# Sequence search for biological patent intelligence

## Derwent SequenceBase helps IP teams to search and analyze biological sequences described in global patent publications

Find, analyze and understand biological sequences data from around the world.

Identify prior art and establish freedom to operate and patentability.

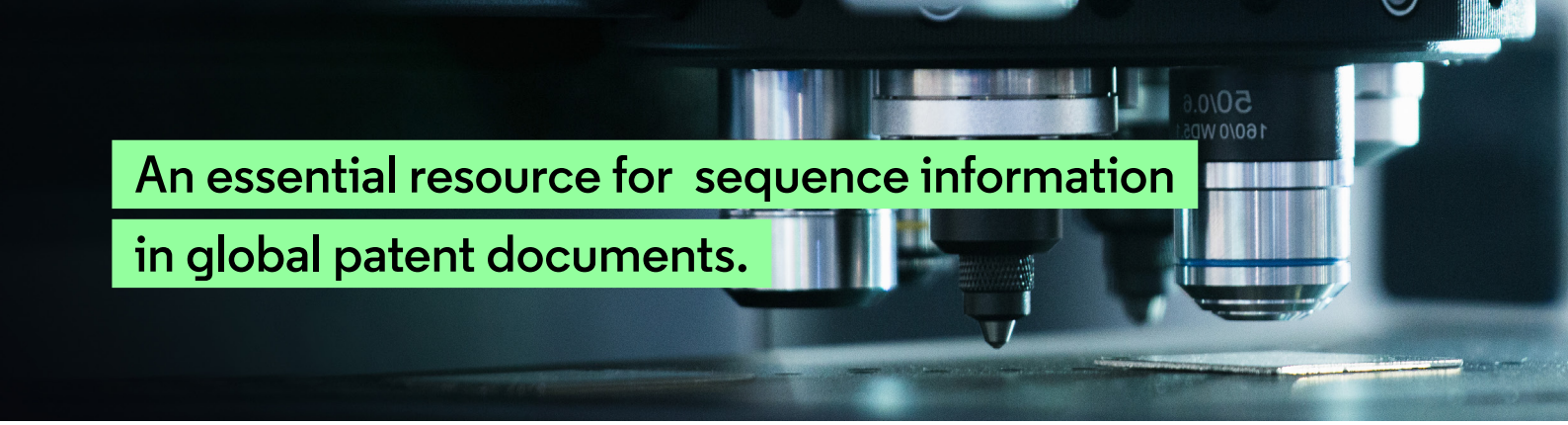
Understand the global patent landscape and competitor activities.

Quickly assess the context of the sequence in the patent.

Surface potential patent infringements.

**Understand the global biological sequence landscape across patent documents**

Clarivate



# An essential resource for sequence information in global patent documents.

## GENESEQ: Spend less time searching and make faster, better informed decisions.

### Comprehensive, timely, annotated coverage

**GENESEQ**

GENESEQ is a proprietary database used to easily search and identify biological sequences (DNA, RNA, protein sequences) covered in patents from 58 issuing authorities. Manually annotated to highlight IP context and biological significance, GENESEQ allows IP professionals and biologic scientists to spend less time searching and understanding sequence data and more time assessing patentability, identifying potential infringing patents and tracking competitor activities.

GENESEQ delivers the entire patent landscape surrounding the biological sequence under investigation with coverage beginning in 1981 from worldwide patent issuing authorities including WO, US, EP, JP, DE, IN, KR, and CN.

**Find context faster with annotated analysis**

Our editorial team of more than 40 biology and life sciences experts provides written summaries that clarify and explain sequence novelty and utility of a given sequence, including:

- Enhanced patent titles that are more intuitive
- Record detail includes organism name, gene/protein name, sequence modification or other highlighted biologically significant regions of the sequence, and associated disease information
- Sequence location within the patent document
- Standardized, full bibliographic data
- Links to identical records from NCBI and SWISSPROT, Gene Ontology (GO) when provided in the patent
- GENESEQ covers 110+ million of sequences from more than hundreds of thousands of unique patents.

Uniquely, we go beyond what is available in electronic sequence submissions to bring you all sequences – even those that are "hard-to-find" in figures and tables ensuring you have the complete picture.

- Nucleic acid sequences 10 or more bases in length
- Amino acid sequences 4 or more residues in length
- All PCR primers and probes of any length

GENESEQ is available through Derwent SequenceBase or can be accessed through an API for organizations that want enterprisewide access to conduct sequence searching and analysis on their in-house platforms.

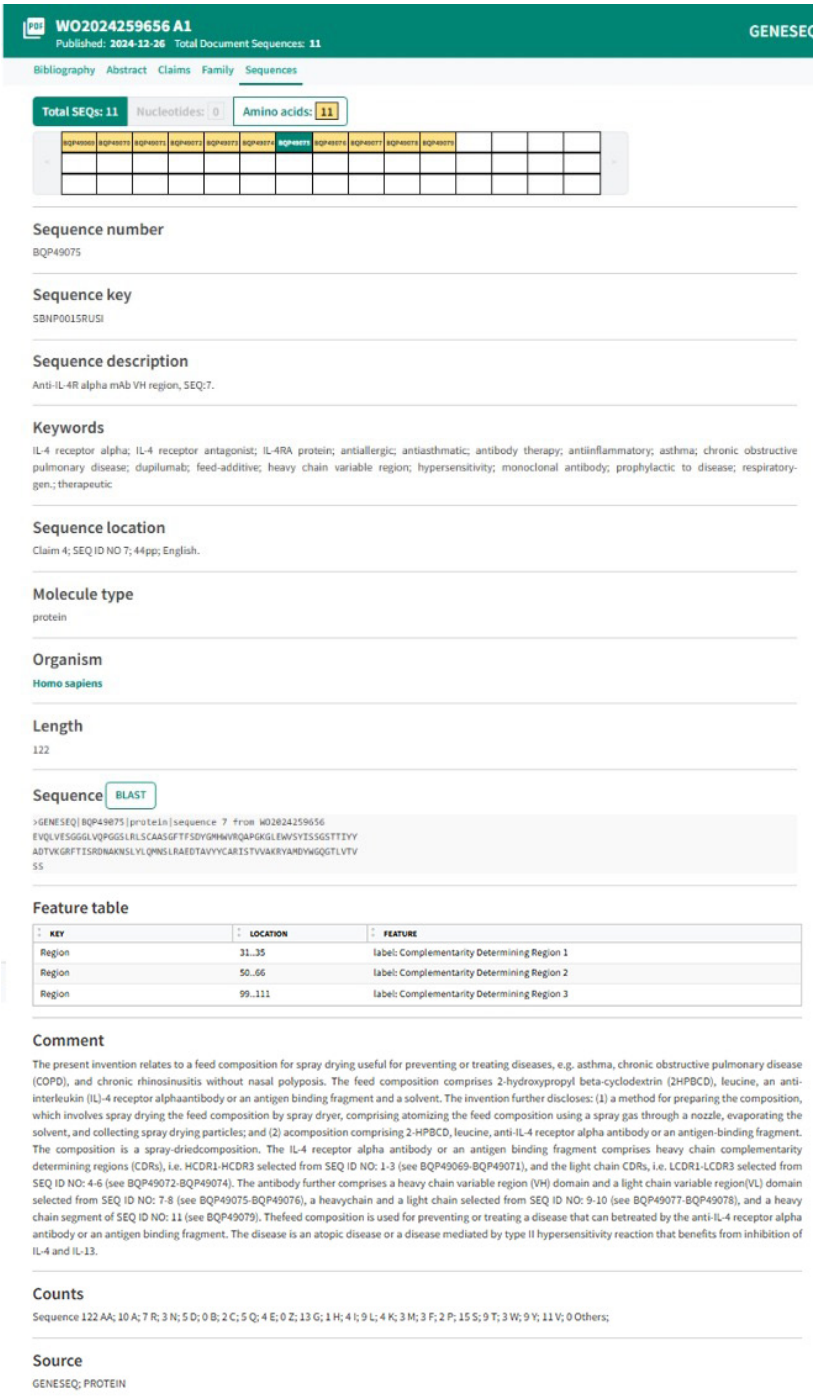
# GENESEQ: Enhanced and curated for more accurate results.

## Example 1: WO2024259656A1

Patent Title: Formulation of an anti-il-4 receptor alpha mab or an antigen-binding fragment thereof comprising 2hpbetacd and leucine

Sequence Listing with 11 sequences provided as part of the specification but NO indication of presence of CDRs.

Example GENESEQ record



WO2024259656 A1
Published: 2024-12-26 Total Document Sequences: 11
GENESEQ

Bibliography Abstract Claims Family Sequences

Total SEQs: 11 Nucleotides: 0 Amino acids: 11

**Sequence number**
BQP49075

**Sequence key**
SBNP0015RUSI

**Sequence description**
Anti-IL-4R alpha mAb VH region, SEQ:7.

**Keywords**
IL-4 receptor alpha; IL-4 receptor antagonist; IL-4RA protein; antiallergic; antiasthmatic; antibody therapy; antiinflammatory; asthma; chronic obstructive pulmonary disease; dupilumab; feed-additive; heavy chain variable region; hypersensitivity; monoclonal antibody; prophylactic to disease; respiratory-gen.; therapeutic

**Sequence location**
Claim 4; SEQ ID NO 7; 44pp; English.

**Molecule type**
protein

**Organism**
Homo sapiens

**Length**
122

**Sequence** BLAST

>GENESEQ|BQP49075|protein|sequence 7 from WO2024259656
EVQLVESGGGLVQPGGSLRLSCAASGFTFSDYGMHWVRQAPGKGLEWVSYISSGSTTIYY
ADTVKGRFTISRDNAKNSLYLQMNSLRAEDTAVYYCARISTVVAKRYAMDYWGQGTLVTV
SS

**Feature table**

| KEY | LOCATION | FEATURE |
|---|---|---|
| Region | 31..35 | label: Complementarity Determining Region 1 |
| Region | 50..66 | label: Complementarity Determining Region 2 |
| Region | 99..111 | label: Complementarity Determining Region 3 |

Manually curated Feature Table highlighting CDRs

**Comment**
The present invention relates to a feed composition for spray drying useful for preventing or treating diseases, e.g. asthma, chronic obstructive pulmonary disease (COPD), and chronic rhinosinusitis without nasal polyposis. The feed composition comprises 2-hydroxypropyl beta-cyclodextrin (2HPBCD), leucine, an anti-interleukin (IL)-4 receptor alphaantibody or an antigen binding fragment and a solvent. The invention further discloses: (1) a method for preparing the composition, which involves spray drying the feed composition by spray dryer, comprising atomizing the feed composition using a spray gas through a nozzle, evaporating the solvent, and collecting spray drying particles; and (2) acomposition comprising 2-HPBCD, leucine, anti-IL-4 receptor alpha antibody or an antigen-binding fragment. The composition is a spray-driedcomposition. The IL-4 receptor alpha antibody or an antigen binding fragment comprises heavy chain complementarity determining regions (CDRs), i.e. HCDR1-HCDR3 selected from SEQ ID NO: 1-3 (see BQP49069-BQP49071), and the light chain CDRs, i.e. LCDR1-LCDR3 selected from SEQ ID NO: 4-6 (see BQP49072-BQP49074). The antibody further comprises a heavy chain variable region (VH) domain and a light chain variable region(VL) domain selected from SEQ ID NO: 7-8 (see BQP49075-BQP49076), a heavychain and a light chain selected from SEQ ID NO: 9-10 (see BQP49077-BQP49078), and a heavy chain segment of SEQ ID NO: 11 (see BQP49079). Thefeed composition is used for preventing or treating a disease that can betreated by the anti-IL-4 receptor alpha antibody or an antigen binding fragment. The disease is an atopic disease or a disease mediated by type II hypersensitivity reaction that benefits from inhibition of IL-4 and IL-13.

**Counts**
Sequence 122 AA; 10 A; 7 R; 3 N; 5 D; 0 B; 2 C; 5 Q; 4 E; 0 Z; 13 G; 1 H; 4 I; 9 L; 4 K; 3 M; 3 F; 2 P; 15 S; 9 T; 3 W; 9 Y; 11 V; 0 Others;

**Source**
GENESEQ; PROTEIN

# Comprehensive global patented sequence coverage

## USGENE

A trusted resource for life science intellectual property professionals, providing unrivaled searchable access to all available peptide and nucleotide sequences from the published applications and issued patents of the United States Patent and Trademark Office.

With over 292 million biological sequences from more than 436K unique U.S. patents and published patent applications, it contains a wealth of essential sequence information not available anywhere else. With coverage beginning in 1981, the database is continually growing with hundreds of documents containing biological sequences added twice per week and within 24 hours of patent publication.

## WOGENE

Providing comprehensive coverage from international published patent applications containing nucleic acid and protein sequences from: the World Intellectual Property Organization, the European Patent Office, the Japanese Patent Office, and the Korean Intellectual Property office.

With coverage dating back to 1979, contains a wealth of essential sequence information not available anywhere else. Additional documents containing biological sequences are added twice per week and within 24 hours of publication.

Coverage in USGENE and WOGENE includes:

- Nucleic acid sequences 10 or more bases in length
- Amino acid sequences 4 or more residues in length
- All PCR primers and probes of any length
- Sequence information, including organism name, molecule type, and modifications

Database annotation includes:

- Patent title
- Sequence identity number (SEQ ID NO) within the sequence listing
- Original source organism
- Standardized bibliographic data

## GenBank

NIH's genetic sequence database is an annotated collection of all publicly available DNA sequences. GenBank is part of the International Nucleotide Sequence Database Collaboration, which comprises the DNA DataBank of Japan (DDBJ), the European Nucleotide Archive (ENA), and GenBank at NCBI.

With over 1 trillion nucleotide bases from more than 486 million individual sequences, GenBank provides access within the scientific community to the most up-to-date and wide-ranging DNA sequence information.

# Derwent SequenceBase

## Complete search algorithm suite to match your use case

### BLAST

Heuristics-based Basic Local Alignment Search Tool from NCBI finds regions of local similarity between sequences.

### MOTIF

Search using regular expressions for specific pattern searching or search specific substitutions.

### Antibody

Find patents using CDRs as queries optionally combined with heavy/ light chains to filter results.
- CDR Variation Tables
- Robust Filtering & Subsearching

### Smith-Waterman

Aligns sequences using dynamic programming to find mathematically optimal local alignment.

### Multiple Sequence Search

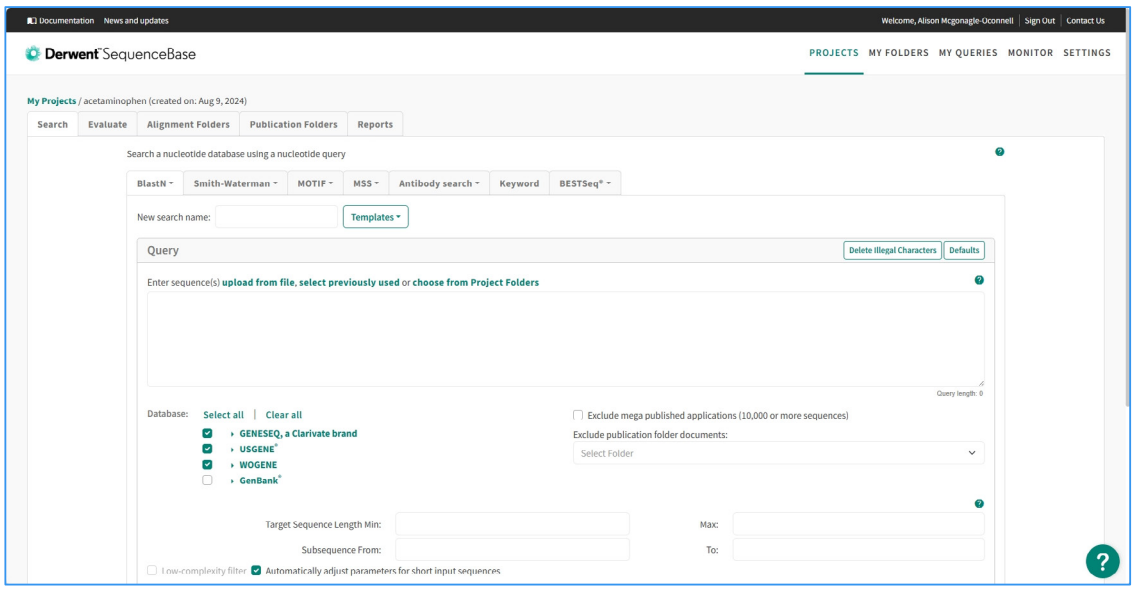
Search for multiple sequences using the Smith-Waterman algorithm, with specialized result analysis capabilities.

### Keyword

Search for sequences by keywords, patents bibliographic information and sequences annotations.

### BestSeq

Proprietary search algorithm based on a speed optimization of Smith-Waterman for patent searching.



# Derwent SequenceBase integrations

## Integrated with these solutions to make it easy to incorporate patent data for biologic sequences into your current workflows

Derwent Innovation – for **patent** search

Integrates full text, enhanced patent and litigation data with sequence patent search results

Cortellis – for life sciences **drug discovery**

Integrates robust, real-time scientific and regulatory data alongside patent sequence information

DartsIP – **litigation intelligence**

Quickly surface litigation surrounding patented sequences

Derwent SequenceBase – **Sequence search** for biological patent intelligence

Connects to a rich, end-to-end ecosystem

BXD1601036403 / 13

## About Clarivate

Clarivate is a leading global provider of transformative intelligence. We offer enriched data, insights & analytics, workflow solutions and expert services in the areas of Academia & Government, Intellectual Property and Life Sciences & Healthcare. For more information, please visit **clarivate.com**.

Discover solutions at:

**clarivate.com**