

The Power of Artificial Intelligence and Route Scouting to Navigate API Complexity

Lonza Small Molecules

As active pharmaceutical ingredients (APIs) continue to grow more complex, process chemists must reconcile more elaborate synthetic pathways with a continued push to reach the clinic rapidly. To achieve efficient API manufacturing, drug sponsors are looking for fast and effective strategies to fulfill raw material needs and avoid supply chain hiccups.

In a recent webinar hosted by Lonza Small Molecules and Elsevier Life Sciences, Dr. Ryan Littich, Head of Advanced Chemistry Technologies, Global R&D, Lonza; and Dr. Juergen Swienty-Busch, Director of Product Management for Chemistry, Elsevier Information Systems, discussed the impact of leveraging predictive retrosynthesis route design and a custom building block library to yield time and cost savings for sponsors developing complex APIs. In the post-webinar Q&A, they tackled questions on the impact of artificial intelligence (AI) on route design, cost-effective modifications, and overall timelines for route scouting.

Q: Can I always expect an answer from Reaxys predictive retrosynthesis?

Juergen Swienty-Busch: Reaxys predictive retrosynthesis has its limitations when it comes to certain compound classes like polymers or biomolecules. In well over 90% of drug-like molecules, Reaxys predictive retrosynthesis will retrieve at least one full predictive route. Most often, far more predictive routes are found and ranked for you to analyze. There is also a mode that provides possible first disconnection steps from where you can further propagate the retrosynthesis plan man-

ually to add additional steps from public or predicted sources. You should see it as a work tool that allows you to investigate certain molecules interactively because the answer is typically quite fast. This is the strength of the system.

Q: How far back in time does Lonza's building block data reach?

Ryan Littich: We cultivate decades worth of information. However, most of our data lies within the last three to four years, dating back to 2018. This is something that we're continuing to build. We are working our way backward and are very interested in learning what we can see in terms of pricing trends with different raw materials.

Q: Have you quantified the degree to which route predictions change on a large scale when certain building block libraries are activated or deactivated?

Littich: The short answer is no. We are working with our colleagues and Elsevier on how to explore that question together. Elsevier has a tremendous perspective on how calculations would change. This is something that we're eager to answer for ourselves, but to date, the answer is we have not quantified the full extent of how a particular target strategy changes as you activate and deactivate specific libraries.

Q: What do I do if I don't get a route for my target molecule?

Swienty-Busch: There are several things that you can do. The Reaxys retrosynthesis tool provides you with a set of parameters that you can tweak. For example, you can adjust the time budget for processing the target molecule, and if you recognize that the first step doesn't give a result, you can extend the time so that it can find better pathways through the predicted steps. You can also do a first disconnect step and get a lot of suggestions on where to start your retrosynthesis planning. You select certain last step suggestions, and from there you click on the starting materials and run a search for making the starting materials. You can do that interactively until you reach a point where you think that you can start it in your wet lab.

Finally, you can also interact with the system as the domain expert: just modify your target molecule in such a way that you make a first disconnection yourself and split the molecule at a bond where you believe that you can handle it, then you get a simpler set of starting materials for the predictive retrosynthesis tool.

Q: How do you keep your raw material database's pricing current?

Littich: This is something that we update on a semi-annual frequency. Every six months, we update with the most recent information and take the opportunity to try to go further backward in time.

Q: Your AI-supported routes to molecules appear to be shorter, but how do they perform against non-AI routes when undertaking experiments in the laboratory? Do you see greater success with respect to successful bond formation with an AI-supported route generation process?

Littich: Generally, predicted route options aren't creating procedural outcomes that are any different than what one would see from previous-generation informatics (which are purely analytical). As a process chemist, you're still judging what would be the best possible approach from additional options which happened to be predicted computationally... and you still have to reduce the best possible approach to practice at the benchtop. There are some interesting reports from 2018 onward, when you started to see colleagues like Elsevier, CAS, and Millipore Sigma begin to address this question. Academics and industrial partners reduced [predicted routes] to practice at the bench, showing you really can find high-yielding, shorter routes that deliver a process you would be happy to move into the clinic and then into commercialization.

It isn't that you're changing what happens at the bench; that remains the same. You [now] have a [additional] sparring partner that says, given this supporting example for a particular [predicted] step, maybe there is merit to examining the same set of reaction conditions in this next context. It is always an SME-driven exercise, and it comes down to what happens in process research, development, and scale-up. That is an excellent question and one that we keep watching and partaking in to answer.

Q: Can you set specifications for the starting materials that the retrosynthesis software goes back to, for example, < \$100 per mole?

Swienty-Busch: Yes, this is possible though not directly flexible in a way that you can enter a certain threshold like dollar amount per mole, gram, or molecular weight. It works in a more indirect way. Reaxys retrosynthesis works with building block libraries, which are nothing more than a list of SMILES strings in a text file. You can slice and dice the libraries as you like and offer them through the interface as selectable parameter options. You can offer libraries which say, all compounds that cost < \$200 per gram, have a low molecular weight, or can be delivered the next day.

Q: Do you have commercial sources for chemicals?

Littich: In Lonza's proprietary database we do have the commercial sources. We have strong intelligence on volumes exchanged and how that's changed over time. It is quite a robust database and [if you contact me], I can share statistics on the number of suppliers represented in the database as it exists right now.

Swienty-Busch: Reaxys has integrated Reaxys commercial substances, which gives you access to several hundred suppliers and a hundred million compounds with several hundred million products. Reaxys predictive retrosynthesis is also linked to that source. You can customize it, as Lonza and other companies have. Contact us and we can help you with that.

Q: Does Lonza offer this as a stand-alone service without the intention to score a development contract?

Littich: Yes, we would do that. Our incentive is to take the same experts who know [the project] really well and show our clients that the legacy of successful development and manufacturing at Lonza is an incentive to stay with us. We always want to move people into the laboratory, but the short answer is yes, absolutely. We are happy to support clients in the design phase and in their supply chain analysis to help them best prepare for entry into the clinic and to speed through pre-IND studies.

Q: Does Reaxys AI use any databases besides the Science Direct database?

Swienty-Busch: Elsevier Science Direct is our flagship product providing users with access to journals and books, but it doesn't have any searchable reactions. Reaxys uses Science Direct data and data from other publishers to create the reaction database. Right now, Reaxys has data from roughly 16,000 titles and covers a hundred million documents from publishers and patent offices around the world. The AI solution learned from the large reaction data set how to do retrosynthesis. In that sense, I would say that predictive retrosynthesis is quite unique based on the data set that we have at hand.

Interested in learning more? You can listen to the full webinar [here](#).

All trademarks belong to Lonza or its affiliates or to their respective third party owners. The information contained herein is believed to be correct and corresponds to the latest state of scientific and technical knowledge. However, no warranty is made, either expressed or implied, regarding its accuracy or the results to be obtained from the use of such information and no warranty is expressed or implied.