



HHS Public Access

Author manuscript

Nat Rev Drug Discov. Author manuscript; available in PMC 2020 February 25.

Published in final edited form as:

Nat Rev Drug Discov. 2019 February ; 18(2): 83–84. doi:10.1038/d41573-018-00007-2.

Mapping biologically active chemical space to accelerate drug discovery

G. Sitta Sittampalam*, Dobrila D. Rudnicki, Danilo A. Tagle, Anton Simeonov, Christopher P. Austin

National Center for Advancing Translational Sciences National Institutes of Health Bethesda, MD, USA.

Abstract

A specialized platform for innovative research exploration — ASPIRE — in preclinical drug discovery could help study unexplored biologically active chemical space through integrating automated synthetic chemistry, high-throughput biology and artificial intelligence technologies.

With increasing understanding of the molecular basis of disease in the last 30 years, a major roadblock to timely translation into new therapies has been the inability to efficiently identify new areas of biologically active small-molecule chemical space¹. Ideally, new chemical probes and drug leads that selectively modulate disease targets and pathways would be produced rapidly and inexpensively, but despite some progress in the past decade², the fundamental challenge of exploring chemical space to define new biology remains largely unsolved. Recently, however, advances in chemistry automation and machine learning/artificial intelligence (AI)³ have raised the prospect of their integration with high-throughput biological screening, assay automation engineering and informatics to enable dramatically more effective, even unsupervised, exploration of biologically active chemical space.

Challenges in chemical space exploration

In its simplest terms, the goal seems straightforward: to define the set of small-molecule chemical structures needed to modulate all biological targets. However, the vast number of chemical structures in drug-like chemical space ($\sim 10^{60}$), and the smaller but still substantial number of biological targets in human and pathogen biological space ($\sim 10^6$), has made progress on this problem painfully slow. Currently, only $\sim 3\%$ of biological space is drugged and a further $\sim 7\%$ is tractable via small-molecule probes¹, while the percentage of drug-like chemical space that has been synthesized is miniscule².

The effort to define biologically active chemical space involves four main disciplines: biology, chemistry, informatics and engineering. In the last three decades, automation and parallelization have radically improved the efficiency of biological testing, informatics analysis and engineering. High-throughput screening (HTS) technologies have dramatically

*gurusingham.sittampalam@nih.gov.

increased the productivity of the bench biologist such that millions of data points can be acquired in a single day. Advances in the capabilities, precision and robustness of engineering technologies at the micro- and macro-level have also enabled increasingly autonomous physiologically relevant biological screening systems. And remarkable advances in computing power and data analysis algorithms have increased the ability to analyse data by orders of magnitude in quantity and quality. These capacities have, in turn, allowed the development of data-driven principles of biological function.

By contrast, the technologies, throughput and reach of synthetic chemistry have remained relatively unchanged over the last several decades, with combinatorial chemistry, microwave synthesis and other technologies having only limited overall impact on the efficiency of chemistry to explore new chemical space (Supplementary Fig. 1). Chemistry has only recently begun adopting automation and AI technologies to facilitate existing chemistries, reaction optimization and nanoscale synthesis and library generation³, and the general practice of chemical synthesis remains largely artisanal, with synthetic throughput of novel bioactive chemicals improved at best by tenfold over the last century. This disparate evolution of the biology and chemistry fields now limits the ability to generate novel chemical probes⁴, pharmacological tools and drugs to modulate undrugged biological space, and thus contributes to translational research inefficiency.

Machine learning and other AI technologies are increasing in use and sophistication, and they learn, interpret and predict outcomes based on vast amounts of data in applications such as facial recognition and driverless vehicles. Similar technologies applied to large genomic, proteomic and clinical data sets are making inroads into biomedical sciences. Furthermore, the development of technologies to integrate machine learning with automated chemical synthesis is currently being funded by the Defense Advanced Research Projects Agency in the “Make-It” programme, which is using both batch and flow chemistry for synthesis of on-demand pharmaceuticals in military field operations. The convergence of nascent automated chemical synthesis technologies, high-throughput biological screening, automation engineering and machine learning/AI make the time right to contemplate a concerted, integrated effort to explore biologically active chemical space. Such an ambitious effort, if successful, could revolutionize chemical genomics and drug discovery in the next decade.

Catalysing platform creation: a workshop

We propose that such an effort would be supported by the establishment of ASPIRE (A Specialized Platform for Innovative Research Exploration) for translational sciences that combines automated chemical synthesis and biological testing (Supplementary Fig. 2). This effort would require deep and multidisciplinary participation of experts in many research institutions including academic, industry and government laboratories around the world. With this in mind, as an initial step, a diverse group of more than 40 such stakeholders were gathered at a 2-day workshop at the NIH in October 2017, along with representatives from funding and regulatory agencies and professional societies (see Supplementary Table 1 for a list of attendees). The aim of the meeting was to identify the current state of the field, major gaps and needs, defined short-term and long-term goals, and necessary stakeholders. The

conclusions of the workshop are summarized in Supplementary Table 2, and the following areas were emphasized.

- The current lack of ‘big data’ in chemistry and the synthetic chemical space explored to date is an impediment to AI/machine learning, because these algorithms use both positive and negative data under a variety of conditions (reagents, yield, reaction time, temperature, pressure, catalyst and more) to learn and predict new synthetic routes.
- New approaches to encapsulation of reaction components in capsules and tablets or 3D-printed reagent chemistries are required for ease of handling in automated systems. These formats should accommodate the use of mixtures of monomers and ligands that allow synthesis without the use of cumbersome glove boxes in automated systems. Such reagent handling technologies are becoming commercially available, but development is needed in this area to enable the use of cost-effective tools across the research community.
- New analytical technologies are crucial, with more versatile electronic laboratory notebook (eLN) technology to seamlessly connect data from synthesis to purification and the final purified and biologically annotated product.
- There is a need to engage synthetic and medicinal chemists to enable adoption of automation and machine learning. Engagement will depend on making a chemist’s daily workflow easier to manage, more productive and efficient via user-friendly automation and machine learning software coupled to more standardized and integrated eLNs. Such systems will allow chemists to redesign their core objectives: improving chemical diversity through synthesis of new libraries, expanding capacity and supporting more efficient exploration of structure–activity relationships⁵.
- Remote access to an automated synthesis and testing facility could have a key role in engaging global chemistry and biology communities, and proof of principle has been demonstrated. Through inexpensive virtual reality technologies and AI, remote molecule design and on-demand synthesis could be used not only by chemists, but also by other related scientific experts (such as biologists, clinicians, pharmacologists and toxicologists). Such global collaborations with medicinal chemists can be productive and rewarding.
- AI and machine learning tools need to be developed specifically for applications in automated chemical synthesis and the connection of known molecular mechanisms to tractable chemical probes and lead molecules for efficient testing of biological and clinical hypotheses.
- An ASPIRE capability would allow expert chemists and biologists to focus on more complex drug discovery tasks and explore solutions to repetitive and mundane steps through automation of synthetic chemistry and bioassays. Such an approach would transform synthetic chemistry to run more autonomously — much as biology and informatics can today — ultimately resulting in a data-driven predictive chemical science with accelerated knowledge acquisition.

Outlook

ASPIRE aims to address two challenges of the current era in biomedical research: to harness new technologies to accelerate understanding of living systems and to fulfil the promise of science to improve the lives of the many patients with untreatable or poorly treatable diseases. As such, it could complement current NIH efforts in unexplored therapeutic space such as the NIH Illuminating the Druggable Genome programme¹.

A pilot for the ASPIRE concept will begin in 2018, as part of the recently announced NIH HEAL (Helping to End Addiction Long-term) initiative on opioid addiction, overdose and pain. Although this pilot will focus on the chemistry and biology of targets for these disorders, the technology platform will be generally applicable, and lessons learned from the pilot will be used to design any further stages of ASPIRE.

Much like other types of space exploration, ASPIRE is an ambitious vision that we hope will spawn new technologies, excite a generation of aspiring scientists and produce solutions to heretofore intractable challenges in science and medicine.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Oprea TI et al. Unexplored therapeutic opportunities in the human genome *Nat. Rev. Drug Discov* 555, 592–592 (2018).
2. Mullard A The drugmaker's guide to the galaxy. *Nature* 549, 445–447 (2017). [PubMed: 28959982]
3. Schneider G Automating drug discovery. *Nat. Rev. Drug Discov* 17, 97–113 (2018). [PubMed: 29242609]
4. Arrowsmith CH et al. The promise and peril of chemical probes. *Nat. Chem. Biol* 11, 536–541 (2015). [PubMed: 26196764]
5. Lowe D AI designs organic syntheses. *Nature* 11, 536–541 (2015).