo, J.; Joyce, L. A.; Kolev, J. N.; Liang, J.; Maloney, K. M.; Mann, B. F.; Marshall, N. M.; McLaughlin, M.; Moore, J. C.; Murphy, G. S.; Nawrat, C. C.; Nazor, J.; Novick, S.; Patel, N. R.; Rodriguez-Granillo, A.; Robaire, S. A.; Sherer, E. C.; Truppo, M. D.; Whittaker, A. M.; Verma, D.; Xiao, L.; Xu, Y. J.; Yang, H., Design of an invitro biocatalytic cascade for the manufacture of islatravir. *Science* **2019**, 366 (6470), 1255-1259.

