Virscidian

In collaboration with

DEEPCURE BIOSERO Waters

The Future of Drug Discovery

DeepCure's AI and Robotics-Driven Pipeline

Introduction

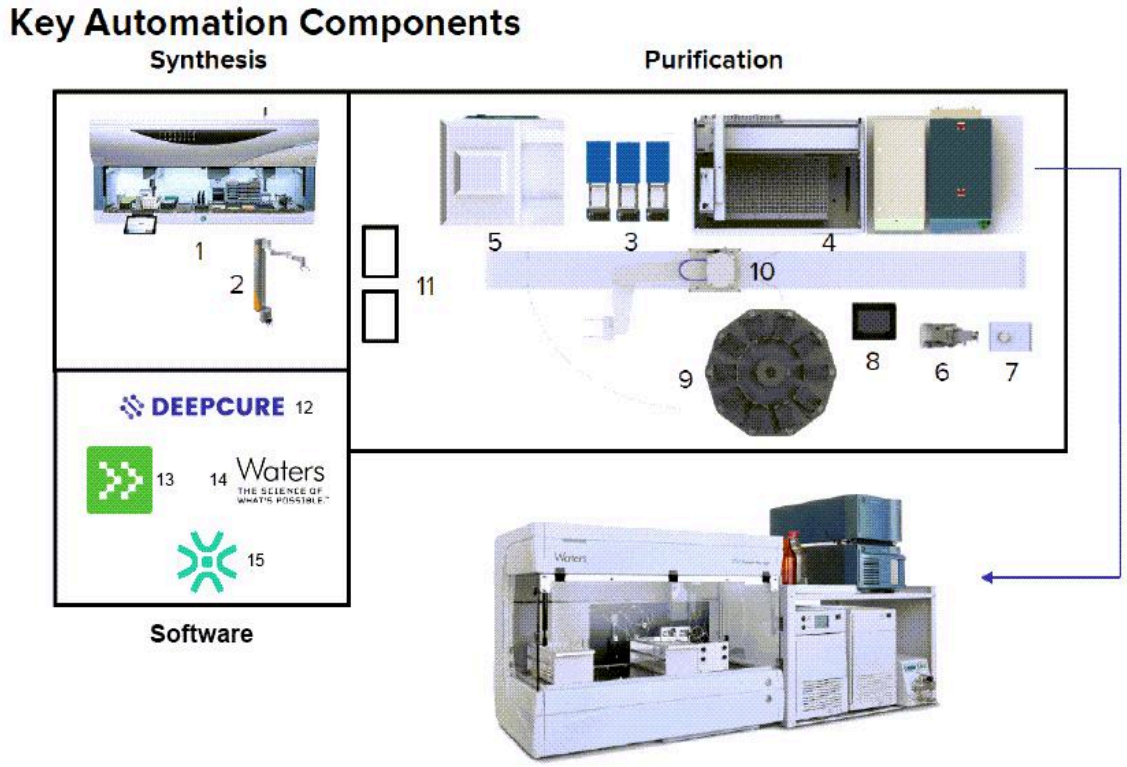
DeepCure was founded in 2018 at MIT by drug discovery scientists, AI engineers, data scientists, and biologists with a goal to use AI to accelerate the drug discovery process and develop high-quality drug candidates faster. Their vision is to eliminate the hit ID and the hit-to-lead stages of drug discovery using AI, allowing them to enter the drug development pipeline at the stage of lead optimization and proceed from there.

DeepCure has a proprietary molecular database, MolDB, containing over one trillion compounds and covering more than 75% of existing medicinal chemistry compounds. They use this database to rapidly predict lead-like compounds and then, their machine learning algorithm designs a synthetic route to the lead compound autonomously. The complete multi-step synthetic route is transferred to their robotic platform at their molecular foundry in Israel where the lead-like compounds are synthesized automatically. Due to the inherent complexity in the structure and synthesis of lead-like compounds, the synthesis platform that DeepCure designed includes compound synthesis as well as automated purification, solvent removal, and compound characterization.

DeepCure has a goal of automatically synthesizing 5,000-10,000 new compounds each month using complex synthetic reactions comprising of up to 9 different steps and dozens of types of reactions. Starting with just the computer-generated list of target compounds as the sole input, they aim to have fully synthesized lead-like compounds that have been fully purified and characterized as the output. The information gleaned from the entire process will be fed back into their algorithms, improving, and expanding their capabilities over time.

Experimental

All the reactions are carried out in 96 well microtiter plates with robotics designed for that format. Using microtiter plates for their automated synthesis does limit the amount of final product that can be collected for any given experiment; however, this was an acceptable compromise given the advantages of this format. A diagram of the hardware layout is shown in Figure 1. On the left is the Tecan Fluent 1080 liquid handler (1) used to dissolve any powders into liquids, mix liquids, and plate compounds. The KX2 Robotic arm (2) passes plates to the handoff nests (11) for the next steps in the process. The additional robotic arms, a Brooks PF400 (10) or a Mecademics Meca500 (6), are able to move the plates to the needed instrumentation in the next step. These include the Hettich CombiDancer for evaporating solvents (5), the Inheco Incubator/shaker (3), a plate hotel (9), a Ziath Mirage barcode reader (8), or a Mettler Toledo balance (7). DeepCure uses a Waters™ LC Prep AutoPurification System (4) and Virscidian® Analytical Studio software for separation, fraction collection, and analysis of their intermediate and final products. This entire process is fully automated, requiring just the target compounds as the input.



1. *Tecan Fluent 1080*
2. *KX2 Robotic Arm*
3. *Inheco Incubator/Shaker*
4. *Waters™ LC Prep AutoPurification System*
5. *Hettich CombiDancer*
6. *Mecademic Meca500*
7. *Mettler Toledo Balance*
8. *Ziath Mirage BC reader*
9. *Carousel*
10. *Brooks PF400 Robotic Arm*
11. *Handoff Nests*
12. *DeepCure Galactus*
13. *Biosero GBG*
14. *Waters MassLynx and FractionLynx*
15. *Virscidian Analytical Studio*

To automate their workflow, DeepCure uses their own proprietary software as well as several commercially available packages. Their Galactus software essentially reverse-engineers each desired lead-like compound and designs the entire multi-step synthesis process required to generate them. Biosero® Green Button Go® (GBG®) Scheduler software schedules and executes all necessary activities across the lab, and Waters MassLynx™ and FractionLynx™ application manager control the LC/MS system that runs all compounds generated by the synthesis and purification process. DeepCure relies on Virscidian Analytical Studio software for automatically processing and analyzing of all the LC/MS data as it is acquired, and automatically generating sample lists for each subsequent stage of analysis based on the data from the previous stage. Analytical Studio is a vendor-agnostic software package designed to automate all the chromatography/mass spectrometry processing steps throughout a workflow, making the same decisions that experienced laboratory scientists would make, only faster and more reproducibly. It is also able to create instrument sample lists for additional procedures in the high throughput workflow. These cutting-edge software tools are critical to the success of DeepCure's ambitious vision.

Workflow in Detail

The steps driven by GBG® in the synthesis and purification workflow are shown in Figure 2. The first step of the process is the user providing a list of target compounds to Galactus. Galactus creates the synthetic route, complete with the weights and volumes of each chemical needed for every step of the synthetic process. Galactus provides this work list to GBG to direct the remainder of the work. The liquid handler plates the necessary chemicals into the wells and moves the crude reaction plate to the nest where it is picked up and moved to the incubator/shaker instrument to begin the reactions. The plate layouts are arranged so that different reactions can be performed simultaneously on the same plate using the same conditions to maximize throughput. After the reactions are complete, the crude reaction plate is moved back to the liquid handler where it is filtered, and an analytical plate is created. The analytical plate is identical in content to the filtered plate except the samples are diluted for analysis by chromatography/mass spectrometry.

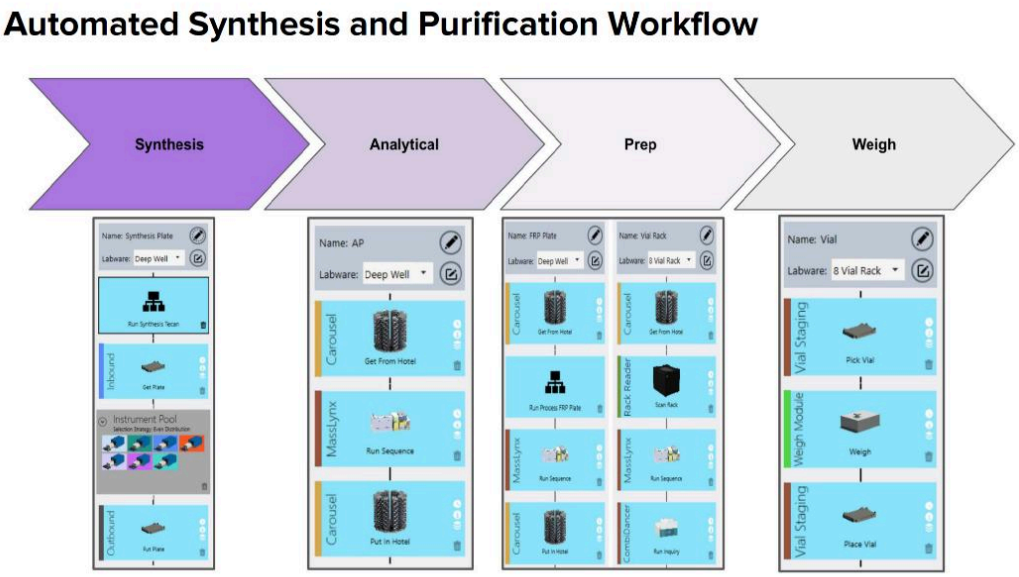


Figure 2. Illustration of the automated synthesis and purification steps driven by GBG® Scheduler software Biosero®.

The analytical plates are transferred to the Waters Prep LC AutoPurification system for compound purification and fraction collection. GBG transfers the master worklist containing the molecular formula for each target compound to Waters AutoLynx application manager. AutoLynx enables entire batches of samples to be submitted to the MassLynx queue for automatic acquisition, data processing, and reporting. Collected mass spectrometry data files from MassLynx are automatically copied by Virscidian's Analytical Studio software and processed using the methods developed for DeepCure by Virscidian's experts.

For DeepCure's purification workflow, Analytical Studio processes the data from the analytical plate, first determining whether the target compound was found in the correct wells. If the target compound is not found that sample is not processed further. If the target compound is detected, the software determines whether the compound is purifiable based on the criteria specified in the method, including peak shape, area counts, co-eluting impurities, and multiple other customized conditions. Analytical Studio will also automatically evaluate the results from various chromatography methods, if they have been acquired, select which method will yield the most purified product, select the appropriate focused gradient for preparative runs, and create the prepLC sample lists containing the suitable methods.

Though Analytical Studio is designed for very high throughput workflows where minimal, if any, data review is performed, it features a user-friendly interface for those who want to review results and confirm that decisions made by the software are in line with decisions the user would make. Figure 3 shows an example of a typical data layout for a purification workflow in Analytical Studio. The user

layout can be easily customized to suit each user's preferences. Here, on the top left is a representation of the 96-well sample plate. If the same sample or plates are run using multiple chromatography methods, each separate method is represented on its own plate. For example, if a plate is acquired using a high pH chromatography method and a low pH method, the plate view panel will show two plates, with each plate displaying results from one chromatography method. Samples can be linked so that when the user clicks on a well in one plate, a dotted line is shown around any of the wells that also contain that compound throughout all phases of the purification and fractionation process.

Analytical Studio automatically selects which method will result in the best fraction(s) being collected. The definition of "best fraction" is adjusted depending on customer requirements and could be the purest fraction or the most volume of compound of interest for example.

The color coding shown in Figure 3 is used throughout Analytical Studio to simplify data review. The dark green indicates that the target compound was found with both a high purity and no impurities eluting nearby, so fraction collection will yield high purity results. The color intensity gives a rough indication of the relative amount of product compared to impurity, with darker colors indicating more of the compound. As the color fades from dark green to light green, this indicates that while the target com pound is still present, the purity is lower so there's likely to be less of that compound collected. The wells shown in red indicate that the target compound is not found. The ones in purple indicate that there are co-eluting or closely eluting impurities, resulting in some risk to fraction collection. In these situations, Analytical Studio will pick the best prep method and fraction collection triggers, but due to the closely eluting impurities, success is not certain.

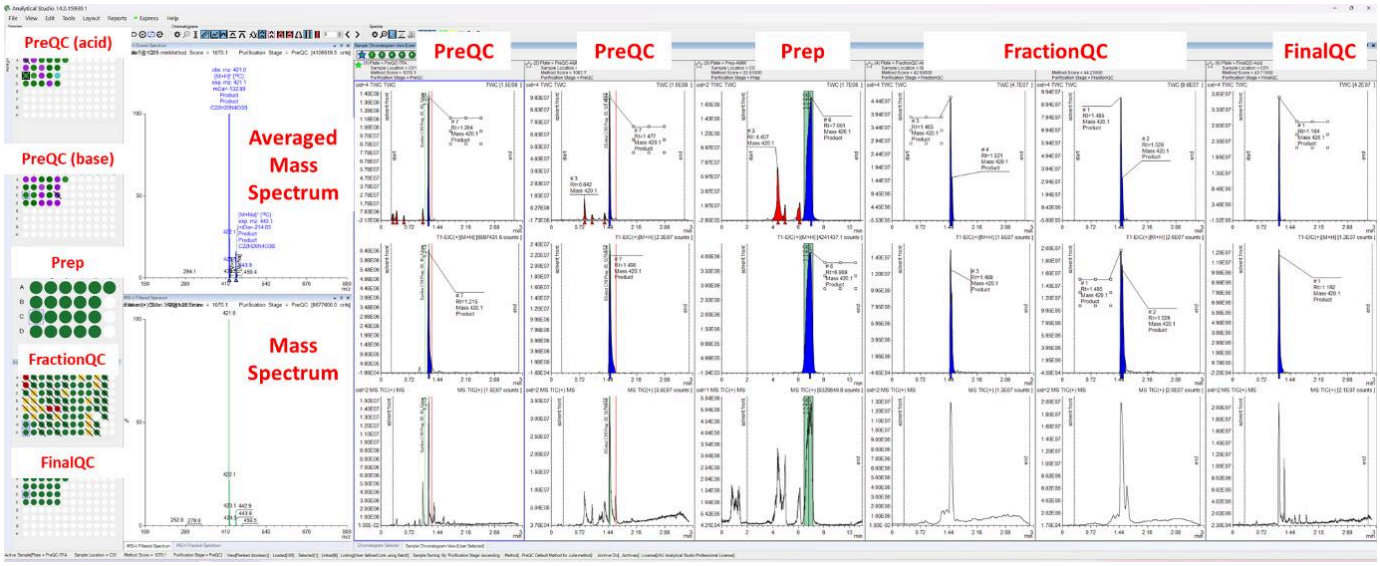


Figure 3. Possible Analytical Studio data layout showing data from all phases of purification process. Data is linked making it simple to follow a given compound from PreQC through to the Final QC stage. The best chromatography method for purification is indicated by the color coding, as is the likelihood of collecting a clean sample of the compound of interest.

Figure 4 shows the mass spectrum and UV traces for a well that indicated there was a risk to clean fraction collection if an acidic mobile phase is used, due to an impurity co-eluting close to the compound of interest. When acquired using a basic mobile phase, no risk to sample collection was found and as such, Analytical Studio recommended proceeding with preparatory chromatography and fraction collection using the basic method, as indicated by the green star in the header.

Figure 5 shows a UV chromatogram for a target compound that was purified using a focused gradient to better separate the compound of interest from a closely eluting impurity. The target compound is shown on the chromatograms in between a green and red line, indicating start and stop of the focused gradient, respectively. The label along the green line indicates that, for the target compound in Figure 5, a focused gradient ramping from 5-35% organic in 12 minutes will be used for the prep method.

Figure 5 is a total wavelength chromatogram showing the beginning and ending points for a focused gradient. The focused gradient, highlighted in green, was automatically selected by Analytical Studio to maximize the purity and/or amount of sample collected in the fraction.

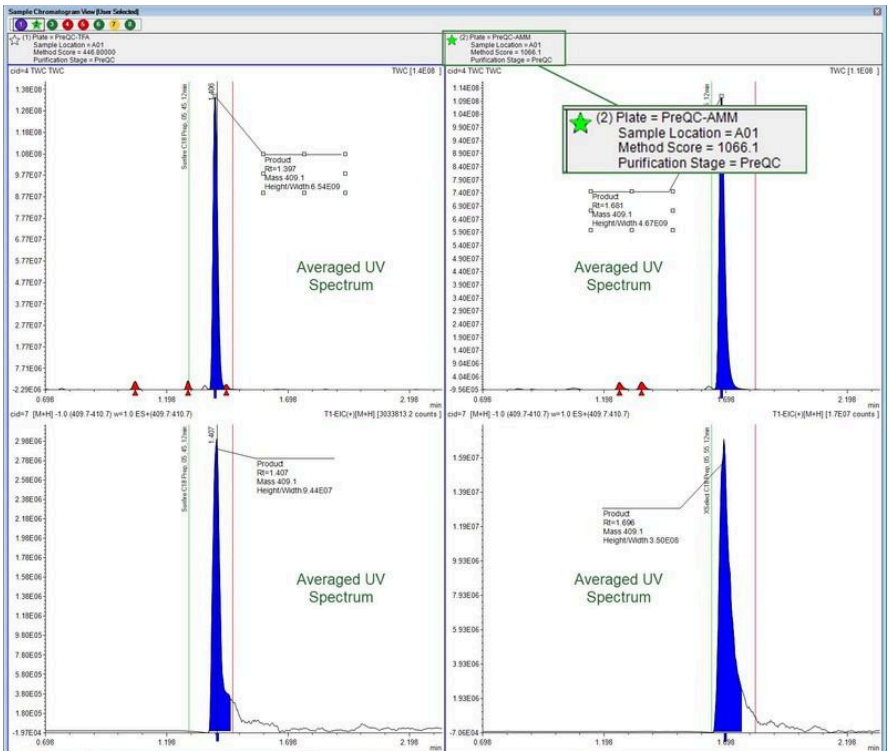


Figure 4 shows both the UV (TWC) and extracted ion (EIC) chromatograms for a compound acquired using both an acidic (left panel) and a basic (right panel) method. Results indicate that the basic method will give the best chance of collecting a high-purity fraction. The best method is indicated by the green star enlarged at the top of the figure.

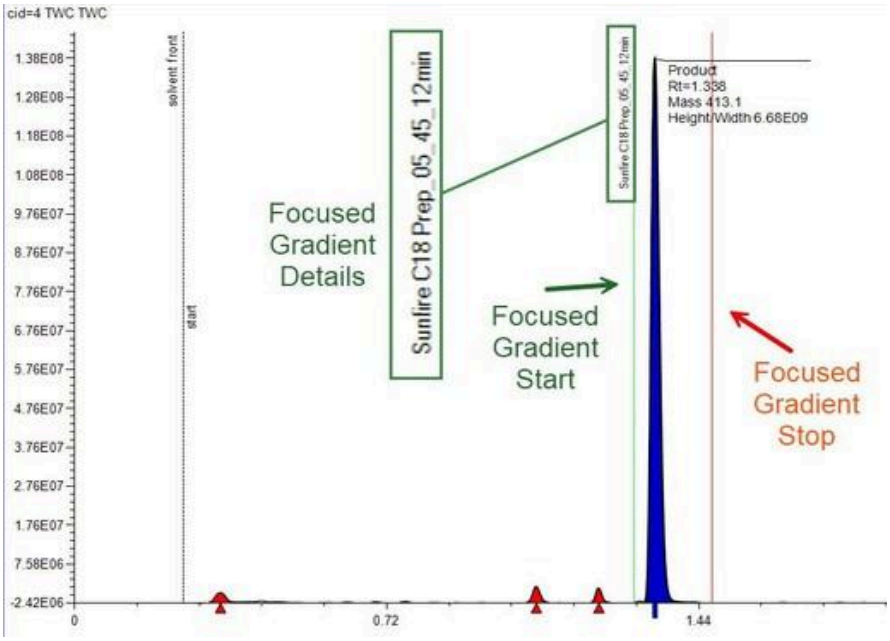


Figure 5 is a total wavelength chromatogram showing the beginning and ending points for a focused gradient. The focused gradient, highlighted in green, was automatically selected by Analytical Studio to maximize the purity and/or amount of sample collected in the fraction.

Sample Attributes		Substances	Expressions					
Container	Sample Name	Product Mas	PreQC-PrepOK	MethodScore	MethodScore	Generic-Prep-Lo Trigge	Fraction Trigger Logi	Generic-Prep-Wavele
[redacted]	[redacted]	515.24	☑	1070.8	70.8	Med_Med	Mass (A or B) and U...	UV220
[redacted]	[redacted]	561.26	☑	1073.2	73.2	Hi_Med	Mass A and UVA	UV340
[redacted]	[redacted]	483.22	☑	1073.3	73.3	Hi_Med	Mass (A or B) and U...	UV340
[redacted]	[redacted]	473.20	☑	1072.3	72.3	Hi_Med	Mass (A or B) and U...	UV340
[redacted]	[redacted]	560.25	☑	1073.8	73.8	Hi_Hi	Mass A and UVA	UV340
[redacted]	[redacted]	645.24	☑	2068.5	68.5	Low_Low	Mass A and UVA	UV220
[redacted]	[redacted]	513.23	☑	1074.9	74.9	Hi_Med	Mass (A or B) and U...	UV340
[redacted]	[redacted]	489.19	☑	2072.8	72.8	Low_Hi	Mass A	UV255
[redacted]	[redacted]	519.19	☑	1070.1	70.1	Med_Med	Mass A and UVA	UV220
[redacted]	[redacted]	541.29	☑	1074.3	74.3	Hi_Hi	Mass A and UVA	UV340
[redacted]	[redacted]	499.21	☑	2074.4	74.4	Low_Med	Mass A and UVA	UV340
[redacted]	[redacted]	583.24	☑	1069.5	69.5	Med_Hi	Mass A and UVA	UV220
[redacted]	[redacted]	555.31	☑	1073.2	73.2	Hi_Med	Mass A and UVA	UV340
[redacted]	[redacted]	546.26	☑	1070.8	70.8	Hi_Med	Mass (A or B) and U...	UV340
[redacted]	[redacted]	536.24	☑	1075.2	75.2	Hi_Med	Mass (A or B) and U...	UV340
[redacted]	[redacted]	553.29	☑	1073.0	73.0	Hi_Med	Mass A and UVA	UV340
[redacted]	[redacted]	513.23	☑	1073.9	73.9	Hi_Med	Mass (A or B) and U...	UV340
[redacted]	[redacted]	523.28	☑	1072.4	72.4	Hi_Med	Mass (A or B) and U...	UV340
[redacted]	[redacted]	599.21	☑	1074.6	74.6	Hi_Med	Mass A and UVA	UV340
[redacted]	[redacted]	560.26	☑	1074.6	74.6	Med_Med	Mass (A or B) and U...	UV340
[redacted]	[redacted]	527.28	☑	1073.7	73.7	Hi_Med	Mass (A or B) and U...	UV340
[redacted]	[redacted]	583.24	☑	1074.7	74.7	Hi_Hi	Mass A and UVA	UV340
[redacted]	[redacted]	535.26	☑	1073.1	73.1	Hi_Hi	Mass A and UVA	UV340

Figure 6. An example of Analytical Studio's On-Screen Report summarizing the trigger logic and thresholds for collecting fractions. These are automatically chosen by Analytical Studio to maximize purity and can be set for each sample individually.

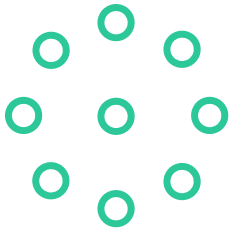
> "Analytical Studio has become a vital component of our automated synthesis and purification processes."
>
> **LUKE NAM, AUTOMATION ENGINEER**
> **DEEPCURE**

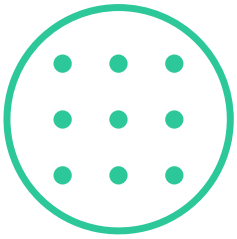
In addition to selecting the best focused gradient, Analytical Studio also determines the triggering method for fraction collection. Figure 6 shows an On-Screen Report, which provides a summary of how fractions will be collected. These triggers are automatically selected for each sample individually yielding far superior results than those achieved when the same triggers even for every sample. Analytical Studio can alter various settings for fraction collection triggers, such as intensity thresholds, which detector should trigger fraction collection, and even specific wavelengths used to trigger fraction collection. The capabilities built into Analytical Studio assured DeepCure that they would be able to keep up with their throughput goals while still generating high quality fractions that can be used for further experiments.

Virscidian DEEPCURE BIOSERO Waters™

Conclusion

DeepCure's unique drug discovery/development model allows them to skip the hit identification and hit-to-lead phases of the drug discovery process and move straight to lead optimization, allowing them to accelerate the timeline of getting new drugs to market. They have developed a proprietary database of lead-like compounds, along with software, that reverse engineers the synthesis of these lead-like compounds. This software, Galactus, generates the step-by-step instructions required to synthesize these complex molecules and integrates with Biosero® GBG® software, allowing them to automate the complete synthesis protocol. The only input by scientists is the list of target compounds to be synthesized. Additionally, DeepCure integrated a fully automated purification and fraction collection procedure into their workflow, necessitated by the complex, multi-step synthesis process required to yield lead-like compounds.





Virscidian

Contact Us Today

US Office (+919) 809-7651
info @virscidian.com
www.virscidian.com

Follow Us

linkedin.com/company/virscidian-llc/

Virscidian's Analytical Studio software is integrated with Waters' LC Prep AutoPurification system and Biosero GBG software and automatically processes the chromatography mass spectrometry data after data collection. Analytical Studio ensures that the target compounds are present in the intended wells and selects the optimum chromatographic methods and fraction collection triggers for each sample, all without any user interference. Additionally, Analytical Studio can determine synthetic reaction yield and how pure the collected fractions are. Thanks to the advanced software employed at DeepCure, they will be custom synthesizing 5,000-10,000 compounds each month and rapidly advancing new leads through the drug development pipeline.

Dotmatics