

BIOSEQUENCE INFORMATION

DGENE

The authoritative
source for your
sequence
searching.
Here.
On STN.

 FIZ KARLSRUHE

FIZ Karlsruhe provides DGENE on STN International

A gem among biotechnology sequence databases, DGENE (Derwent Geneseq) covers nucleic acid and amino acid sequences from patents published by 40 patent offices worldwide. More than half of the sequence data that appears in DGENE is not available in any other public sequence database. It comprises more than 2.9 million records from 1981 to date (status: Aug. 2002) and is updated biweekly. DGENE is offered by FIZ Karlsruhe via the STN International online network.

Stay at the cutting edge of your field of science! Take advantage of the unique features and value-added data provided in this database.

DGENE gives you the competitive edge

- ▶ Find out about the state-of-the-art before commencing your R&D project – avoid duplicate research
- ▶ Gain access to sequence information extracted from original patent documents issued by 40 patent authorities (including the World Intellectual Property Organization, the US, European and Japanese patent offices)
- ▶ Exploit Derwent-enhanced titles, abstracts, and indexing produced by Derwent's bioinformatics experts
- ▶ Make your searching a success by using the powerful search options including advanced homology analysis and straightforward direct match
- ▶ Identify competitors and monitor their patent activities
- ▶ Monitor your patented sequences

Database Content

- ▶ A comprehensive global collection of biosequences published in patents
- ▶ New nucleotide sequences of 10 or more bases in length
- ▶ New peptide and protein sequences of 4 or more amino acid residues in length
- ▶ Sequences shorter than 4 amino acid residues or 10 nucleotides, when they are of crucial importance for the patented invention
- ▶ Polymerase chain reaction (PCR) primers and nucleotide probes of any length
- ▶ Mutated sequences which can be made using a wild-type sequence, even if the mutated sequence does not appear in the specification
- ▶ Known sequences for which a novel use is being claimed
- ▶ Derwent-enhanced titles, abstracts, indexing, and annotations per sequence
- ▶ Patent family bibliographic information from WPINDEX (Derwent World Patents Index)

Database Functionalities

- ▶ Text search in the following fields: title, abstract, keywords, description, and organism name
- ▶ Current awareness searching (SDI)
- ▶ Both exact match and homology searching available
- ▶ Two algorithms for homology calculation: FASTA, BLAST
- ▶ Both online and offline (BATCH) homology searching possible
- ▶ Current awareness sequence homology searching (ALERT)
- ▶ Sequence searching using STN command language or the Sequence Search Assistant in STN on the Web (menu-driven)

Searching biosequences using STN command language

Direct match (Run GETSEQ)

- ▶ Nucleic acid search
- ▶ Protein search
- ▶ Family protein search, which allows chemical family-equivalent substitution of amino acids

Run GETSEQ Options:

/SQEN	- Nucleic exact match
/SQSN	- Nucleic sub-sequence
/SQEP	- Protein exact match
/SQEFP	- Protein exact match (incl. chemical family substitution)
/SQSP	- Protein sub-sequence (DEFAULT)
/SQSFP	- Protein sub-sequence (incl. chemical family substitution)

Upload a query sequence

For uploading a sequence, use the UPLOAD command in STN Express or the UPLOAD SEQUENCE button in the left-hand menu bar in STN on the Web. Check the uploaded query with D LQUE.

Need further explanation of how to use DGENE? Just type "HELP DIRECTORY" when online or consult the DGENE complete help text at www.stn-international.de/training_center/bioseq/dgene_help.pdf.

Find more details on sequence searching in DGENE at www.stn-international.de/service/faq/dgenefaq.pdf.

Similarity (homology) searching (Run GETSIM, Run BLAST)

- ▶ Retrieval of sequences that include the exact or similar sequences
- ▶ Two algorithms: FASTA-based GETSIM and BLAST
- ▶ Assignment of similarity score, similarity percentage (GETSIM), and identity percentage (BLAST) respectively
- ▶ Protein sequences
- ▶ Nucleotide sequences
 - ▶ single strand
 - ▶ complementary strand
 - ▶ both single and complementary strand
- ▶ Translated similarity (search with a protein query sequence against a nucleotide database)
 - ▶ single strand
 - ▶ complementary strand
 - ▶ both single and complementary strand
- ▶ Offline BATCH option for similarity searching
- ▶ ALERT feature for sequence homology-based current awareness

RUN GETSIM and RUN BLAST Options:

/SQP	- Protein similarity (GETSIM and BLAST default)
/SQN	- Nucleotide similarity
/SQN SIN	- Single strand (GETSIM SQN - DEFAULT)
/SQN COM	- Complementary strand
/SQN BOTH	- Single and complementary strands (BLAST SQN - DEFAULT)
/TSQN	- Translated similarity (protein query sequence against nucleotide database)
/TSQN SIN	- Single strand (GETSIM TSQN - DEFAULT)
/TSQN COM	- Complementary strand
/TSQN BOTH	- Single and complementary strands (BLAST TSQN - DEFAULT)

Add "BATCH" to run the search offline, e.g. RUN GETSIM AGCTTGTGCGAA/SQN BOTH BATCH

Add "ALERT" to initiate a current awareness search, e.g. RUN BLAST AGCTTGTGCGAA/SQN ALERT

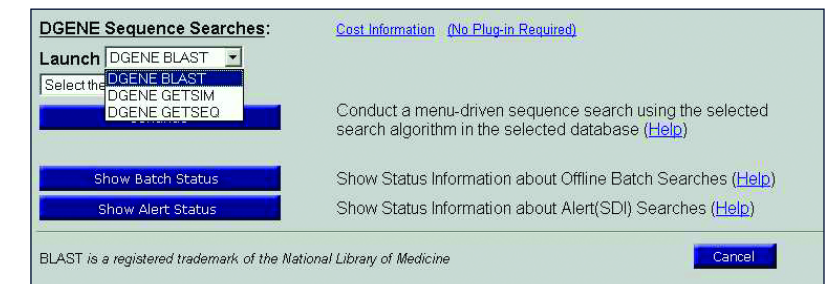
Searching biosequences using Sequence Search Assistant in STN on the Web

Enter the DGENE Sequence Search Assistant via the left-hand menu bar in STN on the Web. Select Search Assistants and then Sequence Assistant. The lower part of the Sequence Search Assistant allows you menu-driven sequence code matching (GETSEQ) and sequence homology searching (GETSIM and BLAST) in DGENE without the need to know the STN command language.

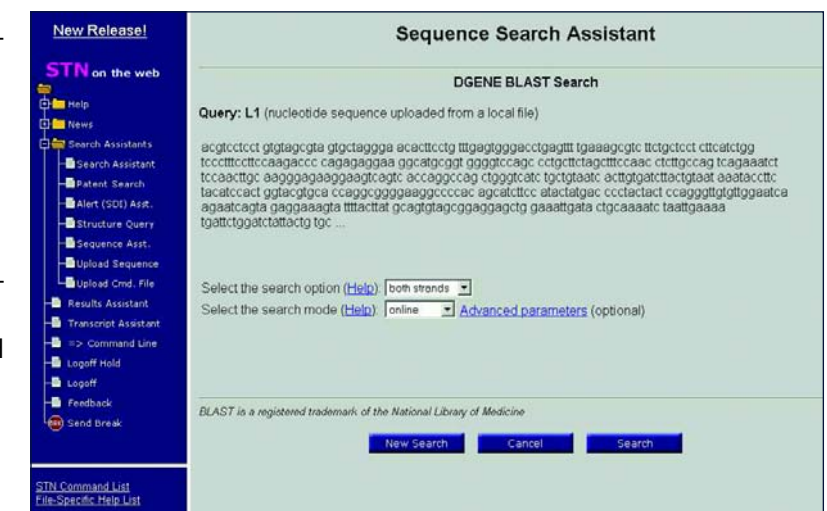
You can conduct a search or check the status and retrieve results of all your offline BATCH searches and ALERTs. To conduct a sequence search, choose the search kind (direct match with GETSEQ or sequence homology with FASTA-based GETSIM or BLAST) in the corresponding pull-down menu. Select the type of search (protein or nucleotide) and make sure that your query sequence corresponds to the type of search you selected. Enter your query sequence manually, upload it from a text file or recall a sequence previously uploaded in the same session.

All search options as given for command-line sequence searching are available in the menu and may be selected accordingly. Conduct your homology sequence search online or offline as a BATCH or initiate a current awareness ALERT. You may also change the advanced parameters for your BLAST homology search. For BLAST advanced settings, please consult NCBI documentation (www.ncbi.nlm.nih.gov/BLAST/).

If you prefer menu-driven searching, the Sequence Search Assistant in STN on the Web walks you through the whole sequence search in DGENE!



No plug-in is required for sequence searching in DGENE.



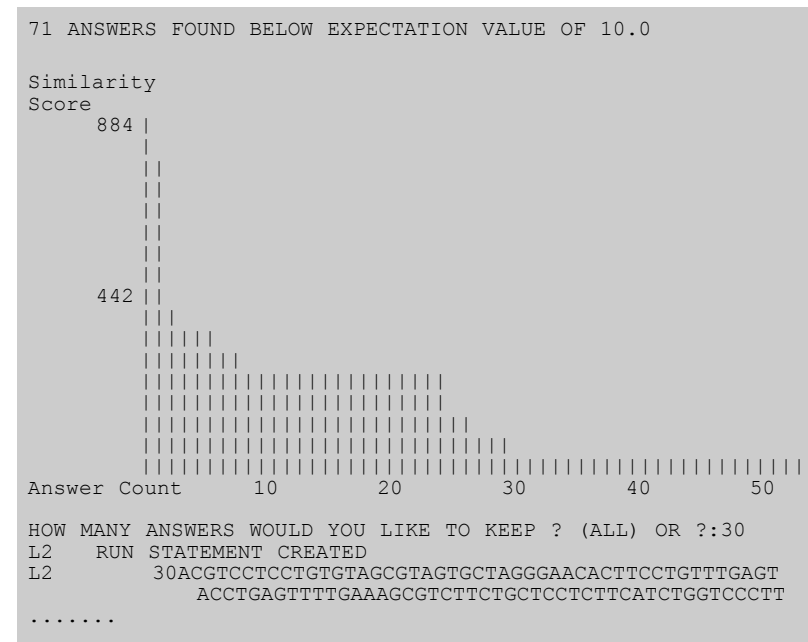
Please keep in mind to use valid sequence formats and to observe maximum sequence lengths for each search option.

Retrieve, process & display your results (command-line mode)

Start the search

```
=> RUN BLAST L1/SQN
```

Graphic display of candidate answers



The number of homologous hit sequences is shown on x-axis in the order of homology indicated by their similarity scores (y-axis).

You have the option to keep the whole answer set or only an individual range of the most homologous hit sequences. Enter the exact number of answers you want to keep for further processing.

Similarity Score

```
=> sor score d
PROCESSING COMPLETED FOR L2
L3 30 SOR L2 SCORE D
```

Answer set can be sorted by similarity score: sort score d (descending)

The Alignment and Similarity Score

A special format (ALIGN) is available to show the alignment between your query and the retrieved sequence. The GETSIM and BLAST alignment output, respectively will be delivered depending on the underlying algorithm of the homology search conducted.

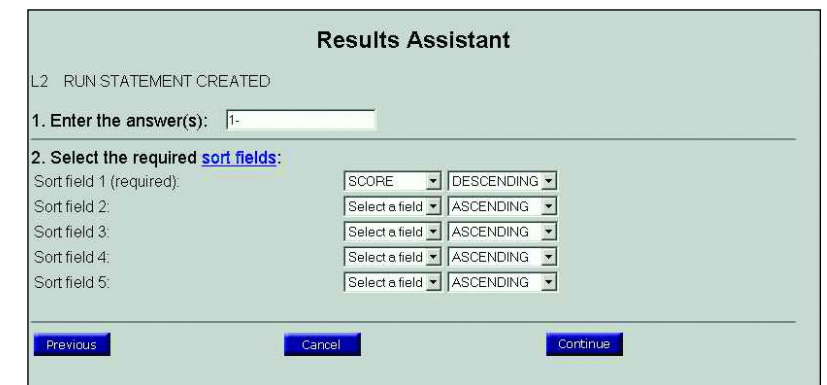
```
=> d 10 score align
```

The align format is FREE!

Retrieve, process & display your results (menu-driven mode)

The Sequence Search and the Results Assistants

The Results Assistant offers the possibility of processing your sequence search results: Sort the hit sequences by descending score or by patent family. To sort your sequence search answer set by patent family, use the corresponding Sort button. A special selection procedure for display of patent-family sorted answer sets is provided in the Results Assistant.



To sort by similarity score, use the Sort by Field Code option. Only the selection of Sort field 1 is required: Select the field SCORE and DESCENDING.

To display results, choose single answers (e.g. 1, 5, 10) or ranges of hits (e.g. 1-10) or a combination of both from the sorted answer set. Select from the menu the format(s) available for display of answers. For details on formats, see the Display format(s) link.

View the results displayed in the selected formats. See patent and sequence information as well as details of alignment and similarity scores. Take advantage of the links to related patent information. The full-text link offers access to the STN Full-Text Solution.

The Other Source (OS) field is linked to the underlying Derwent World Patents Index document: See the DWPI patent document relating to the sequence information retrieved by sequence homology search in DGENE.

A worldwide scientific institution, FIZ Karlsruhe produces, provides and markets scientific and technical information services in print and electronic form. In cooperation with national and international institutions, FIZ Karlsruhe produces databases in the fields of energy, nuclear research and technology, crystallography, plastics, mathematics, computer science and physics. FIZ Karlsruhe also provides a search service for R&D in corporations and institutions.

FIZ Karlsruhe operates STN International in Europe. STN International (The Scientific & Technical Information Network) is the world's premier online service offering access to bibliographic, factual and full-text databases in science and technology. STN's product palette encompasses more than 210 databases with approx. 350 million documents from all fields of science and technology, among them the world's largest and most important patent files as well as subject-related business files.

STN International is jointly operated by FIZ Karlsruhe, Germany; Chemical Abstracts Service (CAS), Columbus OH, USA; and The Japan Science and Technology Corporation (JST), Tokyo.

For questions and details
please contact us

FIZ Karlsruhe
STN Europe
P.O. Box 2465
76012 Karlsruhe, Germany

Tel.: +49 (0) 7247/808-555
Fax: +49 (0) 7247/808-259
E-mail: helpdesk@fiz-karlsruhe.de
www.fiz-karlsruhe.com

or contact your local STN representative:
www.stn-international.de/service/stnagents/stnagent.html

STN Service Centers

FIZ Karlsruhe
STN Europe
P.O. Box 2465
D-76012 Karlsruhe
Germany

E-mail: helpdesk@fiz-karlsruhe.de
www.fiz-karlsruhe.com
Tel.: +49 7247 808-555
Fax: +49 7247 808-259

Chemical Abstracts Service
STN North America
2540 Olentangy River Road
P.O. Box 3012
Columbus, Ohio 43210-0012
U.S.A.

E-mail: help@cas.org
www.cas.org
Free phone +1 800 848 6533
Tel.: +1 614 447 3700
Fax: +1 614 447 3751

The Japan Science and
Technology Corporation (JST)
STN Japan
5-3 Yonbancho Chiyoda-ku
Tokyo 102-0081
Japan

E-mail: helpdesk@mr.jst.go.jp
www.jst.go.jp
Tel.: +81 3 5214 8414
Fax: +81 3 5214 8410

August 2002
© FIZ Karlsruhe 2002