

Navigating In REAL Chemical Space To Find Novel Medicines (Now 3.8 Billion Molecules)

July 27, 2018 by Andrii Buvailo

[Latest update: 27.07.2018]

February 6 2018 will be marked in history as the day when an automobile embarked on a million year trip into far space. On that day, Falcon Heavy, the 230ft megarocket, successfully launched from Cape Canaveral, Florida, bringing on board Elon Musk's personal Tesla Roadster with a crash test dummy called 'Starman' strapped into the driver's seat.

"The imagery of it is something that's going to get people excited around the world." says Elon Musk, a guy who literally revolutionized the area of space travel having developed and successfully tested first reusable rockets in the aerospace history, opening new horizons for cosmos explorations.

While advances in aeronautics are making news headlines daily, a much quieter scientific revolution is also happening in the area of "chemical space" explorations -- the one which might soon uncover novel medicines to cure diseases.



("Starman In Chemical Cosmos", digital painting by Andrii Buvailo)

Computers help guide scientists in ‘chemical cosmos’

To navigate the chemical universe with better hopes, it helps to have a kind of map to know where to “dig for gold”, rather than attempting to search in random regions. This is why computational methods are increasingly adopted in drug discovery programs to provide guidance in chemical space and decrease R&D costs.

A notable example of using computers to assist drug hunters is described in a recent Nature paper “The drug-maker’s guide to the galaxy”. In 2001, chemist Jean-Louis Reymond, at the University of Berne in Switzerland began a to draw a virtual chemical map with the ambition to cover as much of the massive space as possible. It took him sixteen years to amass the largest virtual database of chemical compounds in the world, GDB-17, which included 166 billion small molecules made up of maximum 17 atoms. He then grouped compounds according to 42 characteristics in a multidimensional space in which neighboring molecules have similar properties.

Having a database like this, one of the options to move forward for drug hunters is by using a particular known drug as a starting point -- to identify structurally similar but pharmacologically better versions of the compound in the neighboring regions of chemical space. Reymond and his team used this approach to

identify a shortlist of 344 related compounds and later confirm that two of them could potentially activate the nicotinic acetylcholine receptor, a useful target for disorders involving the nervous system or muscle function, and could be useful for treating muscular atrophy in aging.

An alternative approach includes the use of molecular-docking algorithms to screen vast chemical libraries in silico to find small molecules that bind to a given protein for which its structure is known from X-ray crystallography experiments. Chemists at the University of California, San Francisco, led by Brian Shoichet, used this approach in 2016 to search for a new class of painkiller. The team conducted virtual screening of more than 3 million commercially available compounds with the aim of identifying drug candidates that would selectively activate μ -opioid receptor signaling to relieve pain without disturbing the closely related β -arrestin signaling pathway — the latter being considered to be associated with opioid side effects (lowered breathing rate, constipation). The researchers quickly narrowed down a huge compound library to just 23 likely candidates, 7 of which were found later to possess the desired activity.

Artificial intelligence (AI) is the next step

Latest advances in machine learning (ML) algorithms stimulated a surge of interest in applying artificial intelligence (AI)-driven approaches to chemical space explorations. A wave of AI-driven startups, such as Exscientia, Atomwise, Cloud Pharmaceuticals, and Recursion Pharmaceuticals, to name a few, are utilizing machine's exceptional abilities to cope with big data and find unintuitive novel patterns and insights. Recently, a biotech company Juvenescence Ltd announced about licensing the first purely "AI-born" drug candidate JAI-001 and its analogs from drug discovery startup Insilico Medicine.

"The selection of our first compound family is a landmark event for Juvenescence, and a broader comment on the potential of AI to transform the drug discovery and development industry," said Jim Mellon, Chairman of Juvenescence.

(To read more about key use cases for AI application in pharmaceutical R&D, read Biopharma's Hunt For Artificial Intelligence: Who Does What?)

Finding firm grounds in the world of uncertainty

One of the caveats of computational approaches is their notorious habit of suggesting drug candidates that are nightmares to synthesize in a lab. Chemists have to figure out a recipe for the suggested compound, which can take months or hard and costly work. Since there is not guarantee that the molecule will work once it is made, it represents the risks of wasting time and resources. For example,

Reymond's approach predicts a compound's activity profile correctly only 5–10% of the time, meaning that chemists have to figure out synthesis route (sometimes, extremely difficult, if possible) for up to 20 compounds to end up with just one that acts as expected.

“I would say the bottleneck in our exploration of chemical space is the ability to dare to make compounds,” says Reymond. This problem is a consequence of working with purely virtual chemical spaces -- those which were artificially “designed” without having their synthesizability in mind.



A recently announced collaboration between a chemical supplier Enamine and a cheminformatics software vendor BioSolveIT was a major step towards overcoming synthesizability issues pertaining to chemical space explorations. The newly introduced REAL Space Navigator software allows for ultrafast searches in Readily Accessible (REAL) chemical space containing 650 million compounds, any of which can be synthesized with a guaranteed success rate of 80% and above, within 3-4 weeks.

Update: REAL database has been expanded to include 3.8 billion synthetically accessible molecules.

This high synthesizability rate for all compounds from REAL chemical space is achieved by creating this space through appropriate single step combinations of the 150,000 building blocks available in stock and validated in parallel synthesis conditions while using only 106 most reliable reaction protocols developed and elaborated in numerous laboratory experiments.

The REAL chemical space can be a starting point for not only hit exploration programs but also as a training set for AI-based models to address synthesizability issues, for instance, while performing denovo drug design.

- Insilico Medicine