

# **CAS REGISTRY<sup>SM</sup>: Exact and pattern searching of protein sequences**

November 2008



## Table of Contents

Preface .....	3
Protein sequences in REGISTRY .....	5
Search options .....	6
Amino acid codes .....	7
Display options .....	9
Searching for exact sequence strings .....	10
Searching subsequences .....	11
Searching for functionally similar sequences .....	12
Searching motifs and patterns .....	14
Gaps .....	15
Order of execution of symbols .....	16
Using SEQLINK .....	17
Searching length .....	18
Searching chemical annotation .....	19
Searching for references .....	20
Appendix: Blocking groups for protein sequence annotation .....	22



## Preface

This guide provides an overview and examples of exact and pattern searching of protein sequences in the CAS REGISTRY database on STN.

CAS REGISTRY BLAST<sup>®</sup> similarity searching is available using STN Express<sup>®</sup> or STN<sup>®</sup> on the Web<sup>SM</sup>. For information, refer to the *CAS REGISTRY<sup>SM</sup>: BLAST<sup>®</sup> similarity searching via STN Express<sup>®</sup>* guide available at [www.cas.org](http://www.cas.org).

For information on searching in REGISTRY on STN, please refer to the REGISTRY Database Summary Sheet available at [www.cas.org](http://www.cas.org).



## Protein sequences in REGISTRY

Protein sequence data may be searched and displayed in REGISTRY on STN. References to protein sequences may be searched and displayed in bibliographic databases on STN, e.g., CAPlus<sup>SM</sup>.

Protein sequence information in REGISTRY is compiled by CAS from sequences reported in research articles and patents.

Sequences may be found in REGISTRY for the following classes of proteins and peptides from both journal and patent literature:

- Naturally occurring proteins and peptides
- Sequences deduced from gene translation and reported by the author
- Sequences deduced by gene translation from GenBank<sup>®</sup> (a registered trademark of the U.S. Department of Health and Human Services)
- Chemically modified peptides and proteins
- Genetically engineered and synthetic proteins
- Multichain proteins
- Cyclic peptides
- Fusion proteins
- Peptide metal complexes
- Sequences containing uncommon amino acids, i.e., not genetically encoded
- Partial protein sequences

## Search options

To search for sequence information in REGISTRY, enter the SEARCH (or S) command followed by the search string and a field code.

To search for...	Use this field code	Retrieves	Example
Exact Sequence	/SQEP	Exact match; same length	=> S FCFWKTCT/SQEP
Subsequence	/SQSP	Sequences in which the query sequence may or may not be embedded	=> S LAGLL/SQSP
Exact Family	/SQEFP	Functionally similar amino acids; same length	=> S YGGFL/SQEFP
Subsequence Family	/SQSFP	Functionally similar amino acids; may or may not be embedded	=> S ATCXAWV/SQSFP
Sequence Length	/SQL	Sequences of a certain length	=> S SQL<=10
Annotation	/NTE	Sequences with the search term in the NTE field	=> S MULTICHAIN/NTE

## Amino acid codes

These codes are used for displaying or searching protein sequences with four or more amino acids. Dipeptides and tripeptides are also included in REGISTRY, but may be searched only by name or structure and not by sequence representation.

For common amino acids, either one-letter or three-letter codes may be used. Three-letter codes are used for uncommon amino acids. Enclose three-letter codes or strings of codes in single quotes, e.g.,  
**S 'ASP' 'SER' 'SCY' /SQSP.**

### Common amino acids

1-Letter Code	3-Letter Code	Name
A	Ala	Alanine
B	Asx	Aspartic acid or Asparagine
C	Cys	Cysteine
D	Asp	Aspartic acid
E	Glu	Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
J	Xle	Isoleucine or Leucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
O	Pyl	Pyrrolysine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
U	Scy	Selenocysteine
V	Val	Valine
W	Trp	Tryptophan
X	Xxx	Uncommon or Unspecified
Y	Tyr	Tyrosine
Z	Glx	Glutamic acid or Glutamine

**Note:** The codes B, J, and Z may be used only in subsequence searches (/SQSP and /SQSFP).

## Uncommon amino acids

3-Letter Code	Name	3-Letter Code	Name
Aaa	$\alpha$ -amino acid	Hcy	homocysteine
Aad	2-aminoadipic acid (2-aminohexanedioic acid)	Hhs	homohistidine
Aan	$\alpha$ -asparagine	Hiv	2-hydroxyisovaleric acid
Abu	2-aminobutanoic acid	Hse	homoserine
Aca	2-aminocapric acid (2-aminodecanoic acid)	Hva	2-hydroxypentanoic acid
Agn	$\alpha$ -glutamine	Hyl	5-hydroxylysine
Aib	$\alpha$ -aminoisobutyric acid ( $\alpha$ -methylalanine)	Hyp	4-hydroxyproline
Apm	2-aminopimelic acid (2-aminoheptanedioic acid)	Iva	isovaline
App	$\gamma$ -amino- $\beta$ -hydroxybenzenepentanoic acid	Lac	2-hydroxypropanoic acid (lactic acid)
Asu	2-aminosuberic acid (2-aminooctanedioic acid)	Maa	mercaptoacetic acid
Aze	2-carboxyazetidine	Mba	mercaptobutanoic acid
Bal	$\beta$ -alanine	Mhp	3-hydroxy-4-methylproline
Bas	$\beta$ -aspartic acid	Mpa	mercaptopropanoic acid
Bly	3,6-diaminohexanoic acid ( $\beta$ -lysine)	Nal	3-naphthylalanine
Bua	butanoic acid	Nle	norleucine
Bux	4-amino-3-hydroxybutanoic acid	Nty	nortyrosine
Cap	$\gamma$ -amino- $\beta$ -hydroxycyclohexanepentanoic acid	Nva	norvaline
Cha	3-cyclohexylalanine	Oaa	$\omega$ -amino acid
Cit	N <sup>5</sup> -aminocarbonylornithine	Oic	2-carboxyoctahydroindole
Cya	3-sulfoalanine	Orn	ornithine
Dab	2,4-diaminobutanoic acid	Pen	penicillamine (3-mercaptovaline)
Dpm	diaminopimelic acid	Phg	2-phenylglycine
Dpr	2,3-diaminopropanoic acid	Pip	2-carboxypiperidine
Dsu	2,7-diaminosuberic acid (2,7-diaminooctanedioic acid)	Sar	sarcosine (N-methylglycine)
Edc	S-ethylthiocysteine	Spg	1-amino-1-carboxycyclopentane
Ggu	$\gamma$ -glutamic acid	Sta	statin (4-amino-3-hydroxy-6-methylheptanoic acid)
Gla	$\gamma$ -carboxyglutamic acid	Thi	3-thienylalanine
Glc	hydroxyacetic acid (glycolic acid)	Tic	3-carboxyisoquinoline
Glp	pyroglutamic acid	Tle	3-methylvaline
Har	homoarginine	Tml	$\epsilon$ -N-trimethyllysine
		Tza	3-thiazolylalanine
		Und	undefined
		Wil	$\alpha$ -amino-2,4-dioxypyrimidinepropanoic acid

## Display options

To display answers in REGISTRY, enter the DISPLAY (or D) command followed by the L-number resulting from a search, answer numbers or a range of numbers, and display fields or formats.

### Display fields

Code	Content
RN	CAS Registry Number
CN	Chemical Name
PNTE	Patent Annotation
FS	File Segment
SQL	Sequence Length
NTE	Sequence Annotation
SEQ	Sequence (one-letter codes)
SEQ3	Sequence (three-letter codes)
MF	Molecular Formula
CI	Substance Class Identifier
SR	Source of Registration
LC	CAS Registry Number Locator
DT.CA	CAplus Document Type
RL	CAplus Super Roles
RL.NP	CAplus Super Roles from Non-patents
RL.P	CAplus Super Roles from Patents

### Some display formats

Format	Content
ALL	All available fields, including sequence data and the 10 most recent CA references
SQD	Sequence data, one-letter codes
SQD3	Sequence data, three-letter codes
SQIDE	Sequence data, CN, MF, SR, LC, DT.CA, RL, REF
HIT	All fields containing hit terms
KWIC	All hit terms plus 20 words on either side

## Searching for exact sequence strings

Search for exact sequence strings in the Exact Sequence Search (/SQEP) field. This option is most useful when you need to find analogs differing only in chemical modifications.

### Find analogs of the drug Sandostatin with the sequence FCFWKTCT.

1 Enter REGISTRY.

2 Enter S (SEARCH) and the exact sequence in the /SQEP field. You can use one-letter codes for common amino acids.

An L-number answer set (L1) is created. The number of sequences retrieved (456) is displayed.

3 Display sequence data by entering D (DISPLAY), the L-number, the format, and the answer numbers. The SQD format includes the CAS Registry Number<sup>®</sup> and sequence data using one-letter codes.

The answers have the same sequence and length, but they differ in chemical annotation in the NTE field.

```
=> FILE REG
=> S FCFWKTCT/SQEP
      456 FCFWKTCT/SQEP
      78048 SQL=8
L1      456 FCFWKTCT/SQEP
=> D L1 SQD 5-6

L1 ANSWER 5 OF 456 REGISTRY COPYRIGHT 2008 ACS on STN
RN 1015687-20-3 REGISTRY
FS PROTEIN SEQUENCE; STEREOSEARCH
SQL 8
NTE modified
-----
type          location          description
-----
terminal mod. Phe-1          -          N-acetyl
modification  Thr-8          -          undetermined
                                         modification
-----

SEQ          1 FCFWKTCT
            =====
HITS AT:    1-8

**RELATED SEQUENCES AVAILABLE WITH SEQLINK**

L1 ANSWER 6 OF 456 REGISTRY COPYRIGHT 2008 ACS on STN
RN 1000613-79-5 REGISTRY
FS PROTEIN SEQUENCE; STEREOSEARCH
SQL 8
NTE modified (modifications unspecified)
-----
type          location          description
-----
bridge        Cys-2          - Cys-7    disulfide
                                         bridge
modification  Phe-1          -          undetermined
                                         modification
modification  Lys-5          -          undetermined
                                         modification
-----

SEQ          1 FCFWKTCT
            =====
HITS AT:    1-8

**RELATED SEQUENCES AVAILABLE WITH SEQLINK**
```

## Searching subsequences

To find answers matching the search sequence exactly plus sequences in which the query sequence is part of a longer sequence, search in the Subsequence Search (/SQSP) field.

### Find proteins containing the sequence string GLFGRKTGQAP from the human cytochrome c.

1 Enter REGISTRY.

2 Search the subsequence in the /SQSP field. You can use one-letter codes for common amino acids.

3 Display chemical names (CN), sequence length (SQL), and sequences using one-letter codes (SEQ).

Notice the different chemical names and variable sequence length. The query subsequence is highlighted.

```
=> FILE REG
=> S GLFGRKTGQAP/SQSP
L1          176 GLFGRKTGQAP/SQSP
=> D CN SQL SEQ 3, 14
L1  ANSWER 3 OF 176  REGISTRY  COPYRIGHT 2008 ACS on STN
CN  Cytochrome c (human mutation Gly42Ser)  (CA INDEX NAME)
OTHER NAMES:
CN  3: PN: WO2007018437 SEQID: 3 claimed protein
SQL 105
SEQ  1 MGDVEKGKKI FIMKCSQCHT VEKGGKHKHTG PNLHGLFGRK TGQAPGYSYT
                                     =====
      51 AANKNKGIIW GEDTLMEYLE NPKKYIPGTK MIFVGIKKKE ERADLIAYLK
      101 KATNE
HITS AT: 35-45
**RELATED SEQUENCES AVAILABLE WITH SEQLINK**
L1  ANSWER 14 OF 176  REGISTRY  COPYRIGHT 2008 ACS on STN
CN  Cytochrome c (Macaca sylvanus mitochondria-associated gene
CYCS) (9CI)  (CA INDEX NAME)
OTHER NAMES:
CN  GenBank AAY17034
CN  GenBank AAY17034 (Translated from: GenBank AY918495)
SQL 105
SEQ  1 MGDVEKGKKI FIMKCSQCHT VEKGGKHKHTG PNLHGLFGRK TGQAPGYSYT
                                     =====
      51 AANKNKGITW GEDTLMEYLE NPKKYIPGTK MIFVGIKKKE ERADLIAYLK
      101 KATNE
HITS AT: 35-45
```

## Searching for functionally similar sequences

To search for functionally similar sequences, use the “family” search options:

- Family Exact Sequence Search (/SQEFP)
- Family Subsequence Search (/SQSFP)

In family searches, each common amino acid in the query has to match either the exact amino acid or a functionally similar “equivalent,” as shown in the following table.

<b>Property</b>	<b>Functionally Similar Amino Acids</b>
Neutral-Weakly Hydrophobic	Ala,Gly,Pro,Ser,Thr (A, G, P, S, T)
Hydrophilic-Acid Amine	Asn,Asp,Gln,Glu (N, D, Q, E)
Hydrophilic-Basic	Arg,His,Lys (R, H, K)
Hydrophobic	Ile,Met,Leu,Val (I, M, L, V)
Hydrophobic-Aromatic	Phe,Trp,Tyr (F, W, Y)
Cross-linking	Cys (C)

**Find sequences that are functionally similar to the sequence of synthetic somatostatin (AGCKNFFWKFTSC).**

1 Enter REGISTRY.

2 Search the sequence of somatostatin in the /SQEFP field.

3 Display the names (CN), fields in which hit terms occur (HIT), and chemical annotations (NTE).

The sequence length of answers is the same as the length of the query sequence.

```
=> FILE REGISTRY
=> S AGCKNFFWKFTSC/SQEFP
L1          311 AGCKNFFWKFTSC/SQEFP

=> D CN HIT NTE 6-7

L1  ANSWER 6 OF 311  REGISTRY  COPYRIGHT 2008 ACS on STN
CN  L-Cysteine, N-[[2-[5-[1,3-dihydro-3,3-dimethyl-1-(4-sulfobutyl)-2H-indol-2-ylidene]-1,3-pentadienyl]-3,3-dimethyl-1-(4-sulfobutyl)-3H-indolium-5-yl]carbonyl]-L-alanylglycyl-L-cysteinyl-L-lysyl-L-asparaginyL-L-phenylalanyl-L-phenylalanyl-L-tryptophyl-L-lysyl-L-threonyl-L-phenylalanyl-L-threonyl-L-seryl-, inner salt, cyclic (3→14)-disulfide (9CI) (CA INDEX NAME)
FS  PROTEIN SEQUENCE; STEREOSEARCH
SQL  14

SEQ          1 AGCKNFFWKT FTSC
            =====
HITS AT:     1-14

**RELATED SEQUENCES AVAILABLE WITH SEQLINK**
NTE modified (modifications unspecified)
-----
type          location          description
-----
bridge        Cys-3          - Cys-14      disulfide
               bridge
modification  Ala-1          -              undetermined
               modification
-----

L1  ANSWER 7 OF 311  REGISTRY  COPYRIGHT 2008 ACS on STN
CN  L-Cysteine, N-[[4-(carboxymethoxy)phenyl][4,7,10-tris(carboxymethyl)-1,4,7,10-tetraazacyclododec-1-yl]acetyl]-L-alanylglycyl-L-cysteinyl-N6-[(1,1-dimethylethoxy)carbonyl]-L-lysyl-L-asparaginyL-L-phenylalanyl-L-phenylalanyl-L-tryptophyl-N6-[(1,1-dimethylethoxy)carbonyl]-L-lysyl-L-threonyl-L-phenylalanyl-L-threonyl-L-seryl-, cyclic (3→14)-disulfide (9CI) (CA INDEX NAME)
FS  PROTEIN SEQUENCE; STEREOSEARCH
SQL  14

SEQ          1 AGCKNFFWKT FTSC
            =====
HITS AT:     1-14

**RELATED SEQUENCES AVAILABLE WITH SEQLINK**
NTE modified (modifications unspecified)
-----
type          location          description
-----
bridge        Cys-3          - Cys-14      disulfide
               bridge
-----
```

## Searching motifs and patterns

Complex pattern searching of protein sequences is possible in the /SQSP and /SQSFP subsequence search fields by using the Boolean operators (AND, OR, NOT) as well as special characters and symbols.

Use this symbol...	To...	Retrieves	Example
^	Require the string at the beginning or the end of the sequence	MCGIL at the beginning	=> <b>S ^MCGIL/SQSP</b>
		VCDS at the end	=> <b>S VCDS^/SQSFP</b>
[ ]	Specify alternate residues	LGP followed by either V or L	=> <b>S LGP[VL]/SQSP</b>
[-] or [~]	Exclude a residue or alternate residues	PTGKDEA, PTGKNEA, etc.	=> <b>S PTGK[-H]EA/SQSP</b>
{ } with a number or range	Repeat the preceding string or residue	GGFL, GGFLFL, or GGFLFLFL	=> <b>S GG(FL){1-3}/SQSP</b>
?	Repeat the preceding string or residue zero or one time	FLRRIK or FLRRIRPK	=> <b>S FLRRI(RP)?K/SQSP</b>
*	Repeat the preceding string or residue zero or more times	KLKN, KLKWDN, KLKWDWDN, KLKWDWDWDN, etc.	=> <b>S KLK(WD)*N/SQSP</b>
+	Repeat the preceding string or residue one or more times	AQP, AQPP, AQPPP, etc.	=> <b>S AQP+/SQSP</b>
		AQP, AQPAQP, AQPAQPAQP, etc.	=> <b>S (AQP)+/SQSP</b>
	Specify alternate sequences	ACD or KLM	=> <b>S ACD KLM/SQSP</b>
&	Join together sequence queries	Sequence L1 joined to sequence L3	=> <b>S L1&amp;L3/SQSFP</b>

## Gaps

To specify a gap, use a period (.) for one residue, a colon (:) for zero or one residue, or a period (.) followed by an appropriate repeat expression.

Use this symbol...	To specify...	Retrieves	Example
.	A gap of one residue	SY followed by one residue followed by RPG	=> S SY.RPG/SQSP
.{m} or [m.]	A gap of m residues	SY followed by any two residues followed by RPG	=> S SY.{2}RPG/SQSP
.{m,u} or .{m-u}	A gap of m to u residues	GFF followed by a gap of 2-10 residues followed by LSS	=> S GFF.{2,10}LSS/SQSP
.? or : or .{0,1} or .{0-1}	A gap of zero or one residue	AGA followed by zero or one residue followed by SRI	=> S AGA.?SRI/SQSFP
.* or .{0,} or .{0-}	A gap of zero or more residues	HLC followed by a gap of zero or more residues followed by TYG	=> S HLC.*TYG/SQSP
.+ or .{1,} or .{1-}	A gap of one or more residues	SY followed by any number of residues followed by TH	=> S SY.+TH/SQSP

## Order of execution of symbols

More than one symbol may be used to create complex sequence queries. If you do not use parentheses in sequence queries, the operations are performed in the following order:

1. Repeat symbols ? or \* or +
2. Repeat expressions using curly braces, e.g., {3,6}
3. Concatenation symbol &
4. The vertical bar |

**Find atriopeptin analogs containing RSSCF and QSGLG, separated by a gap of zero or any number of amino acids.**

**1** Enter REGISTRY.

**2** Search the sequence pattern in the /SQSP field. The symbol . \* indicates a gap of any number of amino acids, including zero.

**3** Use the KWIC format to display the hit subsequence in context.

```
=> FILE REGISTRY
=> S RSSCF.*QSGLG/SQSP
L1          553 RSSCF.*QSGLG/SQSP
=> D KWIC 1-3

L1  ANSWER 1 OF 553  REGISTRY  COPYRIGHT 2008 ACS on STN
SEQ  101 PWDSSDRSAL LKSKLRALLT AXRSLRRSSC FGGRMDRIGA QSGLGCNSFR
      =====
HITS AT:  127-145

L1  ANSWER 2 OF 553  REGISTRY  COPYRIGHT 2008 ACS on STN
SEQ  101 PWDSSDRSAL LKSKLRALLT APRSLRRSSC FGGRMDRIGA QSGLGCNSFR
      =====
HITS AT:  127-145

**RELATED SEQUENCES AVAILABLE WITH SEQLINK**

L1  ANSWER 3 OF 553  REGISTRY  COPYRIGHT 2008 ACS on STN
SEQ   1 MKWVSFISLL FLFSSAYSRS LDKRSLRRSS CFGGRMDRIG AQSGLCNSF
      === =====
HITS AT:  28-46
```

## Using SEQLINK

The SEQLINK EXACT command is used to locate additional protein or nucleic acid sequences that match a sequence that has already been retrieved from REGISTRY. SEQLINK is especially useful after searching with a name, name segments, or CAS Registry Numbers.

1 Enter REGISTRY and conduct a search.

2 Enter SEQLINK EXACT L1. L2 contains the CAS Registry Number from L1, plus 55 additional CAS Registry Numbers that have the same sequence as the sequence in L1.

```
=> FILE REGISTRY
=> S 487486-61-3
L1          1 487486-61-3

=> D SQD

L1  ANSWER 1 OF 1  REGISTRY  COPYRIGHT 2008 ACS on STN
RN   487486-61-3  REGISTRY
FS   PROTEIN SEQUENCE
SQL  28

SEQ          1 SLRRSSCFGG RMDRIGAQSG LGCNSFRY

**RELATED SEQUENCES AVAILABLE WITH SEQLINK**

=> SEQLINK EXACT L1
L2          56 SEQLINK EXACT L1

=> D SQD 1

L2  ANSWER 1 OF 56  REGISTRY  COPYRIGHT 2008 ACS on STN
RN   653601-38-8  REGISTRY
FS   PROTEIN SEQUENCE; STEREOSEARCH
SQL  28

NTE  modified (modifications unspecified)
-----
type          ----- location -----      description
-----
bridge        Cys-7          - Cys-23      disulfide
                                   bridge
-----

SEQ          1 SLRRSSCFGG RMDRIGAQSG LGCNSFRY

**RELATED SEQUENCES AVAILABLE WITH SEQLINK**
```

## Searching length

You can refine a sequence search by combining it with a search of sequence length in the /SQL field. You can use the following operators to search sequence lengths.

Use this operator...	To indicate...	Example
>	Greater than	=> S SQL>100
<	Less than	=> S SQL<25
=	Equal to	=> S SQL=15 or 15/SQL
<=	Less than or equal to	=> S SQL<=100
>=	Greater than or equal to	=> S SQL=>120
m-n	Range beginning with m and ending with n	=> S 35-100/SQL

### Find RGDF containing peptides with 10 or fewer amino acids.

1 Enter *REGISTRY* and search the sequence.

2 Search *SQL<=10* to retrieve only sequences with 10 or fewer residues.

```
=> FILE REGISTRY
=> S RGDF/SQSP
L1          12089 RGDF/SQSP
=> S L1 AND SQL=<10
L2          1191 L1 AND SQL=<10
=> D HIT 1-2

L2 ANSWER 1 OF 1191 REGISTRY COPYRIGHT 2008 ACS on STN
SQL 5

SEQ          1 RGDFK
           =====
HITS AT:     1-4

**RELATED SEQUENCES AVAILABLE WITH SEQLINK**

L2 ANSWER 2 OF 1191 REGISTRY COPYRIGHT 2008 ACS on STN
SQL 10

SEQ          1 RGDFEGGGKK
           =====
HITS AT:     1-4

**RELATED SEQUENCES AVAILABLE WITH SEQLINK**
```

## Searching chemical annotation

In the Annotation (/NTE) field, you can search the following types of information for chemically modified sequences:

- Terms for broad classification of the entire protein sequence, e.g., multichain, linear, or cyclic
- Terms for the type of chemical modification, e.g., uncommon amino acid or bridge
- Location of the amino acid where the chemical modification has occurred
- Terms describing the chemical modification, e.g., the name of a blocking group, metal complex, or a bridge. Refer to the Appendix for a list of blocking groups.

In the /NTE field, you can search phrases or single words and combine them by using the Boolean operators (AND, OR, NOT). When you enter terms with punctuation, the phrase is searched. When you enter terms separated by spaces, terms are searched in the same modification, in any order, and any number of words apart. You can use both right and left truncation. A term with left truncation must contain at least four characters, e.g., => **S ?CHLOR?/NTE**.

### Find multichain sequences.

1 Enter **REGISTRY**.

2 Search **MULTICHAIN** in the /NTE field.

```
=> FILE REGISTRY
=> S MULTICHAIN/NTE
L1      21355 MULTICHAIN/NTE
=> D KWIC
L1      ANSWER 1 OF 21355 REGISTRY COPYRIGHT 2008 ACS on STN
NTE     multichain
        modified (modifications unspecified)
```

### Find sequences with the blocking group ethoxycarbonyl, also known as EOC.

1 Search **EOC** in the /NTE field.

```
=> S EOC/NTE
L2      203 EOC/NTE
=> D KWIC
L2      ANSWER 1 OF 203 REGISTRY COPYRIGHT 2008 ACS on STN
NTE     modified (modifications unspecified)
-----
type           ----- location ----- description
-----
modification   Val-1           -           ethoxycarbonyl
                                     <Eoc>
-----
```

## Searching for references

To find references to protein sequences obtained in REGISTRY, use the resulting L-number as a search term in STN databases containing CAS Registry Numbers, e.g., CAPLUS, USPATFULL.

### Find patents on peptides containing RGDF.

1 Search the subsequence in the /SQSP field in REGISTRY.

2 Enter CAPLUS.

3 Combine L1 with PATENT/DT to retrieve patents citing the subsequences.

4 To display bibliographic information and the abstract, enter BIB AB as the display format.

```
=> S RGDF/SQSP
L1      12089  RGDF/SQSP

=> FILE CAPLUS

=> S L1 AND PATENT/DT
L2      1741  L1 AND PATENT/DT

=> D 1 BIB AB

L2      ANSWER 1 OF 1741  CAPLUS  COPYRIGHT 2008 ACS on STN
AN      2008:473431  CAPLUS  Full-text
TI      oncolytic viruses and antiangiogenic agents in the
        treatment of cancer
IN      Karrasch, Matthias; Mescheder, Axel
PA      Medigene AG, Germany
SO      PCT Int. Appl., 69pp.
        CODEN: PIXXD2
DT      Patent
LA      English
FAN.CNT 1
        PATENT NO.      KIND  DATE      APPLICATION NO.  DATE
        -----
PI      WO 2008043576  A1    20080417  WO 2007-EP8930  20071015
        W:  AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH,
                .
                .
                .
PRAI   US 2006-851598P P      20061013
AB      The invention relates to a combination of at least one
        oncolytic virus and at least one antiangiogenic agent
        and to the use of this combination in tumor therapy.
        Intraarterial infusions of oncolytic virus NV1020 to
        a patient with progressive metastatic colorectal
        adenocarcinoma followed by CPT-11 plus cetuximab resulted
        in stabilization of the disease at 6 mo post treatment.
RE.CNT  5      THERE ARE 5 CITED REFERENCES AVAILABLE FOR THIS
        RECORD
        ALL CITATIONS AVAILABLE IN THE RE FORMAT
```

## Find U.S. patents on the peptides containing the residue RGDF.

1 Enter REGISTRY and search the sequence residue in the /SQSP field.

2 Search the REGISTRY sequence search L-number (L1) in USPATFULL to find U.S. patents citing the retrieved sequences.

3 To display the bibliographic information and the abstract, enter BIB AB as the format.

```
=> FILE REGISTRY
=> S RGDF/SQSP
L1          12089 RGDF/SQSP

=> FILE USPATFULL
=> S L1
L2          1710 L1

=> D BIB AB 5
L2 ANSWER 5 OF 1710 USPATFULL on STN
AN 2008:86954 USPATFULL Full-text
TI Nucleic acid and amino acid sequences relating to
   Staphylococcus epidermidis for diagnostics and
   therapeutics
IN Doucette-Stamm, Lynn A., Framingham, MA, UNITED STATES
   Bush, David, Somerville, MA, UNITED STATES
PA Wyeth, Madison, NJ, UNITED STATES (U.S. corporation)
PI US 2008076153 A1 20080327
AI US 2007-882810 A1 20070806 (11)
RLI Division of Ser. Number US 2004-902441, filed on 30
     Jul 2004, PENDING
     Continuation of Ser. Number US 2002-92411, filed on
     7 Mar 2002, ABANDONED
     Division of Ser. Number US 1998-134001, filed on 13
     Aug 1998, GRANTED, Pat.
     Number US 6380370
PRAI US 1997-64964P 19971108 (60)
     US 1997-55779P 19970814 (60)
DT Utility
FS APPLICATION
LREP DRINKER BIDDLE & REATH (DC), 1500 K STREET, N.W.,
     SUITE 1100, WASHINGTON, DC, 20005-1209, US
CLMN Number of Claims: 13
ECL Exemplary Claim: 1
DRWN No Drawings
LN.CNT 2997
CAS INDEXING IS AVAILABLE FOR THIS PATENT.
AB The invention provides isolated polypeptide and nucleic
   acid sequences derived from Staphylococcus epidermidis
   that are useful in diagnosis and therapy of pathological
   conditions; antibodies against the polypeptides; and
   methods for the production of the polypeptides. The
   invention also provides methods for the detection,
   prevention and treatment of pathological conditions
   resulting from bacterial infection.
```

## Appendix: Blocking groups for protein sequence annotation

### BLOCKING GROUPS WITH SHORTCUTS (Alphabetical by Shortcuts)

Shortcut	Blocking Group
Aac	2-amino-2-oxoethyl
2Abz	2-aminobenzoyl
4Abz	4-aminobenzoyl
Ac	acetyl
AcO	acetyloxy
Acm	(acetylamino)methyl
Acr	3-(9-acridinyl)
Adc	tricyclo[3.3.1.1(3,7)]dec-1-yloxy
Aet	2-aminoethyl
All	propenyl
Amoc	(9-anthracenylmethoxy) carbonyl
Aoc	(1,1-dimethylpropoxy) carbonyl
Azoc	[1-methyl-1-[4-(phenylazo)phenyl]ethoxy] carbonyl
Bac	bromoacetyl
Bam	(benzoylamino)methyl
Beoc	(2-bromoethoxy) carbonyl
Bhoc	(diphenylmethoxy) carbonyl
Bic	(5-benzisoxazolylmethoxy) carbonyl
Bmv	1-methyl-3-oxo-3-phenyl-1-propenyl
Bnps	(3-bromo-2-nitrophenyl) thio
BOC	(1,1-dimethylethoxy) carbonyl
Bocae	[(1,1-dimethylethoxy) carbonyl] amino] ethyl
lBop	2-(phenylmethoxy) phenoxy
Bpoc	(1-[1,1'-biphenyl]-4-yl-1-methylethoxy) carbonyl
Br	bromo
Bs	(4-bromophenyl) sulfonyl
Bt	1H-benzotriazol-1-yl
BTC	[(phenylmethyl) thio] carbonyl
Btm	[(phenylmethyl) thio]methyl
i-Bu	2-methylpropyl
t-Bu	1,1-dimethylethyl
Bum	[(2-methyl-1-oxopropyl) amino]methyl
i-BuO	2-methyl-1-oxopropyl
Bz	benzoyl
2BZ	[(2-bromophenyl) methoxy] carbonyl
4BZ	[(4-bromophenyl) methoxy] carbonyl
Bza	1H-benzimidazol-2-yl
Bzh	diphenylmethyl
Bzl	phenylmethyl
Cac	carboxyacetyl
Cbm	aminocarbonyl
Cbs	(4-chlorophenyl) sulfonyl
CBz	(phenylmethoxy) carbonyl
Cdf	chlorodifluoroacetyl
Ceoc	(2-chloroethoxy) carbonyl

## BLOCKING GROUPS WITH SHORTCUTS

Shortcut	Blocking Group
CF3	trifluoromethyl
Chb	(5-chloro-2-hydroxyphenyl)phenylmethylene
Chc	cyclohexylcarbonyl
Chp	cycloheptyl
Chx	cyclohexyl
Chxa	cyclohexylacetyl
Cl	chloro
2-6Clb	(2,6-dichlorophenyl)methyl
Cm	carboxymethyl
Cpc	cyclopentylcarbonyl
Cpe	cyclopentyl
Cpm	cyclopropylmethyl
2CZ	[(2-chlorophenyl)methoxy]carbonyl
4CZ	[(4-chlorophenyl)methoxy]carbonyl
Dbpoc	(2,2-dibromopropoxy)carbonyl
2-4DCZ	[(2,4-dichlorophenyl)methoxy]carbonyl
2-6DCZ	[(2,6-dichlorophenyl)methoxy]carbonyl
Ddz	[1-(3,5-dimethoxyphenyl)-1-methylethoxy]carbonyl
De	2-(diethylamino)ethyl
Dec	1-oxodecyl
Dip	[2-methyl-1-(1-methylethyl)propoxy]carbonyl
Dmoc	[(dimethylamino)oxy]carbonyl
DMB	(3,4-dimethylphenyl)methyl
Dmt	bis(4-methoxyphenyl)methyl
DNP	2,4-dinitrophenyl
DNPS	(2,4-dinitrophenyl)thio
Dpp	diphenoxyphosphinyl
Eac	(ethylamino)carbonyl
Eoc	ethoxycarbonyl
Et	ethyl
F	fluoro
For	formyl
Ft	(1,3-dihydro-1,3-dioxo-2H-isoindol-2-yl)methyl
Glt	4-carboxy-1-oxobutyl
Hex	1-oxohexyl
I	iodo
Ioc	(2-methylpropoxy)carbonyl
Ipa	7-methyl-1-oxooctyl
Ips	(4-iodophenyl)sulfonyl
Kpc	(6-oxo-2-piperidinyl)carbonyl
MOB	(4-methoxyphenyl)methyl
MOS	(4-methoxyphenyl)sulfonyl
Mac	4-methyl-7-amino-coumaryl
Mal	3-carboxy-1-oxo-2-propenyl
Mbh	bis(4-methoxyphenyl)methyl
Me	methyl
MeOe	2-methoxy-2-oxoethyl
Mhoc	[(1-methylcyclohexyl)oxy]carbonyl

## BLOCKING GROUPS WITH SHORTCUTS

Shortcut	Blocking Group
Mmt	(4-methoxyphenyl)diphenylmethyl
Moz	[(4-methoxyphenyl)methoxy]carbonyl
Mpt	dimethylphosphinothioyl
Ms	methylsulfonyl
Msc	[2-(methylsulfonyl)ethoxy]carbonyl
Msi	methylsulfinyl
Msp	4-(methylsulfonyl)phenyl
Mtos	(2,4,6-trimethoxyphenyl)sulfonyl
Mtp	4-(methylthio)phenyl
Mts	(2,4,6-trimethylphenyl)sulfonyl
Mz	[[4-[(4-methoxyphenyl)azo]phenyl]methoxy]carbonyl
N	nitro
N3	azido
Nabs	[4-[(4-hydroxy-1-naphthalenyl)azo]phenyl]sulfonyl
1-Naph	1-naphthalenyl
2-Naph	2-naphthalenyl
Ng	2-methoxy-4-nitrophenyl
Ngu	[imino(nitroamino)methyl]amino
NH2	amino
Nis	(4-nitrophenyl)sulfonyl
Nm	3-nitrophenyl
No	2-nitrophenyl
Np	4-nitrophenyl
Npe	2-nitro-1-phenylethyl
Nps	(2-nitrophenyl)thio
Ns	2-nitro-4-sulfophenyl
O	oxygen
Oct	1-oxooctyl
2OHEt	2-hydroxyethyl
2OHPh	2-hydroxyphenyl
Ole	1-oxo-9-octadecenyl
Pa	1-oxononyl
Pal	1-oxohexadecyl
Pbp	pentabromophenyl
Pcp	pentachlorophenyl
Pfp	pentafluorophenyl
Ph	phenyl
Pht	2-carboxybenzoyl
Pic	4-pyridinylmethyl
2Pip	2-piperidinyl
Pipoc	(1-piperidinyl)oxy carbonyl

## BLOCKING GROUPS WITH SHORTCUTS

Shortcut	Blocking Group
Pnb	(4-nitrophenyl)methyl
PO2	phosphono
Poc	(cyclopentyloxy) carbonyl
Ppt	diphenylphosphinothioyl
Pr	propyl
i-Pr	1-methylethyl
Ptc	(phenylamino)thioxomethyl
Py	2-pyridinyl
3Py	3-pyridinyl
4Py	4-pyridinyl
Pz	[[4-(phenylazo)phenyl]methoxy] carbonyl
Q	quinolinyl
QC	5-chloro-8-quinolinyl
Qu	8-quinolinyl
Qxc	2-quinoxalinylcarbonyl
Sbz	2-sulfobenzoyl
Scm	(carboxymethyl)thio
SO3H	sulfo
Su	2,5-dioxo-1-pyrrolidinyl
Suc	3-carboxy-1-oxopropyl
Tac	[[4-methylphenyl]sulfonyl]amino] carbonyl
Tbs	(1,1-dimethylethyl)dimethylsilyl
TBZ	phenylthioxomethyl
Tcboc	(2,2,2-trichloro-1,1-dimethylethoxy) carbonyl
Tce	2,2,2-trichloroethyl
Tcp	2,4,5-trichlorophenyl
Tec	[2-[(4-methylphenyl)sulfonyl]ethoxy] carbonyl
Teoc	(2,2,2-trichloroethoxy) carbonyl
Tfe	2,2,2-trifluoroethyl
Tfp	2,2,3,3-tetrafluoro-1-oxopropyl
Tmb	(2,4,6-trimethylphenyl)methyl
TNP	2,4,6-trinitrophenyl
Tos	(4-methylphenyl)sulfonyl
Tosa	[(4-methylphenyl)sulfonyl]amino
Trit	triphenylmethyl
Trs	(triphenylmethyl)thio
5Urd	5'-uridylyl
Vi	ethenyl
Xan	9H-xanthen-9-yl
Za	[(phenylmethoxy) carbonyl]amino
Zae	[[phenylmethoxy] carbonyl]amino] ethyl
ZNO2	[(4-nitrophenyl) methoxy] carbonyl
Zoa	[[phenylmethoxy] carbonyl]oxy] acetyl

## BLOCKING GROUPS WITH SHORTCUTS (Alphabetical by Group Name)

Blocking Group Name	Shortcut
acetyl	Ac
(acetylamino)methyl	Acm
acetyloxy	AcO
3- (9-acridinyl)	Acr
amino	NH2
2-aminobenzoyl	2Abz
aminocarbonyl	Cbm
2-aminoethyl	Aet
2-amino-2-oxoethyl	Aac
(9-anthracenylmethoxy) carbonyl	Amoc
azido	N3
1H-benzimidazol-2-yl	Bza
1H-benzotriazol-1-yl	Bt
(5-benzisoxazolylmethoxy) carbonyl	Bic
benzoyl	Bz
(benzoylamino)methyl	Bam
(1- [1,1'-biphenyl]-4-yl-1-methylethoxy) carbonyl	Bpoc
bis(4-methoxyphenyl)methyl	Dmt
bromo	Br
bromoacetyl	Bac
(2-bromoethoxy) carbonyl	Beoc
(3-bromo-2-nitrophenyl) thio	Bnps
[(2-bromophenyl)methoxy] carbonyl	2BZ
(4-bromophenyl) sulfonyl	Bs
carboxyacetyl	Cac
2-carboxybenzoyl	Pht
carboxymethyl	Cm
(carboxymethyl) thio	Scm
4-carboxy-1-oxobutyl	Glt
3-carboxy-1-oxo-2-propenyl	Mal
3-carboxy-1-oxopropyl	Suc
chloro	Cl
chlorodifluoroacetyl	Cdf
(2-chloroethoxy) carbonyl	Ceoc
(5-chloro-2-hydroxyphenyl) phenylmethylene	Chb
[(2-chlorophenyl)methoxy] carbonyl	2CZ
(4-chlorophenyl) sulfonyl	Cbs
5-chloro-8-quinolinyl	QC
cycloheptyl	Chp
cyclohexyl	Chx
cyclohexylacetyl	Chxa
cyclohexylcarbonyl	Chc

## BLOCKING GROUPS WITH SHORTCUTS

Blocking Group Name	Shortcut
cyclopentyl	Cpe
cyclopentylcarbonyl	Cpc
(cyclopentyloxy) carbonyl	Poc
cyclopropylmethyl	Cpm
(2,2-dibromopropoxy) carbonyl	Dbpoc
[(2,4-dichlorophenyl) methoxy] carbonyl	2-4DCZ
2-(diethylamino) ethyl	De
(1,3-dihydro-1,3-dioxo-2H-isoindol-2-yl) methyl	Ft
[1-(3,5-dimethoxyphenyl)-1-methylethoxy] carbonyl	Ddz
[(dimethylamino) oxy] carbonyl	Dmoc
[[ (1,1-dimethylethoxy) carbonyl] amino] ethyl	Bocae
(1,1-dimethylethoxy) carbonyl	BOC
1,1-dimethylethyl	t-Bu
(1,1-dimethylethyl) dimethylsilyl	Tbs
(3,4-dimethylphenyl) methyl	DMB
dimethylphosphinothioyl	Mpt
(1,1-dimethylpropoxy) carbonyl	Aoc
2,4-dinitrophenyl	DNP
(2,4-dinitrophenyl) thio	DNPS
2,5-dioxo-1-pyrrolidinyl	Su
diphenoxyphosphinyl	Dpp
(diphenylmethoxy) carbonyl	Bhoc
diphenylmethyl	Bzh
diphenylphosphinothioyl	Ppt
ethenyl	Vi
ethoxycarbonyl	Eoc
ethyl	Et
(ethylamino) carbonyl	Eac
fluoro	F
formyl	For
9H-xanthen-9-yl	Xan
2-hydroxyethyl	2OHEt
[4-[(4-hydroxy-1-naphthalenyl) azo]phenyl] sulfonyl	Nabs
2-hydroxyphenyl	2OHPh
[imino(nitroamino) methyl] amino	Ngu
iodo	I
(4-iodophenyl) sulfonyl	Ips
2-methoxy-4-nitrophenyl	Ng
2-methoxy-2-oxoethyl	MeOe

## BLOCKING GROUPS WITH SHORTCUTS

Blocking Group Name	Shortcut
[ [4- [(4-methoxyphenyl) azo] phenyl] methoxy] carbonyl	Mz
(4-methoxyphenyl) diphenylmethyl	Mmt
[ (4-methoxyphenyl) methoxy] carbonyl	Moz
(4-methoxyphenyl) methyl	MOB
(4-methoxyphenyl) sulfonyl	OS
methyl	Me
4-methyl-7-amino-coumaryl	Mac
[ (1-methylcyclohexyl) oxy] carbonyl	Mhoc
1-methylethyl	i-Pr
[2-methyl-1- (1-methylethyl) propoxy] carbonyl	Dip
7-methyl-1-oxooctyl	Ipa
1-methyl-3-oxo-3-phenyl-1-propenyl	mv
2-methyl-1-oxopropyl	i-BuO
[ (2-methyl-1-oxopropyl) amino] methyl	Bum
[1-methyl-1- [4- (phenylazo) phenyl] ethoxy] carbonyl	Azoc
[ (4-methylphenyl) sulfonyl] amino	Tosa
[ [ (4-methylphenyl) sulfonyl] amino] carbonyl	Tac
(4-methylphenyl) sulfonyl	Tos
[2- [ (4-methylphenyl) sulfonyl] ethoxy] carbonyl	Tec
(2-methylpropoxy) carbonyl	Ioc
(2-methylpropoxy) methyl	iBom
2-methylpropyl	i-Bu
methylsulfinyl	Msi
methylsulfonyl	Ms
[2- (methylsulfonyl) ethoxy] carbonyl	Msc
4- (methylsulfonyl) phenyl	Msp
4- (methylthio) phenyl	Mtp
1-naphthalenyl	1-Naph
nitro	N
2-nitro-1-phenylethyl	Npe
[ (4-nitrophenyl) methoxy] carbonyl	ZNO2
(4-nitrophenyl) methyl	Pnb
2-nitrophenyl	No
3-nitrophenyl	Nm
4-nitrophenyl	Np
(4-nitrophenyl) sulfonyl	Nis
(2-nitrophenyl) thio	Nps
2-nitro-4-sulfophenyl	Ns
1-oxodecyl	Dec
1-oxohexadecyl	Pal
1-oxohexyl	Hex
1-oxononyl	Pa
1-oxo-9-octadecenyl	Ole

## BLOCKING GROUPS WITH SHORTCUTS

Blocking Group Name	Shortcut
1-oxooctyl	Oct
(6-oxo-2-piperidinyl) carbonyl	Kpc
oxygen	O
pentabromophenyl	Pbp
pentachlorophenyl	Pcp
pentafluorophenyl	Pfp
phenyl	Ph
(phenylamino) thioxomethyl	Ptc
[[4-(phenylazo)phenyl]methoxy] carbonyl	Pz
[(phenylmethoxy) carbonyl] amino	Za
[[ (phenylmethoxy) carbonyl] amino] ethyl	Zae
(phenylmethoxy) carbonyl	CBz
[[ (phenylmethoxy) carbonyl] oxy] acetyl	Zoa
2-(phenylmethoxy) phenoxy	Bop
phenylmethyl	Bzl
[(phenylmethyl) thio] carbonyl	BTC
[(phenylmethyl) thio] methyl	Btm
phenylthioxomethyl	TBZ
phosphono	PO2
2-piperidinyl	2Pip
(1-piperidinyl)oxy carbonyl	Pipoc
propenyl	All
propyl	Pr
2-pyridinyl	Py
4-pyridinylmethyl	Pic
quinolinyl	Q
8-quinolinyl	Qu
2-quinoxalinylyl carbonyl	Qxc
sulfo	SO3H
2-sulfobenzoyl	Sbz
2,2,3,3-tetrafluoro-1-oxopropyl	Tfp
(2,2,2-trichloro-1,1-dimethylethoxy) carbonyl	Tcboc
(2,2,2-trichloroethoxy) carbonyl	Teoc
2,2,2-trichloroethyl	Tce
2,4,5-trichlorophenyl	Tcp
tricyclo[3.3.1.1(3,7)]dec-1-yloxy	Adc
trifluoroacetyl	Tfa
2,2,2-trifluoroethyl	Tfe
trifluoromethyl	CF3

## BLOCKING GROUPS WITHOUT SHORTCUTS

acetylamino  
2-amino-3-(4-hydroxyphenyl)propyl  
aminoiminomethyl  
3-[(4-azidophenyl)azo]  
benzenemethoxy  
4-benzoylbenzoyl  
bis(2-chloroethyl)amino  
bis(ethylamino)methylene  
3-bromo-1-oxo-propyl  
(butylamino)carbonyl  
(butylamino)iminoethyl  
1-(carboxymethyl)cyclohexyl  
(3-carboxy-4-nitrophenyl)thio  
3-carboxypropyl  
[(chloroacetyl)amino]methyl  
1-[(cyclohexylamino)carbonyl]-2-methylpropyl  
(cyclohexyloxy)carbonyl  
1,3-dihydro-1,3-dioxo-2H-H-isoindol-2-yl  
10,11-dihydro-5H-dibenzo[a,d]cyclohepten-5-yl  
[(4,5-dimethoxy-2-nitrophenyl)methoxy]carbonyl  
2-(dimethylamino)ethyl  
[5-(dimethylamino)-1-naphthalenyl]sulfonyl  
[[[(1,1-dimethylethoxy)carbonyl]amino]oxy]acetyl  
[[[(1,1-dimethylethoxy)carbonyl]amino]oxy]-1-oxopropyl  
(1,1-dimethylethyl)thio  
2,2-dimethyl-1-oxopropyl  
(1,6-dioxoheptyl)amino  
3-ethoxy-2-(ethoxycarbonyl)-3-oxopropyl  
(ethylamino)iminoethyl  
(ethylthio)thioxomethyl  
(9H-fluoren-9-ylmethoxy)carbonyl  
2,2,3,3,4,4,4-heptafluoro-1-oxobutyl  
(3a,4,5,6,7,7a-hexahydro-3a,7a-dihydroxy-1H-benzimidazol-2-yl)  
1,3,3a,4,7,7a-hexahydro-1,3-dioxo-4,7-methano-2H-isoindol-2-yl  
3-hydroxy-8-methyl-1-oxononyl  
3-hydroxy-1-oxodecyl  
3-hydroxy-1-oxotetradecyl  
3-hydroxy-1-oxotridecyl  
3-(4-hydroxyphenyl)-1-oxopropyl  
1-iminoethyl  
imino(nitroamino)methyl  
imino[(2,2,2-trifluoroethyl)amino]methyl  
(1-mercaptocyclohexyl)acetyl  
(4-methoxy-2,6-dimethylphenyl)sulfonyl  
4-methoxy-1,4-dioxobutyl  
4-methoxy-4-oxobutyl  
[(4-methoxyphenyl)methyl]sulfinyl  
(4-methoxy-2,3,6-trimethylphenyl)sulfonyl  
[(1-methylethyl)amino]carbonyl  
3-methyl-1-oxobutyl  
2-methyl-1-oxo-2-propenyl  
(4-methylphenyl)methyl  
[(4-methylphenyl)methyl]sulfonyl  
4-(methylsulfonyl)-1-oxobutyl  
4-(methylthio)-1-oxobutyl  
3-(methylthio)propyl  
4-(4-nitrophenoxy)-4-oxobutyl  
(3-nitro-2-pyridinyl)thio  
2-oxo-2-phenylethyl  
1-oxo-3-[(phenylmethyl)thio]propyl  
1-oxo-3-phenyl-2-propenyl  
1-oxo-3-phenylpropyl

1-oxo-3-[4-(sulfooxy)phenyl]propyl  
(pentamethylphenyl)sulfonyl  
[(phenylacetyl)amino]methyl  
4-(phenylazo)benzoyl  
4-(phenylazo)phenyl  
(phenylmethoxy)methyl  
3-phenyl-2-oxaziridinyl  
1-pyrenyl  
2-pyridinylcarbonyl  
(4-pyridinylmethoxy)carbonyl  
(4-pyridinyloxy)carbonyl  
tetrahydro-2H-pyran-2-yl  
1,4,5,6-tetrahydro-2-(nitroamino)-4-pyrimidinyl



A division of the  
American Chemical Society

November 2008  
CAS2537-1108

CAS Customer Care  
Phone: 800-753-4227 (North America)  
614-447-3700 (worldwide)  
Fax: 614-447-3751  
E-mail: [help@cas.org](mailto:help@cas.org)  
Internet: [www.cas.org](http://www.cas.org)