

DNA-encoded Libraries Meet Artificial Intelligence

Sept. 7, 2022 by Andrii Buvailo

The concept of deoxyribonucleic acid (DNA)-encoded chemical libraries was invented three decades ago, but only recently it has become a mainstream screening paradigm in drug discovery -- primarily due to exponential progress in our ability to manipulate and sequence DNA. Another key technology that was instrumental during the rise of DNA-encoded libraries is combinatorial chemistry.

DNA-encoded libraries (DEL) technology provides an opportunity to screen chemical spaces of unprecedented size -- not only billions but trillions of molecules in a single vial. For example, HitGen's libraries contain more than 1 trillion molecules, X-Chem provides a collection of libraries covering over 7.5 billion compounds for screening, while Nuevolution assembled a collection of 40 trillions compounds (in 2019, Nuevolution was acquired by Amgen for \$167 million).

Since DEL technology offers access to essentially the largest chemical space available on the market, this big data technology is a natural fit for big data analytics and modeling technologies offered by the field of artificial intelligence -- and some companies have already been chasing this opportunity.

A notable deal took place in 2020, when Insitro, a leading player in the application of machine learning for drug discovery, founded by Daphne Koller, acquired Haystack Sciences. Haystack's machine learning-based platform combined multiple elements of their DEL technology, including the capability to synthesize broad, diverse, small molecule collections, the ability to execute iterative follow-up, and a proprietary semi-quantitative screening technology, called nDexer™, that generates higher resolution datasets than possible through conventional 'panning' approaches.

In 2020, Google Research published results of their collaborative effort with X-Chem, one of the leading players in the DEL technology space, in an article "Machine learning on DNA-encoded libraries: A new paradigm for hit-finding", where they demonstrate an effective new method for finding biologically active molecules using a combination of physical screening with DNA-encoded small molecule libraries and virtual screening using a graph convolutional neural network (GCNN). This research has led to the creation of the Chemome initiative, a cooperative project between our Accelerated Science team and ZebiAI, a platform that applies massive experimental DNA encoded library data sets to power machine learning for drug discovery.

In its turn, ZebiAI was acquired in 2021 by another notable developer of artificial intelligence-powered drug discovery platform, a clinical stage biotech Relay Therapeutics (NASDAQ: RLAY), where Relay paid \$85 million up-front. This acquisition allowed Relay to incorporate ZebiAI's machine-learning-based DEL technology into their protein targeting platform Dynamo.

READ ALSO: [A Market Review Of DNA-encoded Libraries Technology In Drug Discovery](#)

In October 2021, X-Chem acquired Glamorous AI, a developer of a modular multifaceted artificial intelligence solution for drug discovery RosalindAI, including capabilities of data engineering and featurization, predictive analytics, high-performance computing, and de-novo drug design.

Notably, half a year later, Enko, the crop health company, announced its acquisition of proprietary DNA-encoded libraries and technology enablement from X-Chem. With that deal, Enko announced to be the first company to internalize this technology for agricultural applications. DNA-encoded libraries are a proven tool to kickstart pharmaceutical drug discovery projects because of the efficiency of screening large, diverse chemical spaces. Enko pioneered applying this approach to tackle agriculture challenges. The company said this acquisition is a significant advancement for Enko's ENKOMPASS™ platform, which combines DNA-encoded library screening with machine learning and structure-based design to find novel chemistry and new modes of action to control crop pests and diseases.

HitGen is a Shanghai Stock Exchange listed company, one of the leaders in the DEL technology market. In November 2021, HitGen and Cambridge Molecular announced a strategic partnership, bringing together HitGen's DNA-encoded libraries and DeepDELve 2 – Cambridge Molecular's highly optimized DEL-specific deep learning system. DeepDELve 2 is fully integrated with Enamine REAL space, Mcule, Chemspace, and eleven other compound catalog providers, and can be linked on demand to commercial or internal catalogs of choice. As such, potential ligands are often delivered to a partner's lab in as little as a few days.

Another notable project to implement machine learning on large DNA-encoded libraries, "DEL-ML-CS", was implemented by ChemSpace, a leading chemical provider of small molecules for pharmaceutical research. The team at ChemSpace decided to implement the idea, proposed by Kevin McCloskey et al. in a 2020 paper "Machine Learning on DNA-Encoded Libraries: A New Paradigm for Hit Finding" to combine machine learning and DNA-encoded libraries (DEL) screens. Using a DEL screen, it is possible to access vast chemical space in a single experiment. This helps to get substantial amounts of data faster and with less resources required. This data can be further used to build a relevant machine learning model and then apply it on the ultra-large chemical space of small molecules provided by ChemSpace -- allowing for

a high probability to identify potent hits from the first screen.

Within the scope of this project, ChemSpace provides access to OpenDEL, a HitGen's new encoded library of over 145 million drug-like small molecules that can be screened using only 2 mg of target protein, at costs comparable to HTS on much smaller decks. The OpenDEL library is based on 70 different templates and is designed for high diversity within drug-like space.

The above-mentioned collaborations illustrate a growing interest in the application of artificial intelligence to DNA-encoded chemistry -- for drug discovery and agricultural applications.