Ingenious Brain

Case Study

How Our

Chemical Structure Search Improved Search Recall to Over 95% for a Global Pharma Leader

Objective

A leading pharmaceutical company approached us to conduct a comprehensive chemical structure search for a new series of heterocyclic compounds being developed for oncology applications. Despite using automated patent databases earlier, the client struggled with incomplete and inconsistent search results, primarily due to the complexity of Markush claims with broad variable definitions (R^1, R^2, Ar, Het, etc.) and database format discrepancies.

The client needed a highly accurate, exhaustive, and defensible structure search—one that could uncover hidden prior art, validate structural novelty, and de-risk their IP and R&D investments before clinical progression.

Our Strategic Approach

To overcome these challenges, IeB adopted a hybrid expert-AI methodology—integrating chemistry expertise with advanced computational and database tools—for delivering precision and completeness in structure search results.

Expert-Led Markush Decoding

We manually expanded and validated complex Markush claims, interpreting nested variable dependencies that automated tools typically overlook. This manual decoding ensured that millions of potential permutations were accurately captured.

Hybrid Manual and AI Interpretation

Our experts deployed a dual-layered approach—AI algorithms standardized variable groups while domain experts refined and normalized them—to minimize ambiguity and ensure reliable database querying.

Multi-Database Strategy (MDS)

Instead of relying on a single database, we performed cross-platform searches across STNext, Orbit, and Derwent, reconciling differences in Markush encoding formats to achieve higher recall and reduce dependence on any single source.

Expert-Augmented Substructure Search

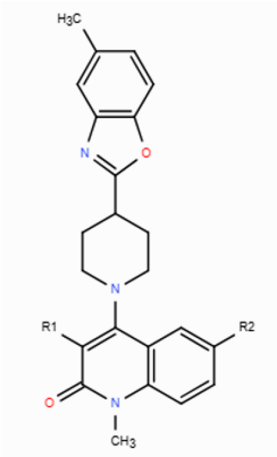
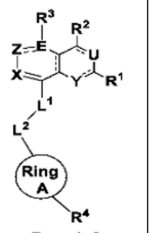
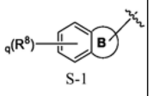
We optimized substructure queries and captured hidden structural variants, reducing false negatives and enabling comprehensive retrieval across chemical analog spaces.

Cross-Source Validation

Each retrieved result underwent manual verification across multiple databases to ensure completeness and consistency—resolving discrepancies that algorithmic systems often miss.

Through this multi-layered workflow, we ensured an exhaustive, defensible, and high-confidence dataset that directly addressed the client's IP strategy challenges.

Snippets

Compound	W02023184327A1 (Prior Art Patent)	IEBS Comment
 Wherein, **R1=methyl, CONH2** **R2=methyl, halogen**	1. **A compound of Formula I**, or a pharmaceutically acceptable salt thereof: wherein: **R 1 is** hydrogen, **halogen (e.g., F**, Cl, or Br) , CN, OH, COOH, CONH 2, NH 2, R A, OR A, COOR A, NH (R A) , N (R A) 2, CONH (R A) , CON (R A) 2, SR A, SOR A, SO 2R A, or P (O) (R A) 2, wherein R A at each occurrence is independently optionally substituted C 1-4 alkyl, optionally substituted C 2-4 alkenyl, optionally substituted C 2-4 alkynyl, optionally substituted C 3-6 cycloalkyl, optionally substituted 5 or 6 membered heteroaryl having 1-4 ring heteroatoms independently selected from O, S, and N, or optionally substituted 4-7 membered heterocyclyl; **R 2 is hydrogen**, halogen (e.g., F, Cl, or Br) , CN, OH, NH 2, R B, OR B, NH (R B) , or N (R B) 2, wherein R B at each occurrence is independently an optionally substituted C 1-4 alkyl, optionally substituted C 2-4 alkenyl, optionally substituted C 2-4 alkynyl, optionally substituted C 3-6 cycloalkyl, or optionally substituted 4-7 membered heterocyclyl; **R 3 is** hydrogen, optionally substituted **C 1-4 alkyl**, optionally substituted C 3-6 cycloalkyl, or optionally substituted 4-7 membered heterocyclyl (e.g., 4-7 membered saturated heterocyclyl) ; **E is N**; **X is** N or **CR 5, wherein R 5** is hydrogen, halogen, CN, **CONH 2**, COOH, CONH (R C) , CON (R C) 2, OR C, NO 2, or COOR C, wherein R C at each occurrence is independently an optionally substituted C 1-4 alkyl; **U is** N or **CR 6, wherein R 6 is hydrogen**, halogen (e.g., F, Cl, or Br) , CN, OH, NH 2, R D, OR D, NH (R D) , N (R D) 2, COOH, CONH 2, COOR D, CONH (R D) , CON (R D) 2, SR D, SOR D, SO 2R D, or P (O) (R D) 2, wherein R D at each occurrence is independently an optionally substituted C 1-4 alkyl, optionally substituted C 2-4 alkenyl, optionally substituted C 2-4 alkynyl, optionally substituted C 3-6 cycloalkyl, optionally substituted 5 or 6 membered heteroaryl having 1-4 ring heteroatoms independently selected from O, S, and N, or optionally substituted 4-7 membered heterocyclyl; **Y is** N or **CR 7, wherein R 7 is hydrogen**, F, Cl, CN, optionally substituted C 1-4 alkyl, or optionally substituted C 1-4 heteroalkyl; **Z is C (O)** , C (=N-OH) , C (=N-O- (C 1-4 alkyl)) , C (=N-NH 2) , C (=N-NH- (C 1-4 alkyl)) , C (=N-N (C 1-4 alkyl) (C 1-4 alkyl)) , SO, SO 2, S (O) (=NH) , or S (O)  Formula I,  S-1	**From a technical perspective it is noted that the Markush scope outlined in the prior art patent application significantly overlaps with the subject compound. To facilitate comparison we have highlighted the text with different colours to emphasize corresponding substitutions in the subject compound.**

HIGH: *Active Patents/Patent Applications with claims that fully or partially overlap with the features of the subject invention will be classified as high risk.*

MEDIUM: *Active Patent/Patent Applications with claims that fully or partially overlap with the features of the subject invention but also have some additional components will be classified as medium risk.*

LOW: *Patent/Patent Applications with claims that fully or partially overlap with the features of the subject invention or have some additional components but are legally not in force will be classified as low risk.*

S. No.	F.NO	Espacenet Link	Google Link	Legal Status	Risk Analysis	IEBS Family Code
1	1	WO2023XXXXXXA1	WO2023XXXXXXA1	PCT-ACTIVE	HIGH	105373269
2	2	EP40XXXXXB1	EP40XXXXXB1	GRANTED	MEDIUM	93814902
3	2	US11XXXXXXB2	US11XXXXXXB2	GRANTED	MEDIUM	93814902

Impact

Our engagement led to transformative outcomes across both strategic and operational fronts:

- Search recall increased from ~40% to over 95%, setting a new internal benchmark for IP search completeness.
- Search turnaround time dropped from 3–4 weeks to just 1–2 weeks, accelerating the client's R&D and IP review cycles.
- Identified critical competitor patents with broad generic claims impacting the client's oncology pipeline strategy.
- Enabled cost-efficient IP intelligence, optimizing both R&D resources and legal expenditure.

Conclusion

Through our expert-augmented, AI-integrated chemical structure search, the client achieved complete coverage across the relevant chemical space—strengthening its freedom-to-operate, R&D direction, and patent filing confidence.

By combining deep chemical expertise, algorithmic precision, and multi-database integration, IeB delivered insights that not only resolved immediate IP uncertainties but also empowered the client to strategically and sustainably refine its oncology portfolio.

Ingenious e-Brain

Ingenious e-Brain is a global research advisory and management consulting firm that helps businesses future-proof their operations by addressing complex challenges with sustainable, strategic, and expert-led solutions. With a global network of over 300 domain experts, analysts, scientists, and consultants across 5 offices in 4 countries—we bring world-class research capabilities and a proven track record of delivering 5,000+ projects across various industries.

With over 13 years of proven excellence, we have successfully tackled business challenges for Fortune 500 and Global 1000 corporations, industry leaders, manufacturing giants, startups, investors, universities, and top companies across domains such as healthcare, sustainability, chemicals, advanced materials, automotive, energy, food & beverage, consumer packaged goods, and high-tech industries, particularly in the field of intellectual property and innovation.

Our services empower organizations to accelerate innovation, optimize R&D portfolios, and navigate complex intellectual property (IP) challenges, all while scaling operations with resilience. We support clients at every stage of the innovation process—from product launches and IP co-creation to market intelligence, consumer sentiment analysis, and gathering actionable customer insights through surveys.

We are located at

India (HQ)
207-208 Welldone TechPark, Sohna Road Sector 48, Gurugram, Haryana 122018

+91 124 429 4218

Delaware (USA)
8 The Green, Suite B, Dover, DE 19901

+1 302 450 1418

California (USA)
99 S Almaden Blvd, Suite 600, San Jose, CA

+1 347 480 2054

London
5, Brayford Square, London, E1 0SG

+447700149056

For enquiries e-mail us at
contact@iebrain.com

Find more about us at
www.iebrain.com

Follow us on