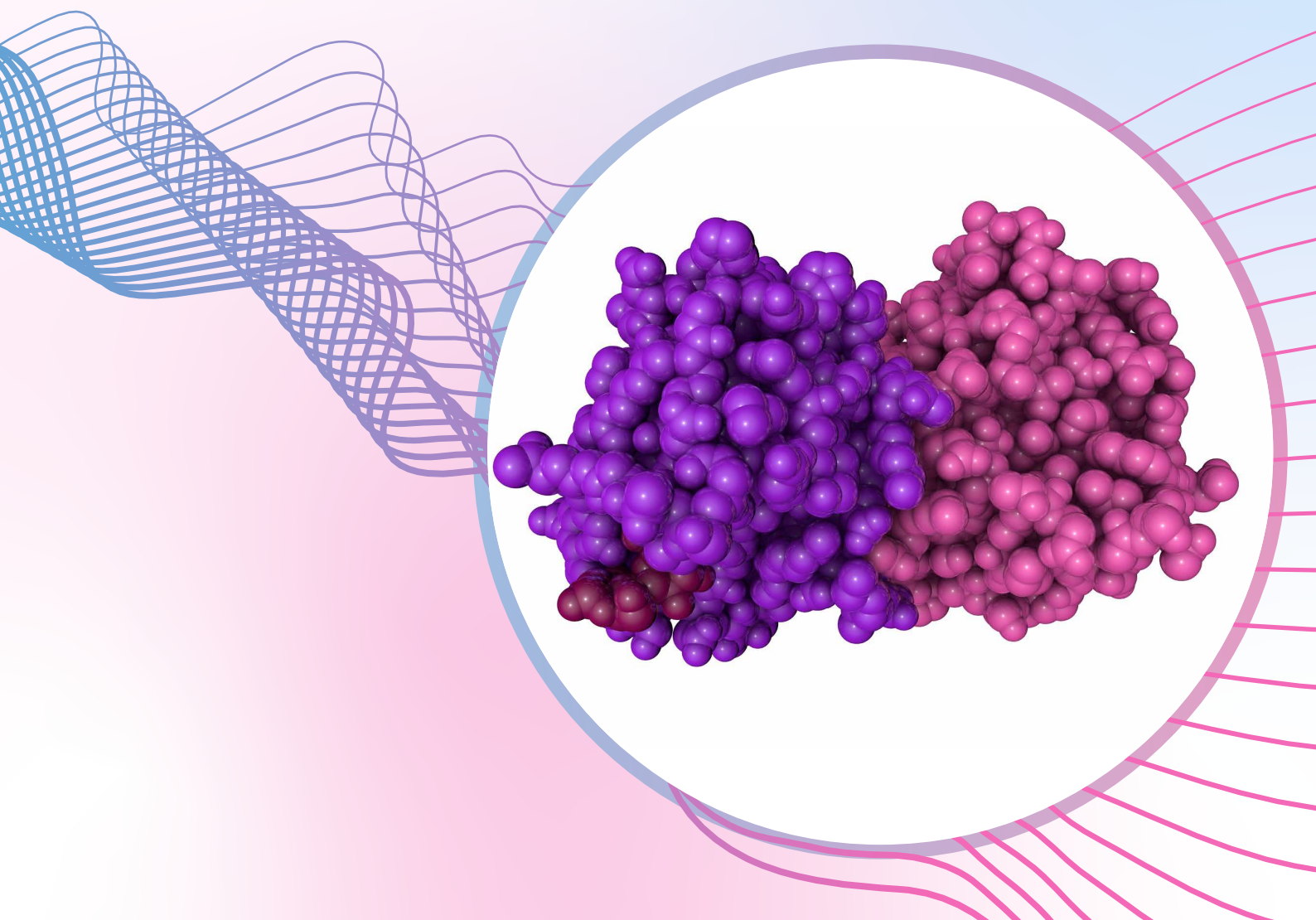


excelra

Supporting KRAS research with GOSTAR data

Extracting patent information with associated structure and activity data to rapidly enhance the customer's KRAS data repository

CASE STUDY





Client

Biopharmaceutical company



Services used

GOSTAR | Data curation services

Context

Kirsten rat sarcoma viral oncogene homolog (KRAS) is a small nucleotide guanosine triphosphate (GTPase) member of the Ras family of oncogenes. KRAS acts as a molecular switch, cycling between inactive (GDP-bound) and active (GTP-bound) states. It converts upstream cellular signals from multiple tyrosine kinases to downstream effectors. This conversion regulates various processes, including cellular proliferation.

Aberrant expression of KRAS accounts for up to 20% of all cancers. Oncogenic KRAS mutations that stabilize GTP binding lead to constitutive activation of KRAS, and downstream signaling appears in 25–30% of lung adenocarcinomas.² Single nucleotide substitutions that result in missense mutations at codons 12 and 13 of the KRAS primary amino acid sequence comprise approximately 40% of these KRAS driver mutations in lung adenocarcinoma.

The KRAS G12D mutation is present in 25% of all pancreatic ductal adenocarcinoma patients, 13.3% of all colorectal carcinoma patients, 10.1% of all rectal carcinoma patients, 4.1% of all non-small cell lung carcinoma patients, and 1.7% of all small cell lung carcinoma patients.

Given its role in malignancy and following the discovery of mutations in tumors, KRAS has attracted a lot of attention in oncology. Many pharmaceutical companies are actively seeking compounds that inhibit KRAS activity and are sufficiently safe and effective to obtain regulatory approval.

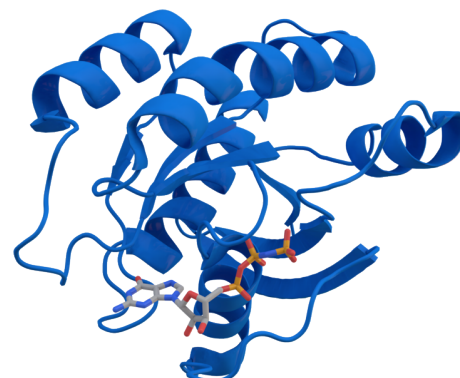


Figure: The crystal structure of Kirsten rat sarcoma viral oncogene homolog (KRAS)¹

Our client's challenge

Our client is a large US biopharmaceutical company engaged in KRAS research. A foundational phase of its R&D activity involves collecting information on existing patents related to KRAS and data on all associated structures and activities. Thousands of patents are published every year relating to KRAS. In 2020, 1,675 patents were published worldwide featuring the term KRAS in either title, abstract, or claim.³ The number increased to 2,050 in 2021⁴ and 2,525 in 2022.⁵ Over the same period, research groups focusing on KRAS increased from 565 in 2020 to 843 in 2022, providing further indication that KRAS has become an increasingly popular research area. Researchers must keep an up-to-date internal repository of data from these patents so they can compare the uniqueness of their solution over competitors and quickly identify avenues of opportunity. But collecting, reviewing, and extracting data from such an extensive collection demands a significant investment of time and resources.

Fortunately, the client was aware of our capabilities in this field. We are the global leader in the manual extraction of SAR data from scientific literature and have been providing data solutions to pharmaceutical companies for 20 years. The client engaged us to extract and validate the required content, and deliver clean, consistent, analysis-ready data.

Our approach

With over 60 PhDs in our data curation team, we have the domain expertise our client demanded to identify the relevant literature, extract the appropriate data, and deliver it in a standardized, analysis-ready format.

Our curation process includes three stages of data extraction: manual curation, review, and quality control. Thanks to our combination of scientific expertise and technical excellence, we were able to collate, prepare and deliver the client's data set in an exceptionally short time. We ensured the manually extracted data included exemplified chemical structures and associated experiments, reported for a variety of assays.

The results

Meeting the client's requirement, we delivered a comprehensive data set that could immediately be added to the internal repository for further review and analysis. Their supplied data included:

- Chemical structure
- InChI
- Compound number or name
- Compound synthesis and associated data
- Compound yield
- Target
- Species
- Cell line
- Endpoint
- Units of measurement
- Measured value
- Title and description of the assay procedure

With our assistance, the client was able to swiftly proceed with the ongoing research program, avoiding the bottleneck of the data collection phase. The efficacy of our manual curation and quality control was greatly appreciated, and the client has repeatedly returned with similar data curation requests.

As the global leader in the manual extraction of SAR data from scientific literature, we regularly replicate these results with some of the world's most successful life science companies. All of our clients understand the value of streamlining their data collection and standardization processes, and we are constantly updating our libraries and expanding our capabilities so more pharmas and biotechs can benefit from the time and cost savings we deliver.

If you have data curation requirements, we can provide what you need. Whatever your objectives, we'll help you achieve them.

For more information about GOSTAR data, contact us

 marketing@excelra.com →

References

- ¹ Thomas Splettstoesser (www.scistyle.com) – Own work, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=53202114>
- ² KRAS Inhibitors | Mirati Therapeutics, Inc. (n.d.). Retrieved from <https://www.mirati.com/science/programs/kras-inhibitors/>
- ³ These results can be found on patent searches with the string TAC: (KRAS) AND PD:[2020-01-01 TO 2020-12-31]
- ⁴ These results can be found on patent searches with the string TAC: (KRAS) AND PD:[2021-01-01 TO 2021-12-31]
- ⁵ These results can be found on patent searches with the string TAC: (KRAS) AND PD:[2022-01-01 TO 2022-12-31]



Where data means more

excelra

BOSTON | UTRECHT | HYDERABAD

Connect with our experts: marketing@excelra.com

www.excelra.com