nature portfolio

https://doi.org/10.1038/d41573-023-00139-0

Supplementary information

The company landscape for artificial intelligence in large-molecule drug discovery

In the format provided by the authors

Analysis of the landscape of AI-driven biotech companies engaged in large-molecule drug discovery

Supplementary Table 1 shows the 82 artificial intelligence (AI)-driven companies included in this analysis organized by year founded (as outlined in Figure 1a). Companies were identified using combined search terms "AI-driven drug discovery" and "antibody", "peptide", "vaccine", "RNA/DNA" within the Pitchbook dataset, and augmented with a manual press search (searches conducted in Q2 2023). In addition, direct competitors to the companies identified were also included in the analysis set, where available and relevant. Information on deals made by these companies with top-20 biopharma companies (defined based on total revenues in 2021 published by FiercePharma) was identified through IQVIA Pharma Deals.

Companies registered in the name of a parent/holding business only (for example, Isomorphic Labs/its parent company Alphabet) were removed to prevent biasing financial analysis. To ensure companies included leveraged AI-driven discovery methodologies, and to mitigate potential database limitations, we applied several filters to the dataset:

- Company websites were reviewed for confirmation of self-reported large-molecule AI or machine learning (ML) activity
- Discontinued assets, assets with no reported development, suspended assets, and assets classified as not applicable (i.e., non-large-molecule therapeutics), were removed
- Small-molecule chemicals, natural products, and cell-based therapies (e.g., CAR-T, stem-cell therapies) were removed, as were in-licensed products.

Specific company examples included in the main text were based on externally published examples or papers referring to the use of specific AI-approaches in large-molecule drug discovery.

Company capitalization analysis

Company capitalization analysis over time was conducted using source data from Pitchbook, with subsequent categorization into several groups based on investment maturity across (1) seed and pre-seed, (2) early-stage venture capital (VC, defined as Series A and B), (3) late-stage VC (defined as Series C and beyond), (4) initial public offering (IPO)/public investment, and separately, (5) debt categories. This analysis is summarized in Figure 1b.

Company portfolio analysis

Supplementary Table 2 summarizes AI-driven drug discovery companies with identified preclinical or clinical large-molecule therapeutics (as outlined in Figure 1c). Each company applies AI in different areas; for example, some focus on identifying novel large-molecule designs applicable to previously undruggable targets (for example, Peptilogics), some focus on identification of novel targets (for example, Deep Genomics) and others combine several AI-assisted technologies to accelerate discovery and reduce the risk of failure during drug development (for example, Sparx Therapeutics and Evaxion).

PharmaProjects (Q2 2023) was used to identify pipeline assets in the portfolio of the included AI-driven drug discovery companies. It should be noted that PharmaProjects is not an exhaustive catalogue of clinical and preclinical development pipelines, and new molecular entities (NMEs) recorded as belonging to an AI-driven drug discovery company may have been in-licensed from other pharmaceutical companies or may have been discovered through non-AI-methods. To mitigate these potential issues, efforts have been made to verify the origins of clinical-phase assets via searches of press releases and company websites.

For all analyses, data has not been reviewed by PitchBook, Citeline PharmaProjects, or IQVIA Pharma Deals analysts.

Supplementary Table 1 | Companies working on AI-driven discovery for large molecules founded by year

Founding year	Companies identified
Pre-2012	AbSci, Arzeda, Evaxion Biotech, iBio, Meridigen, Pepticom, Schrödinger, Second Genome, Syntekabio, ZielBio
2012	AbCellera, Exscientia, G3 Therapeutics, LabGenius
2013	Molcure, Peptilogics
2014	Anima Biotech, Deep Genomics, Resonant Therapeutics, Serotiny, XtalPi
2015	Ardigen, Orionis Biosciences, SEngine Precision Medicine, Systems Oncology
2016	Arbor Biotechnologies, Kintai Therapeutics, Unnatural Products
2017	3T Biosciences, A-Alpha Bio, Abalone Bio, Antiverse, Arontier, Gatehouse Bio, MAbSilico, METiS Therapeutics, Modulus, Neon Biotechnology, Polymaths AI, ProteinQure, RubrYc Therapeutics (acquired), VeriSIM Life
2018	1910 Genetics, Celsius (Biotechnology), Character Biosciences, deepCDR, Dyno Therapeutics, Erasca, Generate Biomedicines, Menten AI, neoX Biotech, Soley Therapeutics, Sparx Therapeutics
2019	Basecamp Research, BigHat Biosciences, Bio Simulytics, Denovo Sciences, Etcembly, Evozyne, Pharm CADD, Ordaos, Stellanova Therapeutics, Valo
2020	Athae Bio, Fresh Wind Biotechnologies, Known Medicine, Nabla Bio, NextPoint (Drug Discovery), Outpace Bio, Patch Biosciences, POLARISqb, Primary Biotech, Promatix, ROME Therapeutics
2021	20n Bio, Ainnocence, Atomic AI, Cradle Bio, Gandeeva Therapeutics, Huashen Zhiyao, Protai
2022	Profluent Bio

Supplementary Table 2 | AI-driven biotech companies with large-molecule drug candidates

Company	Discovery/ preclinical	Clinical				
	No. of assets	No. of assets	Asset name	Asset therapeutic area	Asset target	Phase
Arbor Biotechnologies	4	0	-	-	-	-
Deep Genomics	2	0	-	-	-	-
Evaxion Biotech	2	1	EVAX-01	Oncology (metastatic melanoma)	Unspecified	Phase II
iBio	4	0	-	-	-	-
Peptilogics	1	1	PLG-0206	Infection (antibiotic), periprosthetic joint infection	Unspecified	Phase I
PharmCADD	1	2	EG-COVID-003	COVID-19	Unspecified	Phase I
			EG-COVID-001			Phase II
RubrYc Therapeutics	2	0	-	-	-	-
Second Genome	7	0	-	-	-	-
Sparx Therapeutics	0	1	SPX-101	Oncology (solid tumour)	Claudin-18	Phase I
Stellanova Therapeutics	1	0	-	-	-	-
ZielBio	0	1	ZB-131	Oncology (solid tumours with high cancer-specific plectin levels)	Plectin	Phase II

Further details of reported examples of AI-derived candidates

- **EVAX-01**

  Long, G.V. *et al.* (2022) "Keynote – D36: Personalized immunotherapy with a neoepitope vaccine, EVX-01 and Pembrolizumab in advanced melanoma," *Future Oncology*, 18(31), pp. 3473–3480. Available at: https://doi.org/10.2217/fon-2022-0694

  ClinicalTrials.gov reference: https://clinicaltrials.gov/ct2/show/NCT05309421

- **PLG-0206**

  Huang, D. *et al.* (2021) "The engineered antibiotic peptide PLG0206 eliminates biofilms and is a potential treatment for periprosthetic joint infections," *Antibiotics*, 11(1), p. 41. Available at: https://doi.org/10.3390/antibiotics11010041

  ClinicalTrials.gov reference: https://clinicaltrials.gov/ct2/show/NCT05137314

- **EG-COVID-001/EG-COVID-003**

  Hong, H.C. *et al.* (2021) "An mRNA vaccine against SARS-COV-2: Lyophilized, liposome-based vaccine candidate EG-COVID induces high levels of virus neutralizing antibodies." Available at: https://doi.org/10.1101/2021.03.22.436375

  ClinicalTrials.gov reference: https://clinicaltrials.gov/ct2/show/NCT05188469

- **SPX-101**

  Sparx Group has initiated a phase 1 trial of its first organically developed candidate, SPX-101 News release. 17 January 2022. Available at: https://www.prnewswire.com/news-releases/sparx-group-has-initiated-a-phase-1-trial-of-its-first-organically-developed-candidate-spx-101-301461500.html

  ClinicalTrials.gov reference: https://clinicaltrials.gov/ct2/show/NCT05231733

- **ZB-131**

  ZielBio Announces first patient dosed in phase 1/2 clinical trial of ZB131, its novel monoclonal antibody targeting cancer specific plectin. News release. 15 February 2022. Available at: https://prn.to/3JAk2vf

  ClinicalTrials.gov reference: https://clinicaltrials.gov/ct2/show/NCT05074472

	Disease understanding & Target identification	Hit/Lead generation		Hit/Lead selection and optimisation	Development transition
	Antibody target identification	**Library generation**	**Functional and structural screening**	**Lead optimisation**	**Preclinical optimisation**
Traditional wet lab research activities (non-exhaustive)	• Validate target antigen • Enhance antigen immunogenicity • Produce and purify antigen	• Select and immunise host (e.g., mouse) • Prepare library (e.g., hybridoma, phage / yeast display)	• Screen and select antibodies (e.g., FACS, SPR) • Characterise reactivities • Solve the crystal structure	• Chimerisation and humanisation • Affinity maturation (e.g., ELISA, SPR) • Efficacy and safety	• Optimise expression and purification • Create stable cell lines • Scale production
Traditional *in silico* methods	Homology modelling of initial starting structures to assist the screening and selection of antibodies	Mutational scanning to create combinatorial libraries via peptide-target interaction (e.g., Rosetta-Antibody-Design, FoldX)	Docking and molecular dynamics simulations to predict binding and specificity of antibody-target interactions (e.g., Gromacs, Haddock, Glide, Rosetta)	Electrostatic surface prediction (e.g., CamSol, abYsis, Schrödinger)	Up-/downstream process simulation to optimise manufacturing and purification, leading to improved yields and reduced costs (e.g., OpenFOAM)
Advanced AI tools	Gene-disease associations via GWAS or knowledge graphs (e.g., Sei) Antigen structure prediction as basis for epitope identification or virtual screening (e.g., AlphaFold, ESMFold, DLAB)	Generative language models for expanding existing or de novo libraries with desired properties, such as specificity (e.g., IgLM)	Deep learning-assisted repertoire analysis and selection based on properties such as stability, humanness (e.g., DeepImmune)	Antibody structure optimisation to build structure around predicted epitope (e.g., Constrained Hallucinations, Ig-VAE) or optimise for solubility	Gene expression optimisation (e.g., Enformer) Developmentability detection and thresholds (e.g., therapeutic antibody profiling)

Supplementary Figure 1 | Discovery pipeline for antibody therapeutics, highlighting where traditional *in silico* methods and AI tools have accelerated progression. FACs, fluorescence-activated cell sorting; ELISA, enzyme-linked immunosorbent assay; SPR, surface plasmon resonance; GWAS, genome-wide association study; DLAB, deep-learning for antibodies.