

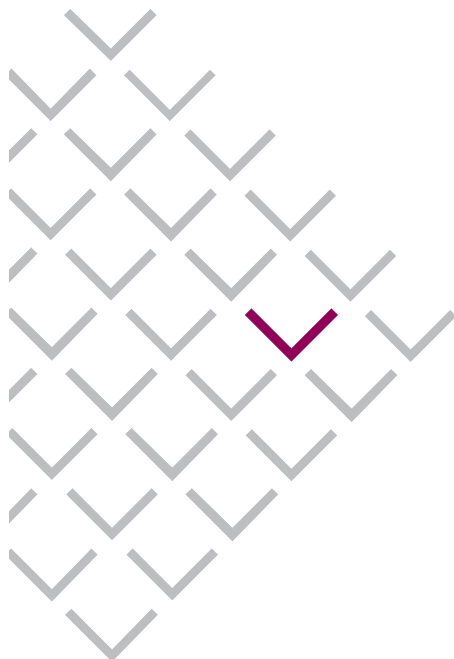


Chemical Abstracts Service
provides access to STN International
in North America

STNNotes

AUGUST 2002

No. 31



In response to customer requests for more detailed information on new and enhanced system features, we have created STNNotes. STNNotes give you the in-depth technical details you need for efficient use of STN. We hope you find this information useful. Please let us know how we can continue to improve in meeting your technical information needs.

Enhanced Sequence Searching in REGISTRY with SEQLINK

With SEQLINK EXACT you can find additional protein or nucleic acid sequences that have the same backbone sequence as a CAS Registry Number that you already have.

With the addition of a new command SEQLINK EXACT, sequence searching in REGISTRY or ZREGISTRY is more flexible and comprehensive.

CAS Sequence Registration and Indexing

REGISTRY contains CAS Registry Numbers for sequences indexed for CA and CPlus, plus the nucleotide sequences from GenBank. The CAS Registry Number[®] for each sequence is linked to the original source of registration or the reference from which it was indexed. For example, a separate CAS Registry Number is assigned to each GenBank accession number. Each sequence indexed by CAS from literature or patents is also identified by a separate CAS Registry Number, regardless of its sequence structure. In addition, chemically modified sequences receive their own CAS Registry Numbers even though their backbone sequence may be identical to another record in REGISTRY.



For comprehensive retrieval of CAplus references to a given sequence structure, CAS Registry Numbers for all records with the same sequence must be used.

With SEQLINK EXACT you can easily find additional protein or nucleic acid sequences that have the same sequence as a protein or nucleic acid sequence that you have already retrieved from REGISTRY/ZREGISTRY. Using SEQLINK EXACT, you may start with the CAS Registry Number for a sequence (or a set of sequences) of any length and easily retrieve all REGISTRY records for sequences with the same backbone structure(s) as those in the starting set.

Using SEQLINK EXACT

You may use SEQLINK EXACT with any of the following terms:

- CAS Registry Number for a single sequence
- E-number containing the CAS Registry Number for a single sequence
- REGISTRY L-number answer set
- ANALYZE L-number set containing CAS Registry Numbers

Use SEQLINK EXACT when you have obtained the CAS Registry Number for a sequence, e.g., via a GenBank name search or by extracting a CAS Registry Number from a CAplus answer set, and wish to locate all other REGISTRY records that have the same sequence.

Find references discussing sequences with the same structure as the sequence identified as GenBank AY089963.

```
=> S GenBank AY089963/CN
L1          1 GENBANK AY089963/CN

=> D L1 SQIDE

L1  ANSWER 1 OF 1  REGISTRY  COPYRIGHT 2002 ACS
RN  422624-05-3  REGISTRY
CN  GenBank AY089963 (9CI)  (CA INDEX NAME)
FS  NUCLEIC ACID SEQUENCE
SQL 1182
NA  276 a   229 c   261 g   416 t

SEQ      1 atggagtctc tgctctctag ttctctctt gtttccgctg...
        51 ttgttgaag aagcagaatc taaagctcca ctctttatca...
        :
        :
        1151 cagagtactt gctgttacct ttttgaagt ga
MF  Unspecified
CI  MAN
SR  GenBank
LC  STN Files:   CAPLUS, GENBANK
**RELATED SEQUENCES AVAILABLE WITH SEQLINK**
        1 REFERENCES IN FILE CAPLUS (1967 TO DATE)

=> SEQLINK EXACT L1
L2          7 SEQLINK EXACT L1

=> D 1-7

L2  ANSWER 1 OF 7  REGISTRY  COPYRIGHT 2002 ACS
RN  434253-22-2  REGISTRY
CN  GenBank AX428158 (9CI)  (CA INDEX NAME)
FS  NUCLEIC ACID SEQUENCE
MF  Unspecified
CI  MAN
SR  GenBank
LC  STN Files:   GENBANK
        :
        :
        :

L2  ANSWER 7 OF 7  REGISTRY  COPYRIGHT 2002 ACS
RN  303071-58-1  REGISTRY
CN  DNA (Arabidopsis gene ATPT2
    dimethylallyltransferase cDNA)(9CI) (CA INDEX NAME)
OTHER NAMES:
CN  1: PN: W00063391 SEQID: 1 claimed DNA
CN  1: PN: W00179472 SEQID: 1 claimed DNA
CN  DNA (Arabidopsis thaliana dimethylallyltransferase
    isoenzyme ATPT2 cDNA)
CN  DNA (Arabidopsis thaliana prenyltransferase
    isoenzyme ATPT2 cDNA)
FS  NUCLEIC ACID SEQUENCE
MF  Unspecified
CI  MAN
SR  CA
LC  STN Files:   CA, CAPLUS
        :
        :
        :
```

Locate the sequence of interest.

Note that the sequence is too large to run as an exact sequence search.

Use SEQLINK EXACT to locate all sequences with the same sequence as that shown for L1.

Some of the retrieved sequences have references in CAplus and some appear only in GenBank.

Unlike the exact sequence search options, /SQEP, /SQEFP and /SQEN, which accept only a single sequence structure, SEQLINK EXACT may be used on a set of CAS Registry Numbers representing different sequences. SEQLINK EXACT finds the additional sequence records, if any exist, for each sequence in an answer set. In addition, there are no restrictions on the length of the sequence that may be searched in SEQLINK EXACT.

```

=> FILE CAPLUS; S L2
L3          4 L2

=> S L1
L4          1 L1

=> D L3 HITRN 1-4

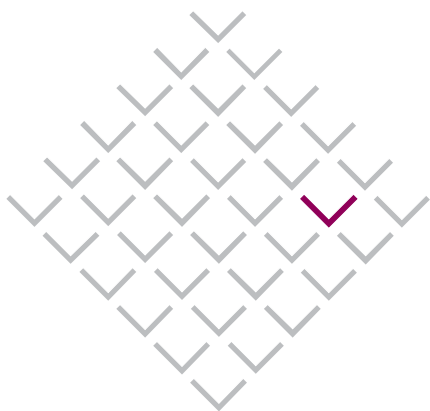
L3 ANSWER 1 OF 4 CAPLUS COPYRIGHT 2002 ACS
IT INDEXING IN PROGRESS
IT 422624-05-3, GenBank AY089963
RL: BSU (Biological study, unclassified); PRP
   (Properties); BIOL
   (Biological study)
   (nucleotide sequence; isolation and
   characterization of homogentisate phytyltransferase
   genes from Synechocystis sp. PCC 6803 and
   arabidopsis)

L3 ANSWER 2 OF 4 CAPLUS COPYRIGHT 2002 ACS
IT 413635-13-9
RL: AGR (Agricultural use); BSU (Biological study,
   unclassified); PRP
   (Properties); BIOL (Biological study); USES (Uses)
   (nucleotide sequence; cDNAs encoding
   prenyltransferase and tocopherol cyclase and their
   use in improving tocopherol synthesis in transgenic
   plants)

L3 ANSWER 3 OF 4 CAPLUS COPYRIGHT 2002 ACS
IT 303071-58-1
RL: BOC (Biological occurrence); BSU (Biological
   study, unclassified); BUU
   (Biological use, unclassified); PRP (Properties);
   BIOL (Biological study);
   OCCU (Occurrence); USES (Uses)
   (nucleotide sequence; nucleic acid sequences
   encoding plant and Synechocystis proteins involved
   in tocopherol synthesis)

L3 ANSWER 4 OF 4 CAPLUS COPYRIGHT 2002 ACS
IT 303071-58-1
RL: AGR (Agricultural use); BOC (Biological
   occurrence); BSU (Biological
   study, unclassified); PRP (Properties); BIOL
   (Biological study); OCCU
   (Occurrence); USES (Uses)
   (nucleotide sequence; protein and cDNA sequences
   of Arabidopsis dimethylallyltransferase and uses
   thereof on altering tocopherol synthesis in plants)
  
```

The SEQLINK EXACT answer set (L2) retrieves 3 additional references that would not have been retrieved if you searched only the CAS Registry Number associated with the starting GenBank name (L1).



Cost

There are no extra charges for using the SEQLINK EXACT command in REGISTRY or ZREGISTRY.

For more information:

Enter HELP SEQLINK at an arrow prompt for online information on using the SEQLINK EXACT command.