Advances in Computer-Aided Synthesis Planning for Medicinal Chemistry

Managing Virtual Compound Libraries and *de novo* Design with Practical Constraints

Brian Atwood, Maud Parrot, Matthew Medcalf, Rohit Arora, Hamza Tajmouati, Yann Gaston-Mathé, Nicolas Do Huu, Quentin Perron

Defining “Synthesizability”

As a starting point, we propose the following classifications, in order of decreasing accuracy and / or utility, of computer-aided synthesis planning (CASP) tools:

1. Predicts a retrosynthesis for a compound, then delivers the synthesized compound.
2. A full synthesis plan, comprised of the entire synthesis tree, planned conditions and literature references, and vendor information for all of the starting materials and reagents. The synthesis plan itself should match the price, reaction, and other constraints relevant to the specific application.
3. A retrosynthesis tree without the above level of detail or conformity to constraints.
4. A numerical value predicting the overall likelihood that a certain molecule is synthesizable, potentially with a gradient of values to distinguish easy to synthesize versus hard to synthesize from an arbitrary perspective.
5. A binary prediction: a molecule can be synthesized or cannot be synthesized.

Benchmarking CASP Tools

Modern CASP tools are capable of providing synthesizability predictions between (2) and (4), with most at the level of (3). The accuracy of these tools is now generally at the level of being indistinguishable from a human proposed synthesis plan or retrosynthesis tree, with the exception of highly complex targets such as natural products. Below are results from our own Turing test conducted with our CASP tool Spaya, and an earlier example from the literature. (see figure 2)

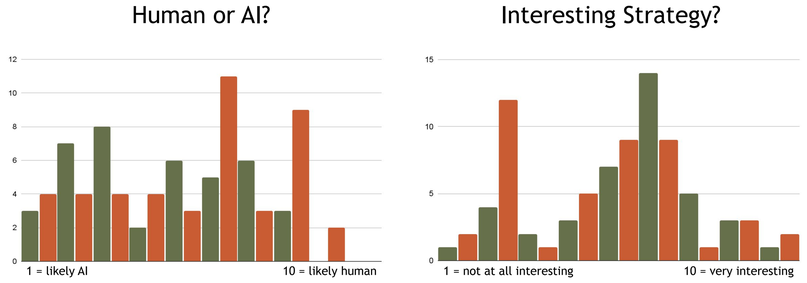
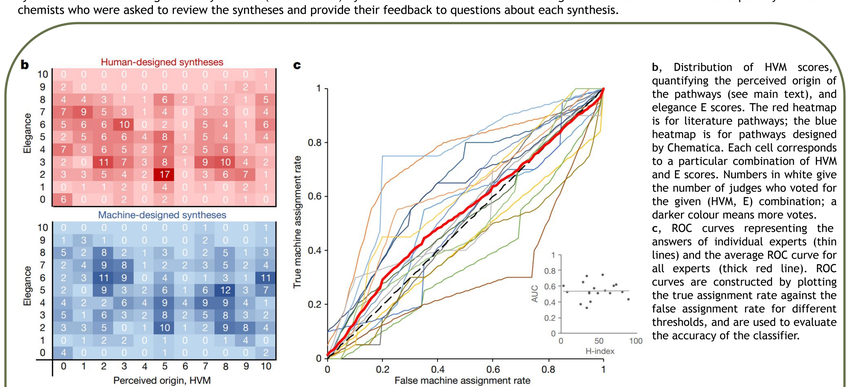


Figure 1. Spaya Turing test results. Two different synthesis schemes for 14 different small molecule drugs were generated in Spaya, one exactly matching the literature-reported synthesis and one from the best-rated AI-designed retrosynthesis which was **distinct** from the literature-reported synthesis. Sets containing randomly chosen (AI or human) syntheses for all 14 molecules were generated and distributed to expert synthetic chemists who were asked to review the syntheses and provide their feedback to questions about each synthesis.



b, Distribution of HVM scores, quantifying the perceived origin of the pathways (see main text), and elegance E scores. The red heatmap is for literature pathways; the blue heatmap is for pathways designed by Chematica. Each cell corresponds to a particular combination of HVM and E scores. Numbers in white give the number of judges who voted for the given (HVM, E) combination; a darker colour means more votes.
c, ROC curves representing the answers of individual experts (thin lines) and the average ROC curve for all experts (thick red line). ROC curves are constructed by plotting the true assignment rate against the false assignment rate for different thresholds, and are used to evaluate the accuracy of the classifier.

Figure 2, from: *Nature* 588, 83-88 (2020). https://doi.org/10.1038/s41586-020-2855-y
Inspiration for our Turing test, from the work of Mikulak-Klucznik, B., Gołębiowska, P., Bayly, A.A. *et al.* Computational planning of the synthesis of complex natural products.

For CASP tools within the same classification of capabilities, additional methods to benchmark accuracy and performance have been used:

- Top-k exact match accuracy can be used to evaluate whether the ground truth reactants fall into the list of top-k predicted reactants for single reaction steps.
- Benchmark datasets evaluate if retrosynthesis trees can be predicted by CASP tools and if the tools can correctly predict the ground truth within their results.
- Search algorithm path length and overall search times can be used to quantify the efficiency of the algorithms.

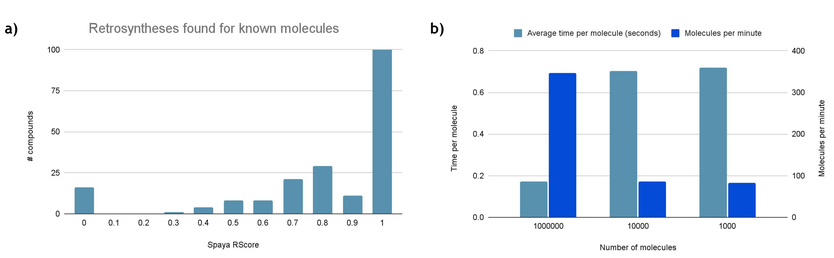
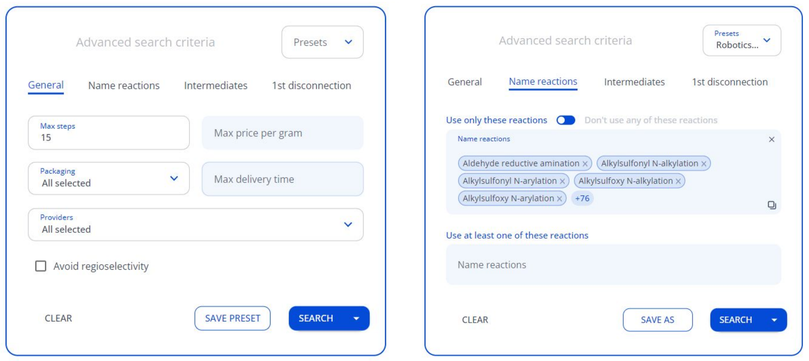


Figure 3. Benchmarking for Spaya. a) Number of solved retrosynthesis routes for known molecules b) Spaya API throughput in molecules per minute and average time per molecule

Matching Synthesis Needs and Constraints

The Spaya retrosynthesis software allows additional constraints to be placed on the system to reflect practical needs in synthesis planning:



We have used these constraints to construct preset parameter sets for synthesis planning under some hypothetical scenarios. We have found that these constraints drastically impact the performance of the algorithms as well as the nature of the results. In each case, valid results are obtained, however the results for the same molecule under different constraints often do not overlap at all.
The parameter sets we commonly use are:

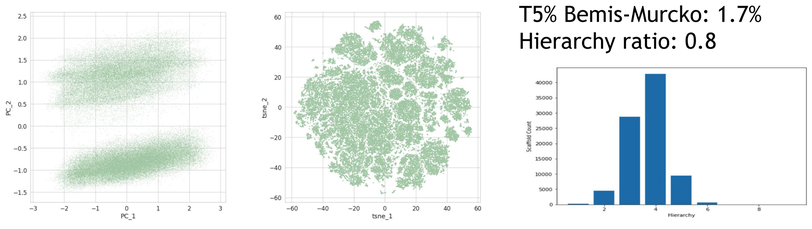
- **Depth** - starting material price: <$10/g, catalog: off-the-shelf only, providers: non-aggregators only, delivery time: <7 days to ship
- **Robotics** - starting material price: <$1000 per g, catalog: any, providers: any, delivery time: <3 days to ship, max number of steps: 5, named reactions: validated list of robotics-compatible reactions only
- **API** - starting material price: any, catalog: any, delivery time: any, max number of steps: 5, named reactions: forbid heavy-metal catalyzed/containing reactions, intermediates: forbid known genotoxic substructures and challenging-to-remove intermediates
- **Green** - starting material price: any, catalog: any, delivery time: any, named reactions: validated list of reactions compatible with green chemistry principles, intermediates: routes must pass through crystalline intermediates

Impact of Advanced Settings

Impact on Virtual Libraries

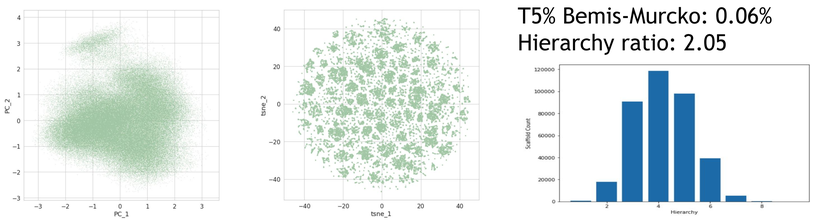
Libraries for hit discovery are often physical libraries of synthesized molecules or virtual libraries of easily synthesized molecules. These are highly valuable resources in modern drug discovery. Advances in screening platform technologies, both physical and virtual, continue to improve our access to large libraries containing diverse scaffolds.
From DELs to ultra-large virtual catalogs, the number of realistically screenable compounds is easily in the 10s of billions of compounds. DELs and many of the ultra-large virtual catalogs utilize reliable reactions and large building block catalogs to arrive at the staggering number of compounds available. There is, however, a tradeoff between reaction reliability and diversity of molecule scaffolds. A similar concern was raised regarding the lack of diversity of combinatorial chemistry for “one-bead-one-compound” strategies. On the other hand, generative AI tools have many methods available to generate highly diverse libraries. Generative chemistry suffers from well documented issues with the synthesizability of generated compounds, in some cases providing nonsensical structures.

Virtual Library Enumeration



T5% Bemis-Murcko: 1.7%
Hierarchy ratio: 0.8

Makya Generative AI



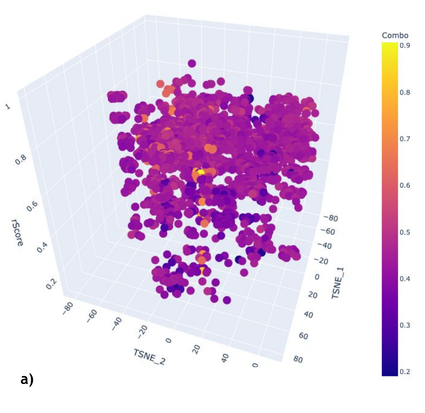
T5% Bemis-Murcko: 0.06%
Hierarchy ratio: 2.05

Figure 4. Analysis of the chemical spaces from a virtual library enumeration used to build a commercial ultra-large virtual library versus a generative AI approach.

Guiding *de novo* design with Spaya

Our approach to address the synthesizability issue in generative chemistry has been to couple generation with full retrosynthesis prediction through our CASP platform Spaya.

Spaya can be used retrospectively to triage virtual libraries based on synthesizability:



Spaya can be used during the reinforcement learning process to drive generation to synthesizable chemical space:

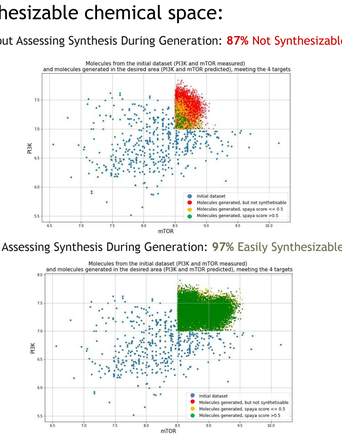


Figure 5. **a)** 3D scatter plot of generated molecules with synthesizability score along the z-axis and color based on an activity model. **b)** impact of Spaya score on molecule generation in Makya.

Impact on Synthesis with CROs

Our technology stack allows us to drive design towards certain end goals regarding compound synthesis. We have used this to drive down CRO costs for molecule synthesis by targeting lower costs designs with our generative AI while maintaining predicted activity. In one hit discovery project, our collaborators obtained price estimates for their initial set of 20 molecules at approximately $20,000 per molecule with a total delivery time of 1 year. We used Spaya and Makya to design similar compounds available at much lower cost and provided them with a new listing of 20 compounds. They entered a contract with a CRO to have the molecules synthesized at $2,000 per molecule and within 3 months received 10 compounds with 3 showing activity at the desired target.