ENHANCING SEARCHING CAPABILITIES WITH INCHI STRINGS AND INCHI KEYS IN CAS STNEXT

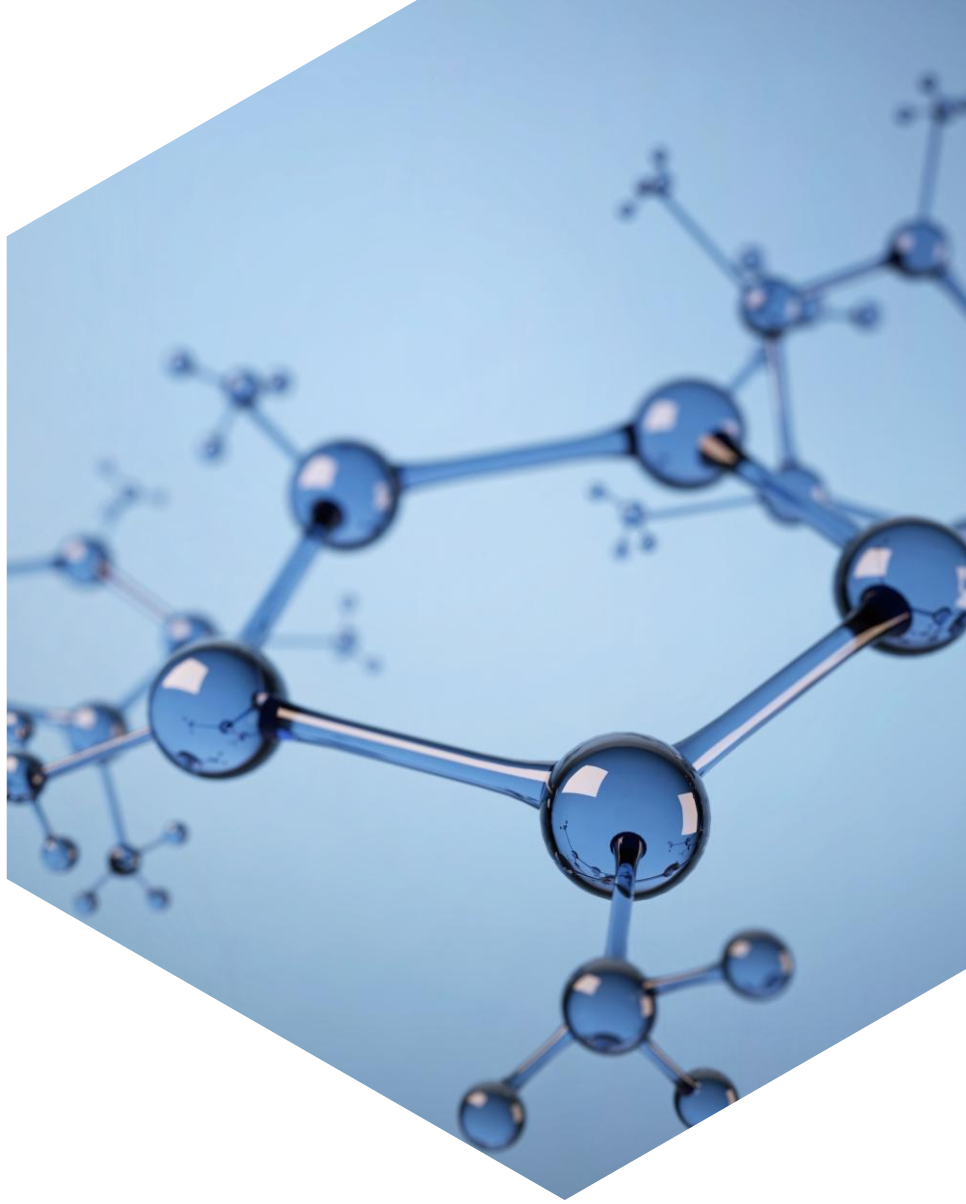
Jim Brown – FIZ Karlsruhe

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

AGENDA

- What is InChI?
- Implementation of InChI in STNext databases
- Examples
- Considerations



FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

WHAT IS INCHI?

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

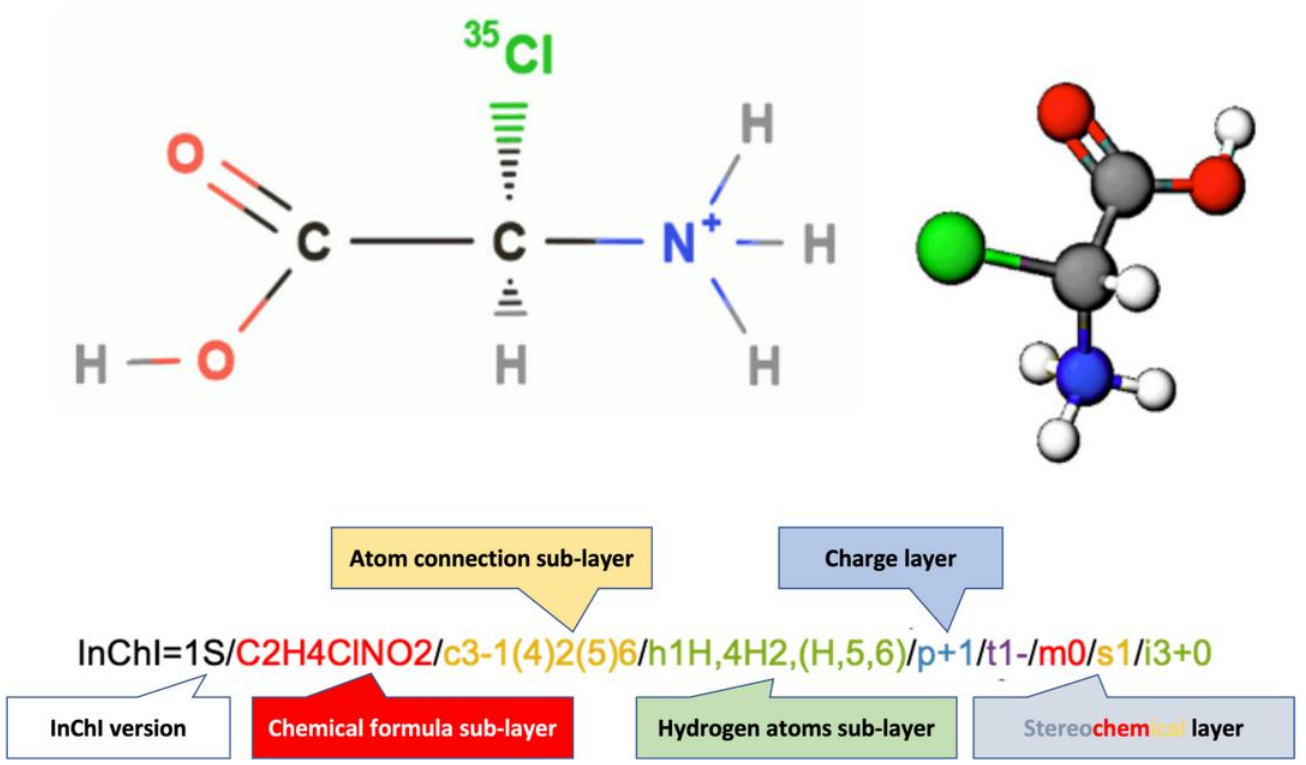
CAS
A division of the American Chemical Society

InChI – International Chemical Identifier

- Unique identifier for chemical structures developed to support the communication about specific chemical compounds.
- Developed by IUPAC and InChI Trust; released 2005
- Standardized, non-proprietary digital identifier generated from the structure by the freely available, open source InChI algorithm

InChI – layered structure

Standard InChI defines chemical structures by four levels of information - connectivity, isotopy, stereochemistry and electricity.



Source https://www.inchi-trust.org/

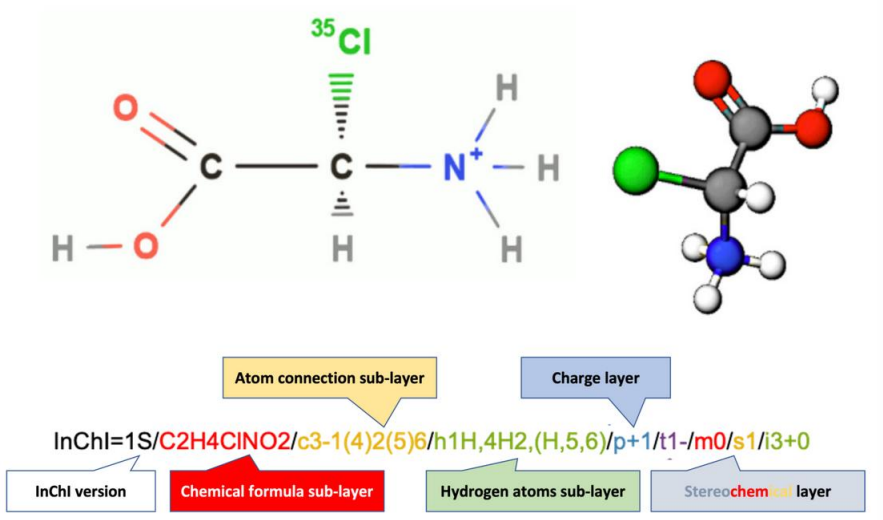
FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

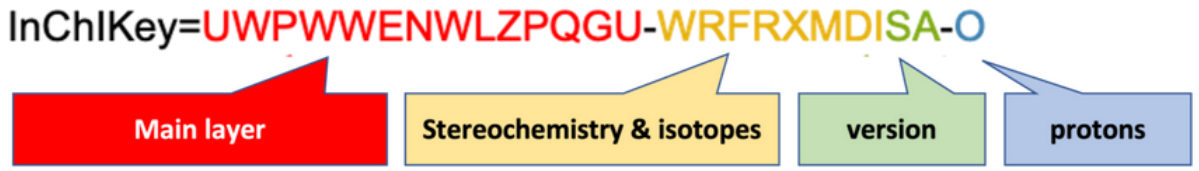
Standard InChIKey

Fixed-length, browser-searchable string derived from standard InChI.

Standard InChI



Standard InChIKey



Source https://www.inchi-trust.org/

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

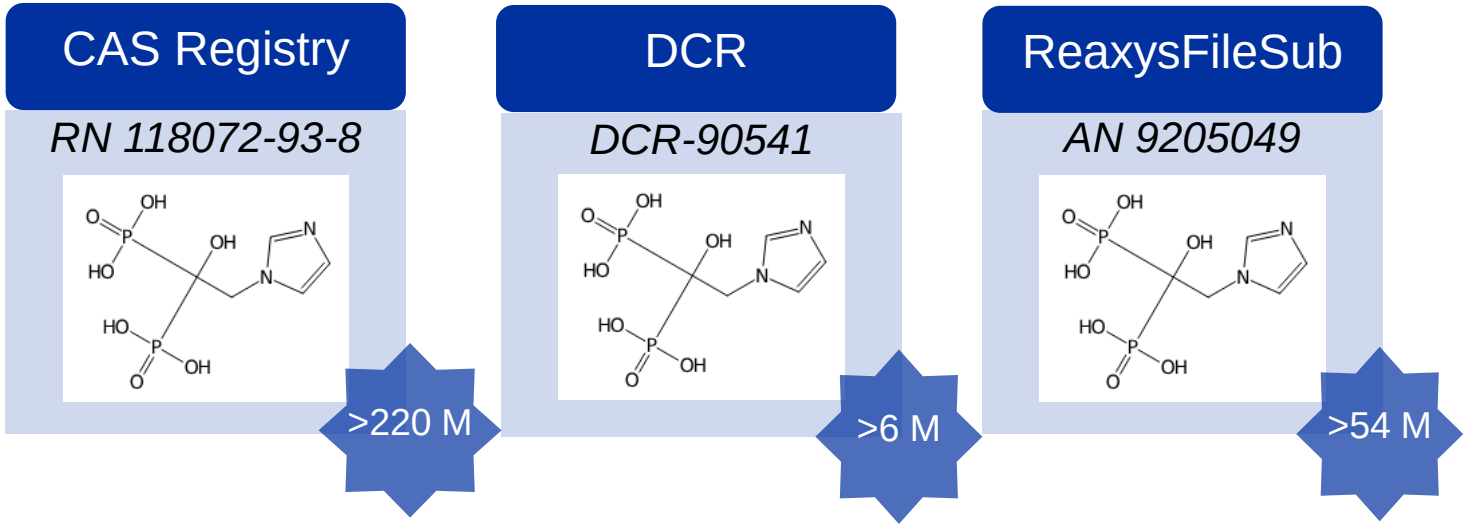
IMPLEMENTATION OF INCHI IN STNEXT DATABASES

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

STNext - InChI data in STNext databases

Major structure databases provide >280 million structure records with InChIs



INCH InChI=1S/C5H10N2O7P2/c8-5(15(9,10)11,16(12,13)14)3-7-2-1-6-4-7/h1-2,4,8H,3H2,(H2,9,10,11)(H2,12,13,14)

INKY XRASPMIURGNCCH-UHFFFAOYSA-N

INCH – InChI (IUPAC International Chemical Identifier); INKY - InChIKey

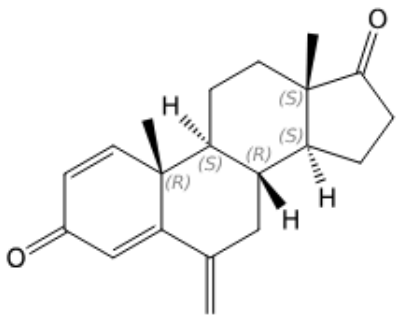
FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

CAS Registry and DCR with InChIs and InChIKeys

IUPAC International Chemical Identifiers widely used since 20 years

RN 107868-30-4 REGISTRY
ED Entered STN: 02 May 1987
CN Androsta-1,4-diene-3,17-dione, 6-methylene- (CA INDEX NAME)
OTHER CA INDEX NAMES:
CN 6-Methyleneandrosta-1,4-diene-3,17-dione
MF C20 H24 O2
INCH InChI=1S/C20H24O2/c1-12-10-14-15-4-5-18(22)20(15,3)9-7-16(14)19(2)8-6-13(21)11-17(12)19/h6,8,11,14-16H,1,4-5,7,9-10H2,2-3H3/t14-,15-,16-,19+,20-/m0/s1
INKY BFYIZQONLCFLEV-DAELLWKTSA-N



Absolute stereochemistry shown

InChIs are available in all standard display formats

InChI (INCH)

- Unique identifier created from chemical structure
- Depends on structure indexing rules of database producer

InChiKey (INKY)

- Hashed version of the full InChI
- Standardized to a length of 27 characters
- Facilitates database and Web searches

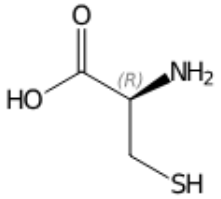
InChIKey retrieval capabilities

Layered design fully implemented on STNext

XUJNEKJLAYXESH-REOHCLBHSA-N

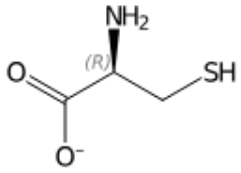
Core structure *Stereochemistry/Isotopes* *Protonation state*

```
=> S XUJNEKJLAYXESH-REOHCLBHSA-N/INKY

L1             2 XUJNEKJLAYXESH-REOHCLBHSA-N/INKY

=> S XUJNEKJLAYXESH-REOHCLBHSA/INKY

L2            11 XUJNEKJLAYXESH-REOHCLBHSA/INKY

=> S XUJNEKJLAYXESH/INKY

L3            92 XUJNEKJLAYXESH/INKY
```



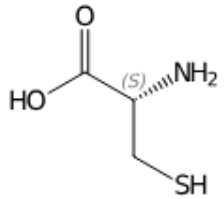
RN 52-90-4 L-cysteine

XUJNEKJLAYXESH-REOHCLBHSA-N



RN 19237-87-7 L-cysteine, ion

XUJNEKJLAYXESH-REOHCLBHSA-M



RN 921-01-7 D-cysteine

XUJNEKJLAYXESH-UWTATZPHSA-N

Value of InChI Keys on STNext

Useful cross database reference keys

- **Analyzing and sorting structure search results by InChIKeys**
  - Especially useful for evaluating substructure search results
- **Selective search**
  - For a specific stereoisomer or a specific isotope-labeled form of a chemical compound
  - Independent from chemical name variations and availability of stereospecific structure search

Collecting substance data from various resources

  - From STNext substance databases
  - From Internet resources
- **Identifying duplicate structures across databases**
  - Evaluation of multifile structure search results and looking at unique chemical compounds

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Searching InChI Keys on STNext

- **InChI Keys can be searched at three different levels**
  - Entire InChI Key
  - First two parts
  - First part only
- **InChI codes (or strings) cannot be searched, only displayed**

Searching InChI Keys on STNext

- **InChI Keys can be searched at three different levels**
  - Entire InChI Key – Core structure + Stereochemistry/Isotopes + Protonation site
  - First two parts – Core Structure + Stereochemistry/Isotopes
  - First part only – Core structure

```
=> S XUJNEKJLAYXESH-REOHCLBHSA-N/INKY

L1             2 XUJNEKJLAYXESH-REOHCLBHSA-N/INKY

=> S XUJNEKJLAYXESH-REOHCLBHSA/INKY

L2            11 XUJNEKJLAYXESH-REOHCLBHSA/INKY

=> S XUJNEKJLAYXESH/INKY

L3            92 XUJNEKJLAYXESH/INKY
```

Search run in CAS REGISTRY.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

EXAMPLES

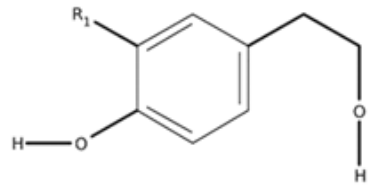
FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Example 1

Analysis of an answer set for core structures

```
=>
Uploading structure file: …
```




```
R-Group Definitions
R1: O,S,H,F,Cl
Node Attributes
Ring Nodes : 1  2  3  4  5  6
Chain Nodes : 7  8  9  10  11  12
L1      STRUCTURE UPLOADED

=> S L1 CSS FUL
L2            8 SEA CSS FUL L1
```

R-group defined as 5 different atoms

Closed substructure search for structure with a variable group in DCR.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Example 1 (cont.)

Analysis of an answer set for core structures

```
=> ANA L2 INKY LEN14

L3          ANALYZE L2 1- INKY LEN 14 :        6 TERMS


=> D 1-6

L3          ANALYZE L2 1- INKY LEN 14 :        6 TERMS

TERM #    # OCC   # DOC   % DOC INKY
------ ------- ------ ------ --------------
     1        2      2  28.57 JUUBCHWRXWPFFH
     2        2      2  28.57 YCCILVSKPBXVIP
     3        1      1  12.50 PQWLPSXCNUCESA
     4        1      1  14.29 TVZJWJYHPVHFGY
     5        1      1  14.29 VPWOUSYKQMPMMJ
     6        1      1  14.29 WYPBWLOUTCEOJL
*********   END OF L3    *********
```

ANALYZE of answer set for core structures by using the 14-character INKY string

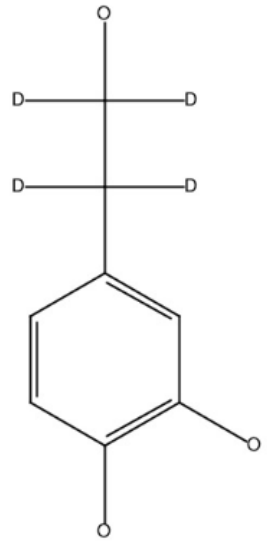
- InChIKeys (/INKY) are indexed and can also be analyzed in three different lengths which represent different degrees of precision.
- 14 characters: core structure (connectivity and bonds)
- 25 characters: core, stereochemistry, isotopes
- 27 characters: fully defined structure; core, stereochemistry, isotopes and charge.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

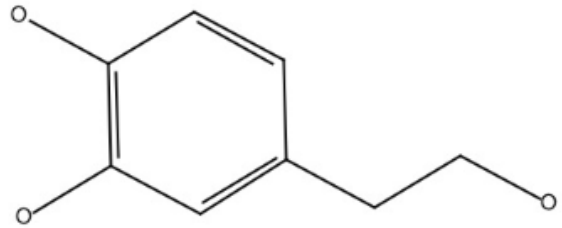
Example 1 (cont.)

```
=> S JUUBCHWRXWPFFH/INKY

L4            2 JUUBCHWRXWPFFH/INKY

=> D 1-2


L4     ANSWER 1 OF 2 DCR  COPYRIGHT 2025 CLARIVATE on STN.
AN     DCR-7765322   DCR
DCSE   7765322-0-0-0
CN.P   4-[2-hydroxy(1,1,2,2-2H4)ethyl]benzene-1,2-diol
```




```
MF     C8 H10 O3
INCH   InChI=1S/C8H10O3/c9-4-3-6-1-2-7(10)8(11)5-6/h1-2,5,9-11H,3-4H2/i3D2,4D2
INKY   JUUBCHWRXWPFFH-KHORGVISSA-N
ED     Entered STN: 26 Feb 2025
       Last updated on STN: 26 Feb 2025
```

```
L4     ANSWER 2 OF 2 DCR  COPYRIGHT 2025 CLARIVATE on STN.
AN     DCR-93253   DCR
DCSE   93253-0-0-0
CN.P   DIHYDROXYPHENYLETHANOL
CN.S   4-(2-Hydroxy-ethyl)-benzene-1,2-diol
SY     3,4-DIHYDROXYPHENYLETHANOL; 3-HYDROXYTYROSOL; DIHYDROXYPHENYLETHANOL;
       DOPET; HYDROXYTYROSOL
```



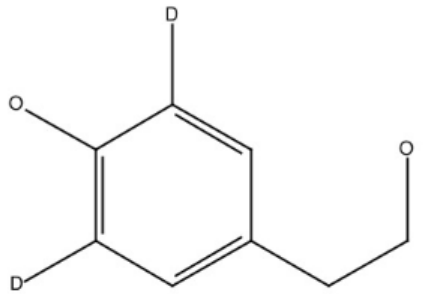
```
MF     C8 H10 O3
INCH   InChI=1S/C8H10O3/c9-4-3-6-1-2-7(10)8(11)5-6/h1-2,5,9-11H,3-4H2
INKY   JUUBCHWRXWPFFH-UHFFFAOYSA-N
ED     Entered STN: 2 Apr 2001
       Last updated on STN: 10 Apr 2025
       Update DWPI Cross Ref.: 20 Jun 2025
```


FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

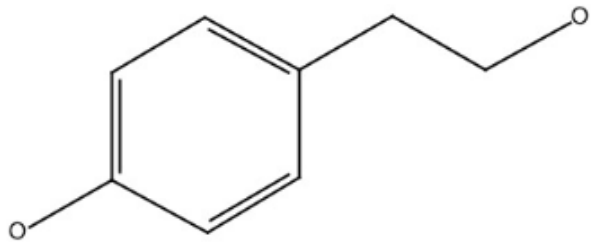
Example 1 (cont.)

```
=> S YCCILVSKPBXVIP/INKY

L5          2 YCCILVSKPBXVIP/INKY

=> D 1-2


L5     ANSWER 1 OF 2 DCR  COPYRIGHT 2025 CLARIVATE on STN.
AN     DCR-7623334   DCR
DCSE   9344-0-0-1
CN.P   TYROSOL (BIS-2H)
CN.S   4-(2-HYDROXYETHYL)(2,6-XFFFDH2)PHENOL
SY     TYROSOL (BIS-2H)
```




```
CMT    isotope labeled, dideuterated
MF     C8 H10 O2
INCH   InChI=1S/C8H10O2/c9-6-5-7-1-3-8(10)4-2-7/h1-4,9-10H,5-6H2/i3D,4D
INKY   YCCILVSKPBXVIP-NMQOAUCRSA-N
ED     Entered STN: 11 Dec 2024
       Last updated on STN: 26 Dec 2024
```

```
L5     ANSWER 2 OF 2 DCR  COPYRIGHT 2025 CLARIVATE on STN.
AN     DCR-9344   DCR
DCSE   9344-0-0-0
CN.P   TYROSOL
CN.S   4-(2-Hydroxy-ethyl)-phenol
SY     4-HYDROXYPHENETHYL ALCOHOL; HYDROXYETHYLPHENOL-2-4;
       HYDROXYETHYLPHENOL-2-PARA; HYDROXYPHENETHYL-ALCOHOL-4;
       HYDROXYPHENETHYL-ALCOHOL-PARA; HYDROXYPHENYLETHANOL-2-PARA;
       HYDROXYPHENYLETHANOL-4-2; HYDROXYPHENYLETHANOL-4-BETA;
       PARA-HYDROXYPHENYLETHANOL-BETA; TYROSOL; TYROSOL-PARA
```




```
MF     C8 H10 O2
INCH   InChI=1S/C8H10O2/c9-6-5-7-1-3-8(10)4-2-7/h1-4,9-10H,5-6H2
INKY   YCCILVSKPBXVIP-UHFFFAOYSA-N
ED     Entered STN: 13 Aug 1999
       Last updated on STN: 10 Apr 2025
       Update DWPI Cross Ref.: 18 Jun 2025
```


FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Example 2

Targeted search for a specific stereoisomer and its preparation

```
=> S D-PENICILLAMINE/CN

FILE 'DCR'
L6              0 D-PENICILLAMINE/CN

FILE 'REGISTRY'
L7              1 D-PENICILLAMINE/CN

FILE 'REAXYSFILESU'
L8              7 D-PENICILLAMINE/CN
```

Search for the chemical name of the stereoisomer D-penicillamine finds

- 0 hits in DCR
- 1 hit in REGISTRY
- 7 hits in REAXYSFILESUB

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Example 2 (cont.)

ReaxysFileSub records

```
L8    ANSWER 1 OF 7 REAXYSFILESU COPYRIGHT 2025 ELSEVIER INC. on STN.
AN     11346972   REAXYSFILESU
CN     d-penicillamine; disodium EDTA; OTTENS Mixed Berry; phosphoric
       acid; propylparaben; sodium methylparaben; sodium saccharine;
       water; purified; mixture of
SD     mixture (composition partially given)
…
L8    ANSWER 2 OF 7 REAXYSFILESU COPYRIGHT 2025 ELSEVIER INC. on STN.
AN     11346973   REAXYSFILESU
CN     d-penicillamine; disodium EDTA; phosphoric acid; propylparaben;
       sodium methylparaben; sodium saccharine; water; purified;
       mixture of
SD     mixture (composition partially given)
…
```

Mixtures of D-penicillamine

Multicomponent substances with D-penicillamine include the name of the component D-penicillamine in the chemical name.

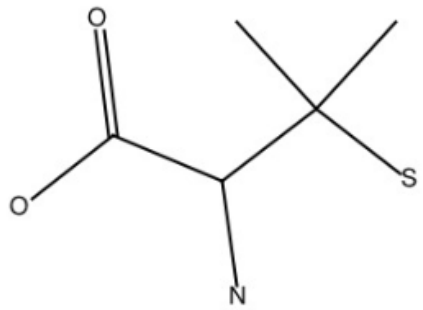
FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

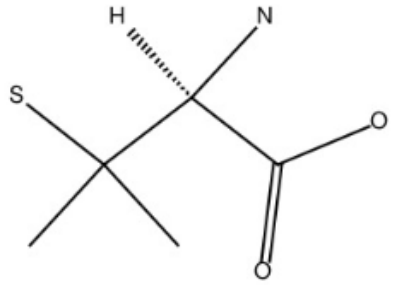
Example 2 (cont.)

3 ReaxysFileSub records of interest

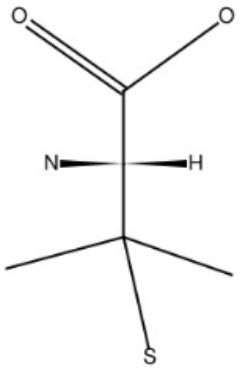
L13 ANSWER 5 OF 7 REAXYSFILESU COPYRIGHT 2025 ELSEVIER INC. on STN.
AN 4291203 REAXYSFILESU
CN **D-penicillamine**; 2-Amino-3-mercapto-3-methyl-butyric acid
SD acyclic
MF C5 H11 N O2 S
CMF C5 H11 N O2 S
LSF C5H11NO2S
INCHI VVNCNSJFMMFHPL-UHFFFAOYSA-N
AINCHI VVNCNSJFMMFHPL-QDQILVOLCP
MW 149.214
MARKREF.CNT 0
REC 13
ED Entered STN: 13 Jul 2020
Last updated on STN: 19 Jan 2024



L13 ANSWER 6 OF 7 REAXYSFILESU COPYRIGHT 2025 ELSEVIER INC. on STN.
AN 1722375 REAXYSFILESU
RN 52-67-5
CN 3,3-dimethyl-D-cysteine; **D-penicillamine**; penicillamine; 3B2; -thiovaline; DPA
SD acyclic
MF C5 H11 N O2 S
CMF C5 H11 N O2 S
LSF (CH3)2(HS)CCH(NH2)COOH
INCHI VVNCNSJFMMFHPL-VKHMYHEASA-N
AINCHI VVNCNSJFMMFHPL-GFBCBKCJDA
MW 149.214
MARKREF.CNT 2
REC 4713
ED Entered STN: 14 Jul 2020
Last updated on STN: 20 Jun 2025



L13 ANSWER 7 OF 7 REAXYSFILESU COPYRIGHT 2025 ELSEVIER INC. on STN.
AN 1722374 REAXYSFILESU
RN 1113-41-3
CN L-penicillamine; **D-penicillamine**; penicillamine
SD acyclic
MF C5 H11 N O2 S
CMF C5 H11 N O2 S
LSF HSC(CH3)2CH(NH2)COOH
INCHI VVNCNSJFMMFHPL-GSVOUGTGSA-N
AINCHI VVNCNSJFMMFHPL-ZKQKRMTKDJ
MW 149.214
MARKREF.CNT 0
REC 893
ED Entered STN: 14 Jul 2020
Last updated on STN: 20 Jun 2025



FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

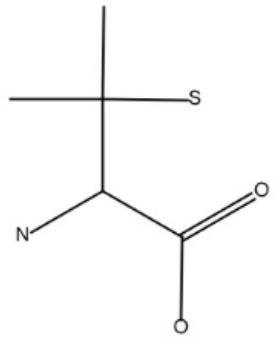
CAS
A division of the American Chemical Society

Example 2 (cont.)

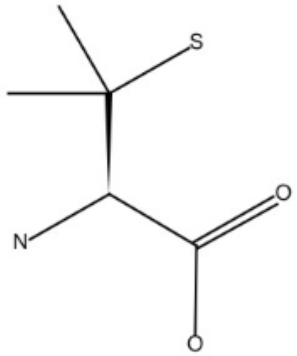
DCR records for InChI Keys from ReaxysFileSub

In this record, info on stereochemistry is given in the comment field.

```
=> S (VVNCNSJFMMFHPL-UHFFFAOYSA-N OR VVNCNSJFMMFHPL-VKHMYHEASA-N OR VVNCNSJFMMFHPL-GSVOUGTGSA-N)/INKY

             1 VVNCNSJFMMFHPL-UHFFFAOYSA-N/INKY
             1 VVNCNSJFMMFHPL-VKHMYHEASA-N/INKY
             1 VVNCNSJFMMFHPL-GSVOUGTGSA-N/INKY
L14          3 (VVNCNSJFMMFHPL-UHFFFAOYSA-N OR VVNCNSJFMMFHPL-VKHMYHEASA-N OR
               VVNCNSJFMMFHPL-GSVOUGTGSA-N)/INKY

=> D 1-3


L14    ANSWER 1 OF 3 DCR  COPYRIGHT 2025 CLARIVATE on STN.
AN     DCR-1150233   DCR
DCSE   8198-0-0-0
CN.P   PENICILLAMINE
CN.S   2-Amino-3-mercapto-3-methyl-butyric acid
SY     PENICILLAMINE
```

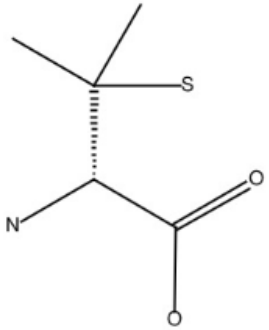



```
L14    ANSWER 2 OF 3 DCR  COPYRIGHT 2025 CLARIVATE on STN.
AN     DCR-961512   DCR
DCSE   8198-2-0-0
CN.P   L-PENICILLAMINE
CN.S   (R)-2-Amino-3-mercapto-3-methyl-butyric acid
SY     L-PENICILLAMINE; PENICILLAMINE-L
```




```
MF     C5 H11 N O2 S
INCH   InChI=1S/C5H11NO2S/c1-5(2,9)3(6)4(7)8/h3,9H,6H2,1-2H3,(H,7,8)/t3-/m1/s1
INKY   VVNCNSJFMMFHPL-GSVOUGTGSA-N
ED     Entered STN: 11 Oct 2004
       Last updated on STN: 20 Jun 2025
       Update DWPI Cross Ref.: 18 Jun 2025
```

```
L14    ANSWER 3 OF 3 DCR  COPYRIGHT 2025 CLARIVATE on STN.
AN     DCR-8198   DCR
DCSE   8198-1-0-0
CN.P   PENICILLAMINE
SY     PENICILLAMINE
```




```
CMT    D-isomer
MF     C5 H11 N O2 S
INCH   InChI=1S/C5H11NO2S/c1-5(2,9)3(6)4(7)8/h3,9H,6H2,1-2H3,(H,7,8)/t3-/m0/s1
INKY   VVNCNSJFMMFHPL-VKHMYHEASA-N
ED     Entered STN: 3 Jun 1999
       Last updated on STN: 13 Apr 2022
       Update DWPI Cross Ref.: 20 Jun 2025
```

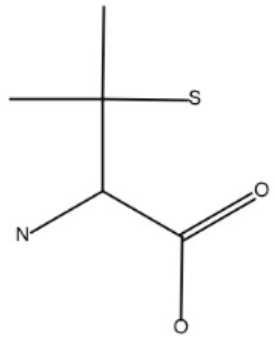

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

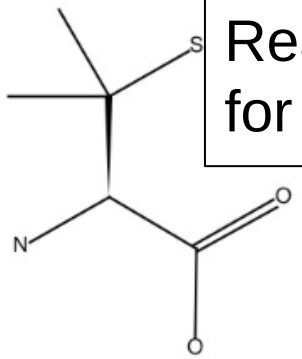
Example 2 (cont.)

DCR records for InChI Keys from ReaxysFileSub

```
=> S (VVNCNSJFMMFHPL-UHFFFAOYSA-N OR VVNCNSJFMMFHPL-VKHMYHEASA-N OR VVNCNSJFMMFHPL-GSVOUGTGSA-N)/INKY

             1 VVNCNSJFMMFHPL-UHFFFAOYSA-N/INKY
             1 VVNCNSJFMMFHPL-VKHMYHEASA-N/INKY
             1 VVNCNSJFMMFHPL-GSVOUGTGSA-N/INKY
L14          3 (VVNCNSJFMMFHPL-UHFFFAOYSA-N OR VVNCNSJFMMFHPL-VKHMYHEASA-N OR
               VVNCNSJFMMFHPL-GSVOUGTGSA-N)/INKY

=> D 1-3


L14    ANSWER 1 OF 3 DCR  COPYRIGHT 2025 CLARIVATE on STN.
AN     DCR-1150233   DCR
DCSE   8198-0-0-0
CN.P   PENICILLAMINE
CN.S   2-Amino-3-mercapto-3-methyl-butyric acid
SY     PENICILLAMINE
```



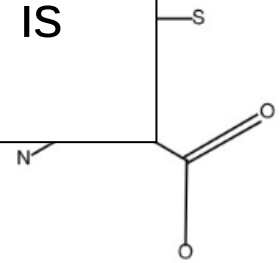

```
L14    ANSWER 2 OF 3 DCR  COPYRIGHT 2025 CLARIVATE on STN.
AN     DCR-961512   DCR
DCSE   8198-2-0-0
CN.P   L-PENICILLAMINE
CN.S   (R)-2-Amino-3-mercapto-3-methyl-butyric acid
SY     L-PENICILLAMINE; PENICILLAMINE-L
```




```
MF     C5 H11 N O2 S
INCH   InChI=1S/C5H11NO2S/c1-5(2,9)3(6)4(7)8/h3,9H,6H2,1-2H3,(H,7,8)/t3-/m1/s1
INKY   VVNCNSJFMMFHPL-GSVOUGTGSA-N
ED     Entered STN: 11 Oct 2004
       Last updated on STN: 20 Jun 2025
       Update DWPI Cross Ref.: 18 Jun 2025
```

The corresponding ReaxysFileSub record is for D-Penicillamine.

```
L14    ANSWER 3 OF 3 DCR  COPYRIGHT 2025 CLARIVATE on STN.
AN     DCR-8198   DCR
DCSE   8198-1-0-0
CN.P   PENICILLAMINE
SY     PENICILLAMINE
```




```
CMT    D-isomer
MF     C5 H11 N O2 S
INCH   InChI=1S/C5H11NO2S/c1-5(2,9)3(6)4(7)8/h3,9H,6H2,1-2H3,(H,7,8)/t3-/m0/s1
INKY   VVNCNSJFMMFHPL-VKHMYHEASA-N
ED     Entered STN: 3 Jun 1999
       Last updated on STN: 13 Apr 2022
       Update DWPI Cross Ref.: 20 Jun 2025
```


FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Example 2 (cont.)

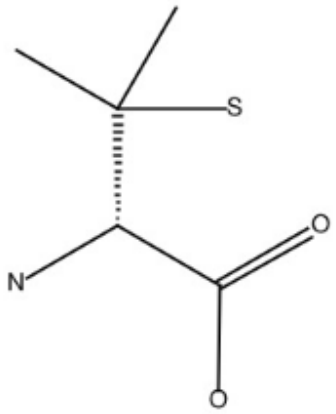
Corresponding DWPI search

```
=> FILE WPINDEX
```

```
=> S L14(T)(PRD OR P)/RL

          882 L14
      1161220 PRD/RL
      1374864 P/RL
L15         45 L14(T)(PRD OR P)/RL
```

```
=> D BIB HITSTR 1-10
```

```
L15  ANSWER 2 OF 45  WPINDEX COPYRIGHT 2025   CLARIVATE on STN
AN   2024-B87529 [2024094]   WPINDEX Full-text
TI   Synthesizing D-penicillamine, involves taking penicillin G potassium salt
     as raw material, performing condensation and ring opening with
     mono-substituted hydrazine in mixed solvent under action of catalyst,
     adding acetic acid, acidifying, adding n-heptane, crystallizing, obtaining
     D-penicillamine
DC   B02; J04
IN   CUI C; HU Y; YANG M; YU S
PA   (CASA-N) CAS ALDONE DALIAN PHARM TECHNOLOGY CO
CYC  1
PI   CN 118851961     A  20241029 (2024094)* ZH
     CN 118851961     B  20250124 (2025014)  ZH
ADT  CN 118851961 A CN 2024-10889777 20240703; CN 118851961 B CN 2024-10889777
     20240703
FDT  CN 118851961 B Previous Publ CN 118851961 A
PRAI CN 2024-10889777       20240703
AN.S DCR-8198
CN.P PENICILLAMINE
MF   C5 H11 N O2 S
STR
```



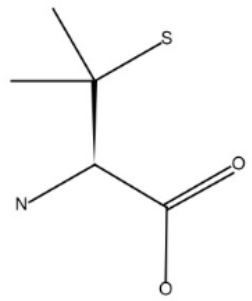
FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

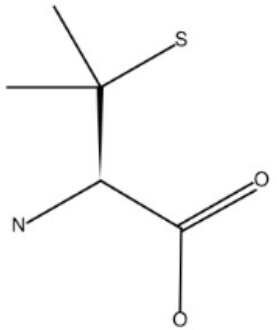
Example 2 (cont.)

Corresponding DWPI search

L15 ANSWER 3 OF 45 WPINDEX COPYRIGHT 2025 CLARIVATE on STN
AN 2024-B4750R [2024095] WPINDEX Full-text
TI Preparing L-penicillamine, compises e.g. nitrating 3,3-dimethylacrylate, subjecting to Michael addition with thio reagent in presence of alkaline catalyst, reducing nitro group to amino group to obtain racemic penicillamine, resolving racemic by ester hydrolase, and performing alkaline hydrolysis
DC B05; D16
IN CHEN H; DONG C; HUANG J; MAN Y; WEI S; WEI Y; XIAO Y; XIONG Z; YUAN L; ZHANG G; ZHANG H
PA (ASYM-C) TIANJIN ASYMCHEM PHARM CO LTD
CYC 1
PI CN 118834148 A 20241025 (2024095)* ZH
CN 118834148 B 20250121 (2025010) ZH
ADT CN 118834148 A CN 2024-11316816 20240920; CN 118834148 B CN 2024-11316816 20240920
FDT CN 118834148 B Previous Publ CN 118834148 A
PRAI CN 2024-11316816 20240920
AN.S DCR-961512
CN.P L-PENICILLAMINE
CN.S (R)-2-Amino-3-mercapto-3-methyl-butyric acid
MF C5 H11 N O2 S
STR



L15 ANSWER 6 OF 45 WPINDEX COPYRIGHT 2025 CLARIVATE on STN
AN 2023-991165 [2023081] WPINDEX Full-text
TI Producing L-penicillamine useful e.g. as synthetic raw material for pharmaceuticals involves cyclizing D-penicillamine, acting D-amino acid oxidase on cyclic D-penicillamine compound, reducing thiazoline compound and ring-opening compound
DC B05; D16
IN IWAMOTO Y; IWASAKI A
PA (KANF-C) KANEKA CORP
CYC 138
PI WO 2023176487 A1 20230921 (2023081)* JA 68[0]
ADT WO 2023176487 A1 WO 2023-JP7888 20230302
PRAI JP 2022-184513 20221118
JP 2022-42906 20220317
AN.S DCR-961512
CN.P L-PENICILLAMINE
CN.S (R)-2-Amino-3-mercapto-3-methyl-butyric acid
MF C5 H11 N O2 S
STR



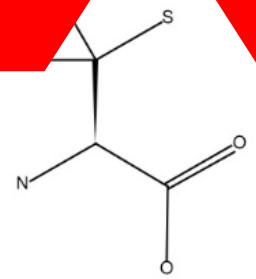
FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

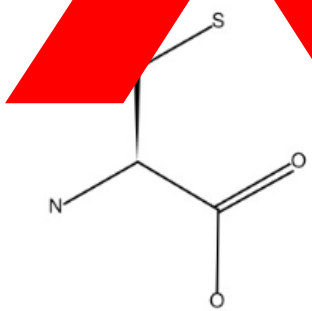
Example 2 (cont.)

Corresponding DWPI search

L15 ANSWER 3 OF 45 WPINDEX COPYRIGHT 2025 CLARIVATE on STN
AN 2024-B4750R [2024095] WPINDEX Full-text
TI Preparing L-penicillamine, compises e.g. nitrating 3,3-dimethylacrylate, subjecting to Michael addition with thio reagent in presence of alkaline catalyst, reducing nitro group to amino group to obtain racemic penicillamine, resolving racemic by ester hydrolase, and performing alkaline hydrolysis
DC B05; D16
IN CHEN H; DONG C; HUANG J; MAN Y; WEI S; WEI Y; XIAO Y; XIONG Z; YUAN L; ZHANG C; ZHANG H
PA (AS ASYMCHEM PHARM CO LT
CYC 1
PI CN 1188 241025 (2024
CN 1188347 50121 (202
ADT CN 1188341 316816 CN 118834148 B CN 2024-11316816
20240920
FDT CN 118834148 B 8 A
PRAI CN 2024-11316816
AN.S DCR-961512
CN.P L-PENICILLAMINE
CN.S (R)-2-Amino-3-merc ric acid
MF C5 H11 N O2 S
STR



L15 ANSWER 6 OF 45 WPINDEX COPYRIGHT 2025 CLARIVATE on STN
AN 2023-991165 [2023081] WPINDEX Full-text
TI Producing L-penicillamine useful e.g. as synthetic raw material for pharmaceuticals involves cyclizing D-penicillamine, acting D-amino acid oxidase on cyclic D-penicillamine compound, reducing thiazoline compound and ring-opening compound
DC B05; D16
IN IWAMOTO Y; IWASAKI A
PA (KANF-C) KANEKA CORP
CYC 138
PI WO 202317648 0921 (2023081)
ADT WO 2023176487 TP7888 202
PRAI JP 2022-184513 118
JP 2022-42906 7
AN.S DCR-961512
CN.P L-PENICILLAMINE
CN.S (R)-2-Amino-3-mercapto c acid
MF C5 H11 N O2 S
STR



These records were captured because the DCR record for L-penicillamine was included in the strategy.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Example 2 (cont.)

CAS REGISTRY records for InChI Keys from ReaxysFileSub

```
=> FILE REGISTRY
```

```
=> S (VVNCNSJFMMFHPL-UHFFFAOYSA-N OR VVNCNSJFMMFHPL-VKHMYHEASA-N OR VVNCNSJFMMFHPL-GSVOUGTGSA-N)/INKY

             1 VVNCNSJFMMFHPL-UHFFFAOYSA-N/INKY
             2 VVNCNSJFMMFHPL-VKHMYHEASA-N/INKY
             1 VVNCNSJFMMFHPL-GSVOUGTGSA-N/INKY
L16          4 (VVNCNSJFMMFHPL-UHFFFAOYSA-N OR VVNCNSJFMMFHPL-VKHMYHEASA-N OR
               VVNCNSJFMMFHPL-GSVOUGTGSA-N)/INKY

=> D 1-4
```

```
L16  ANSWER 1 OF 4  REGISTRY  COPYRIGHT 2025 ACS on STN
RN   2920635-83-0  REGISTRY
ED   Entered STN:  18 Apr 2023
CN   INDEX NAME NOT YET ASSIGNED
FS   STEREOSEARCH
MF   (C5 H11 N O2 S)x
INCH InChI=1S/C5H11NO2S/c1-5(2,9)3(6)4(7)8/h3,9H,6H2,1-2H3,(H,7,8)/t3-/m0/s1
INKY VVNCNSJFMMFHPL-VKHMYHEASA-N
CI   PMS
PCT  Polyamide, Polyamide formed
SR   CA
LC   STN Files:   CA, CAPLUS
```

```
L16  ANSWER 2 OF 4  REGISTRY  COPYRIGHT 2025 ACS on STN
RN   1113-41-3  REGISTRY
ED   Entered STN:  16 Nov 1984
CN   L-Valine, 3-mercapto-  (CA INDEX NAME)
OTHER CA INDEX NAMES:
CN   3-Mercapto-L-valine
CN   Valine, 3-mercapto-, L- (8CI)
OTHER NAMES:
CN   (+)-Penicillamine
CN   (2R)-2-Amino-3-methyl-3-sulfanylbutanoic acid
CN   (R)-2-Amino-3-mercapto-3-methylbutanoic acid
CN   (R)-Penicillamine
CN   L-Penicillamine
CN   NSC 241261
FS   STEREOSEARCH
DR   16414-53-2, 73706-37-3
MF   C5 H11 N O2 S
INCH InChI=1S/C5H11NO2S/c1-5(2,9)3(6)4(7)8/h3,9H,6H2,1-2H3,(H,7,8)/t3-/m1/s1
INKY VVNCNSJFMMFHPL-GSVOUGTGSA-N
```


FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Example 2 (cont.)

CAS REGISTRY records for InChI Keys from ReaxysFileSub

```
L16  ANSWER 3 OF 4  REGISTRY  COPYRIGHT 2025 ACS on STN
RN   52-67-5  REGISTRY
ED   Entered STN:  16 Nov 1984
CN   D-Valine, 3-mercapto-  (CA INDEX NAME)
OTHER CA INDEX NAMES:
CN   3-Mercapto-D-valine
CN   Valine, 3-mercapto-, D- (8CI)
OTHER NAMES:
CN   (-)-Penicillamine
CN   (2S)-2-Amino-3-methyl-3-sulfanylbutanoic acid
CN   (2S)-2-Amino-3-methyl-3-sulfanylbutanoic acid
CN   (2S)-2-Azaniumyl-3-methyl-3-sulfanylbutanoate
CN   (S)-3,3-Dimethylcysteine
CN   (S)-Penicillamine
CN   β-Thiovaline
CN   2-Amino-3-mercapto-3-methylbutanoic acid
CN   Artamin
CN   Cuprenil
CN   Cuprimine
CN   Cupripen
CN   D-(-)-Penicillamine
CN   D-3-Mercaptovaline
CN   D-Mercaptovaline
CN   D-Penamine
CN   D-Penicillamine
INCH InChI=1S/C5H11NO2S/c1-5(2,9)3(6)4(7)8/h3,9H,6H2,1-2H3,(H,7,8)/t3-/m0/s1
INKY VVNCNSJFMMFHPL-VKHMYHEASA-N
```

```
L16  ANSWER 4 OF 4  REGISTRY  COPYRIGHT 2025 ACS on STN
RN   52-66-4  REGISTRY
ED   Entered STN:  16 Nov 1984
CN   Valine, 3-mercapto-  (CA INDEX NAME)
OTHER CA INDEX NAMES:
CN   3-Mercaptovaline
CN   DL-Valine, 3-mercapto-
CN   Valine, 3-mercapto-, DL- (8CI)
OTHER NAMES:
CN   (±)-Penicillamine
CN   β,β-Dimethylcysteine
CN   β-Mercaptovaline
CN   β-Thiovaline
CN   2-Amino-3-mercapto-3-methylbutanoic acid
CN   2-Amino-3-mercapto-3-methylbutyric acid
CN   2-Amino-3-methyl-3-sulfanylbutanoic acid
CN   2-Azaniumyl-3-methyl-3-sulfanylbutanoate
CN   Cysteine, 3,3-dimethyl-
CN   DL-β-Mercaptovaline
CN   DL-3-Mercaptovaline
CN   dl-DMC
CN   DL-Penicillamine
CN   DMC
CN   NSC 22880
CN   NSC 44656
DR   5139-02-6
MF   C5 H11 N O2 S
INCH InChI=1S/C5H11NO2S/c1-5(2,9)3(6)4(7)8/h3,9H,6H2,1-2H3,(H,7,8)
INKY VVNCNSJFMMFHPL-UHFFFAOYSA-N
```


FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

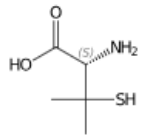
CAS
A division of the American Chemical Society

Example 2 (cont.)

Corresponding HCAplus search

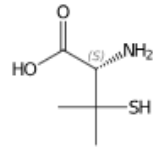
```
=> S L16/PREP

         6863 L16
      9261531 PREP/RL
L17        302 L16/PREP
               (L16 (L) PREP/RL)

=> D BIB HITSTR 1-10


L17  ANSWER 1 OF 302  HCAPLUS  COPYRIGHT 2025 ACS on STN
PatentPak PDF | PatentPak PDF+ | PatentPak Interactive
AN   2025:1078465  HCAPLUS Full-text
DN   191:9138
TI   Green preparation D-penicillamine using acidic cationic resin
IN   Chen, Xiaodong; Zhou, Liang; Chu, Changhu; Wu, Teng; Liu, Xuejun
PA   Shanghai Neutan Pharma Research and Development Co., Ltd., Peop. Rep.
     China
SO   Faming Zhuanli Shenqing, 7pp.
     CODEN: CNXXEV
DT   Patent
LA   Chinese
FAN.CNT 1
PPPI
     PATENT NO.          KIND  DATE      LANGUAGE   PatentPak
     ----------------    ----  --------  ---------- ------------------------
     CN 119977889         A    20250513  Chinese    PDF | PDF+ | Interactive
PI
     PATENT NO.          KIND  DATE        APPLICATION NO.          DATE
     ----------------    ----  --------    ---------------------    --------
     CN 119977889         A    20250513    CN 2023-11495766         20231110
PRAI CN 2023-11495766          20231110
PSPI
     PATENT NO.          KIND  STATUS          STATUS DATE
     ----------------    ----  --------------  -----------
     CN 119977889         A    Alive           20250522
```

```
IT   52-67-5 P
     RL: IMF (Industrial manufacture); SCLM (Substance claimed); SPN (Synthetic
     preparation); THU (Therapeutic use); BIOL (Biological study); PREP
     (Preparation); USES (Uses)
        (green prepn. D-penicillamine using acidic cationic resin)
RN   52-67-5  HCAPLUS
CN   D-Valine, 3-mercapto-  (CA INDEX NAME)
```



Absolute stereochemistry shown

```
L17  ANSWER 2 OF 302  HCAPLUS  COPYRIGHT 2025 ACS on STN

AN   2024:2594088  HCAPLUS Full-text
TI   Self-assembly of differently charged trimesic based lithocholic
     amphiphiles and their assessment on antimicrobial and biostimulant
     properties
AU   Ezhumalai, Nishanthi; Panchalingam, Santhiya; Govindaraju, Kasivelu;
     Kannan, Malaichamy; Kasthuri, Jayapalan; Rajendiran, Nagappan
CS   Department of Polymer Science, Guindy Campus, University of Madras, Tamil
     Nadu, 600025, India
SO   Colloids and Surfaces, B: Biointerfaces (2025), 246, 114391
     CODEN: CSBBEQ; ISSN: 0927-7765
DOI  10.1016/j.colsurfb.2024.114391
PB   Elsevier B.V.
DT   Journal; (online computer file)
LA   English
IT   INDEXING IN PROGRESS
IT   52-67-5 P
     RL: AGR (Agricultural use); BPN (Biosynthetic preparation); PRP
     (Properties); RCT (Reactant); RGT (Reagent); BIOL (Biological study);
     PREP (Preparation); RACT (Reactant or reagent); USES (Uses)
        (trimesic based lithocholic ester functionalized zwitterionic;
        self-assembly of differently charged trimesic based lithocholic
        amphiphiles and their assessment on antimicrobial and biostimulant
        properties)
RN   52-67-5  HCAPLUS
CN   D-Valine, 3-mercapto-  (CA INDEX NAME)
```



Absolute stereochemistry shown

```
OSC.G   2     THERE ARE 2 CAPLUS RECORDS THAT CITE THIS RECORD (2 CITINGS)
RE.CNT  50    THERE ARE 50 CITED REFERENCES AVAILABLE FOR THIS RECORD
              ALL CITATIONS AVAILABLE IN THE RE FORMAT
```


FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Example 2 (cont.)

Corresponding HCAplus search

L17 ANSWER 5 OF 302 HCAPLUS COPYRIGHT 2025 ACS on STN
PatentPak PDF | PatentPak PDF+ | PatentPak Interactive
AN 2024:2295125 HCAPLUS Full-text
DN 188:355886
TI Preparation of L-penicillamine
IN Xiong, Zhengchang; Xiao, Yi; Zhang, Haiming; Yuan, Long; Dong, Changming; Wei, Yiran; Chen, Hui; Wei, Shanshan; Zhang, Guangwei; Man, Yi; Huang, Jie
PA Tianjin Asymchem Pharmaceutical Co., Ltd., Peop. Rep. China
UO ASYMCHEM INC.
UOS Asymchem
SO Faming Zhuanli Shenqing, 26pp.
CODEN: CNXXEV
DT Patent
LA Chinese
FAN.CNT 1
PPPI

PATENT NO.	KIND	DATE	LANGUAGE	PatentPak
CN 118834148	A	20241025	Chinese	PDF | PDF+ | Interactive
CN 118834148	B	20250121	Chinese	PDF

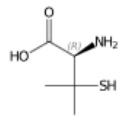
PI

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
CN 118834148	A	20241025	CN 2024-11316816	20240920
CN 118834148	B	20250121		

PRAI CN 2024-11316816 20240920
PSPI

PATENT NO.	KIND	STATUS	STATUS DATE
CN 118834148	A	Alive	20241031
CN 118834148	B	Alive	20241031

IT 1113-41-3 P
RL: SCLM (Substance claimed); SPN (Synthetic preparation); PREP (Preparation)
(prepn. of L-penicillamine)
RN 1113-41-3 HCAPLUS
CN L-Valine, 3-mercapto- (CA INDEX NAME)



Absolute stereochemistry shown
Rotation (+)

L17 ANSWER 7 OF 302 HCAPLUS COPYRIGHT 2025 ACS on STN
PatentPak PDF | PatentPak PDF+ | PatentPak Interactive
AN 2024:1464016 HCAPLUS Full-text
DN 187:205681
TI Cellulose ether and preparation method thereof
IN Xing, Jianlei; Jiang, Shuzheng; Yu, Xiaoliang; Jiao, Ningning; Xie, Haoran
PA Wuqiao Qiyuan Cellulose Co., Ltd., Peop. Rep. China
SO Faming Zhuanli Shenqing, 9pp.
CODEN: CNXXEV
DT Patent
LA Chinese
FAN.CNT 1
PPPI

PATENT NO.	KIND	DATE	LANGUAGE	PatentPak
CN 118290656	A	20240705	Chinese	PDF | PDF+ | Interactive
CN 118290656	B	20241101	Chinese	PDF

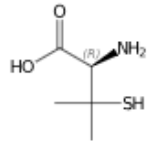
PI

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
CN 118290656	A	20240705	CN 2024-10492788	20240423
CN 118290656	B	20241101		

PRAI CN 2024-10492788 20240423
PSPI

PATENT NO.	KIND	STATUS	STATUS DATE
CN 118290656	A	Alive	20240718
CN 118290656	B	Alive	20240718

IT 1113-41-3 DP, reaction products with dialdehyde CM-cellulose
RL: IMF (Industrial manufacture); SCLM (Substance claimed); TEM (Technical or engineered material use); PREP (Preparation); USES (Uses)
(prepn. method of cellulose ether)
RN 1113-41-3 HCAPLUS
CN L-Valine, 3-mercapto- (CA INDEX NAME)



Absolute stereochemistry shown
Rotation (+)

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Example 2 (cont.)

Corresponding HCAplus search

L17 ANSWER 5 OF 302 HCAPLUS COPYRIGHT 2025 ACS on STN
PatentPak PDF | PatentPak PDF+ | PatentPak Interactive
AN 2024:2295125 HCAPLUS Full-text
DN 188:355886
TI Preparation of L-penicillamine
IN Xiong, Zhengchang; Xiao, Yi; Zhang, Haiming; Yuan, Long; Dong, Changming; Wei, Yiran; Chen, Hui; Wei, Shanshan; Zhang, Guangwei; Man, Yi; Huang, Jie
PA Tianjin Asymchem Pharmaceutical Co., Ltd., Peop. Rep. China
UO ASYMC
UOS Asymche
SO Faming Zh 26pp.
CODEN: CNXX
DT Patent
LA Chinese
FAN.CNT 1
PPPI

PATENT NO.		GE	PatentPak
CN 118834148	A	nese	PDF | PDF+ | Interactive
CN 118834148		se	PDF

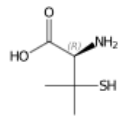
PI

PATENT NO.		TION NO.	DATE
CN 118834148		16816	20240920
CN 118834148		20250121	

PRAI CN 2024-11 20240920
PSPI

PATENT	IND	STATUS	
CN	A	Alive	202
CN 118834148	B	Alive	20241031

IT 1113-41-3 P
RL: SCLM (Substance claimed); SPN (Synthetic preparation); PREP (Preparation)
(prepn. of L-penicillamine)
RN 1113-41-3 HCAPLUS
CN L-Valine, 3-mercapto- (CA INDEX NAME)



Absolute stereochemistry shown
Rotation (+)

L17 ANSWER 7 OF 302 HCAPLUS COPYRIGHT 2025 ACS on STN
PatentPak PDF | PatentPak PDF+ | PatentPak Interactive
AN 2024:1464016 HCAPLUS Full-text
DN 187:205681
TI Cellulose ether and preparation method thereof
IN Xing, Jianlei; Jiang, Shuzheng; Yu, Xiaoliang; Jiao, Ningning; Xie, Haoran
PA Wuqiao Qiyuan Cellulose Co., Ltd., Peop. Rep. China
SO Faming Zhuanli Shenqing, 9pp.
CODEN: CNXXEV
DT Patent
LA Chinese
FAN.CNT 1
PPPI

PATENT NO.	KIND	DATE	LANGUAGE	PatentPak
CN 118290656		0240705	Chin	PDF+ | Interactive
CN 118290656		41101	Ch	

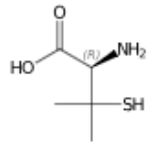
PI

PATENT NO.	K	ION NO.	DATE
CN 118290656	A	024-10492788	20240423
CN 118290656	B		

PRAI CN 2024-10492788
PSPI

PATENT NO.	KI	TUS DATE
CN 118290656		18
CN 118290656		ive

IT 1113-41-3 DP, r ucts with dia ellulose
RL: IMF (Indu acture); SCLM (Su med); TEM (Technical or engineer use); PREP (Preparati ses)
(prepn. method of cellulose ether)
RN 1113-41-3 HCAPLUS
CN L-Valine, 3-mercapto- (CA INDEX NAME)



Absolute stereochemistry shown
Rotation (+)

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Example 2 (cont.)

Rerun DCR/WPINDEX search using just one InChI Key

```
 => FILE DCR
```

```
=> S VVNCNSJFMMFHPL-VKHMYHEASA-N/INKY

L18          1 VVNCNSJFMMFHPL-VKHMYHEASA-N/INKY
```

```
 => FILE WPINDEX
```

```
 => S L18(T)(PRD OR P)/RL

           538 L18
       1161220 PRD/RL
       1374864 P/RL
L19          30 L18(T)(PRD OR P)/RL
```

Original search retrieved 45 records.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Example 2 (cont.)

Rerun CAS REGISTRY/HCAplus search using just one InChI Key

```
=> FILE REGISTRY
```

```
=> S VVNCNSJFMMFHPL-VKHMYHEASA-N/INKY

L20             2 VVNCNSJFMMFHPL-VKHMYHEASA-N/INKY
```

```
=> FILE HCAPLUS
```

```
=> S L20/PREP

          5952 L20
       9261531 PREP/RL
L21         206 L20/PREP
                (L20 (L) PREP/RL)
```

Monomer and polymer have the same InChI Key due to the monomer-based indexing.

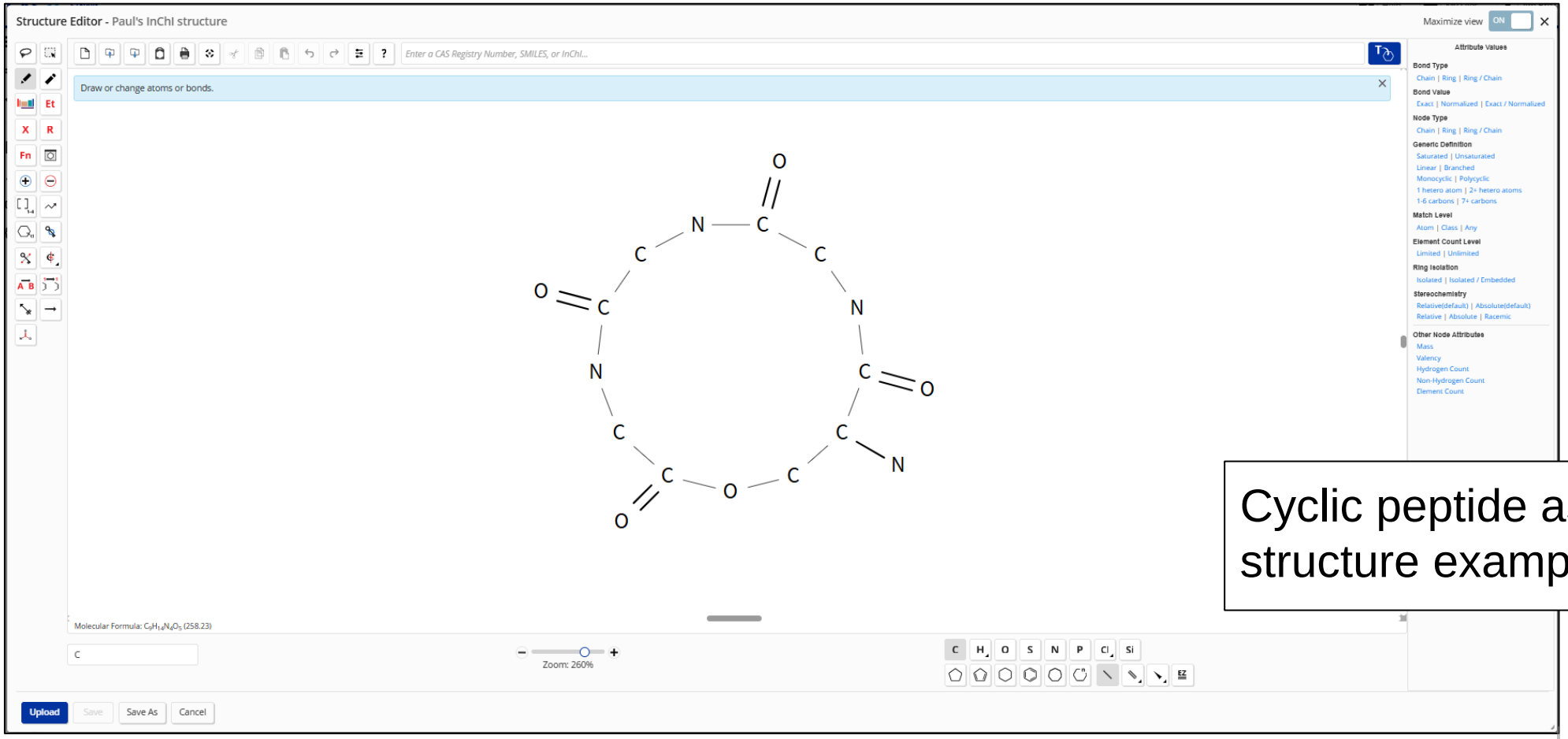
Original search retrieved 302 records.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Example 3

Find unique substance records across REGISTRY, DCR & REAXYSFILESUB



Cyclic peptide as the structure example.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Example 3 (cont.)

Find unique substance records across REGISTRY, DCR & REAXYSFILESUB

```
=> S L1 SSS FULL

FULL SEARCH INITIATED 14:11:43
FULL SCREEN SEARCH COMPLETED -    12917 TO ITERATE

100.0% PROCESSED      12917 ITERATIONS                          1029 ANSWERS
SEARCH TIME: 00.00.01

L2         1029 SEA SSS FUL L1

=> ANA L2 INKY 1-

L3          ANALYZE L2 1- INKY :    1022 TERMS

=> D

L3          ANALYZE L2 1- INKY :    1022 TERMS

TERM #   # OCC  # DOC  % DOC INKY
------ ------- ------ ------ ---------------
     1       2      2   0.19 OXKLHOSMVDGTBS-UHFFFAOYSA-N
     2       2      2   0.19 QSPKBORZTSNKNP-MMMVMMOASA-N
     3       2      2   0.19 RBHIOSHGUQQZQF-MMMVMMOASA-N
     4       2      2   0.19 SGGJJTTZBRPIKP-UHFFFAOYSA-N
     5       1      1   0.10 ABCJMLIOAYXJJP-HFWKHEOBSA-N
     6       1      1   0.10 ACYGACFVUBXFPG-JFAYZGHXSA-N
     7       1      1   0.10 ADNGAVBKDGODMA-UGFNBYPTSA-N
     8       1      1   0.10 ADVMDYNGUKYGDK-FAFIKCBISA-N
     9       1      1   0.10 AHTNGRTYKCUSSD-TWMHPYOZSA-N
    10       1      1   0.10 AHUYNAIWJQEFGX-FFBZIVEISA-N
1012 MORE TERMS WITH AN OCCURRENCE COUNT OF 1
```

Search is run in CAS REGISTRY.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Example 3 (cont.)

Find unique substance records across REGISTRY, DCR & REAXYSFILESUB

```
=> S L1 SSS FULL

FULL SEARCH INITIATED 14:18:11
FULL SCREEN SEARCH COMPLETED -        0 TO ITERATE

  0.0% PROCESSED    6263191 ITERATIONS                              141 ANSWERS
SEARCH TIME: 00.00.01

L4          141 SEA SSS FUL L1

=> ANA L4 INKY 1-

L5          ANALYZE L4 1- INKY :     139 TERMS

=> D

L5          ANALYZE L4 1- INKY :     139 TERMS

TERM #   # OCC  # DOC  % DOC INKY
------ ------- ------ ------ ---------------
     1       2      2   1.42 HJEBQUNJIQPONN-ANVYOYFISA-N
     2       2      2   1.42 HJEBQUNJIQPONN-KTURZQIXSA-N
     3       1      1   0.71 AENANIALMPCGPR-OTRWZRNCSA-N
     4       1      1   0.71 AERVTQZWEKLTOH-KUPAKDCRSA-N
     5       1      1   0.71 AJIQHWJOMQUKGQ-FMCGYTSFSA-N
     6       1      1   0.71 AVVXSZZSJQUPHA-QWSOZRLTSA-N
     7       1      1   0.71 AXMVMQQOBZPNAY-FFRZOQKGSA-N
     8       1      1   0.71 BEJRLMHCSIOJDI-IDPDROOQSA-N
     9       1      1   0.71 BFEKVXPJQSHMSY-YJZKUPLCSA-N
    10       1      1   0.71 BJWIDOUFUOSKJW-ABBHOYPZSA-N
129 MORE TERMS WITH AN OCCURRENCE COUNT OF 1
```

Search is run in DCR.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Example 3 (cont.)

Find unique substance records across REGISTRY, DCR & REAXYSFILESUB

```
=> S L1 SSS FULL

FULL SEARCH INITIATED 14:23:04
FULL SCREEN SEARCH COMPLETED -        0 TO ITERATE

  0.0% PROCESSED   54922368 ITERATIONS                          1201 ANSWERS
SEARCH TIME: 00.00.02

L6         1201 SEA SSS FUL L1

=> ANA L6 INKY 1-

L7          ANALYZE L6 1- INKY :    1167 TERMS

=> D

L7          ANALYZE L6 1- INKY :    1167 TERMS

TERM #   # OCC  # DOC  % DOC INKY
------ ------- ------ ------ ---------------
     1       3      3   0.25 DXMWUZNSJCZZLR-LOYRBTAXSA-N
     2       3      3   0.25 ZRZMGPQJQHXOHS-AFRLDPKMSA-N
     3       2      2   0.17 HCIYRJZEGWHZMT-PUDJQCGISA-N
     4       2      2   0.17 LMBFAGIMSUYTBN-MPZNNTNKSA-N
     5       2      2   0.17 MEAXCSPGSRHSOY-CNZUMQHHSA-N
     6       2      2   0.17 NUNXIPWJBFWITR-UPQQPLLJSA-N
     7       2      2   0.17 PZQLEZBFSMMAHY-LNHLIAMYSA-N
     8       2      2   0.17 QFEWDDBQTSSPCY-BZKBKYEGSA-N
     9       2      2   0.17 QHJMQPOXERDDJU-YYMGAGDFSA-N
    10       2      2   0.17 QSPKBORZTSNKNP-MMMVMMOASA-N
    11       2      2   0.17 RDSXDSVHWBEZKH-ZTHUPLLRSA-N
    12       2      2   0.17 RGMTULDDXQGTSL-PUDJQCGISA-N
    13       2      2   0.17 UARCSGORGPZIBZ-ZTHUPLLRSA-N
    14       2      2   0.17 YHMOPLGVQZVMEM-APIHUWKFSA-N
    15       2      2   0.17 YYIFRWLBTUAIOT-CNIKCXRFSA-N
```

Search is run in ReaxysFileSub.

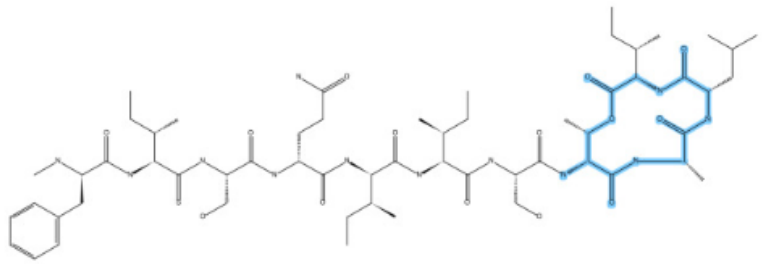
FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Example 3 (cont.)

Find unique substance records across REGISTRY, DCR & REAXYSFILESUB

```
=> S L6 NOT (L3 OR L5)


L9            QUE  TERMS FROM L3 WITH NO HITS:   469 TERMS


L11           QUE  TERMS FROM L5 WITH NO HITS:   130 TERMS
L12       645 L6 NOT (L8 OR L10)

=> D


L12    ANSWER 1 OF 645 REAXYSFILESU COPYRIGHT 2025 ELSEVIER INC. on STN.
AN     63390503   REAXYSFILESU
CN     Ile(sub)2(/sub)(sup)S(/sup)-Leu(sub)10(/sub)-teixobactin
MF     C58 H96 N12 O14 S
CMF    C58 H96 N12 O14 S
LSF    C58H96N12O14S
INCHI  OHQJALSJOZTACI-ATPZDXELSA-N
MW     1217.54
MARKREF.CNT 0
REC    1
ED     Entered STN: 13 Jun 2025
       Last updated on STN: 13 Jun 2025
```



L3 = REG set ANALYZEd for InChI Keys.
L5 = DCR set ANALYZEd for InChI Keys.

Unique record from ReaxysFileSub.

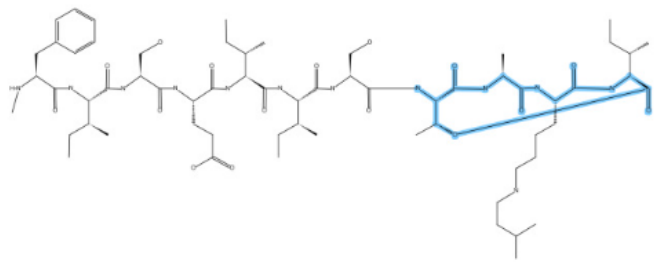
FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Example 3 (cont.)

Find unique substance records across REGISTRY, DCR & REAXYSFILESUB

```
=> S L4 NOT (L3 OR L7)

L14             QUE  TERMS FROM L3 WITH NO HITS:  1011 TERMS

L16             QUE  TERMS FROM L7 WITH NO HITS:  1158 TERMS
L17         128 L4 NOT (L13 OR L15)

=> D

L17    ANSWER 1 OF 128 DCR  COPYRIGHT 2025 CLARIVATE on STN.
AN     DCR-7664927   DCR
DCSE   7664927-1-0-0
```



```
MF     C63 H106 N12 O16
INCH   InChI=1S/C63H106N12O16/c1-14-35(7)48(72-56(83)44(64-13)31-41-23-19-18-20-
       24-41)59(86)69-45(32-76)57(84)68-43(26-27-47(78)79)55(82)71-50(37(9)16-
       3)61(88)73-49(36(8)15-2)60(87)70-46(33-77)58(85)75-52-40(12)91-
       63(90)51(38(10)17-4)74-54(81)42(67-53(80)39(11)66-62(52)89)25-21-22-29-
       65-30-28-34(5)6/h18-20,23-24,34-40,42-46,48-52,64-65,76-77H,14-17,21-
       22,25-33H2,1-13H3,(H,66,89)(H,67,80)(H,68,84)(H,69,86)(H,70,87)(H,71,82)
       (H,72,83)(H,73,88)(H,74,81)(H,75,85)(H,78,79)/t35-,36-,37-,38-,39-
       ,40+,42-,43-,44-,45-,46-,48-,49-,50-,51-,52-/m0/s1
INKY   NGFYWFKKUYEPFF-RDZWRUAUSA-N
ED     Entered STN: 10 Jan 2025
       Last updated on STN: 10 Jan 2025
       Update DWPI Cross Ref.: 10 Jan 2025
```

L3 = REG set ANALYZEd for InChI Keys.
L7 = ReaxysFileSub set ANALYZEd for InChI Keys.

Unique record from DCR.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

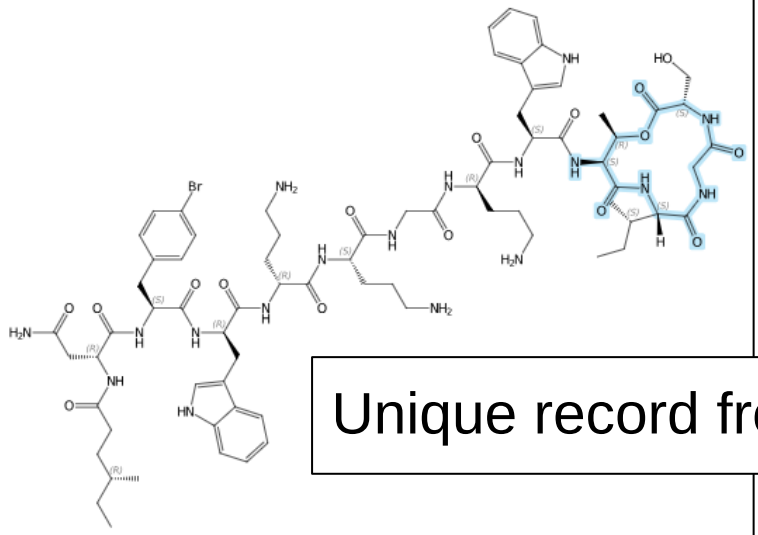
CAS
A division of the American Chemical Society

Example 3 (cont.)

Find unique substance records across REGISTRY, DCR & REAXYSFILESUB

```
=> S L2 NOT (L5 OR L7)


L19             QUE  TERMS FROM L5 WITH NO HITS:    128 TERMS


L21             QUE  TERMS FROM L7 WITH NO HITS:    614 TERMS
L22         471 L2 NOT (L18 OR L20)

=> D


L22  ANSWER 1 OF 471  REGISTRY  COPYRIGHT 2025 ACS on STN
RN   3082422-60-1  REGISTRY
ED   Entered STN:  09 May 2025
CN   INDEX NAME NOT YET ASSIGNED
FS   PROTEIN SEQUENCE; STEREOSEARCH
MF   C74 H105 Br N18 O16
```

```
INCH InChI=1S/C74H105BrN18O16/c1-6-40(3)22-27-60(96)85-57(34-59(79)95)70(104)89-
     54(31-43-23-25-46(75)26-24-43)68(102)91-55(32-44-35-80-49-17-10-8-15-
     47(44)49)69(103)88-53(21-14-30-78)67(101)87-51(19-12-28-76)65(99)82-37-
     61(97)84-52(20-13-29-77)66(100)90-56(33-45-36-81-50-18-11-9-16-
     48(45)50)71(105)93-64-42(5)109-74(108)58(39-94)86-62(98)38-83-
     72(106)63(41(4)7-2)92-73(64)107/h8-11,15-18,23-26,35-36,40-42,51-58,63-
     64,80-81,94H,6-7,12-14,19-22,27-34,37
     5H3,(H2,79,95)(H,82,99)(H,83,106)(H,8
     8,103)(H,89,104)(H,90,100)(H,91,102)(
     ,51+,52-,53-,54+,55-,56+,57-,58+,63+,
INKY SDGMFKCJCIHVNW-SEEVIRBPSA-N
SR   CA
LC   STN Files:   CAPLUS, TOXCENTER

**RELATED SEQUENCES AVAILABLE WITH SEQLINK
```



Absolute stereochemistry shown

```
1 REFERENCES IN FILE CAPLUS (1907 TO DATE)
```

L5 = DCR set ANALYZEd for InChI Keys.
L7 = ReaxysFileSub set ANALYZEd for InChI Keys.

Unique record from CAS REGISTRY.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

CONSIDERATIONS

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Considerations

Identification of same substances

- Works well for the majority of substances
- Restrictions are based on
  - limitations of the InChI system
  - different indexing by database producers

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

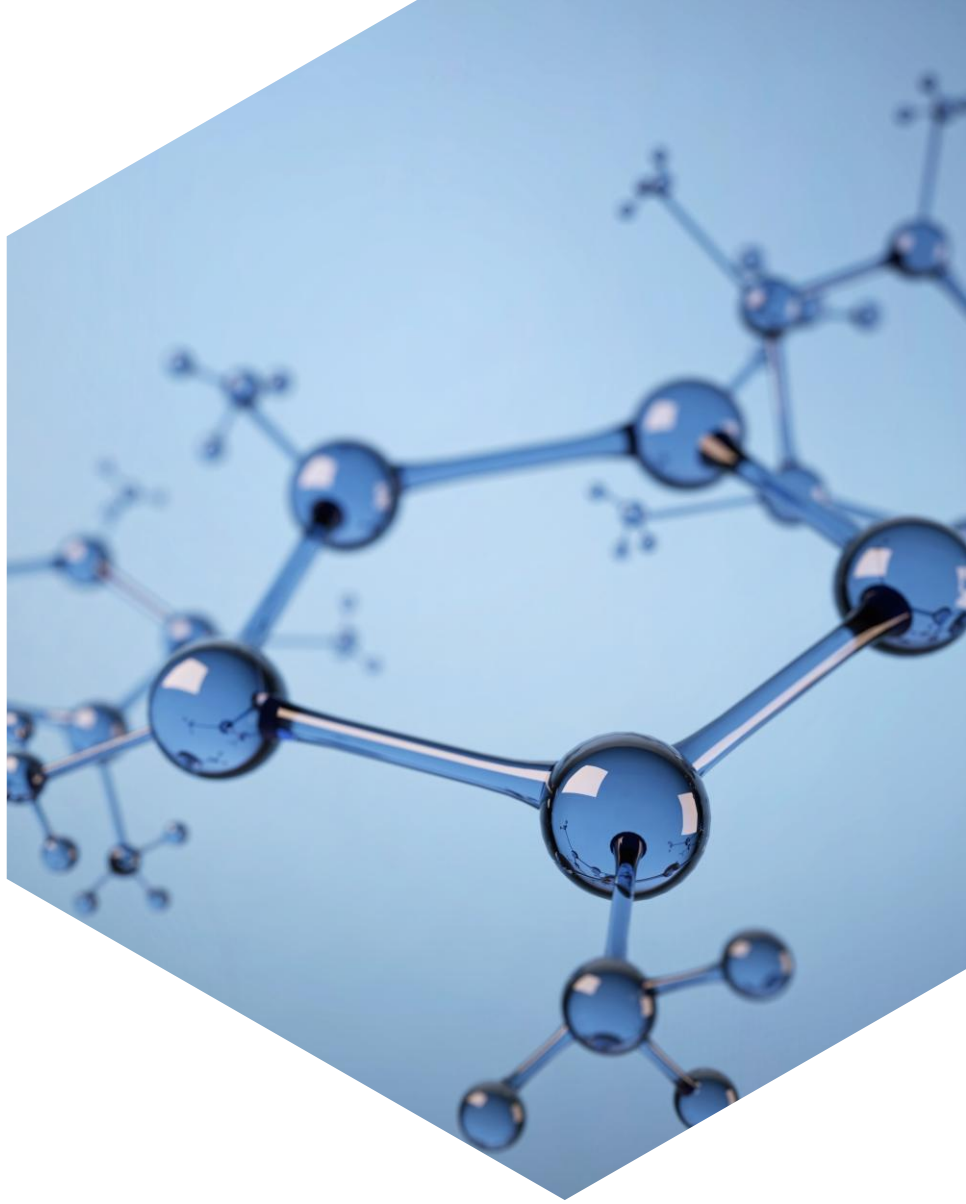
Considerations

Limitation of the InChI system

- Technical limitation of 1023 atoms; not suitable for very large compounds
- InChI needs enhancements for less well defined chemical entities; e.g.,
  - Variable structures
  - Polymers
  - Complex organometallics
  - Biologicals
  - Nanomaterials
  - Mixtures

Summary

- InChI Keys are unique substance identifiers that can be used across multiple databases
- Implementation of InChI in STNext databases
  - CAS REGISTRY, DCR, ReaxysFileSub
- Consider source of information, difference in indexing systems between database producers

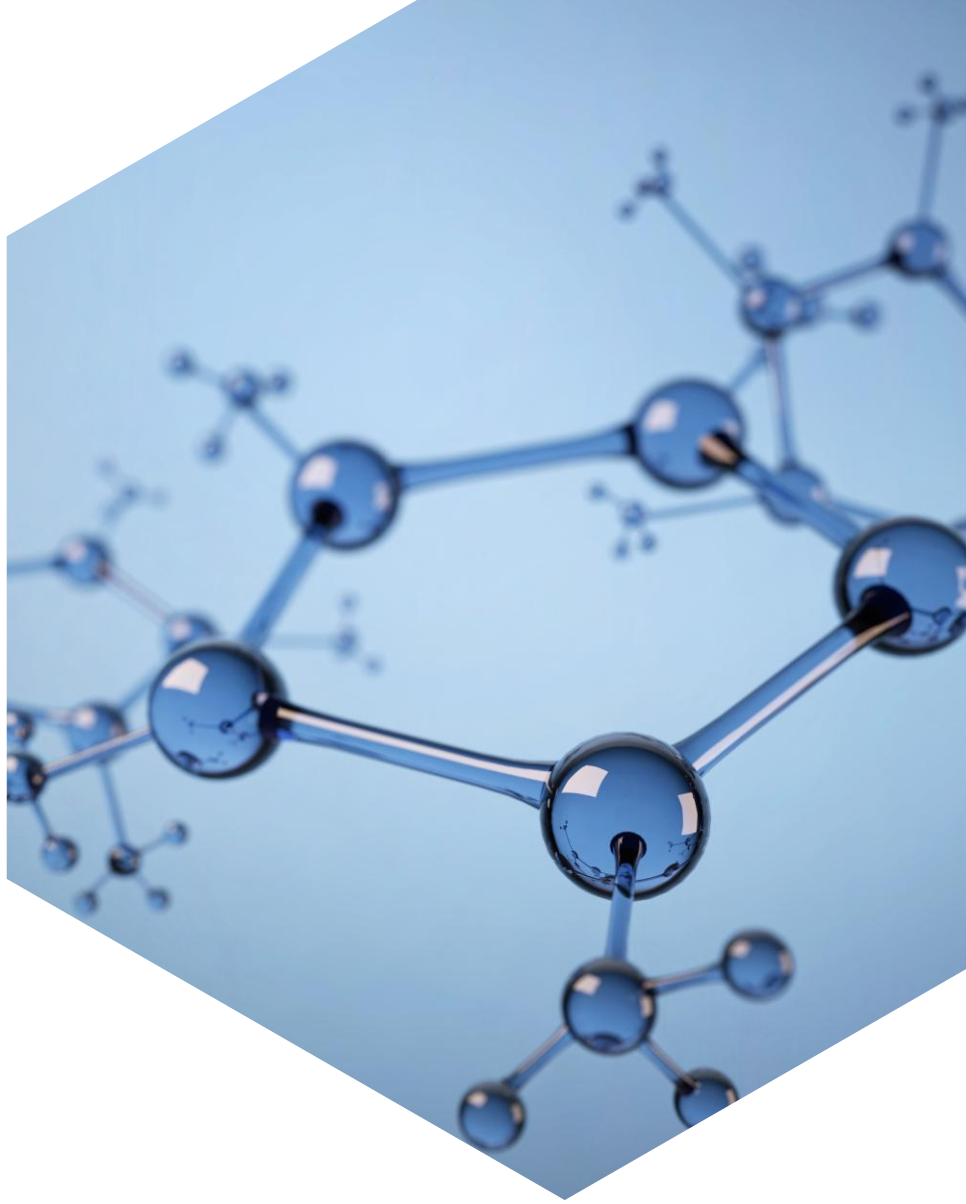


FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Special thanks

- Claudia Dauvermann-Gotsche
- Paul Peters



FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Thank you

Jim Brown
Senior Product Specialist IP
jim.brown@fiz-k.com

CONTACT

CAS
help@cas.org
cas.org

EMEA Help
EMEAhelp@cas.org

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society