AI Drug Discovery & Development Summit

patsnap

Patents In, SAR Out: Grounding Agents in Chemistry

Mark Davies
Head of Life Sciences Product

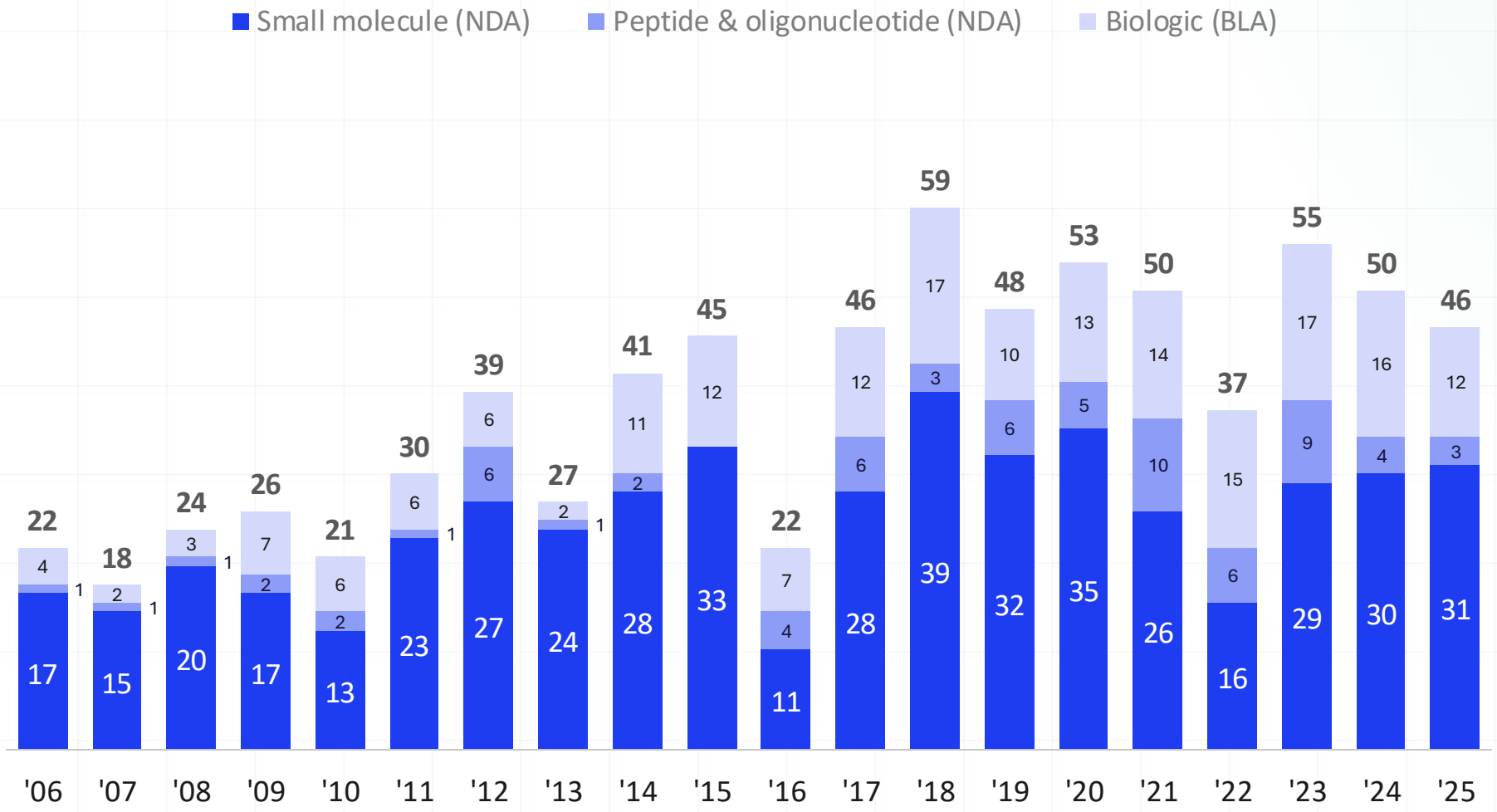
Boston, Oct 2026

patsnap

A long time ago in a biotech far, far away*....

*London actually

Small molecules make up 65% of novel approvals since 2006



31

Small molecule approvals in 2025

65%

Share of novel approvals, 2006 to 2025

Source: Novel Drug Approvals at FDA, de la Torre & Albericio 2026 Molecules, Wang et al. 2026 Molecules

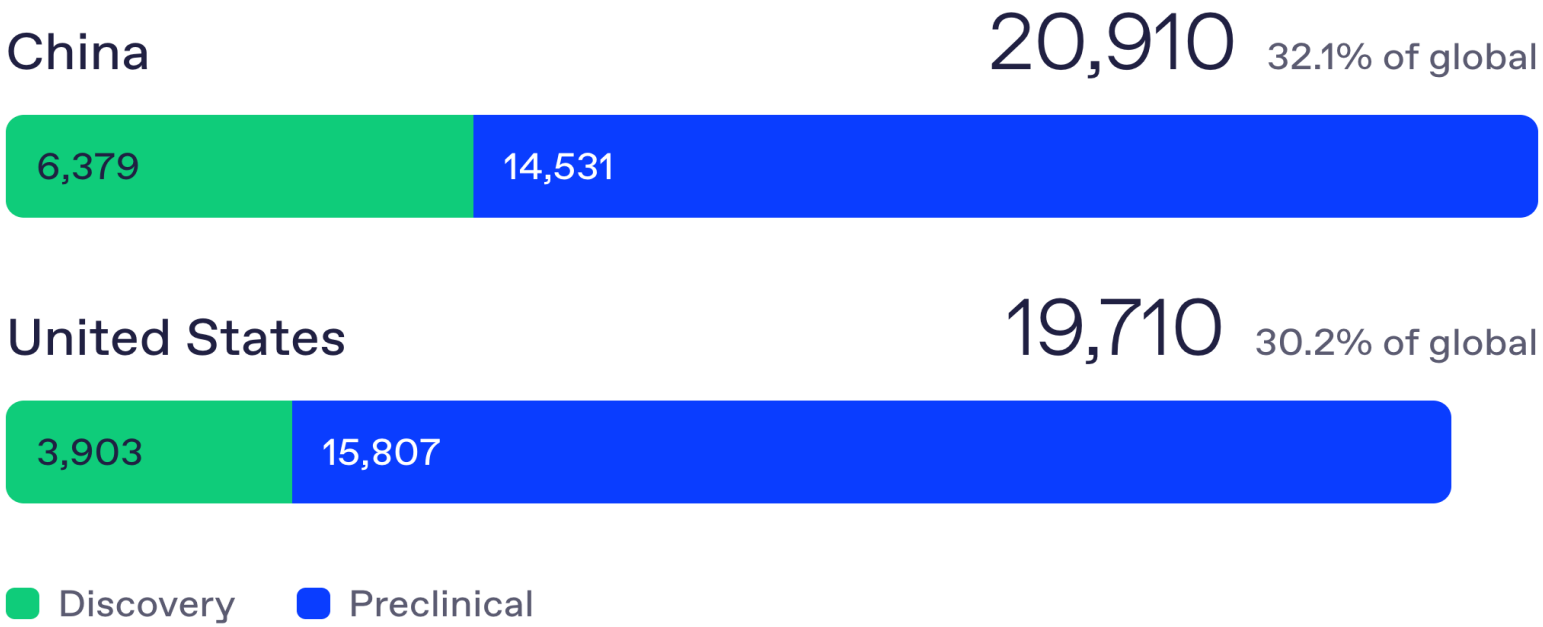
The emerging player: China

1 in 3 early-stage drug assets now originates in China

CHINA'S SHARE

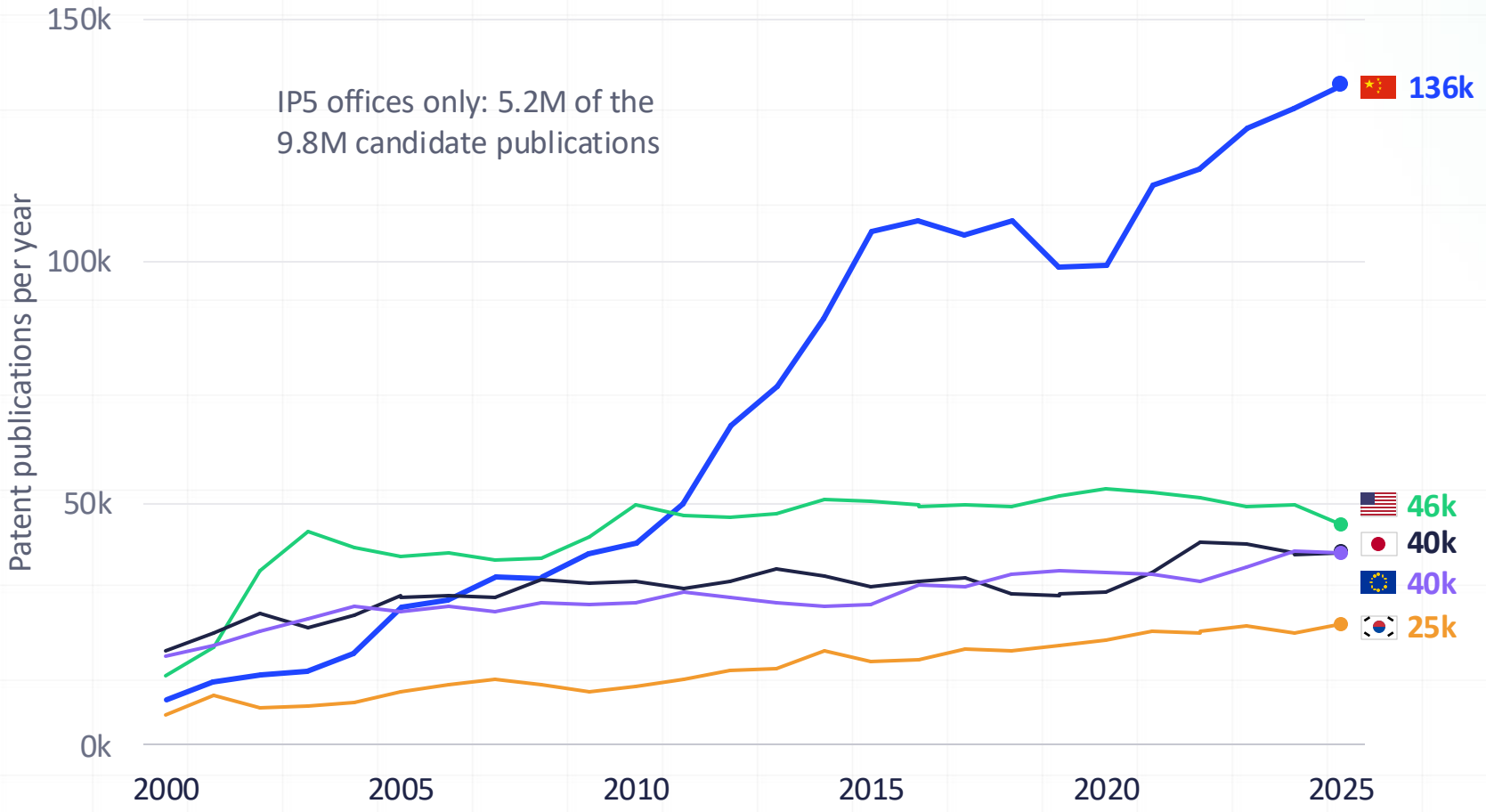
32.1%

Of 65,155 Discovery and Preclinical drugs worldwide



Source: Patsnap Synapse. Drugs whose current global highest development stage is Discovery or Preclinical by originator country; snapshot as of 13 Oct 2026

China now publishes 3x the small-molecule patents of the US



9.8M

Candidate small-molecule patents, all jurisdictions

1.52M

Estimated to hold SAR data

71M

Projected activity records to extract

Source: Patsnap Life Sciences analysis of IP5 publications, 2000 to 2025, matching A61K, A61P, C07D, C07C, C07F or C07H in IPC (incl. IPCR) or CPC.

AI-enabled drug discovery with publicly disclosed cycle time

Company	Asset[1]	Target	Target novelty	Modality	Indication	Furthest clinical phase[2]	Cycle time measurement[3]	Discovery cycle time,[4] months	Comparison with benchmark, approximate % faster
GV20 Therapeutics	GV20-0251	IGSF8	First in class	Monoclonal antibody	Solid tumors	Phase 1/2	Target ID to IND	36	25–50[5]
Iambic Therapeutics	IAM1363	HER2	Fast follow	Small molecule	HER2-driven solid tumors	Phase 1	Hit ID to FIH	24	50–65
Absci	ABS-101	TL1A	Fast follow	Monoclonal antibody	Inflammatory bowel disease	Phase 1	Hit ID to FIH	24	50–65
Generate: Biomedicines	GB-0669	SARS-CoV-2 S2 stem-helix	First in class	Monoclonal antibody	COVID-19 prophylaxis	Phase 1	Hit ID to FIH	17	65–75
Insilico Medicine	ISM3312	3CLPro	Fast follow	Small molecule	COVID-19	Phase 1	Hit ID to PCC	28	15–45
Recursion	REC-1245	RBM39	Fast follow	Molecular glue degrader	Biomarker-enriched solid tumors and lymphoma	Phase 1/2	Target ID to PCC	18	45–65[5]
Insilico Medicine	Rentosertib (INS018_055)	TNIK	First in class	Small molecule	Fibrotic disease	Phase 3	Target ID to PCC	18	45–65[5]
Schrödinger	SGR-1505	MALT1	Fast follow	Small molecule	Hematological malignancies	Phase 1	Hit ID to PCC	10	70–80
Recursion (Exscientia)	EXS21546	A2AR	Fast follow	Small molecule	Oncology	Phase 1/2	Hit ID to PCC	9	70–80
Benchmark: large pharma median[6] (range reflecting target novelty and modality variation)							**Hit ID to FIH**	32–50 (PCC), 16–20 (FIH)	

[1]Nonexhaustive set of publicly disclosed AI-enabled programs with reported discovery cycle times. Inclusion is based on data availability and visibility, and the sample may be biased toward successful, well-publicized, and more advanced programs, with limited representation of undisclosed or unsuccessful efforts. [2]Either current most advanced status or the highest stage reached before discontinued. [3]FIH = first-in-human; ID = identification; IND = investigational new drug; PCC = preclinical candidate. [4]Based on company disclosures and press releases. [5]Not an apples-to-apples benchmark, because of different starting or end points from the benchmark measures (ie, target ID vs hit ID). [6]McKinsey research productivity benchmark.
Source: Company websites; Pharmaprojects; press search

McKinsey & Company

~18months

Avg discovery cycle time for small molecules

58% ↑

Faster than large pharma benchmark

AI-enabled drug discovery: what is holding us back?

A decade of AI in drug discovery, and the evidence of clinically relevant impact is “disappointingly limited.”

01 Chemical data is sparse, local and biased

4 standard ADME datasets share 0.1% of compounds and 0.6% of scaffolds.

02 Chemistry has no AlphaFold shortcut

Small molecules span ~10^{60} of space with no equivalent. SMILES is a notation, not a sequence.

03 The bottleneck is reduction to practice

The data already exists. It has never been made usable.



nature reviews drug discovery

Perspective

Check for updates

Artificial intelligence in drug discovery — what it is, where we stand and the path forward

Abstract

Artificial intelligence (AI) in drug discovery has attracted increasing interest over the past decade. It is now time for a critical review of progress in the field: where did we advance — and where are we yet to see impact — when it comes to what matters in drug discovery, which is to deliver safer and more efficacious medicines to patients faster? Although a wide variety of AI methods have been developed, applied and benchmarked, evidence of their clinically relevant impact is, so far, disappointingly limited. In this Perspective we discuss potential reasons, including an insufficient focus on clinical translation during model development, difficulties with applying AI algorithms on conditional life science data, and insufficient problem definitions and the resulting underspecification of computational models for real-world use cases. ‘Technology push’ compared with ‘science pull’ is also likely to be an underlying factor, as well as the substantial time required to operationalize technical capabilities into systems that are sufficiently scaled and accessible for users. We provide recommendations for the development of AI in drug discovery with the aim of increasing its translational relevance. For example, benchmarking studies of AI tools in drug discovery need to move on from model validation and instead focus on their ability to improve decision making.

Sections

Introduction

Where would improvements matter most?

Challenges posed by biological data

Challenges posed by chemical data

Generating data for AI applications

Model validation by itself is not sufficient

Where are we?

Where do we go from here?

Human psychology and extrinsic motivators

Conclusion

Bender et al. 2026, *Nature Reviews Drug Discovery*
doi.org/10.1038/s41573-026-01496-2

Poll #1:

What’s your biggest bottleneck with patent chemistry?

A

Getting the data out

Structures and activities locked in images and tables

B

Knowing what to trust

No way to tell a good value from a bad one

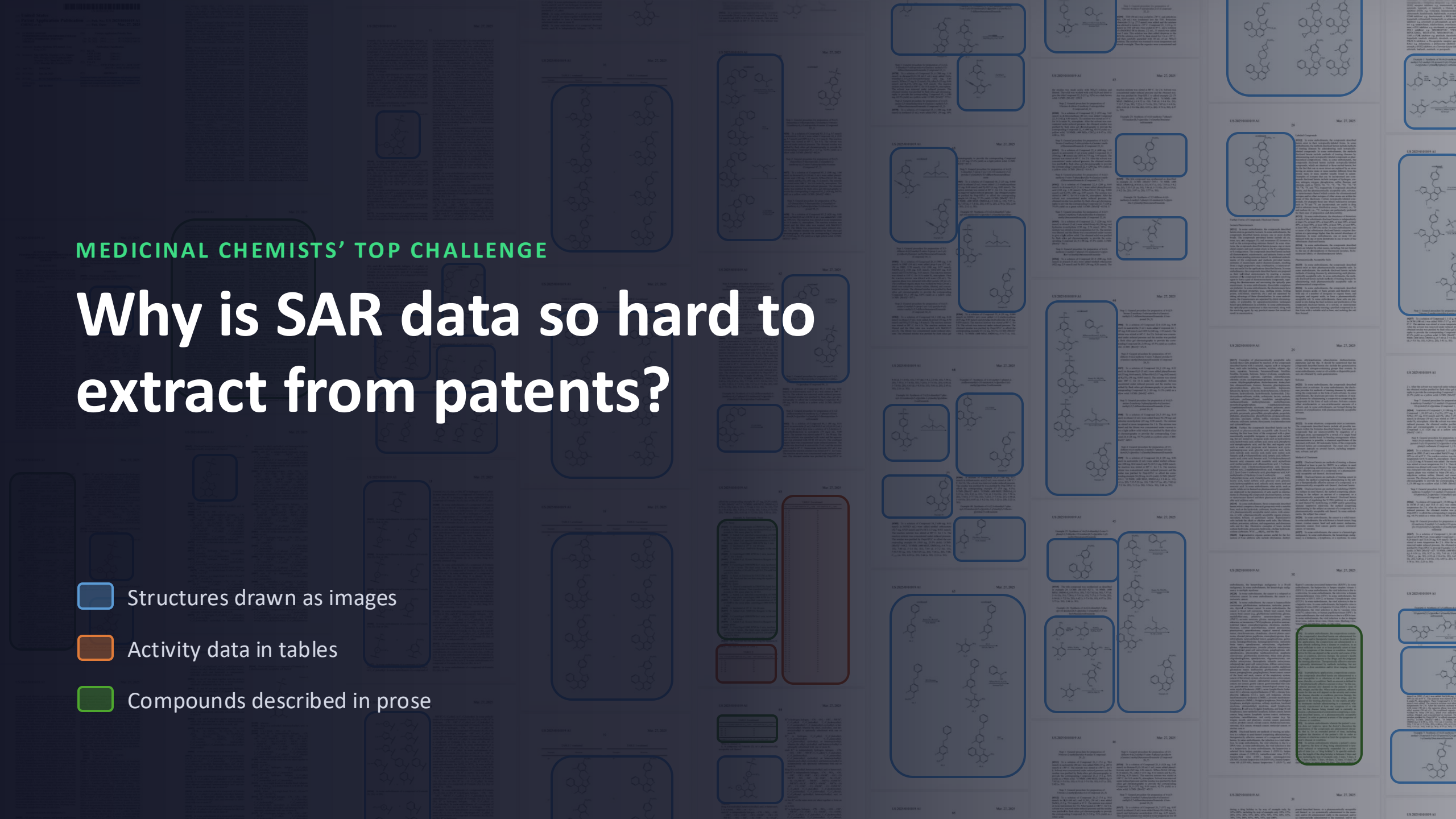
C

Getting it where you work

Curated data that never reaches your tools or models

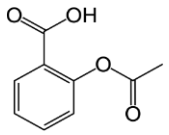
MEDICINAL CHEMISTS' TOP CHALLENGE

Why is SAR data so hard to extract from patents?



6 common examples of chemistry data locked in patents

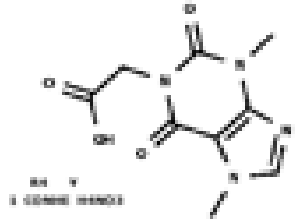
01 Structures come in many formats



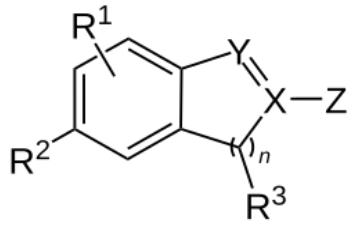
2-acetoxybenzoic acid
o-acetylsalicylic acid
acetylsalicylic acid
acetyl salicylate

Cmpd1

02 Image quality varies wildly



03 Markush structures



04 Semi-quantitative activity values

Compounds	MV4–11 GI_{50}	Kasumi–1 GI_{50} (nM)
Compound 17	+++	+++
Compound 18	+++	+++
Compound 23	+++	N/A
Compound 24	++	N/A
Compound 25	+++	N/A

05 Assay metadata distributed throughout patent

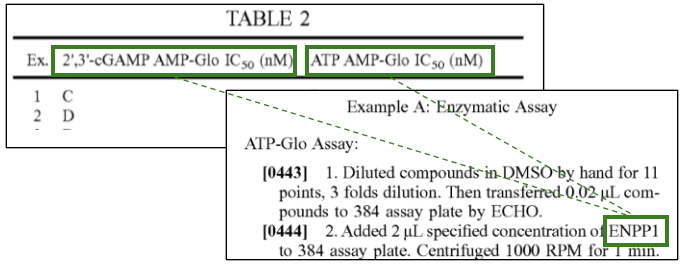


TABLE 2

Ex. 2',3'-cGAMP AMP-Glo IC_{50} (nM) ATP AMP-Glo IC_{50} (nM)

1 C
2 D

Example A: Enzymatic Assay

ATP-Glo Assay:

[0443] 1. Diluted compounds in DMSO by hand for 11 points, 3 folds dilution. Then transferred 0.02 μL compounds to 384 assay plate by ECHO.

[0444] 2. Added 2 μL specified concentration of ENPP1 to 384 assay plate. Centrifuged 1000 RPM for 1 min.

06 Translation corrupts the data

Eine Mischung aus 6-(Difluoπτiethyl)-N-[(1 R,2S)-2,6-dimethyl-2,3-dihydro-1 H- inden-1-yl]-1 ,3,5-triazine-2,4-diamin (0,25 g; 0,0008 Mol) und

Solving patent chemistry & data extraction at scale and speed

WHERE OUR AI, MEDICINAL CHEMISTRY & DATA TEAMS INVEST

1

SOTA OCSR, NER & OCR

Structure, entity and text recognition tuned for patent-grade images, table, figures and languages

2

Domain-focused skills & tools

Purpose-built to analyze and process chemistry, chemical names and experimental (SAR) data, not generic text

3

117 patent authorities, daily refresh

Direct access to global patent source data, refreshed with new filings every day

The SAR Extractor Agent

Extract chemistry and SAR data from patents, at scale and speed.

97%	OCSR precision
89.5%	Structure-field precision · 71.4% end-to-end
97.5%	Activity-value precision · 75.8% recall
117	Patent authorities, with daily updates

Powered by Eureka by patsnap

HOW THE SAR EXTRACTOR AGENT WORKS

5 steps from patent to validated, confidence-scored SAR data

1

LLM

Extract SAR data

Identify and pull 10 key fields for the small-molecule drug and its target

2

PATSNAP DATA

Fetch substance

Use the substance code to fetch the compound image or IUPAC name in the patent

3

OCSR / NER

Transform to structure

Convert the substance image or name into a machine-readable chemical structure

4

LLM

Validate & confirm

Confirm the structure and activity data are consistent

5

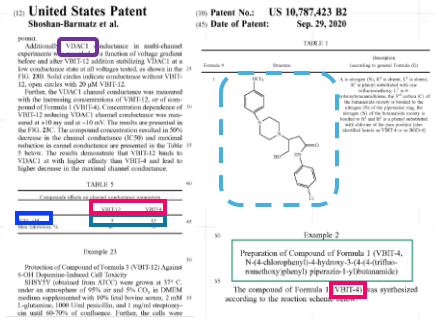
EUREKA SAR

Confidence scoring

Assign a confidence score to every structure and activity

INPUT

Patent — image + text



OUTPUT

Validated, confidence-scored SAR data



OCSR = optical chemical structure recognition · NER = named-entity recognition

Powered by Eureka by patsnap

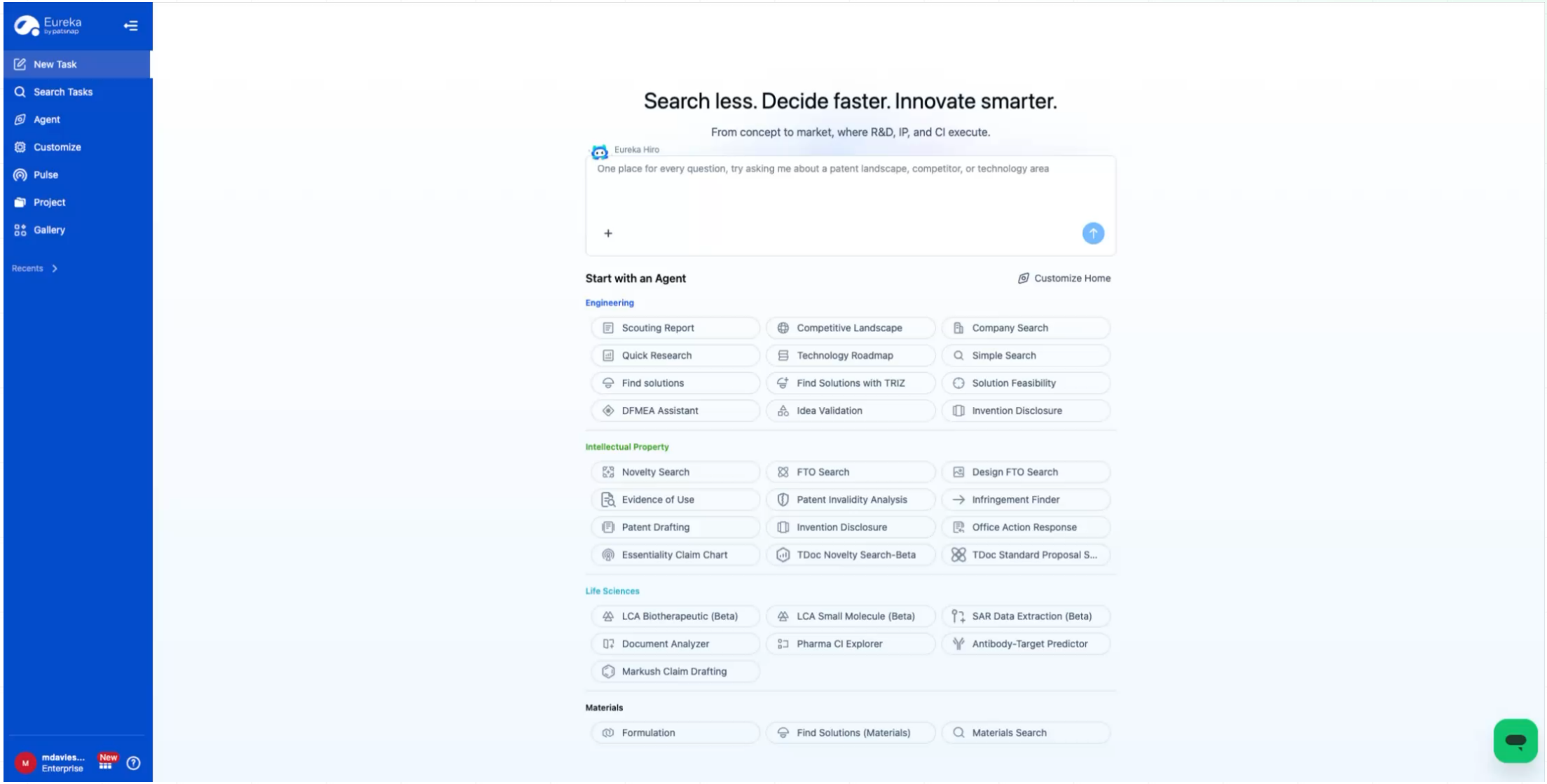
patsnap

SAR EXTRACTOR

Let's see it in action.

Powered by Eureka by patsnap

Patent in. SAR table out.



Lead Compound Analyzer:
From complex patent to ranked clinical leads

HOW THE AGENT PRIORITIZES COMPOUNDS

1

Map every compound

Structures mapped for the patent's exemplified compounds, with valid SMILES

2

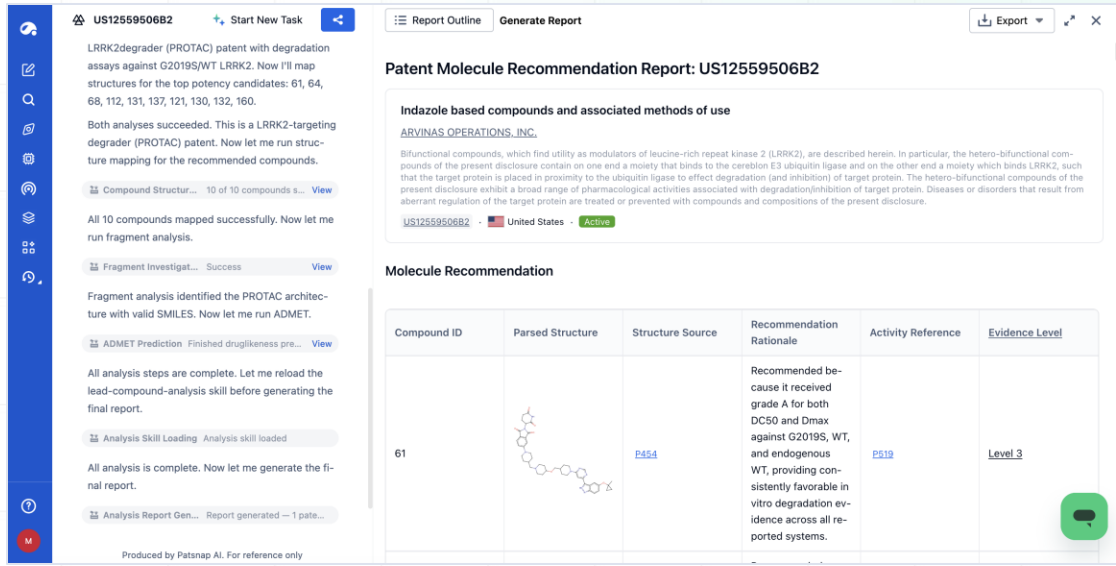
Profile fragments & ADMET

Frequent fragments identified and ADMET properties predicted for top candidates

3

Rank leads, with sources

Ranked recommendations with rationale linked back to the patent's assay data



Outputs Ranked leads with rationale, structures and SMILES

Powered by Eureka by patsnap

CUSTOMER USE CASE · GLOBAL PHARMA

From competitor patent to design decision

1 TRACK

Synapse by patsnap · Open Platform by patsnap

Target watch & structure alerts

Watch drug development related to a target. Get alerted when new filings claim similar scaffolds.

IP team

2 EXTRACT

Eureka by patsnap

Lead Compound Analyzer

Lead compounds, scaffolds and ADMET data pulled from each new competitor filing.

Med chem team

3 BENCHMARK

Eureka by patsnap

SAR Extractor

Competitor activity data, tabulated to benchmark internal candidates.

Med chem team

4 ASSESS

Chemical by patsnap · Analytics by patsnap

Is my compound claimed?

Internal candidates checked against competitors' patented compounds, with the wider landscape mapped.

IP & legal

Open Platform by patsnap

MCP + REST API

The same data, inside the AI tools you already use.

open.patsnap.com

Claude · ChatGPT · Copilot Studio · Codex · Your agents

patsnap

Patsnap Life Sciences Product Suite

Eureka by patsnap

AI agents, incl. SAR Extractor and Lead Compound Analyzer

Synapse by patsnap

146k+ drugs, **48k+** targets, **1.3M+** clinical trial records

Chemical by patsnap

228M+ chemical structures

Bio by patsnap

1.05B+ biological sequences

Analytics by patsnap

Search & analysis across **215M+** patents from **174** jurisdictions

Open Platform by patsnap

MCP servers and REST APIs for your AI tools

SAR Agent: 3 Key Takeaways & 1 Action

01

There is a shortage of readable chemistry

10 years of AI, and clinical impact remains “disappointingly limited.”

What’s missing is readable chemistry, not better model architectures.

02

The data is in patents, and it is now extractable

1.52 million candidate patents. 71 million activity records.

Every value scored for confidence and linked back to the page it came from.

03

It must arrive where chemists already work

Available for your pipeline today: CSV, Excel and SDF

Q4 2026: MCP delivery for curated chemistry inside the tools you already have open.

ACTION POINT

Try it yourself

Bring a patent you already know the answer to, and check what comes back.

SEE IT LIVE

Thank you!

www.patsnap.com

patsnap

Make every pharma decision count

Patsnap Life Sciences: Connected intelligence for confident R&D, IP, and portfolio decisions

Our global footprint

Headquartered in Singapore, with
over 1,000 employees worldwide



Backed by world-class investors

SoftBank Vision Fund · vertex VENTURES · SEQUOIA · NUS National University of Singapore

Built for the world's most ambitious innovators

18,000+ Customers **60** Countries **17** Industries

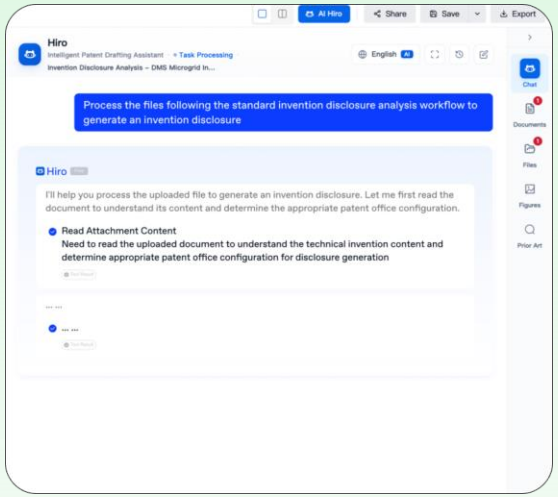
Trusted by: **33%** of the world's top 2,500 R&D spenders
28 of the top 30 biopharma companies globally

Alnylam PHARMACEUTICALS · astellas · Boehringer Ingelheim · Bristol Myers Squibb
Chiesi · danaher · GSK · Johnson&Johnson · novo nordisk
NOVARTIS · BAYER · sanofi · Takeda · teva
CSL Behring · AbCellera · Apellis · ARCUS BIOSCIENCES · nanoform small is powerful®
Dyne THERAPEUTICS · Genmab · immatics · Kindeva DRUG DELIVERY · Isomorphic Labs
genfit · ORUM THERAPEUTICS · SILENCE THERAPEUTICS · HOUSTON Methodist LEADING MEDICINE · St. Jude Children's Research Hospital

patsnap

Domain-specific intelligence for every discipline

Intellectual Property



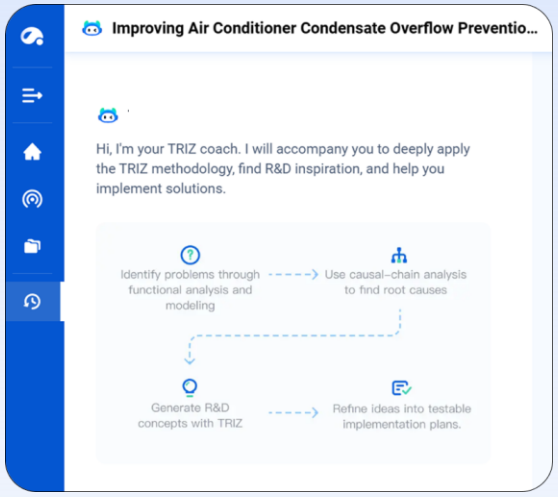
- Manage your patent strategy in one place. Search prior art, draft filings, respond to office actions, and identify commercial opportunity

Life Sciences



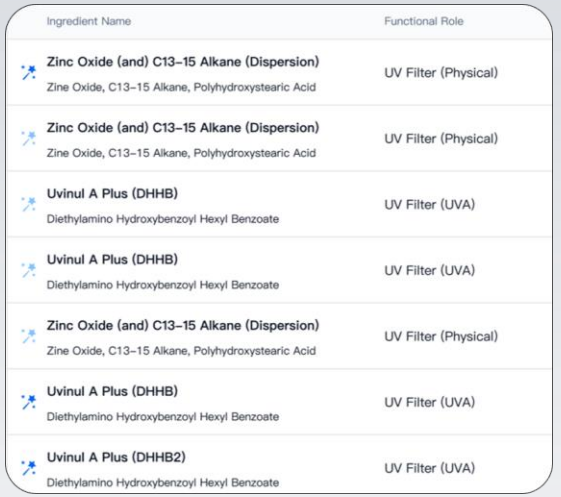
- Decide what's worth the lab. Validate targets, mechanisms, and translational risk before you spend

Engineering



- Cut through technical complexity to find solutions you can evaluate, refine and ship

Materials



- Find viable materials faster. Match property targets to real candidates and risks

One Platform Embedded End to End

SaaS, agents, and MCPs – from early discovery through clinical and commercialization, built for your daily workflows

	Intellectual Property (IP) • Freedom-to-Operate • Patent Landscaping	Research & Dev (R&D) • Target Discovery • Molecule Prioritization	Competitive Intelligence (CI) • Pipeline Surveillance • Licensing & Deal Tracking • Emerging Technology Trends
MCPs Plug & Play for Enterprise-Wide Adoption	**Core Patent MCP**	**Bio & Small Molecule MCP**	**Pharma Intelligence MCP**
Purpose-Built Agents & Skills Specialized Assistants Engineered to Automate Custom Workflow Tasks	**Novelty Search** **AI Drafting**	**SAR Extractor** **Lead Compound Analyzer**	**Hiro LS Researcher** **Eureka Scout**
SaaS Intelligence Ready-to-Use Search, Analytics, and AI-Powered Workflows	Analytics by patsnap	Synapse by patsnap	Bio by patsnap Chemical by patsnap
LS Data & Model Foundry Domain-Specific, FAIR-Principled Data Curation	**Patent** · **Scientific Literature** · **Clinical Trials** · **Deals & News** · **Conference Abstracts** **OCR** · **OCSR** · **NER** · **NOR** · **NLP**		

AI Maturity

70% Reduction in prior art review time utilizing AI search workflows

Real-Time active visibility into global competitor activity

Maximized R&D efficiency through AI-driven molecule prioritization

patsnap

patsnap

Why should you trust SAR Extractor?

Lead Compound Analyzer

Prioritize

Pinpoint likely clinical leads, ranked, with source-linked reasoning.

- ✓ Small molecule, ADC, RNA, PROTAC and more
- ✓ ADMET, frequent fragments, SMILES

→

SAR Data Extraction

Go deep

Every compound and activity value, structured and machine-ready.

- ✓ Structures, IDs and quantitative bioactivity
- ✓ Export to XLS, CSV and SDF

200x

Faster dataset curation from large patents

97%

OCSR model precision

84.5%

End-to-end precision