

USGENE® (USPTO Genetic Sequence Database)

Subject Coverage	Peptide and nucleic acid sequence data			
File Type	Bibliographic			
Features	Alerts (SDIs)	Weekly		
	CAS Registry Number® Identifiers	<input type="checkbox"/>	Page Images	<input type="checkbox"/>
	Keep & Share	<input checked="" type="checkbox"/>	SLART	<input checked="" type="checkbox"/>
	Learning Database	<input type="checkbox"/>	Structures	<input type="checkbox"/>
Record Content	<ul style="list-style-type: none"> • All available peptide and nucleic acid sequences from the published applications and issued patents of the United States Patent and Trademark Office (USPTO). • Extensive bibliographic and text search options, including publication title, abstract, patent assignees at issue, full inventor names, plus the complete set of publication, application, priority, and parent case WIPO/PCT numbers and dates. 			
File Size	<ul style="list-style-type: none"> • More than 74.2 million records (07/2020) • More than 52.0 million nucleic acid sequences (07/2020) • More than 22.1 million protein sequences (07/2020) • More than 200,000 U.S. publications from 1981 to date 			
Coverage	1981-present			
Updates	Weekly. Typically available within 3 days of publication by the USPTO.			
Language	English			
Database Producer	SequenceBase Corporation 3 Dellview Drive Edison, NJ 08820-2545 USA			
	Copyright Holder			
Database Supplier	FIZ Karlsruhe STN Europe P.O. Box 2465 76012 Karlsruhe Germany Phone: +49-7247-808-555 Fax: +49-7247-808-259 Email: helpdesk@fiz-karlsruhe.de			

Sources

Published applications and issued patents of the USPTO.

User Aids

- Online Helps (HELP DIRECTORY lists all help messages available)
 - STNGUIDE
 - USGENE Workshop Manual:
<http://www.stn-international.com/en/training-center/documentation/usgene-workshop-manual>
-

Clusters

- ALLBIB
 - AUTHORS
 - BIOSCIENCE
 - CORPSOURCE
 - HPATENTS
 - MEDICINE
 - PATENTS
 - PHARMACOLOGY
- STN Database Clusters information:
<http://www.stn-international.de/en/customersupport/customer-support#cluster+%7C+subjects+%7C+features>
-

Search and Display Field Codes

Fields that allow left truncation are indicated by an asterisk (*).

General Search Fields

Search Field Name	Search Code	Search Examples	Display Codes
Basic Index* (contains single words from the title (TI), abstract (AB), organism species (ORGN), molecule type (MTY), and description (DESC) fields)	None or /BI	S PLANT ?GENE? AND RNA S DNA DEPENDENT	TI, AB, ORGN, MTY, DESC
Abstract	/AB	S ANAPHYLATOXIN/AB	AB
Accession Number	/AN	S 4305872.2/AN	AN
Application Country (WIPO code and text)	/AC	S US/AC AND L1	AI
Application Date (1)	/AD	S JAN 2000/AD	AI
Application Number (2)	/AP	S US1980-162082/AP	AI
Application Year (1)	/AY	S 2002/AY	AI
Description	/DESC	S GHRH/DESC	DESC
Document Type (code and text)	/DT	S PATENT/DT	DT
Entry Date (1)	/ED	S ED>=JAN 2007	ED
Exemplary Claim	/ECLM (or /MCLM)	S NANOCRYSTAL/ECLM	ECLM
Expiration Date (calculated) (1)	/XPD	S JUN 2022-DEC 2022/XPD	XPD
Expiration Year (1)	/XPY	S 2020-2022/XPY	XPD
Feature Table *	/FEAT	S ?COMBINAT?/FEAT S (RNA AND BINDING)/FEAT	FEAT
File Segment	/FS	S NUCLEIC/FS S PROTEIN/FS	FS
Inventor	/IN (or /AU)	S MANDREKAR MICHELLE?/IN	IN
Molecule Type	/MTY	S RNA/MTY	MTY
Note	/NTE	S DISCLAIMER/NTE	NTE
Organism	/ORGN	S CRICETULUS AUREUS/ORGN	ORGN
Patent Assignee (3)	/PA (or /CS)	S AMGEN/PA	PA
Patent Country (WIPO code and text)	/PC	S US/PC	PI
Patent Kind Code	/PK	S USB2/PK	PI
Patent Number (2)	/PN (or /PATS)	S US6686189/PN	PI
Patent Sequence Location	/PSL	S CLAIM/PSL	PSL
Patent Term Adjustment (number of days) (1)	/PTA	S 100-150/PTA	PTA
Priority Country	/PRC	S JP/PRC	PRAI
Priority Date (1)	/PRD	S 20070202/PRD	PRAI
Priority Number (2)	/PRAI (or PRN)	S US2002-500660/PRN	PRAI
Priority Year (1)	/PRY	S 2006/PRY	PRAI

USGENE**General Search Fields (cont'd)**

Search Field Name	Search Code	Search Examples	Display Codes
Publication Date (1)	/PD	S PD=4 MAR 2003 S 20030403/PD	PI
Publication Year (1)	/PY	S 2003/PY	PI
Related Application Country (WIPO code and text)	/RLC	S WO/RLC	RLI
Related Application Date (1)	/RLD	S 20010313/RLD	RLI
Related Application Number (2)	/RLN (or /RLI)	S WO2000-EP4786/RLN	RLI
Related Application Year (1)	/RLY	S 2000-2001/RLY	RLI
Sequence Identity Number (1)	/SEQN	S 337/SEQN	SEQN
Sequence Length (1)	/SQL	S 150-175/SQL	SQL
Sequence Number Count (1)	/SEQC	S 3-4/SEQC	SEQC
Sequence Source	/SSO	S NCBI/SSO	SSO
Title	/TI	S HYBRIDIZATION ASSAY#/TI	TI
Update Date (1,4)	/UP	S UP=30 JUL 2010	UP

(1) Numeric search field that may be searched using numeric operators or ranges.

(2) Either STN or Derwent format may be used.

(3) Search with implied (S) proximity is available in this field.

(4) Updates of records are available since 30 July 2010.

Super Search Fields

Enter a super search code to execute a search in one or more fields that may contain the desired information. Super search fields facilitate crossfile and multifile searching. EXPAND may not be used with super search fields. Use EXPAND with the individual field codes instead.

Search Field Name	Search Code	Fields Searched	Search Examples	Display Codes
Application Number Group	/APPS	/AP, /RLN, /PRN	S US2001-809522/APPS	AI, RLI, PRAI

Sequence Similarity Searching (BLAST/GETSIM)

The GETSIM and BLAST® run packages are available to search the USGENE database for protein and nucleotide sequence data by similarity (homology). BLAST is provided in USGENE with the permission of the National Center for Biotechnology Information (NCBI) of the National Library of Medicine (NLM). GETSIM is provided in USGENE by FIZ Karlsruhe GmbH, and is based upon the FASTA algorithm.

To initiate a BLAST or GETSIM search the following search codes have to be specified: SQP for searching peptide sequences (default), SQN for nucleotide sequences, or TSQN for searching peptide sequences translated from USGENE nucleotide sequences. The GETSIM or BLAST search can be run in offline BATCH mode or used as the basis of a current-awareness ALERT. The offline search mode offers an email notification option which allows users to see when batch search results are available for download. When using the SQN option, it is possible to specify whether single (SIN), complementary (COM), or BOTH strands should be searched. The options can be specified together with the search code, e.g., /SQN COM. If no search option is given, SIN (single) will be used by default for GETSIM, and BOTH (both) will be used by BLAST. Note that for the TSQN option generally both strands will be searched, i.e., for a single polypeptide query, the TSQN option will cover all six possible translations (three reading frames of both the single and the complementary nucleotide sequences). Nucleotide and protein sequences can be subjected to a similarity search in various ways. A query can be prepared with the query command and saved beforehand, it can be entered directly on the command line using RUN GETSIM/BLAST, or it may be uploaded from an ASCII file using the UPLOAD command. You may also use the Sequence Query Upload Wizard from STN Express version 8.3+. A diagram is generated that shows the similarity between the retrieved sequences and the query. The x-axis represents the number of answers with a specific degree of similarity (represented by y-axis). In addition, two values are given, the query self score value defining the maximum score value possible when the query is aligned to itself, and the score value of the best answer of the retrieved answer set. You have three possibilities to select the result answer set.

You can either:

- 1) Keep the complete answer set (ALL).
- 2) Keep a subset of the complete answer set by specifying a smaller number of top scoring answers.
- 3) Specify the minimum percentage of the self score value, to keep a subset of the complete answer set, where the answers have a better score than your chosen minimum percentage of the query self score value.

The generated L-number contains all answers or the specified subset of answers, but they are sorted by descending accession number. This L-number may be re-arranged by descending similarity score or descending percent identity. Just type "SOR SCORE D" to sort by descending similarity score or "SOR IDENT D" to sort by descending percent identity and the corresponding L-number at an arrow prompt.

It is possible to see the alignment between the retrieved sequence and the query sequence with the display format ALIGN (for GETSIM or for BLAST). The top line is the query sequence and the bottom line the hit sequence. The BLAST ALIGN format follows the standard convention for NCBI alignment displays. The GETSIM ALIGN format uses two dots to represent identical nucleotides/peptides, a blank if there is no match, and one dot to indicate a chemical "family" match. Gaps inserted in the query or answer sequence for alignment purposes are shown with an underscore.

In addition to the sequence alignment, a special format SEQO is provided in USGENE. The display format SEQO shows the corresponding original sequence which might include the nucleotide sequence of a USGENE record together with the protein sequence it expresses as given by the patent applicant.

GETSIM / BLAST: Types of Searches

Description	Search Code	Search Example (4)
Peptide Homology Nucleotide Homology Single Strand Complementary Strand Both Strands Translated Peptide Homology	/SQP /SQN /TSQN	RUN BLAST L1 /SQP RUN BLAST L1 /SQN RUN GETSIM L1 /SQN SIN (1) RUN GETSIM L1 /SQN COM RUN BLAST L1 /SQN BOTH (2) RUN BLAST L1 /TSQN RUN GETSIM L1 /TSQN
Offline BATCH search	/SQP BATCH /SQN BATCH /TSQN BATCH	RUN BLAST L1 /SQP BATCH RUN GETSIM L1 /TSQN BOTH BATCH RUN BLAST L1/TSQN BATCH RUN BLAST L1/SQP COM ALERT
Current-awareness ALERT (3)	/SQP ALERT /SQN ALERT /TSQN ALERT	RUN BLAST L1/SQP ALERT RUN GETSIM L1 /TSQN ALERT

(1) GETSIM default setting

(2) BLAST default setting

(3) Homology ALERT search, which runs every update of the database (once a week)

(4) Where L1 is a sequence query generated using the UPLOAD or QUERY command

Advanced User Option for BLAST

For the experienced user of BLAST, a variety of options is available via the STN command line. Altering these parameters will have a profound effect on the outcome of the search. FIZ Karlsruhe strongly recommends that users are completely familiar with NCBI documentation before embarking on customizing any of these settings. For further information: http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs.

The advanced user options are specified with a single letter code preceded by a hyphen and followed by a blank and the required value, e.g., RUN BLAST L1 /SQN -E 0.1.

Advanced User Options

Option	Switch	Values
1. Filter	-f	T (True), F (False), C (coiled-coil). Default value is T. If T is set, for peptides the SEG and for nucleotides the DUST filter is employed. C represents the "coiled coil" filter.
2. Expectation Value	-e	Floating point number. Default is 10.
3. Word Size	-w	11 (default) or 7-23 for nucleotides 3 (default) or 2 for peptides
4. Strand	-s	1 (sin), 2 (com), or 3 (both; default)
5. Matrix	-m	BLOSUM62 (default), BLOSUM80, BLOSUM45, PAM30, or PAM70
6. Gap Penalty	-g	11 (peptides) (default) 5 (nucleotides) (default)
7. Gap Extension	-x	1 (peptides) (default) 2 (nucleotides) (default)
8. Penalty for nucleotide mismatch	-q	-3 (default)
9. Reward for nucleotide match	-r	1 (default)

BLAST Matrix settings (for option 5.)

Please note that for a certain matrix only a restricted set of possible gap and gap extension values is possible. The settings available to each matrix are summarized in the table below. Default settings are indicated in the table. Any different combinations will be rejected by the system and a warning message issued.

Matrix	Gap	Gap Extension
BLOSUM62	9	2
	8	2
	7	2
	12	1
	11	1 (default)
	10	1
BLOSUM80	8	2
	7	2
	6	2
	11	1
	10	1 (default)
BLOSUM45	9	1
	13	3
	11	3
	12	3
	9	3
	15	2 (default)
	14	2
	13	2
	12	2
	19	1
	18	1
17	1	
PAM30	7	2
	6	2
PAM70	5	2
	10	1
	8	1
	9	1 (default)
	8	2
	7	2
	6	2
	11	1
	10	1 (default)
	9	1

8

USGENE

Example: Online BLAST homology search for a nucleotide sequence and subsequent refinement with text terms and dates

=> **upload**

```
IS THIS DATA A QUERY, OR FOR A RUN PACKAGE? Q/R/(END):r
ENTER NAME OF RUN PACKAGE, END OR (?):blast
START LOCAL KERMIT TRANSMIT PROCESS
```

```
UPLOAD SUCCESSFULLY COMPLETED
L1 GENERATED
```

=> **d l1 lque**

```
L1 ANSWER 1 USGENE COPYRIGHT 2007 SEQUENCEBASE CORP on STN
LQUE gtatatataa cgtgatgagc gtacgggtgc ggagacgcac cggagcgcctcgcgccagccgc
cgctccaagc ccctgagggt tccggggacc acaatgaacaagttgctgtg ctgcgcgctc
gtgtttctgg acatctccat taagtggaccaccaggaaa cgtttcctcc aaagtacctt
cattatgacg aagaaacctctcatcagctg ttgtgtgaca aatgtcctcc tggtagctac
ctaaaacaacactgtacagc aaagtggaag accgtgtgcy ccccttgccc
tgaccactactacacagaca gctggcacac cagtgcgag tgtctatact
gcagccccgtgtgcaaggag ctgcagtagc tcaagcagga gtgcaatcgc
accacaaccgcgtgtgca atgcaaggaa gggcgctacc ttgagataga
gttctgcttgaaacatagga gctgccctcc tggatttgga gtgggtgcaag
ctggaacccagagcgaat acagtttgca aaagatgtcc agatgggttc
ttctcaaagtagacgtcatc taaagcacc ttagagaaaac acacaaattg
cagtgtctttggtctcctgc taactcagaa aggaaatgca acacacgaca
acatatgttccggaaacagt gaatcaactc aaaaatgtgg aatagatgtt
accctgtgtgaggaggcatt cttcagggtt gctgttccta caaagtttac
gcctaactggcttagtgtct tggtagacaa ttgctctggc accaaagtaa
acgcagagagtgtagagagg ataaaacggc aacacagctc acaagaacag
actttccagctgctgaagtt atggaaacat caaaaacaaag cccaagatat
agtcaagaagatcatccaag atattgacct ctgtgaaaac agcgtgcagc
ggcacattggacatgctaac ctacacctcg agcagcttcg tagcttgatg
gaaagcttaccgggaaagaa agtgggagca gaagacattg aaaaaacaat
aaaggcatgcaaacccagtg accagatcct gaagctgctc agtttggtgc
gaataaaaaatggcgaccaa gacacctga agggcctaag gcacgcacta
aagcactcaaagagctacca ctttcccaa actgtcactc agagtctaaa
gaagaccatcaggttccttc acagcttcac aatgtacaaa ttgtatcaga
agttatTTTTtagaaatgata ggtaaccagg tccaatcagt aaaaataagc
tgcttataactggaaatggc cattgagctg tttcctcaca attggcgaga tcccatggatgataa
```

=> **run blast L1/sqn**

BLAST Version 2.2

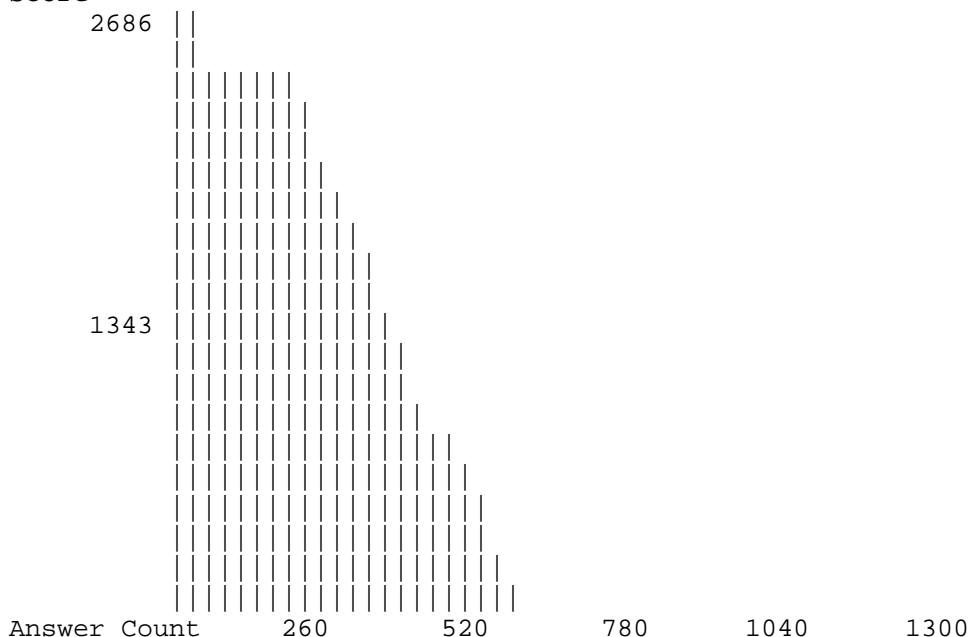
The BLAST software is used herein with permission of the National Center for Biotechnology Information (NCBI) of the National Library of Medicine (NLM). See also, Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." *Nucleic Acids Res.* 25:3389-3402

.....

1281 ANSWERS FOUND BELOW EXPECTATION VALUE OF 10.0

QUERY SELF SCORE VALUE IS 2686
 BEST ANSWER SCORE VALUE IS 2686

Similarity
 Score



ENTER EITHER THE NUMBER OF ANSWERS YOU WISH TO KEEP
 OR ENTER MINIMUM PERCENT OF SELF SCORE FOLLOWED BY %
 (BEST ANSWER PERCENTAGE IS 100%)

ENTER (ALL) OR ? :80%

L2 RUN STATEMENT CREATED

L2 226 GTATATATAACGTGATGAGCGTACGGGTGCGGAGACGCACCGGAGCGCTC
 GCCCAGCCGCGCTCCAAGCCCCTGAGGTTTCCGGGGACCACAATGAACA
 AGTTGCTGTGCTGCGCGCTCGTGTCTGGACATCTCCATTAAGTGGACC
 ACCCAGGAAACGTTTCTCCTCAAAGTACCTTCATTATGACGAAGAAACCTC

 AGACGTACCACTTTCCAAAAGTGTCACTCAGAGTCTAAAGAAGACCATC
 AGGTTTCCTTACAGCTTCAATGTACAAATTGTATCAGAAGTTATTTTT
 AGAAATGATAGGTAACCAGGTCCAATCAGTAAAAATAAGCTGCTTATAAC
 TGGAAATGGCCATTGAGCTGTTTCTCCTCACAATTGGCGAGATCCCATGGAT
 GATAA/SQN. -E 10.0

Answer set arranged by accession number; to sort by descending
 similarity score, enter at an arrow prompt (=>) "sor score d".

=> **sor score d**

PROCESSING COMPLETED FOR L2

L3 226 SOR L2 SCORE D

=> **d trial score align 1 226**

L3 ANSWER 1 OF 226 USGENE COPYRIGHT 2008 SEQUENCEBASE CORP on STN
 TI OSTEOPROTEGERIN (PublishedApplication)

10

USGENE

MTY Nucleic acid
SQL 1355
SCORE 2686 100% of query self score 2686
BLASTALIGN

Query = 1355 letters
Length = 1355
Score = 2686 bits (1355), Expect = 0.0
Identities = 1355/1355 (100%)
Strand = Plus / Plus

```
Query: 1 gtatatataacgtgatgagcgtacgggtgaggagacgcaccggagcgctcgcccagccgc
      |
Sbjct: 1 gtatatataacgtgatgagcgtacgggtgaggagacgcaccggagcgctcgcccagccgc

Query: 61 cgctccaagcccctgaggtttccggggaccacaatgaacaagttgctgtgctgcgcgctc
      |
Sbjct: 61 cgctccaagcccctgaggtttccggggaccacaatgaacaagttgctgtgctgcgcgctc

Query: 121 gtgtttctggacatctccattaagtggaccaccaggaaacgtttcctccaaagtacctt
      |
Sbjct: 121 gtgtttctggacatctccattaagtggaccaccaggaaacgtttcctccaaagtacctt
      .
      .
Query: 1261 ggtaaccagggtccaatcagtaaaaataagctgcttataactggaaatggccattgagctg
      |
Sbjct: 1261 ggtaaccagggtccaatcagtaaaaataagctgcttataactggaaatggccattgagctg

Query: 1321 ttcctcacaattggcgagatcccatggatgataa 1355
      |
Sbjct: 1321 ttcctcacaattggcgagatcccatggatgataa 1355
```

L3 ANSWER 226 OF 226 USGENE COPYRIGHT 2008 SEQUENCEBASE CORP on STN
TI Albumin fusion proteins (PublishedApplication)

MTY DNA
SQL 1140
SCORE 2252 83% of query self score 2686
BLASTALIGN

Query = 1355 letters
Length = 1140
Score = 2252 bits (1136), Expect = 0.0
Identities = 1139/1140 (99%)
Strand = Plus / Plus

```
Query: 157 gaaacgtttcctccaaagtaccttcattatgacgaagaaacctctcatcagctggtgtg
      |
Sbjct: 1 gaaacgtttcctccaaagtaccttcattatgacgaagaaacctctcatcagctggtgtg

Query: 217 gacaaatgtcctcctggtacctacctaacaactgtacagcaaagtggaagaccgtg
      |
Sbjct: 61 gacaaatgtcctcctggtacctacctaacaactgtacagcaaagtggaagaccgtg

Query: 277 tgcgccccttgccctgaccactactacacagacagctggcacaccagtgacgagtgtcta
      |
Sbjct: 121 tgcgccccttgccctgaccactactacacagacagctggcacaccagtgacgagtgtcta
      .
      .
      .
```

```
Query: 1177 actcagagtctaaagaagaccatcagggttccttcacagcttcacaatgtacaaattgtat
          |||
Sbjct: 1021 actcagagtctaaagaagaccatcagggttccttcacagcttcacaatgtacaaattgtat

Query: 1237 cagaagttatTTTTtagaaatgataggtaccagggtccaatcagtaaaaataagctgctta
          |||
Sbjct: 1081 cagaagttatTTTTtagaaatgataggtaccagggtccaatcagtaaaaataagctgctta
```

=> s l3 and ((tnf## or tumor necrosis factor?)(2a)(receptor? or binding protein?))/bi,eclm and ay>1990 and granted/ss0

L5 10 L4 AND ((TNF## OR TUMOR NECROSIS FACTOR?)(2A)(RECEPTOR? OR BINDING PROTEIN?))/BI,ECLM AND AY>1990 AND GRANTED/SSO

=> d bib ab eclm align 1-

```
L5 ANSWER 1 OF 10 USGENE COPYRIGHT 2008 SEQUENCEBASE CORP on STN
AN 7094564.1 DNA (genomic) USGENE
TI Human tumor necrosis factor receptor (Patent)
IN Greene John (Gaithersburg, MD); Fleischmann Robert D. (Gaithersburg, MD)
PA Human Genome Sciences Inc (Rockville MD)
PI US 7094564 B1 20060822
AI US 1995-469637 19950606
PRAI US 1995-469637 19950606
     WO 1995-US3216 19950315
XPD 20230822 (calculated)
PSL Claim 1; SEQ ID NO 1
DESC DNA (genomic); sequence 1 of 10
DT Patent
AB A human TNF receptor and DNA (RNA) encoding such receptor and a procedure
   for producing such receptor by recombinant techniques is disclosed. Also
   disclosed are methods for utilizing such receptor for screening for
   antagonists and agonists to the receptor and for ligands for the receptor.
   Also disclosed are methods for utilizing such agonists to inhibit the
   growth of tumors, to stimulate cellular differentiation, to mediate the
   immune response and anti-viral response, to regulate growth and provide
   resistance to certain infections. The use of the antagonists as a
   therapeutic to treat autoimmune diseases, inflammation, septic shock, to
   inhibit graft-host reactions, and to prevent apoptosis is also disclosed.
   Also disclosed are diagnostic methods for detecting mutations in the
   nucleic acid sequence encoding the receptor and for detecting altered
   levels of the soluble receptor in a sample derived from a host.
ECLM US7094564 B1: What is claimed is:1. An isolated polynucleotide
     comprising the polynucleotide recited in SEQ ID NO:1.
```

BLASTALIGN

```
Query = 1355 letters
Length = 1527
Score = 2561 bits (1292), Expect = 0.0
Identities = 1302/1304 (99%), Gaps = 1/1304 (0%)
Strand = Plus / Plus
```

```
Query: 50 cgcccagccgccgc-tccaagcccctgaggtttccggggaccacaatgaacaagttgctg
          |||
Sbjct: 1 cgcccagccgccgcctccaagcccctgaggtttccggggaccacaatgaacaagttgctg

Query: 109 tgctgcgcgctcgtgtttctggacatctccattaagtggaccaccaggaaacgtttcct
          |||
Sbjct: 61 tgctgcgcgctcgtgtttctggacatctccattaagtggaccaccaggaaacgtttcct
```

USGENE

```

Query: 169  ccaaagtaccttcattatgacgaagaaacctctcatcagctggtgtgtgacaaatgtcct
          |||
Sbjct: 121  ccaaagtaccttcattatgacgaagaaacctctcatcagctggtgtgtgacaaatgtcct
          .
          .
Query: 1129 atgcacgcactaaagcactcaaagacgtaccactttccaaaactgtcactcagagtcta
          |||
Sbjct: 1081 atgcacgcactaaagcactcaaagacgtaccactttccaaaactgtcactcagagtcta

Query: 1189 aagaagaccatcagggttccttcacagcttcacaatgtacaaattgtatcagaagttattt
          |||
Sbjct: 1141 aagaagaccatcagggttccttcacagcttcacaatgtacaaattgtatcagaagttattt

Query: 1249 ttagaaatgataggtaccaggtccaatcagtaaaaataagctgcttataactggaaatg
          |||
Sbjct: 1201 ttagaaatgataggtaccaggtccaatcagtaaaaataagctgcttataactggaaatg

Query: 1309 gccattgagctgtttcctcacaattggcgagatcccatggatga 1352
          |||
Sbjct: 1261 gccattgagctgtttcctcacaattggcgagatcccatggatga 1304

```

.....

```

L5  ANSWER 10 OF 10  USGENE COPYRIGHT 2008 SEQUENCEBASE CORP on STN
AN  6613544.5  DNA      USGENE
TI  Osteoprotegerin (Patent)
IN  Boyle William J. (Moorpark, CA); Lacey David L. (Thousand Oaks, CA);
    Calzone Frank J. (Pine Crest Circle, CA); Chang Ming-Shi (Newbury Park,
    CA)

```

```

PA  Amgen Inc (Thousand Oaks CA)
PI  US 6613544      B1    20030902
AI  US 1995-577788  19951222
DT  Patent

```

```

AB  The present invention discloses a novel secreted polypeptide, termed
    Osteoprotegerin, which is a member of the tumor necrosis factor receptor
    superfamily and is involved in the regulation of bone metabolism. Also
    disclosed are nucleic acids encoding Osteoprotegerin, polypeptides,
    recombinant vectors and host cells for expression, antibodies which bind
    Osteoprotegerin, and pharmaceutical compositions. The polypeptides are
    used to treat bone diseases characterized by increased resorption such as
    osteoporosis.

```

```

ECLM US6613544 B1: What is claimed is:1. An isolated nucleic acid encoding a
    polypeptide comprising the amino acid sequence from residues 1 to 401 or
    from residues 22 to 401 as shown in FIG. 9B (SEQ ID NO:6).

```

```

BLASTALIGN
Query  = 1355 letters
Length = 1355
Score  = 2686 bits (1355), Expect = 0.0
Identities = 1355/1355 (100%)
Strand = Plus / Plus

```

```

Query: 1      gtatatataacgtgatgagcgtacgggtgcgagacgcaccggagcgcctcgcccagccgc
          |||
Sbjct: 1      gtatatataacgtgatgagcgtacgggtgcgagacgcaccggagcgcctcgcccagccgc

Query: 61     cgctccaagcccctgaggtttccggggaccacaatgaacaagttgctgtgctgcgcgctc
          |||
Sbjct: 61     cgctccaagcccctgaggtttccggggaccacaatgaacaagttgctgtgctgcgcgctc

```

```
Sbjct: 1141 aagcactcaaagacgtaccactttcccaaaactgtcactcagagtctaaagaagaccatc
Query: 1201 aggttccttcacagcttcacaatgtacaaattgtatcagaagttatTTTTAGAAATGATA
|
|
|
Sbjct: 1201 aggttccttcacagcttcacaatgtacaaattgtatcagaagttatTTTTAGAAATGATA
Query: 1261 ggtaaccaggtccaatcagtaaaaaataagctgcttataactggaaatggccattgagctg
|
|
|
Sbjct: 1261 ggtaaccaggtccaatcagtaaaaaataagctgcttataactggaaatggccattgagctg
Query: 1321 tttcctcacaattggcgagatcccatggatgataa 1355
|
|
|
Sbjct: 1321 tttcctcacaattggcgagatcccatggatgataa 1355
```

Searching Sequence Data with the GETSEQ Run Package

Sequence information (amino acid and nucleic acid sequences) may be retrieved by using a variety of search fields available with the GETSEQ run package. The query may be first created with the QUERY command, and subsequently searched through the GETSEQ run package specifying the query L-number (e.g., RUN GETSEQ L9, if L9 represents the sequence query). The L-number may also derive from a previous sequence search in another STN database with biosequence search capabilities, e.g., the CAS REGISTRYSM, DGENE. The query may also be directly entered within the GETSEQ package at a colon prompt after GETSEQ has been initialized with the RUN command (i.e., RUN GETSEQ). Offline sequence searching is also available for GETSEQ searches.

Sequence Search Terms: Sequence Code Match

Terms	Query Examples
One-letter codes for common amino acid (1) Three-letter codes for common amino acids (1) One-letter codes for nucleic acids (2)	QUE LAGLL/SQSP QUE 'MET-GLY-LEU-TRP-TRP-ARG'/SQSFP QUE ACTACCTTCAAATACTAC/SQEN QUE ACTACCTTCAAATACTAC/SQSN

- (1) Enter 'HELP AAC' at an arrow prompt to display a table of the one- and three-letter codes for common amino acids.
(2) Enter 'HELP NUC' at an arrow prompt to display a table of the codes for nucleic acids.

Types of Sequence Searches: Sequence Code Match

Sequence data for nucleic acid and protein sequences are displayed in the SEQ field with one-letter codes and the SEQ3 field with three-letter codes for proteins only.

Type	Definition	Search Code	Query Examples
Sequence Exact Protein	Search for sequences that match the query. (2)	/SQEP	QUE AKRSSKM/SQEP QUE 'ALA-LYS-ARG-SER-SER-LYS'/SQEP
Sequence Exact Family, Protein	Search for sequences that match the query and those in which family-equivalent substitution of the query amino acids occur. (1,2)	/SQEFP	QUE SETLR/SQEFP QUE 'SER-GLU-THR-LEU-ARG'/SQEFP
Subsequence, Protein	Search for exact answers plus sequences in which the query sequence is embedded. (2)	/SQSP	QUE SKGYF/SQSP QUE 'SER-LYS-GLN-TYR-PHE'/SQSP

Types of Sequence Searches: Sequence Code Match (cont'd)

Type	Definition	Search Code	Query Examples
Subsequence Family, Protein	Search for exact sequences, subsequences, and answers in which family-equivalent substitution of the query amino acids occurs. (1,2)	/SQSFP	QUE SYVVE/SQSFP QUE 'SER-TYR-VAL-VAL-GLU'/SQSFP
Sequence Exact, Nucleic Acid	Search for sequences that match the query. Ambiguity codes for nucleic acids are allowed. (2)	/SQEN	QUE ACTACCTTCAAATACTAC/SQEN
Subsequence Nucleic Acid	Search for exact answers, plus sequences in which the query sequence is embedded. Ambiguity codes for nucleic acids are allowed. (2)	/SQSN	QUE ACTACCTTCAAATACTAC/SQSN

(1) The families of amino acid equivalents retrieved in protein family searches are:

- P, A, G, S, T (neutral, weakly hydrophobic)
- Q, N, E, D, B, Z (hydrophilic, acid amine)
- H, K, R (hydrophilic, basic)
- F, Y, W (hydrophobic, aromatic)
- L, I, V, M (hydrophobic)
- C (cross-link forming)

(2) Variability symbols are allowed.

Variability Symbols for Sequence Code Match Searches (1, 2)

Symbol(s)	Function	Query Examples
[]	to specify alternate residues	QUE LGP[VL]/SQSP QUE LGP['VAL'LEU'LYS']/SQSP
[-]	to exclude a specific residue or alternate residues	QUE LGP[-H]/SQSP QUE LGP[-'HIS']/SQSPSP QUE LGP[-HL]/SQSP
{m}	to repeat the preceding sequence or sequence query (L#) m times	QUE (FL){2}/SQSP QUE L4{2}/SQSP QUE (CTG){2}/SQSN QUE TAA(TAAA){2}/SQSN
{m,u} or {m-u}	to repeat the preceding sequence or sequence query (L#) m to u times	QUE GG(FL){1,2}/SQSP QUE L3{1,3}/SQSP QUE (CTG){1,3}/SQSN
? or {0,1} or {0-1}	to repeat the preceding sequence or sequence query (L#) zero or	QUE FLRRI(RP)?K/SQSP QUE FLRRI(RP){0,1}K/SQSP QUE L1{-1}NN/SQSP QUE L1{0,1}NN/SQSP QUE CAT(CGA){0,1}GGAC/SQSN
* or {0,} or {0-}	to repeat the preceding sequence or sequence query (L#) zero or more times	QUE KLK(WD){0,}N/SQSP QUE KLK(WD)*N/SQSP QUE L1{0-}NN/SQSP QUE L1{0,}NN/SQSP QUE CAT(CTG){0,}TATT/SQSN

Variability Symbols for Sequence Code Match Searches (1, 2) (cont'd)

Symbol(s)	Function	Query Examples
+ or {1,} or {1-}	to repeat the preceding sequence or sequence query (L##) one or more times	QUE KLK(DLE){1,}/SQSP QUE KLK(DLE)+/SQSP QUE L2{1-}/SQSP QUE L2{1,}/SQSP QUE CAT(CTG){1,}TATT/SQSN
&	to join together sequence expressions or queries (L#s)	QUE L1&L3/SQSFP QUE L2&L5{1,3}/SQSP

(1) In addition, the caret (^) and the vertical bar (|) may be used. The caret is used at the beginning or at the end of a sequence to search for that sequence at the beginning or end of sequence field. The vertical bar is the symbol for alternation, i.e., it is used to separate alternate sequence queries.

(2) For more information on specifying variability in sequence code match queries, enter HELP SQQ at an arrow prompt (=>).

Specifying Gaps in GETSEQ Sequence Queries

Symbol(s)	Function	Query Examples
.	a gap of one residue	QUE SY.RPG/SQSP QUE SY..RPG/SQSP QUE AAG...TGC/SQSN
.{m} or [m.]	a gap of m residues	QUE SY.{2}RPG/SQSP QUE SY[2.]RPG/SQSP
.{m,u} or .{m-u}	a gap of m to u residues	QUE GFF.{2,10}LSS/SQSP QUE GFF.{2-10}LSS/SQSP QUE AAG.{2,5}TGC/SQSN
: or .? or .{0,1} or .{0-1}	a gap of zero or one residues	QUE AGA:SRI/SQSFP QUE AGA.?SRI/SQSFP QUE AGA.{0,1}SRI/SQSFP QUE AGA.{0-1}SRI/SQSFP
.* or .{0,} or .{0-}	a gap of zero or more residue	QUE HLC.*TYG/SQSP QUE HLC.{0,}TYG/SQSP QUE HLC.{0-}TYG/SQSP QUE AAGGCAGATG.*GCAA/SQSN
.+ or .{1,} or .{1-}	a gap of one or more residues	QUE SY.+TH/SQSP QUE SY.{1,}TH/SQSP QUE SY.{1-}TH/SQSP QUE TCCTG.+GTGG/SQSN

DISPLAY and PRINT Formats

Any combination of formats may be used to display or print answers. Multiple codes must be separated by spaces or commas, e.g., D L1 1-5 TI AU. The fields are displayed or printed in the order requested.

Hit-term highlighting is available for all fields. Highlighting must be ON during SEARCH to use the HIT, KWIC, and OCC formats.

Format	Content	Examples
AB	Abstract	D AB
AI (AP) (1)	Application Information	D AI
AIO (2)	Application Information, Original	D AIO
AN	Accession Number	D AN
CLM	Claims	D CLM
DESC	Description	D DESC
DT (TC)	Document Type	D DT
ECLM (MCLM)	Exemplary Claim	D ECLM
ED	Entry Date	D ED
FEAT	Feature Table	D FEAT
FS (2)	File Segment	D FS
IDENT (2,3)	Percent Identity	D IDENT
IN (AU)	Inventor	D IN
MTY	Molecule Type	D MTY
NTE	Note	D NTE
ORGN	Organism Name	D ORGN
PA (CS)	Patent Assignee	D PA
PI (PN, PATS) (1)	Patent Information	D PI
PRAI (PRN) (1)	Priority Information	D PRAI
PSL	Patent Sequence Location	D PSL
PTA (2)	Patent Term Adjustment (number of days)	D PTA
RLI (RLN) (1)	Related Application Number	D RLI
RLIO (2)	Related Application Information, Original	D RLIO
SCORE (2,4)	Similarity Score	D SCORE
SEQ	Sequence (1-letter codes)	D SEQ
SEQ3	Sequence (3-letter codes)	D SEQ3
SEQC (2)	Sequence Number Count	D SEQC
SEQN (2)	Sequence Identity Number	D SEQN
SEQO	Original Sequence (alignment of nucleotide sequence and peptide sequence it expresses when given)	D SEQO
SQL	Sequence Length	D SQL
SSO	Sequence Source	D SSO
TI	Title	D TI
UP (5)	Update Date	D UP
XPD	Expiration Date	D XPD
ALIGN (4)	Alignment between query and retrieved sequence in a sequence search (RUN GETSIM, RUN GETSEQ or RUN BLAST)	D ALIGN
ALL (1)	AN (MTY), TI, IN, PA, PI, AI, RLI, PRAI, XPD, NTE, PSL, DESC, DT, AB, CLM, SSO, ORGN, SQL, SEQ, FEAT	D ALL
IALL (1)	ALL, indented with text labels	D IALL
APPS (1)	AI, RLI, PRAI	D APPS
BIB (STD) (1)	AN (MTY), TI, IN, PA, PI, AI, RLI, PRAI, XPD, NTE, PSL, DESC, DT (BIB is the default)	D BIB
IBIB (1)	BIB, indented with text labels	D IBIB
BRIEF (1)	AN (MTY), TI, IN, PA, PI, AI, RLI, PRAI, XPD, NTE, PSL, DESC, DT, AB, ECLM, SSO, ORGN, SQL, SEQ, FEAT	D BRIEF
IBRIEF (1)	BRIEF, indented with text labels	D IBRIEF
CFAM (1)	Condensed family format (from INPADOCDB)	D CFAM
FAM (1)	AN, table of patent family information (from INPADOCDB)	D FAM
FASTA	FASTA format	D FASTA
FASTA2	FASTA format, header truncated	D FASTA2
LS (1)	Legal Status (from INPADOCDB)	D LS

DISPLAY and PRINT Formats (cont'd)

Format	Content	Examples
LS2 (1) SQIDE SQ3IDE SCAN (6) TRIAL(TRI, SAM, FREE)	Legal Status (from INPADOCDB), detailed version with display headers TI, SQL, SEQ, FEAT TI, SQL, SEQ3, FEAT TI (random display without answer numbers) TI, DESC, MTY, SQL	D LS2 D SQIDE D SQ3IDE D SCAN D TRIAL
HIT KWIC OCC	Hit term(s) and field(s) Up to 50 words before and after hit term(s) (KeyWord-In-Context) Number of occurrences of hit term(s) and field(s) in which they occur	D HIT D KWIC D OCC

- (1) By default, patent numbers, application and priority numbers are displayed in STN Format. To display them in Derwent format, enter SET PATENT DERWENT at an arrow prompt. To reset display to STN format, enter SET PATENT STN.
- (2) Custom display only
- (3) Use Run BLAST first. See page 3, Similarity Search.
- (4) Use RUN GETSIM, RUN GETSEQ, or RUN BLAST first. See page 3, Similarity Search.
- (5) Updates of records are available since 30 July 2010.
- (6) SCAN must be specified on the command line, e.g., D SCAN or DISPLAY SCAN.

SELECT, ANALYZE, and SORT Fields

The SELECT command is used to create E-numbers containing terms taken from the specified field in an answer set.

The ANALYZE command is used to create an L-number containing terms taken from the specified field in an answer set.

The SORT command is used to rearrange the search results in either alphabetic or numeric order of the specified field(s).

Field Name	Field Code	ANALYZE/ SELECT (1)	SORT
Abstract	AB	Y	N
Accession Number	AN	Y	Y
Application Country	AC	Y	N
Application Date	AD	Y	N
Application Information, Original	AIO	Y	N
Application Number	AP (AI)	Y	N
Application Year	AY	Y	N
Description	DESC	Y	N
Document Type	DT (TC)	Y	Y
Entry Date	ED	Y	Y
Expiration Date	XPD	Y	N
Expiration Year	XPY	Y	N
Feature Table	FEAT	Y	N
File Segment	FS	Y	Y
Inventor	IN (AU)	Y	N
Molecule Type	MTY	Y	Y
Note	NTE	Y	N
Organism Name	ORGN	Y	Y
Patent Assignee	PA (CS)	Y	Y
Patent Country	PC	Y	Y
Patent Information	PI (PN)	Y	Y
Patent Kind Code	PK	Y	Y
Patent Sequence Location	PSL	Y	N
Patent Term Adjustment	PTA	Y	N

SELECT, ANALYZE, and SORT Fields (cont'd)

Field Name	Field Code	ANALYZE/ SELECT (1)	SORT
Percent Identity	IDENT	N	Y
Priority Country	PRC	Y	N
Priority Date	PRD	Y	N
Priority Information	PRAI (PRN)	Y	N
Priority Year	PRY	Y	N
Publication Date	PD	Y	Y
Publication Year	PY	Y	Y
Related Application Country	RLC	Y	N
Related Application Date	RLD	Y	N
Related Application Information Original	RLIO	Y	N
Related Application Number	RLN (RLI)	Y	N
Related Application Year	RLY	Y	N
Sequence (1-letter codes)	SEQ	Y (2)	N
Sequence (3-letter codes)	SEQ3	Y (2)	N
Sequence Identity Number	SEQN	Y	N
Sequence Length	SQL	Y	Y
Sequence Number Count	SEQC	Y	N
Sequence Source	SSO	Y	N
Similarity Score	SCORE (3)	N	Y
Title	TI	Y (default)	Y
Update Date	UP	Y (4)	Y

(1) HIT may be used to restrict terms extracted to terms that match the search expression used to create the answer set, e.g., SEL HIT TI.

(2) Appends /SQSP to the terms created by SELECT.

(3) Used with a L-number created by BLAST or GETSIM

(4) Updates of records are available since 30 July 2010.

Sample Records**DISPLAY ALL**

```

AN 7553483.1 Protein USGENE
TI Chemokine mutants acting as chemokine antagonists (Patent)
IN Proudfoot Amanda (Chens sur Lemans, FR); Kosco-Vilbois Marie (Minzier,
FR); Shaw Jeffrey (Geneva, CH)
PA Merck Serono SA (Coinsins Vaud CH)
PI US 7553483 B2 20090630
US 20050220757 A1 20051006
WO 2003051921 A 20030626
AI US 2002-499100 20021216
RLI WO 2002-EP14325 20021216
PRAI EP 2001-761 20011217
XPD 20221216 (calculated)
NTE Subject to any Disclaimer, the term of this patent is extended or
adjusted under 35 USC 154(b) by 433 days.
PSL Claim 1; SEQ ID NO 1
DESC Artificial Protein; Mutant chemokine; sequence 1 of 13
DT Patent
AB Mutants of specific CC-chemokines containing a non-conservative
substitution in a conserved consensus sequence act as CC-chemokine
antagonists, and can be effectively used in the treatment of autoimmune
and inflammatory diseases, cancers, and viral or bacterial infections.
Particularly preferred are the RANTES/CCL5 mutants having the amino acid
sequence of SEQ ID NO: 1, 2, 3, or 4.
CLM US7553483 B2: 1. An isolated mutant of CCL5 (RANTES) comprising SEQ ID
NO: 1, 4, 5, 6 or 7, wherein said mutant antagonizes in vivo cellular
recruitment caused by CCL5.

```

2. An isolated polypeptide comprising a heterologous polypeptide sequence and a polypeptide selected from the group consisting of SEQ ID NO: 1, 4, 5, 6 and 7.
3. The polypeptide according to claim 2, wherein said heterologous polypeptide sequence comprises an amino acid sequence selected from one or more of the following protein sequences: extracellular domains of membrane-bound protein, immunoglobulin constant regions, multimerization domains, extracellular proteins, signal peptide-containing proteins, or export signal-containing proteins.
4. A pharmaceutical composition comprising a pharmaceutically acceptable excipient and a mutant of CCL5 (RANTES) comprising SEQ ID NO: 1, 4, 5, 6 or 7, wherein said mutant antagonizes in vivo cellular recruitment caused by CCL5.
5. A method of making a pharmaceutical composition comprising combining a pharmaceutically acceptable excipient and a mutant of CCL5 (RANTES) according to claim 1.
6. The isolated mutant according to claim 1, wherein said mutant comprises SEQ ID NO: 1.
7. The isolated mutant according to claim 1, wherein said mutant comprises SEQ ID NO: 4.
8. The isolated mutant according to claim 1, wherein said mutant comprises SEQ ID NO: 5.
9. The isolated mutant according to claim 1, wherein said mutant comprises SEQ ID NO: 6.
10. The isolated mutant according to claim 1, wherein said mutant comprises SEQ ID NO: 7.
11. The isolated polypeptide according to claim 2, wherein said isolated polypeptide comprises a heterologous polypeptide sequence and SEQ ID NO: 1.
12. The isolated polypeptide according to claim 2, wherein said isolated polypeptide comprises a heterologous polypeptide sequence and SEQ ID NO: 4.
13. The isolated polypeptide according to claim 2, wherein said isolated polypeptide comprises a heterologous polypeptide sequence and SEQ ID NO: 5.
14. The isolated polypeptide according to claim 2, wherein said isolated polypeptide comprises a heterologous polypeptide sequence and SEQ ID NO: 6.
15. The isolated polypeptide according to claim 2, wherein said isolated polypeptide comprises a heterologous polypeptide sequence and SEQ ID NO: 7.
16. The pharmaceutical composition according to claim 4, wherein said mutant comprises SEQ ID NO: 1.
17. The pharmaceutical composition according to claim 4, wherein said mutant comprises SEQ ID NO: 4.
18. The pharmaceutical composition according to claim 4, wherein said mutant comprises SEQ ID NO: 5.
19. The pharmaceutical composition according to claim 4, wherein said mutant comprises SEQ ID NO: 6.

USGENE

20. The pharmaceutical composition according to claim 4, wherein said mutant comprises SEQ ID NO: 7.

SSO PROTEIN; USPTO; GRANTED

ORGN Artificial sequence

SQL 68

SEQ

```
1 spysdttpc cfayiarplp rahikeyfyt snkcsnpavv fvtrknrsvc
51 anpekkwvre yinslems
```

FEATURE TABLE:

Key	Location
	mutant chemokine

DISPLAY FASTA

=> D FASTA

FASTA:

```
>USGENE|20100017904.32958|Protein|sequence 32958 from US20100017904
mgevvatweateggagvkgpvvvtgasgflgswlvmkllqagyvratvrdpanvvtkplldlpgater
lslwkadladegsfddairgctgvfhvatpmdfeskdpenevikptvegmmmsimrackeagtvrriivfts
sagtvnieerqrpyvdqdnwsdvdvcqrvmktgwmvfvskslaekaamayaaehgldfisiiptlvvgpf
lsagmpsslitalalvtgnehysilkqvqfvhlldldahlflehpaaagryvcsshdatihglaaml
rdrypeydiperfpgieddlqpvvhfsskklldhgfthfkytvedmfdaairmcrekgliplataggral
```

DISPLAY FASTA2

=>D FASTA2

FASTA2:

```
>USGENE|Protein
mgevvatweateggagvkgpvvvtgasgflgswlvmkllqagyvratvrdpanvvtkplldlpgater
lslwkadladegsfddairgctgvfhvatpmdfeskdpenevikptvegmmmsimrackeagtvrriivfts
sagtvnieerqrpyvdqdnwsdvdvcqrvmktgwmvfvskslaekaamayaaehgldfisiiptlvvgpf
lsagmpsslitalalvtgnehysilkqvqfvhlldldahlflehpaaagryvcsshdatihglaaml
rdrypeydiperfpgieddlqpvvhfsskklldhgfthfkytvedmfdaairmcrekgliplataggral
```

In North America

CAS
STN North America
P.O. Box 3012
Columbus, Ohio 43210-0012 U.S.A.

CAS Customer Center:
Phone: 800-753-4227 (North America)
614-447-3700 (worldwide)
Fax: 614-447-3751
Email: help@cas.org
Internet: www.cas.org

In Europe

FIZ Karlsruhe
STN Europe
P.O. Box 2465
76012 Karlsruhe
Germany
Phone: +49-7247-808-555
Fax: +49-7247-808-259
Email: helpdesk@fiz-karlsruhe.de
Internet: www.stn-international.com

In Japan

JAIICI (Japan Association for
International Chemical Information)
STN Japan
Nakai Building
6-25-4 Honkomagome, Bunkyo-ku
Tokyo 113-0021, Japan
Phone: +81-3-5978-3601 (Technical Service)
+81-3-5978-3621 (Customer Service)
+81-3-5978-3600
Email: support@jaici.or.jp (Technical Service)
customer@jaici.or.jp (Customer Service)
Internet: www.jaici.or.jp