FROM THE ANALYST'S COUCH

AI in small-molecule drug discovery: a coming wave?

Madura K. P. Jayatunga, Wen Xie, Ludwig Ruder, Ulrik Schulze and Christoph Meier

Credit: Ayman-Alakhras/iStock/Getty Images Plus

Artificial intelligence (AI) offers the potential to transform drug discovery. Over the last few years, AI-enabled drug discovery has grown substantially through technological progress, such as the use of neural networks to design molecules and the application of knowledge graphs to understand target biology.

Several AI-native drug discovery companies have progressed molecules into clinical trials, in some cases reporting greatly accelerated timelines and reduced costs, raising high expectations in the R&D community. In addition, many established pharmaceutical companies have formed discovery partnerships with AI companies to explore the technology. Despite this progress, it is still early days for AI in drug discovery, with many open questions about its impact and future potential.

We see several dimensions for AI to create value in drug discovery, including greater productivity (faster speed and/or lower cost), broader molecular diversity and improved chances of clinical success. Here, we present an analysis of the impact of AI along these dimensions using publicly available data. We focused mainly on small-molecule drug discovery, for which AI approaches are relatively more established.

Impact in small-molecule drug discovery
Pipeline growth. We focused our analysis on 24 'AI-native' drug discovery companies, for which AI is central to their discovery strategy (see Supplementary information for a list and analysis strategy). For a subset of 20 of these companies, we were able to reconstruct their pipelines between 2010 and 2021 using public databases. During this time, AI drug discovery companies had rapid pipeline growth, with an average annual growth rate of around 36%. This is driven mainly by assets and programmes at the discovery and preclinical stage (FIG. 1a), reflecting the early-stage nature of AI-native companies. Today the combined pipeline of these 20 AI companies contains ~160 disclosed discovery programmes and preclinical assets and about 15 assets in clinical development.

For comparison, the combined in-house-originated pipeline of the top 20 pharma companies contains ~330 disclosed discovery programmes and preclinical assets, and ~430 assets in phase I clinical development (using the same public data sources and excluding partnered assets or programmes; FIG. 1b). So, AI companies appear to have a combined pipeline equivalent to 50% of the in-house discovery and preclinical output of 'big pharma'. Even if we assume under-reporting of discovery programmes and preclinical assets by pharma companies and over-reporting by AI companies, this seems an impressive picture. Nevertheless, it remains to be seen how many of the AI-enabled preclinical programmes reach the clinical trial stage, and how successful AI-derived assets will be in clinical trials.

Pipeline composition of AI drug discovery companies. We further analysed the current pipelines of the full list of 24 AI-native drug discovery companies with regards to therapeutic areas and target classes. Detailed target information was available for only about a quarter of AI-enabled R&D programmes and assets, but analysis of this partial dataset suggests that AI-native drug discovery companies often focus on well-established target classes (FIG. 2a). For example, more than 60% of all disclosed targets of AI companies are enzymes such as kinases, and other well-known drug target classes such as

G-protein-coupled receptors also make up a high proportion.

This strong emphasis on well-established targets as appropriate testing grounds could be driven by multiple factors, including a desire to de-risk internal pipelines by focusing on targets with validated biology, to prove the viability of their technology platforms and to address well-known challenges such as selectivity issues for well-characterized targets with rich data (often including structural information). In contrast, top-20 pharma companies tend to have pipelines that balance both emerging and established target classes (FIG. 2a).

Despite these trends, there are some reported examples of potential first-in-class AI-derived compounds for novel targets, including protein tyrosine phosphatase SHP2, DNA helicase WRN and paracaspase MALT1, for which AI-derived compounds are among the first for which first-in-human studies or studies to enable an investigational new drug (IND) application have been initiated (see Supplementary information for details).

In terms of therapy area, most of the disclosed AI discovery programmes and assets are in the oncology and central nervous system areas, probably due to the high unmet medical need and many well-characterized targets (FIG. 2b).
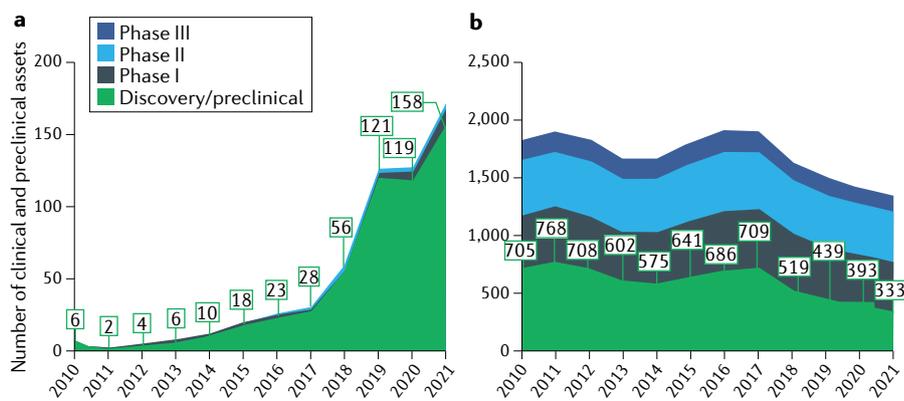
Fig. 1 | **Number of annual R&D programmes and assets over time, showing the growth of AI-enabled drug discovery. a** | AI-native drug discovery companies. **b** | For comparison, top-20 pharma companies. See Supplementary information for details.
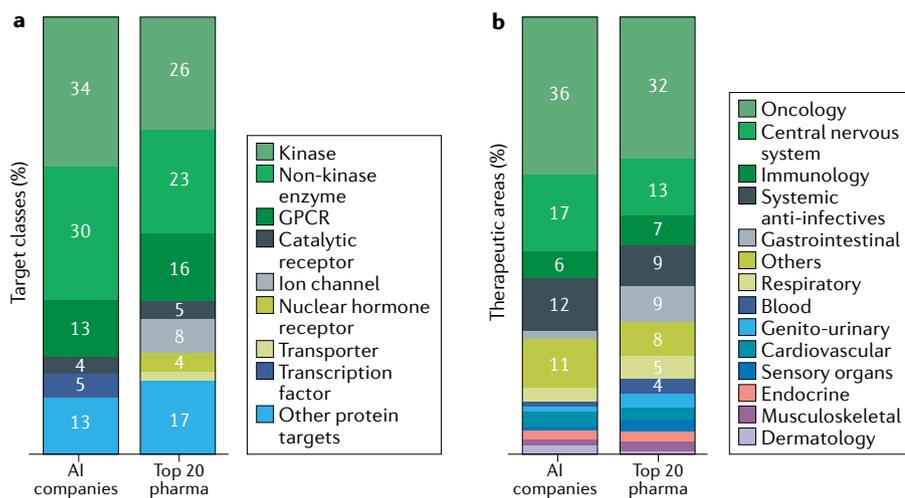
Fig. 2 | **AI drug discovery companies focus on well-established target classes and therapeutic areas.** Target classes (**a**) and therapeutic areas (**b**) of reported assets of AI-native companies and assets discovered in-house by the top 20 pharma companies, where available. See Supplementary information for details.

Chemical structures and properties of AI-derived molecules. Publicly available data on the chemical structures of AI-derived assets is currently limited. As a result, a systematic statistical analysis is not feasible at the moment. However, an analysis of examples with some disclosed data may provide a glimpse of the future.

One such example is TYK2 inhibitors. TYK2 is a member of the Janus kinase (JAK) family, which have multiple existing inhibitors, including 10 marketed products. One common issue with these molecules is their limited selectivity for a single JAK isoform, which affects their safety profile. AI-enabled discovery efforts have recently identified an asset with a novel, allosteric mode of action that appears to be at least 20-fold selective for TYK2 over other members of the JAK family and therefore might have a more favourable safety profile (see Supplementary information for details).

Interestingly, when comparing the structure of AI-derived, TYK2-selective inhibitors with classically discovered, less-selective JAK inhibitors in chemical space, we do not observe striking differences (Supplementary Fig. 1a). Rather, AI-derived, TYK2-selective inhibitors seem to expand into under-represented areas of chemical space.

Some data for assets targeting serotonin receptors have also been disclosed. Here, AI-enabled discovery has produced chemistry comparable to classically discovered molecules. Two AI-derived small molecules targeting serotonin receptors — a 5-HT$_{1A}$ agonist and a bispecific 5-HT$_{1A}$ agonist and 5-HT$_{2A}$ antagonist — have recently entered the clinic (see Supplementary information).

Chemical space analysis, based on structures published in patents, suggests that these molecules occupy similar chemical space to previously published drugs (Supplementary Fig. 1b,c). Such results might be a reflection of the data they have been generated and trained with (see Related links).

Taken together, these examples indicate that AI-enabled strategies can discover molecules comparable to classical discovery efforts, with the potential to explore adjacent chemical space.

Discovery timelines of AI-derived molecules. One of the greatest hopes for AI-enabled drug discovery is an acceleration of discovery timelines — for example, rapid target identification and validation, or fewer and faster cycles of molecule design and optimization.

While it is notoriously difficult to measure discovery timelines using publicly available data, we were able to reconstruct the approximate timelines for selected pharma–AI partnerships and discovery programmes. Based on the timing of patents, publications and public announcements, we find multiple AI-enabled programmes completing the entire discovery and preclinical journey in less than four years (Supplementary Fig. 2). Such initial data points compare favourably to historical timelines in the industry of five to six years and seem particularly impressive given that AI is still nascent in discovery and likely to accelerate further as AI companies mature.

Conclusions and outlook

Drug discovery is a multi-dimensional, multi-step search and optimization problem. AI — with its powerful new tools solving complex problems — has the potential to play an important role in drastically improving this process. Our analysis indicates early signs of a fast-approaching, AI-enabled wave with the potential to radically change drug discovery.

However, the impact we observed for AI varies for different dimensions. We identified signs of increased early discovery efficiency and productivity; AI companies, most of which started less than ten years ago, have achieved a substantial fraction of the preclinical output of the top 20 pharma companies. We also already see examples of novel chemistry for major targets and a potential early glimpse at increased molecular diversity targeting new biological mechanisms. Lastly, we found initial evidence of potential acceleration of discovery timelines.

For other dimensions, it is too early to draw conclusions. For example, the impact on cost is currently difficult to assess, although we believe that scaling AI systematically across R&D could provide major cost improvements. Most importantly, it remains to be seen whether this AI discovery wave continues and translates into clinical success and better medicines for patients.

If it does, AI-enabled drug discovery could prove a game-changer for pharmaceutical R&D, especially for small-molecule drug discovery, potentially allowing it to 'catch up' with other modalities that typically have faster discovery timelines, such as monoclonal antibodies. This will impact how research and discovery organizations should be organized and governed to unlock AI's full potential.

Madura K. P. Jayatunga[1], Wen Xie[2], Ludwig Ruder[3], Ulrik Schulze[3] and Christoph Meier[1]✉

[1]Boston Consulting Group, London, UK.

[2]Boston Consulting Group, Beijing, China.

[3]Boston Consulting Group, Zurich, Switzerland.

✉*e-mail: meier.chris@bcg.com*

RELATED LINKS
Lowe, D. AI-generated clinical candidates, so far: https://www.science.org/content/blog-post/ai-generated-clinical-candidates-so-far
Wills, T. Assessing structural novelty of the first AI-designed drug candidates to go into human clinical trials: https://www.cas.org/resources/blog/ai-drug-candidates