BEST PRACTICES FOR ADVANCED STRUCTURE SEARCHING IN CAS STNEXT

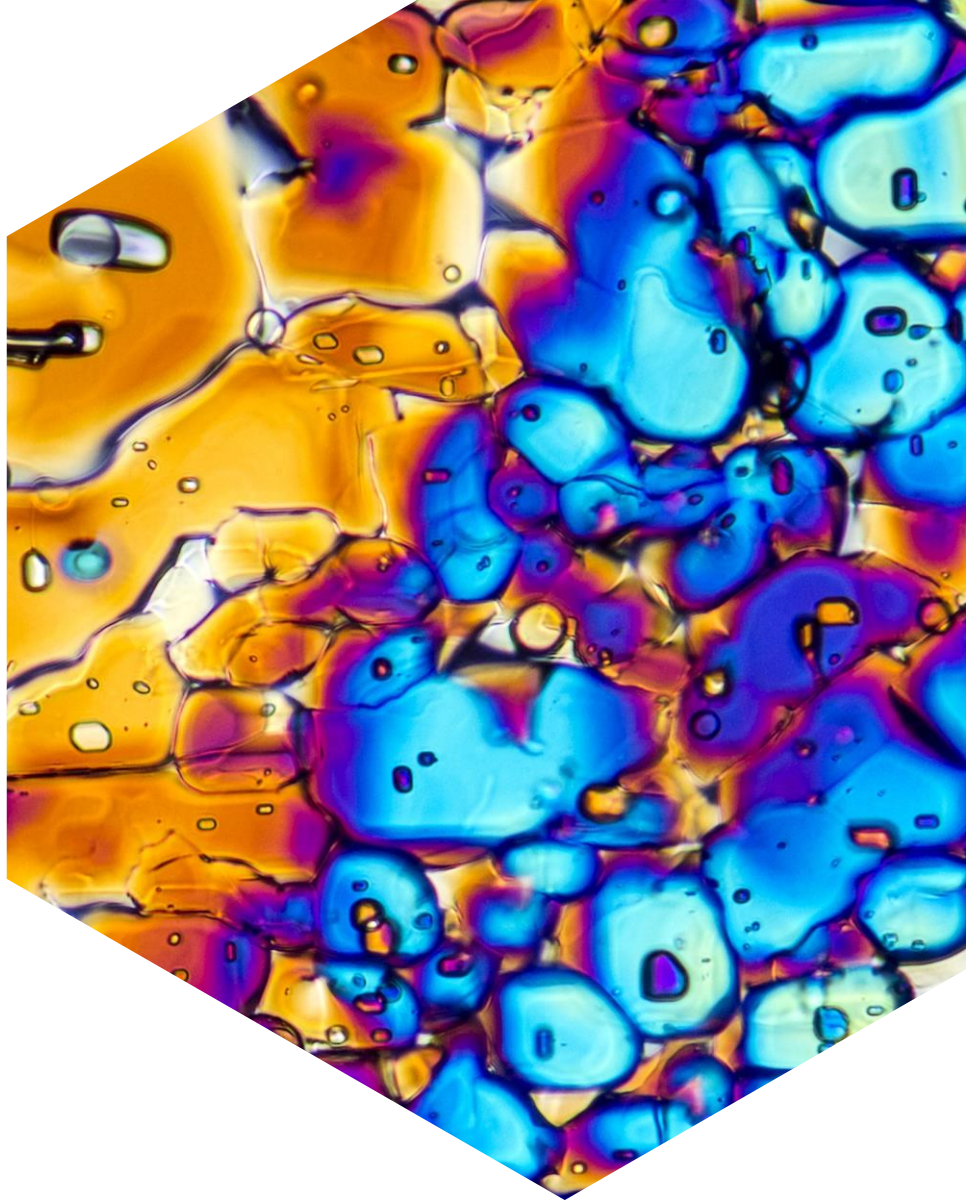
Paul Peters – CAS, Jim Brown – FIZ Karlsruhe

CAS
A division of the
American Chemical Society

Agenda

Strategies for complex chemical structures

- How to search for macrocyclic compounds with variable ring sizes across multiple databases
- Deduplicate patent results
- When to include stereochemistry in Derwent databases
- How to search for radicals or ions

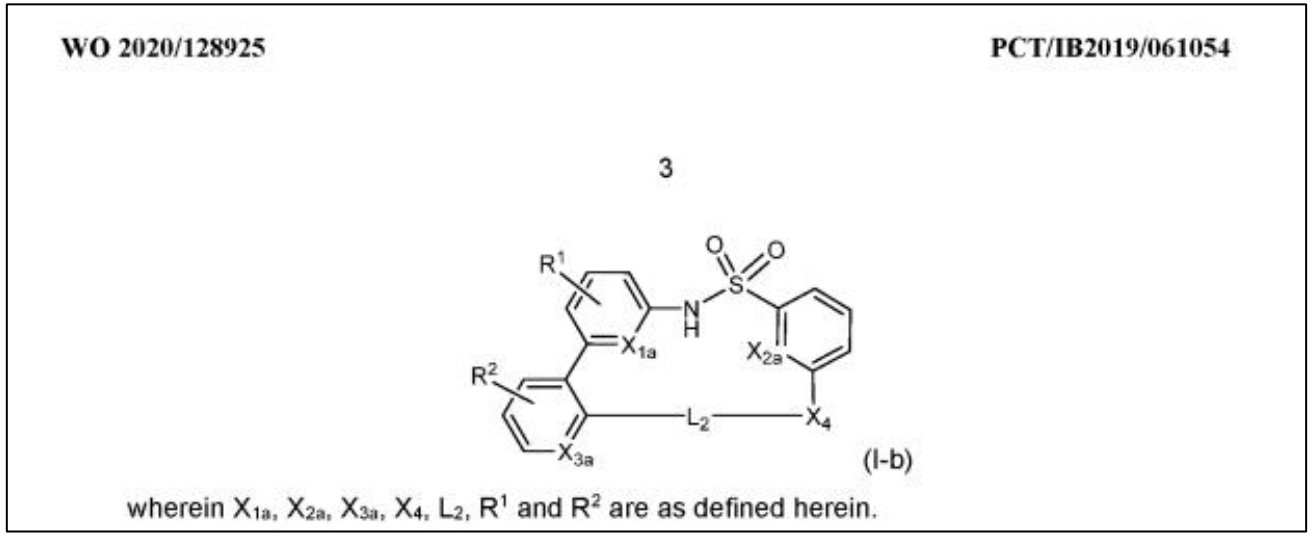


CAS
A division of the American Chemical Society

Searching Macrocyclic compounds

Challenges of adding variability into the ring structure

- Options for Ring System Data in CAS REGISTRY, but not in MARPAT
- Watch out for *structure too large*
- Use Non-H attachment to indicate further ring nodes
- Keeping ring with X3a completely open in terms of shape and elements

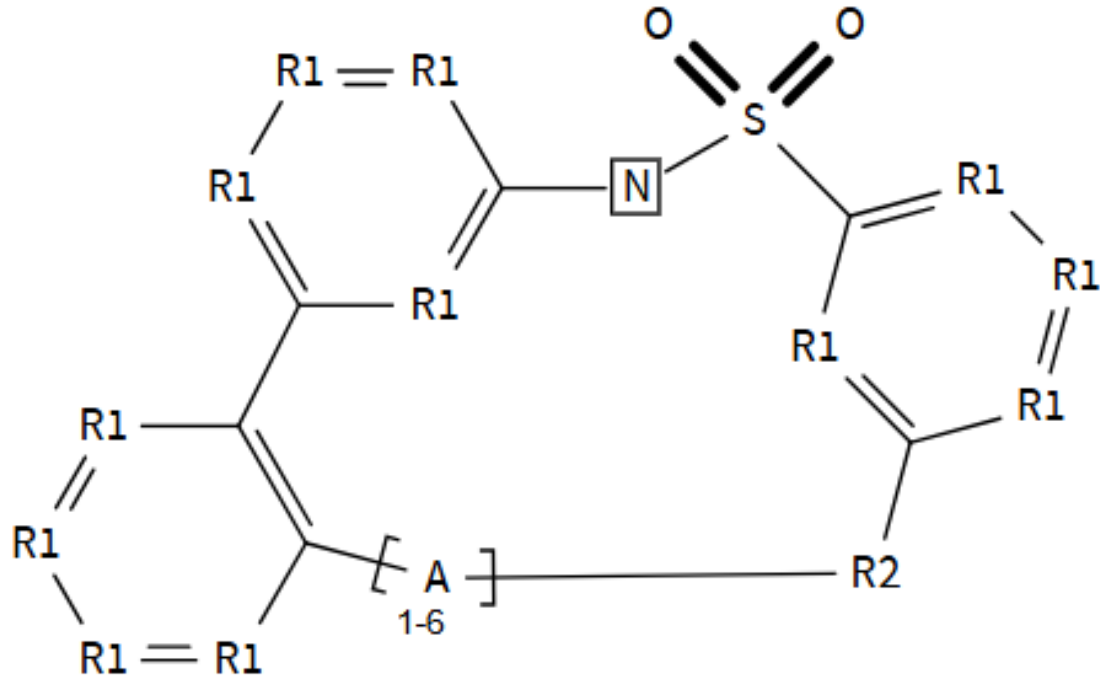


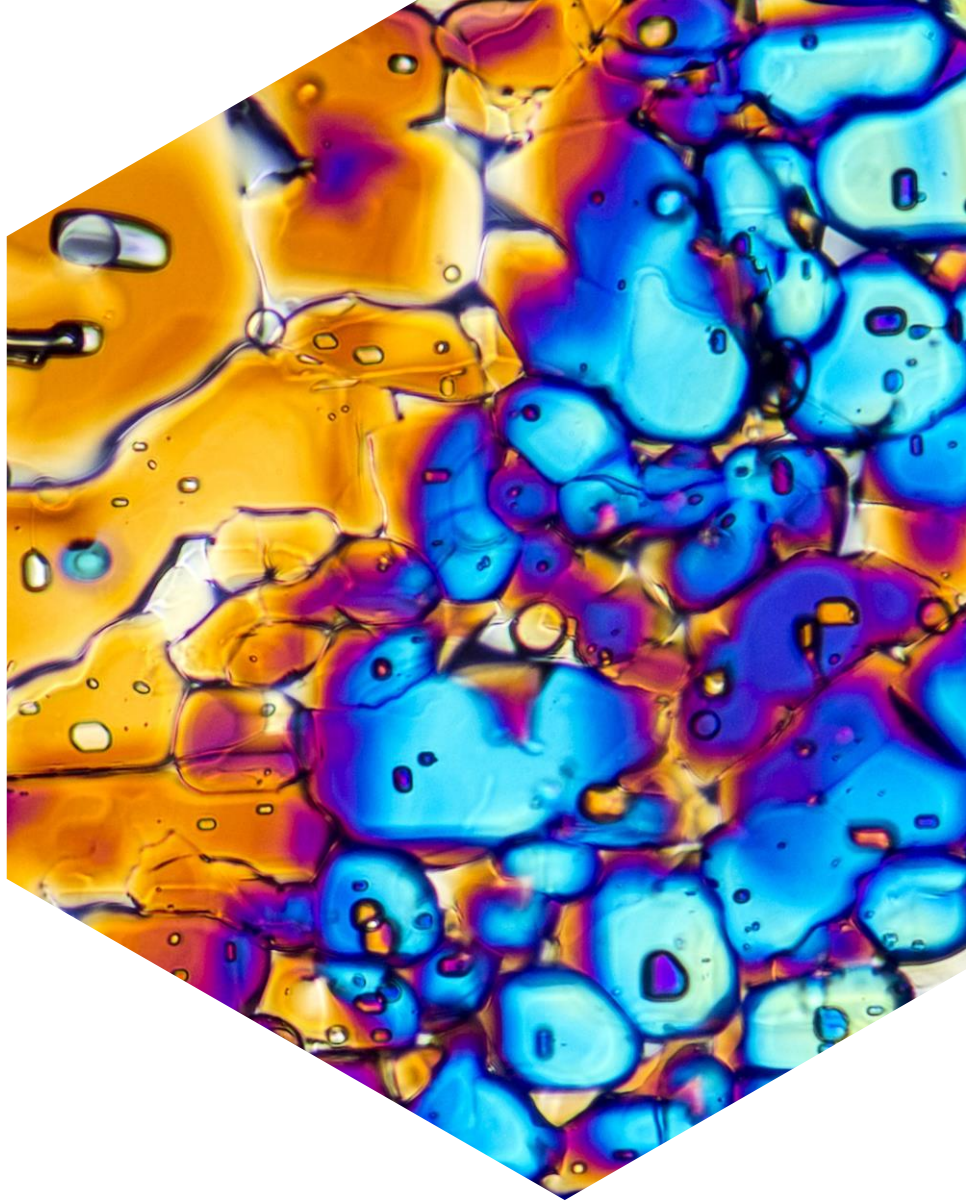
CAS
A division of the American Chemical Society

Trying to draw as defined

Retrieves 148 substances in REGISTRY

But MARPAT gives *Str too large* message



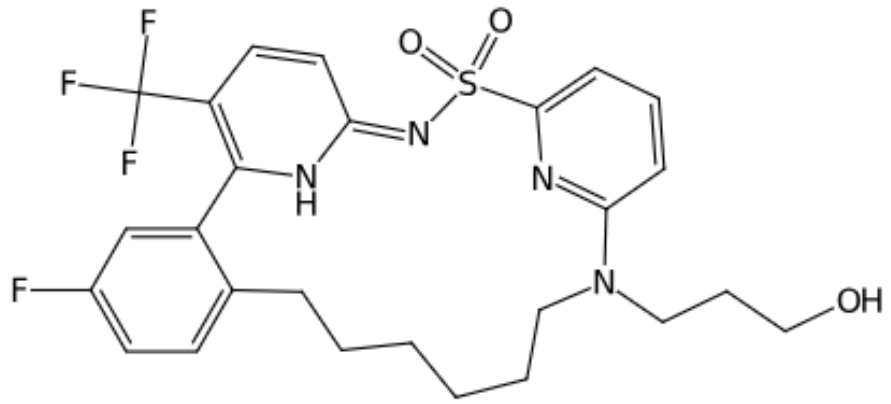


CAS
A division of the American Chemical Society

Ring system data for macrocycles

The ring size of the smallest ring (/SZS) shows the value 17

RN 2446430-34-6 REGISTRY
IN INDEX NAME NOT YET ASSIGNED
MF C26 H28 F4 N4 O3 S



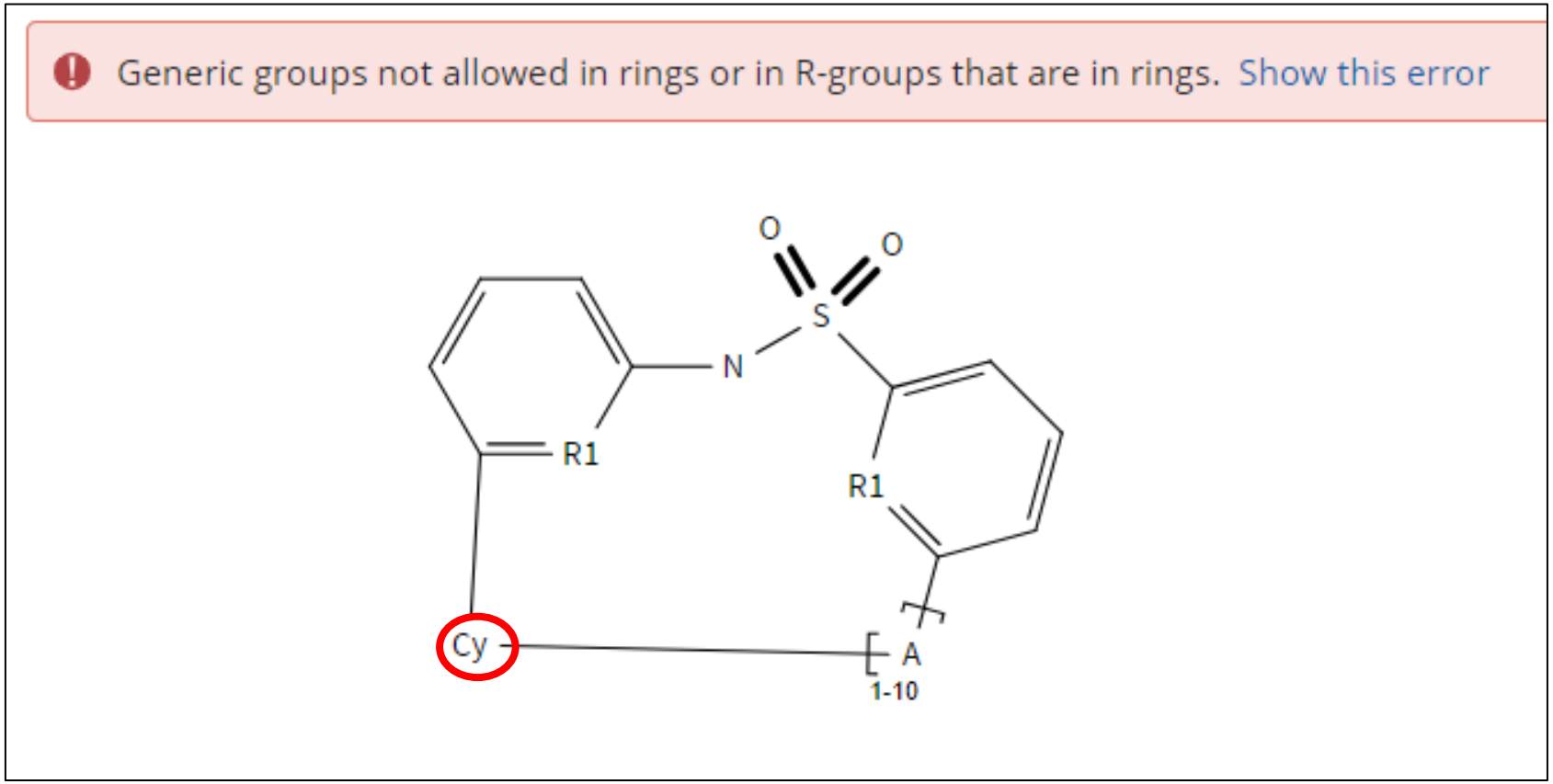
Ring System Data

Elemental Analysis EA	Elemental Sequence ES	Size of the Rings SZ	Ring System Formula RF	Ring Identifier RID	RID Occurrence Count
C5N-C5N-C6-C12N4S	NC5-NC5-C6-NSCNCNC9NC	6-6-6-17	C22N4S	206213.1.2	1

CAS
A division of the American Chemical Society

Embedded generic node error

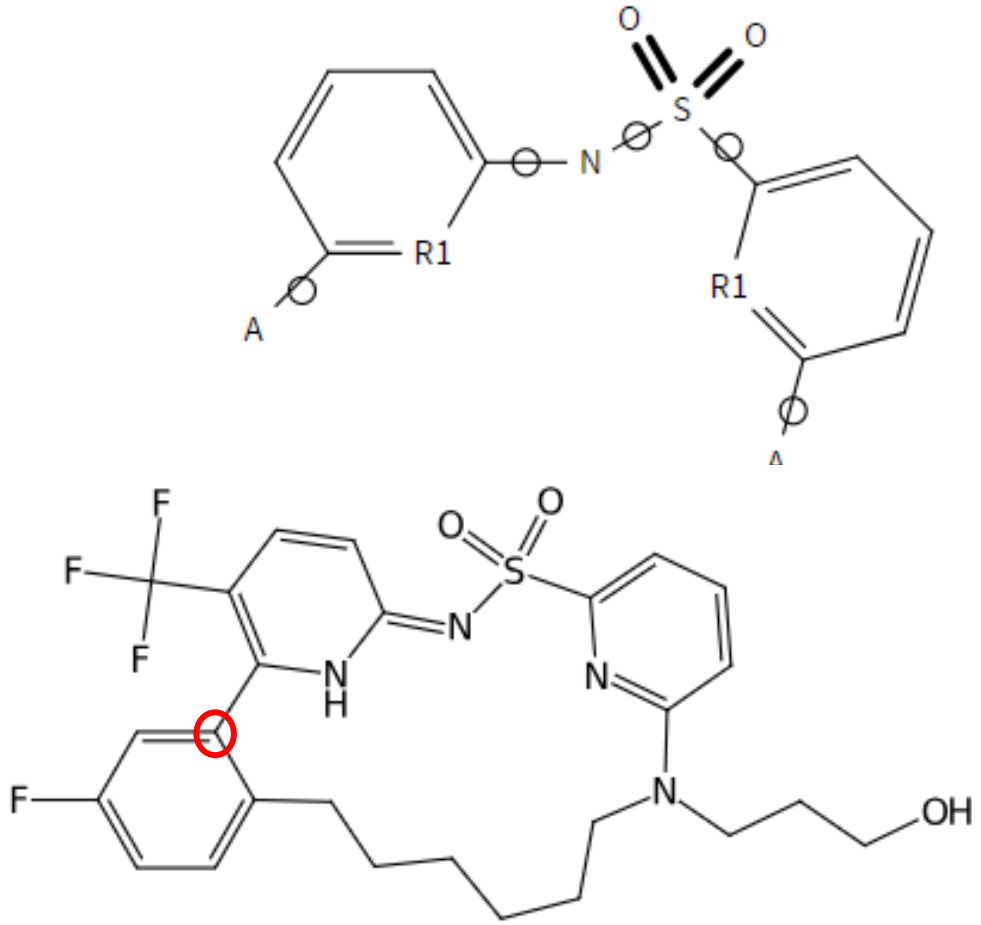
Cannot define a generic ring inside a ring



CAS
A division of the American Chemical Society

Approach in REGISTRY

- Indicating all ring bonds
- R1=C/N
- A is any element except H
- Smallest ring size should be at least 13
- How to indicate that the A node on the left is the start of any ring?

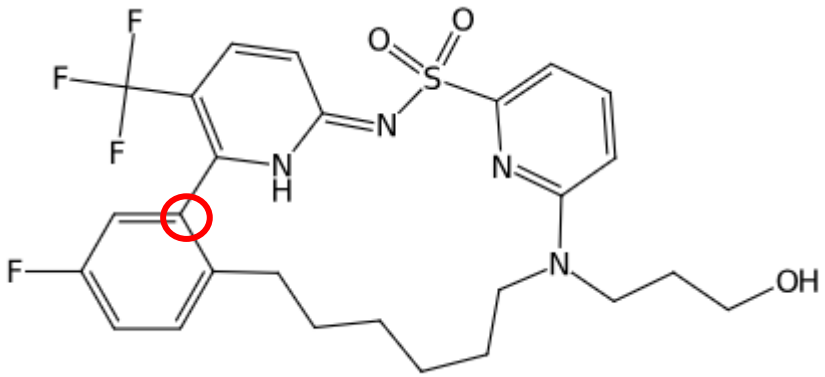


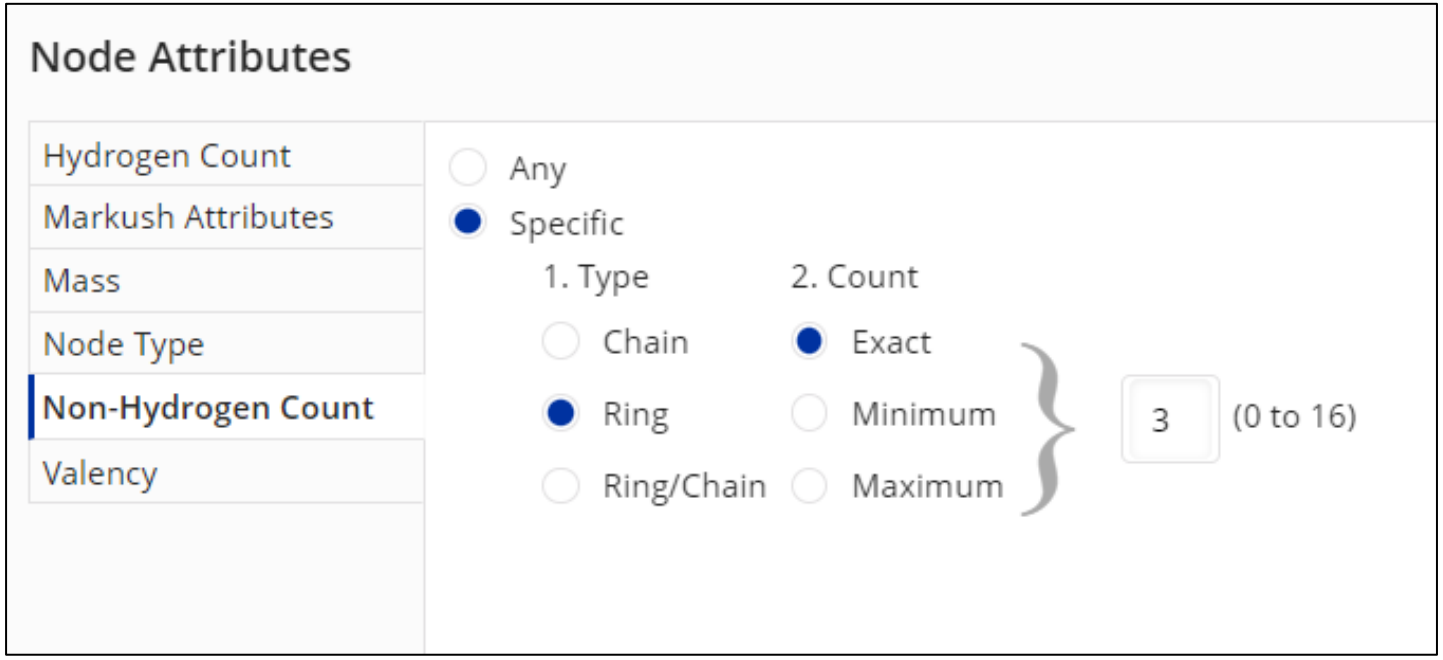
CAS
A division of the American Chemical Society

Non-H attachment solves this issue

Indicates the number of connections at this location

- This atom is connected to 3 other ring atoms
- By specifying that that A must have a non-H attachment of Exact Ring 3, there must be a new ring starting at that position





CAS
A division of the American Chemical Society

Running the search in CAS REGISTRY

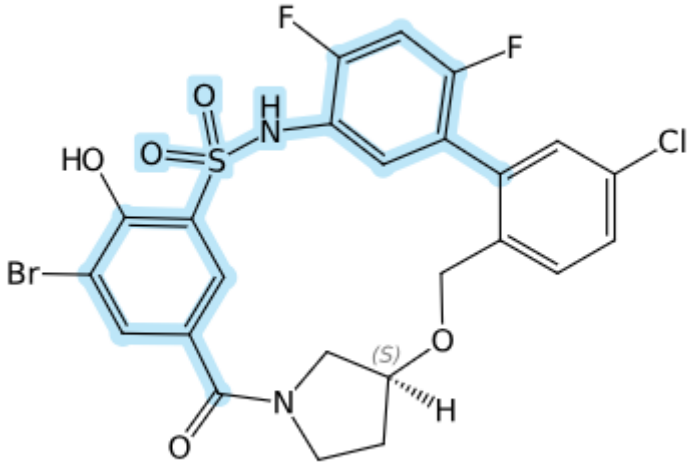
Most of the retrieved substances already have larger rings

```
=> s l5 sss full

FULL SEARCH INITIATED 02:30:47
100.0% PROCESSED     112596 ITERATIONS                          549 ANSWERS
SEARCH TIME: 00.00.01

L6           549 SEA SSS FUL L5

=> s l6 and szs>=13

        1493610 SZS>=13
L7          539 L6 AND SZS>=13
```

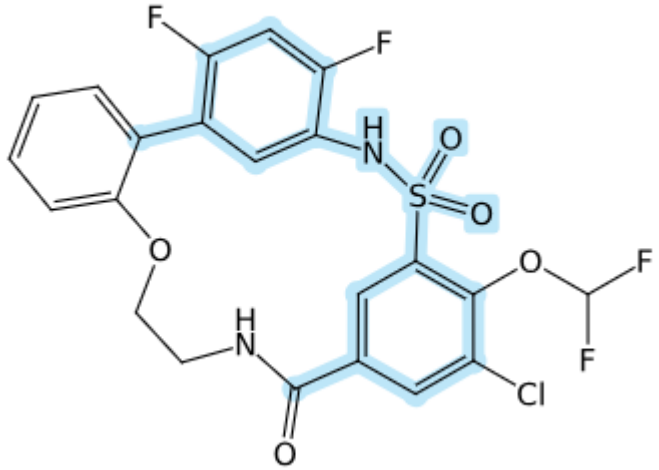

CAS
A division of the American Chemical Society

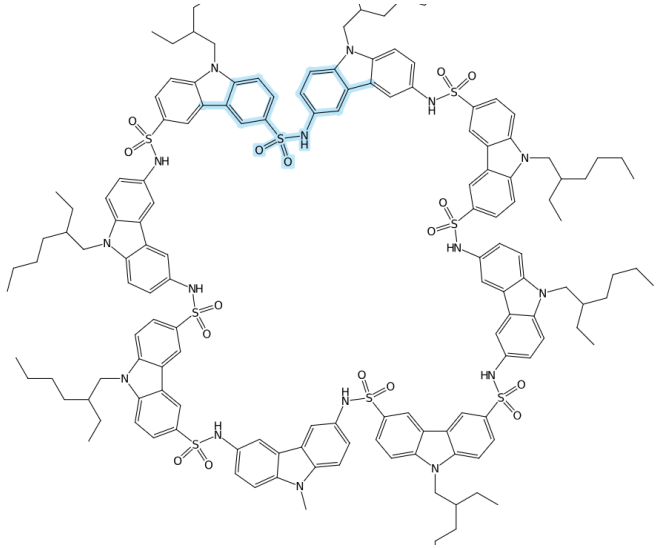
Examples of retrieved exact structures

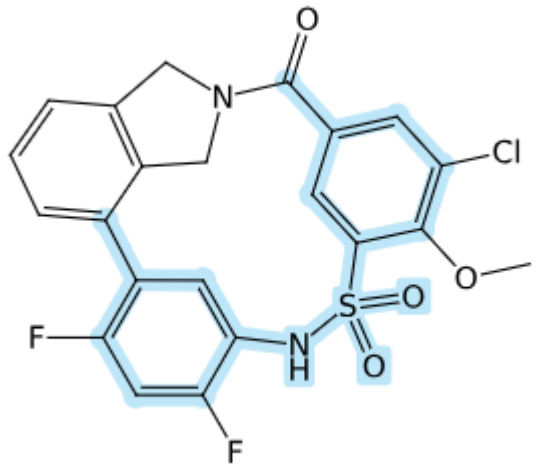
Largest ring system had 64 atoms, could limit to the number of rings in the ring system



Absolute stereochemistry shown







L6	549	S L5 SSS FULL
L7	539	S L6 AND SZS>=13
L8	528	S L7 NOT NRRS>6

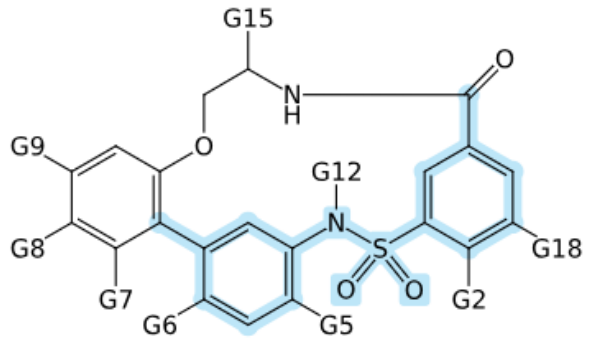
CAS
A division of the American Chemical Society

Search works well in MARPAT

Non-H attachment is well-recognized

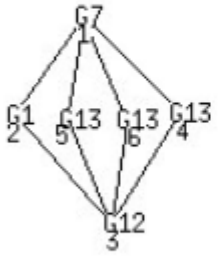
```
=> s l5 sss ful

FULL SEARCH INITIATED 02:39:47
100.0% PROCESSED       3911 ITERATIONS                    5 ANSWERS
SEARCH TIME: 00.00.01

L9             5 SEA SSS FUL L5
```

MSTR 1A Assembled



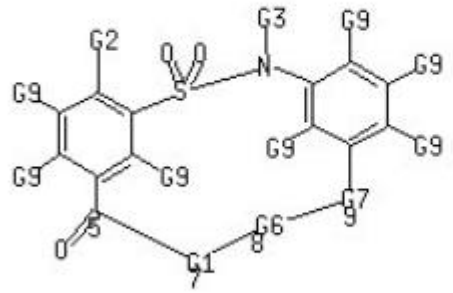
Patent location: claim 1

MSTR 1A



G1 = 36-1 33-3

MSTR 1A

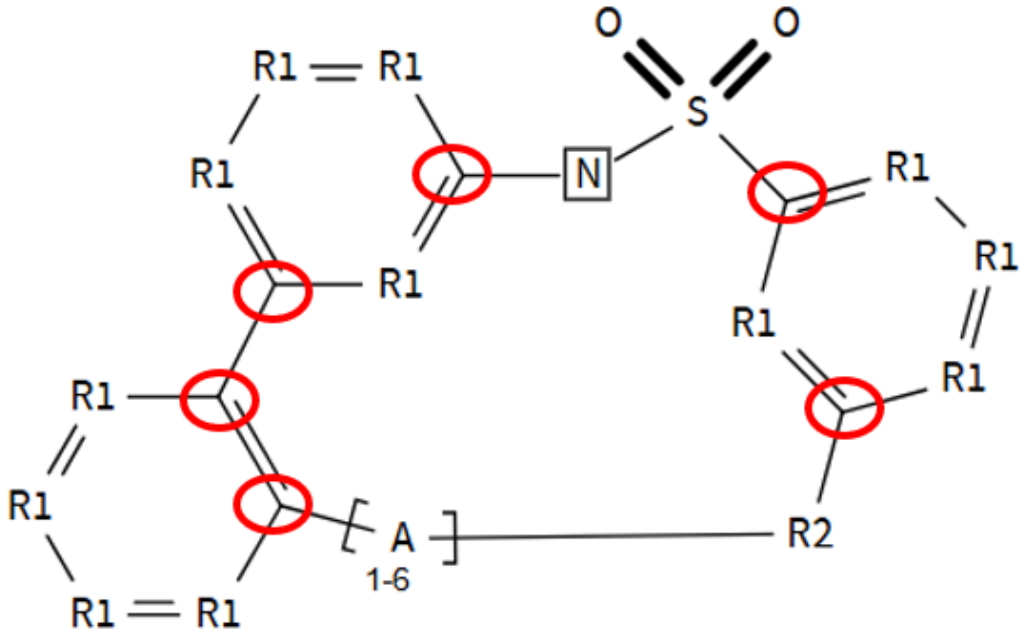


G7 = o-C6H4 (opt. substd. by G15)

CAS
A division of the American Chemical Society

Alternative strategy: Positions marked have Exact 3 Ring connections

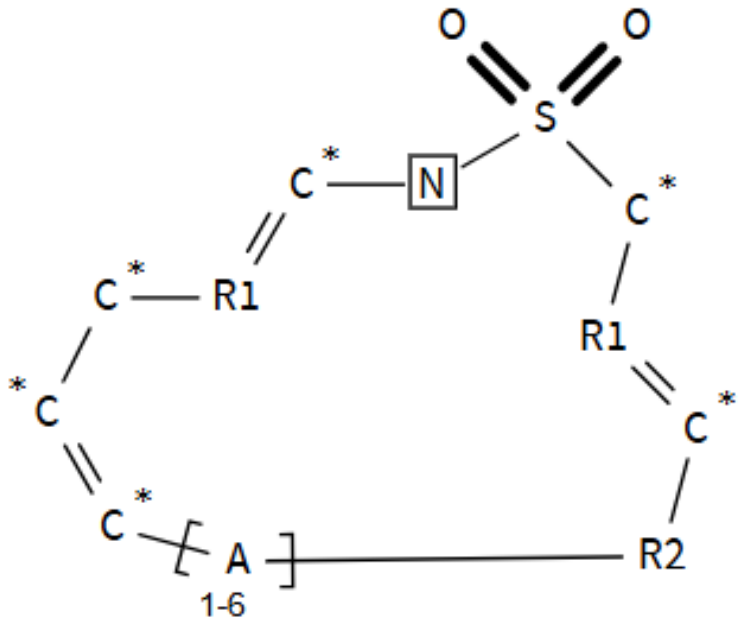
Remove the R1 group nodes completely that are not part of the inner ring

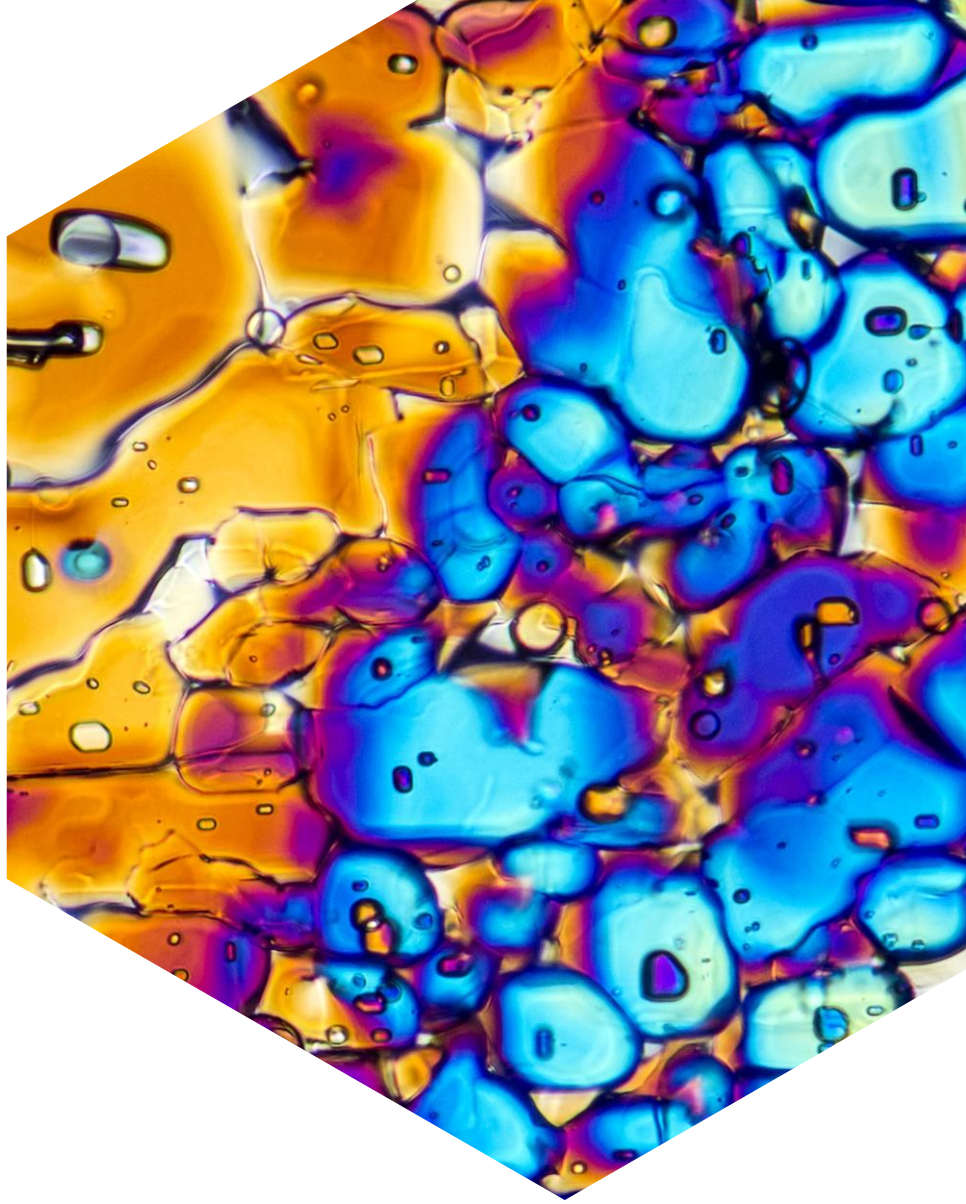


CAS
A division of the American Chemical Society

Inner ring structure search

- Carbon nodes with * have non-H attachment of Exact 3 Ring
- Gives the same 148 substance records in REGISTRY, but still too large for MARPAT

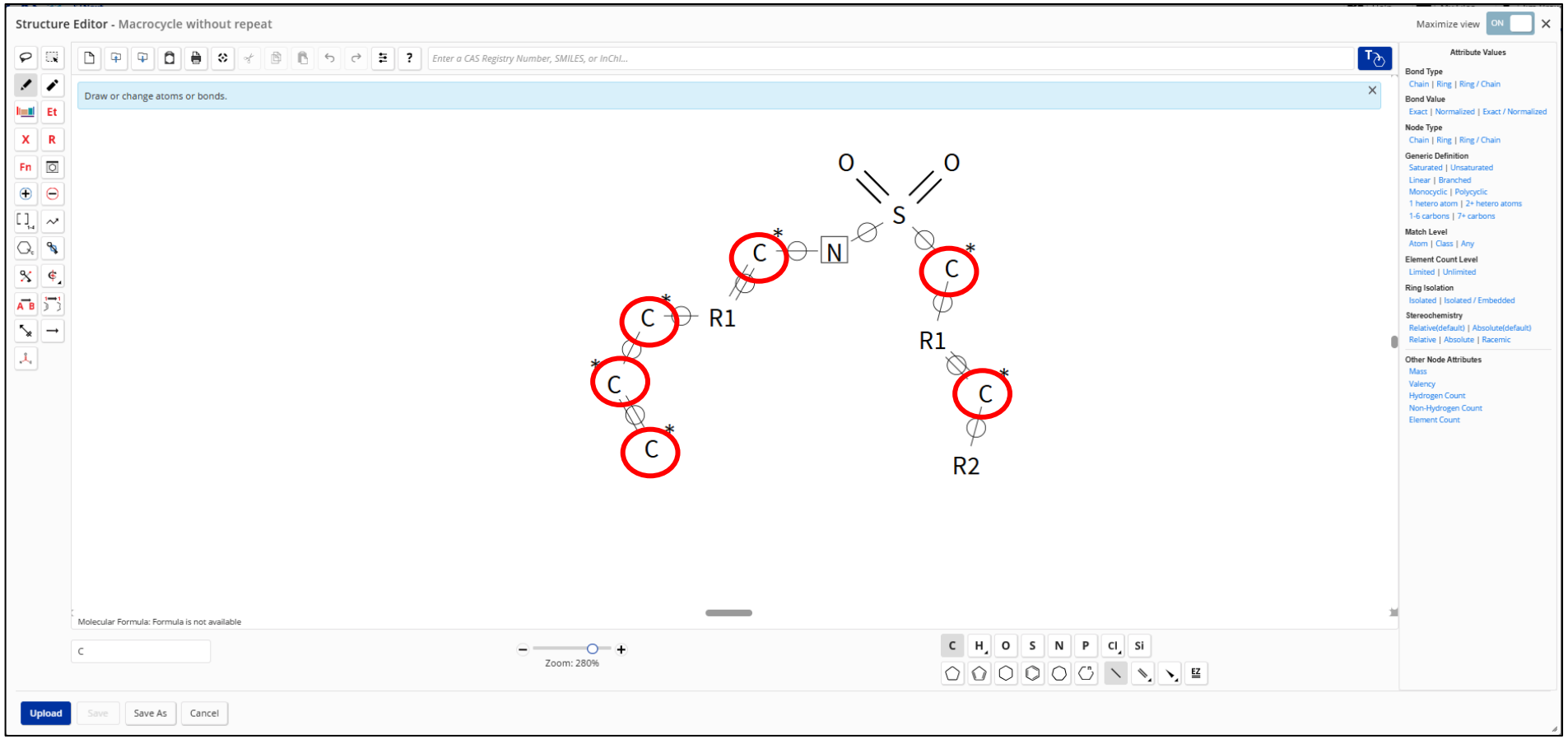




CAS
A division of the American Chemical Society

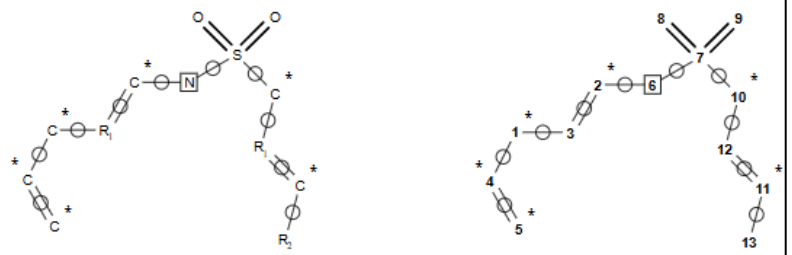
The issue is the repeating group

Let’s just remove it, but leave the bonds “ring” and the non-H attachments in place.



The issue is the repeating group

Uploading structure file: Macrocycle without repeat



R-Group Definitions
R1: C,N
R2: N,O
Node Attributes
Ring Nodes : 1 2 4 5 6 7 10 11
Chain Nodes : 8 9
Non H Count
Ring, Exact (3) : 1 2 4 5 10 11
Bond Attributes
Ring Bonds : 1-4 2-3 2-6 3-1 4-5 6-7 7-10 11-12 11-13 12-10
Chain Bonds : 7-8 7-9
Exact/Normalized Bonds : 1-4 2-3 2-6 3-1 4-5 6-7 7-8 7-9 7-10 11-12 11-13 12-10
Markush Attributes
Match Level (ATOM) : 1 2 4 5 6 7 10 11
Match Level (CLASS) : 8 9
Element Count Level (LIMITED) : 1 2 4 5 6 7 8 9 10 11

L1 STRUCTURE UPLOADED

=> **S L1 SSS FULL**

FULL SEARCH INITIATED 16:46:40
FULL SCREEN SEARCH COMPLETED - 387537 TO ITERATE

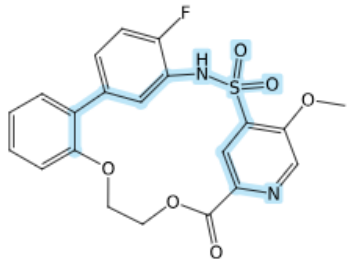
100.0% PROCESSED 387537 ITERATIONS 171 ANSWERS
SEARCH TIME: 00.00.01

L2 171 SEA SSS FUL L1

=> **D SQIDE**

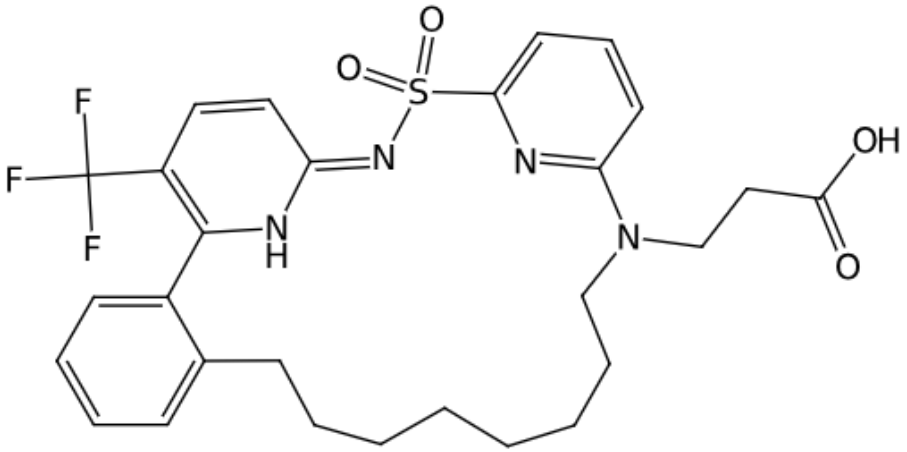
L2 ANSWER 1 OF 171 REGISTRY COPYRIGHT 2025 ACS on STN
RN **3009138-74-0** REGISTRY
ED Entered STN: 30 Nov 2023
CN INDEX NAME NOT YET ASSIGNED
MF C21 H17 F N2 O6 S
SR CA
LC STN Files: CA, CAPLUS, TOXCENTER
DT.CA CAplus document type: Patent

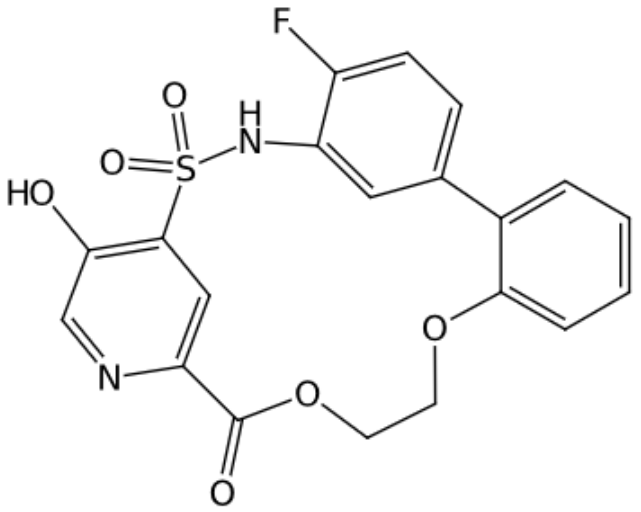
RL.P Roles from patents: PREP (Preparation); RACT (Reactant or reagent)

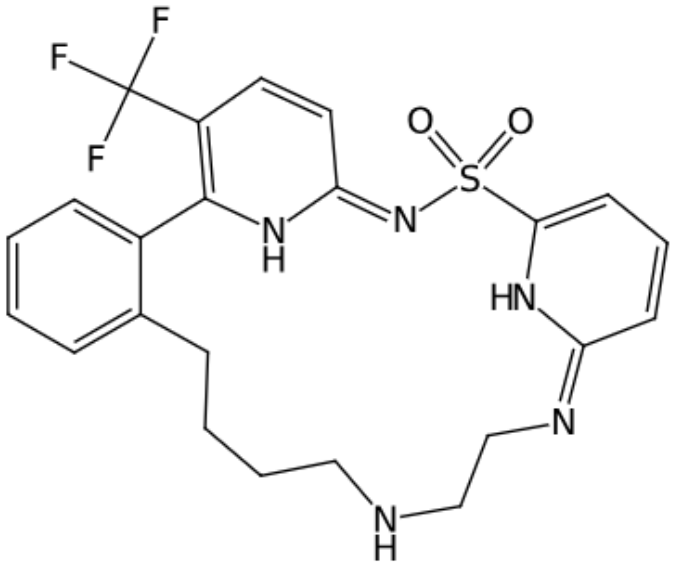


CAS
A division of the American Chemical Society

Some additional results for this query



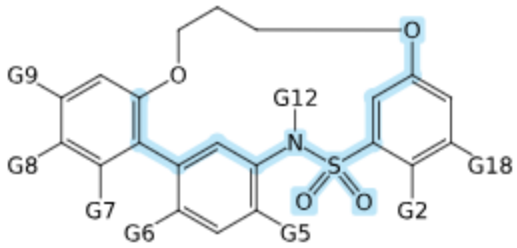




• x HCl

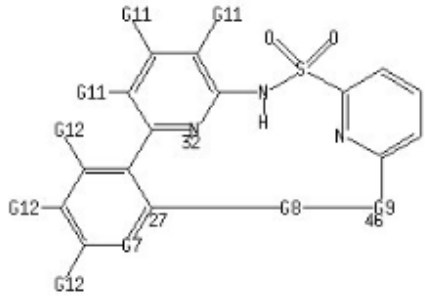
The search also runs in MARPAT

```
=> S L1 SSS FULL

FULL SEARCH INITIATED 16:57:48
FULL SCREEN SEARCH COMPLETED -     3089 TO ITERATE

100.0% PROCESSED       3089 ITERATIONS                          2 ANSWERS
SEARCH TIME: 00.00.01

L3            2 SEA SSS FUL L1

=> D FQHIT


L3   ANSWER 1 OF 2  MARPAT  COPYRIGHT 2025 ACS on STN

  MSTR 1B Assembled
```




```
Patent location:          claim 1
Note:                     or pharmaceutically acceptable salts
Note:                     substitution is restricted
Note:                     additional derivatization also claimed
Stereochemistry:          or stereoisomers
```

```
=> D FQHIT 2


L3   ANSWER 2 OF 2  MARPAT  COPYRIGHT 2025 ACS on STN

  MSTR 1


 G1

G1      = 32
```



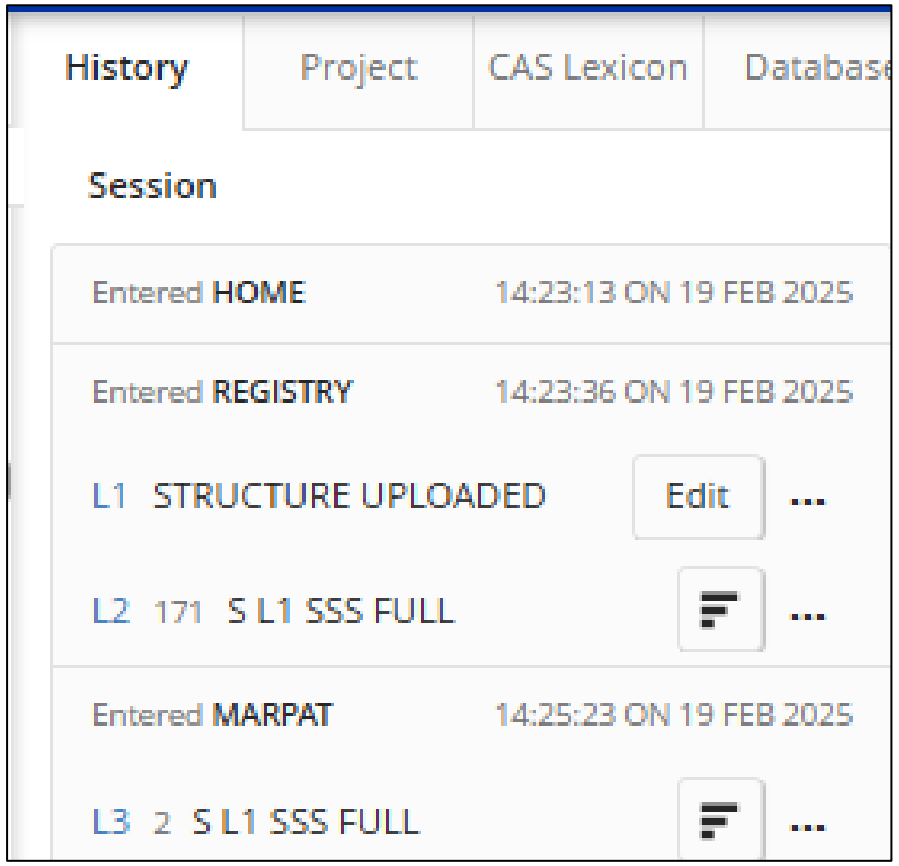

```
G9      = 56
```


```
Patent location:          claim 1
Note:                     additional derivatization also claimed
Note:                     or pharmaceutically acceptable salts
```

CAS
A division of the American Chemical Society

Bringing results into HCAplus

Limit to patent records




```
=> FILE HCAPLUS; S L2; S L3

L4              4 L2


L5              2 L3

=> S (L4 OR L5) AND P/DT

      21193982 P/DT
L6            3 (L4 OR L5) AND P/DT
```


CAS
A division of the American Chemical Society

Display results for this query

File HCAPLUS

=> D BIB HITSTR 1-3

L6 ANSWER 1 OF 3 HCAPLUS COPYRIGHT 2025 ACS on STN
PatentPak PDF | PatentPak PDF+ | PatentPak Interactive
AN 2023:2302644 HCAPLUS Full-text
DN 185:11477
TI Preparation of macrocyclic compounds as ATP citrate lyase inhibitors
IN Pinkosky, Stephen Lawrence; Davenport, Adam James; Gleave, Laura Jane; Southey, Michelle Wing Yan; Ursinyova, Nina Connelly; Walker, Patrick Ross; Yarnold, Christopher John
PA Esperion Therapeutics, Inc., USA
UO ESPERION THERAPEUTICS, INC.
UOS ESPERION THERAPEUTICS, INC.
SO PCT Int. Appl., 303pp.
CODEN: PIXXD2
DT Patent
LA English
FAN.CNT 1
PPPI

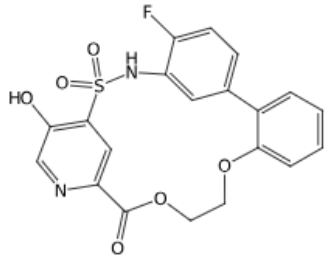
PATENT NO.	KIND	DATE	LANGUAGE	PatentPak
WO 2023215220	A1	20231109	English	PDF | PDF+ | Interactive
AU 2023264491	A1	20241031	English	PDF
CN 119137107	A	20241213	Chinese	PDF
KR 2025004873	A	20250108	Korean	PDF

PI

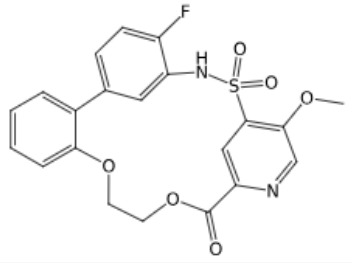
PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2023215220	A1	20231109	WO 2023-US20546	20230501
AU 2023264491	A1	20241031	AU 2023-264491	20230501
SG 11202407284	A	20241129	SG 2024-11202407284	20230501
CN 119137107	A	20241213	CN 2023-80038044	20230501
KR 2025004873	A	20250108	KR 2024-7039451	20230501
IN 202417089629	A	20241220	IN 2024-17089629	20241119

OS MARPAT 185:11477

IT 3009137-41-8 P
RL: PAC (Pharmacological activity); SPN (Synthetic preparation); THU (Therapeutic use); BIOL (Biological study); PREP (Preparation); USES (Uses)
(prepn. of macrocyclic compds. as ATP citrate lyase inhibitors)
RN 3009137-41-8 HCAPLUS
CN INDEX NAME NOT YET ASSIGNED



IT 3009138-74-0 P
RL: RCT (Reactant); SPN (Synthetic preparation); PREP (Preparation); RACT (Reactant or reagent)
(prepn. of macrocyclic compds. as ATP citrate lyase inhibitors)
RN 3009138-74-0 HCAPLUS
CN INDEX NAME NOT YET ASSIGNED



CAS
A division of the American Chemical Society

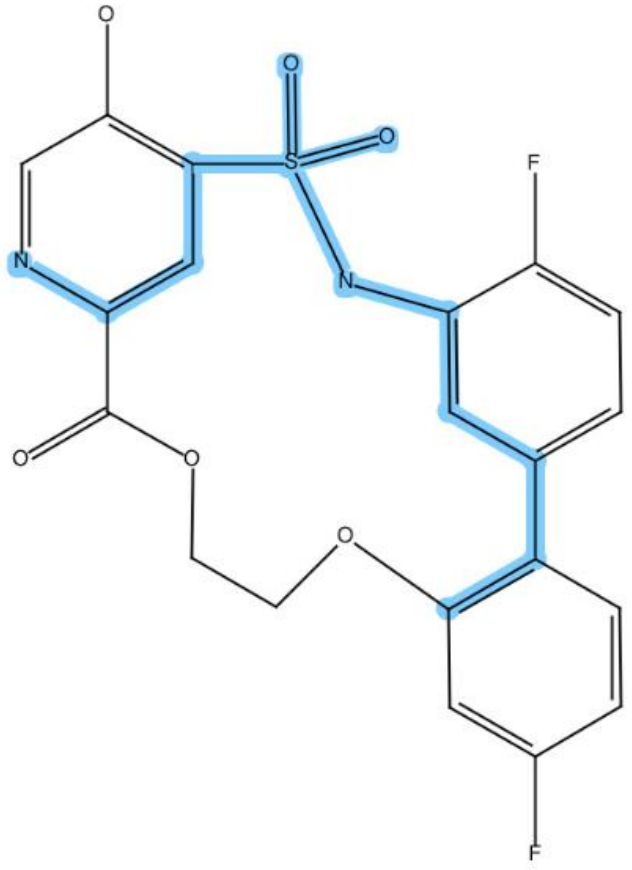
Repeat strategy on Derwent databases

Structure works across all structure searchable databases

```
=> FILE DCR; S L1 SSS FULL
```

```
L7                1 SEA SSS FUL L1

=> D
```

L7 ANSWER 1 OF 1 DCR COPYRIGHT 2025 CLARIVATE on STN.
AN DCR-6815203 DCR
DCSE 6815203-0-0-0



CAS
A division of the American Chemical Society

Repeat strategy in Derwent databases

Structure works across all structure searchable databases

CAS
A division of the American Chemical Society

Display results for this query

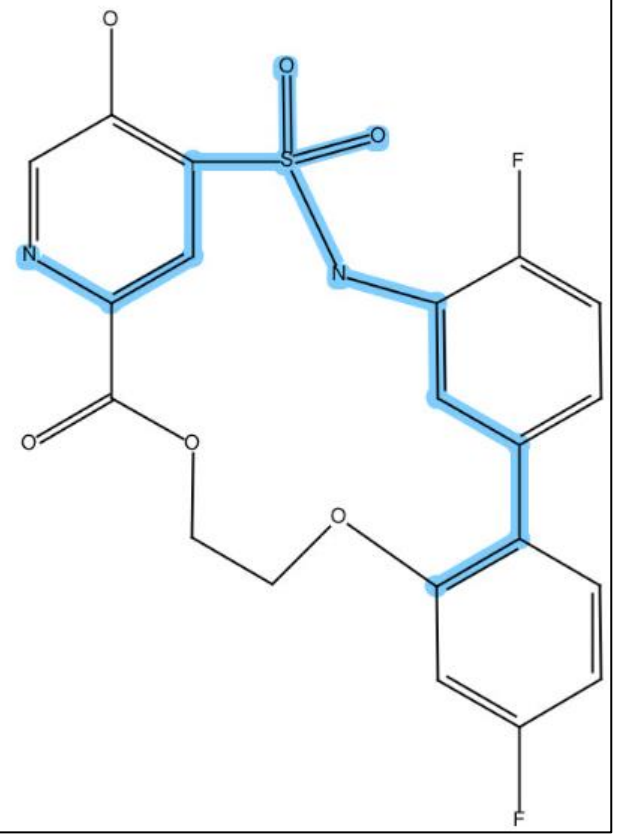
HITSTR for DCR results; AHITSTR for assembled DWPIM results

File WPINDEX

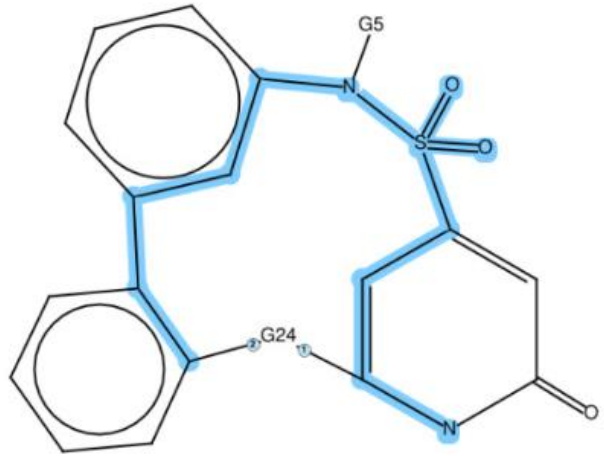
=> D BIB HITSTR AHITSTR

L11 ANSWER 1 OF 8 WPINDEX COPYRIGHT 2025 CLARIVATE on STN
AN 2023-B6665B [2024001] WPINDEX Full-text
TI New 1-benzyl-3-phenylbenzene compound used in pharmaceutical composition for treating liver condition e.g. NAFLD or NASH, type-2 diabetes, inflammation, chronic kidney disease, autoimmunity and cancer
DC B02
IN DAVENPORT A J; GLEAVE L J; GLEBE L J; PINKOSKI S L; PINKOSKY S L; SAUDI A M N; SOUTHEY M W Y; URSINOVA N C; URSINYOVA N C; WALKER P R; YARNOLD C J
PA (ESPE-N) ESPERION THERAPEUTICS INC
CYC 139
PI WO 2023215220 A1 20231109 (2024001)* EN 303[0]
AU 2023264491 A1 20241031 (2024089) EN
CN 119137107 A 20241213 (2024103) ZH
IN 202417089629 A 20241220 (2025004) EN
SG 11202407284 A 20241129 (2025005) EN
KR 2025004873 A 20250108 (2025007) KO
ADT WO 2023215220 A1 WO 2023-US20546 20230501; AU 2023264491 A1 AU 2023-264491 20230501; CN 119137107 A CN 2023-80038044 20230501; AU 2023264491 A1 PCT Application WO 2023-US20546 20230501; CN 119137107 A PCT Application WO 2023-US20546 20230501; IN 202417089629 A PCT Application WO 2023-US20546 20230501; SG 11202407284 A PCT Application WO 2023-US20546 20230501; KR 2025004873 A PCT Application WO 2023-US20546 20230501; CN 119137107 A PCT Nat. Entry CN 2023-80038044 20241101; KR 2025004873 A KR 2024-7039451 20230501; SG 11202407284 A SG 2024-11202407284 20230501; SG 11202407284 A PCT Nat. Entry SG 2024-11202407284 20241017; IN 202417089629 A IN 2024-17089629 20241119; KR 2025004873 A PCT Nat. Entry KR 2024-7039451 20241127
FDT AU 2023264491 A1 Based on WO 2023215220 A; CN 119137107 A Based on WO 2023215220 A; IN 202417089629 A Based on WO 2023215220 A; SG 11202407284 A Based on WO 2023215220 A; KR 2025004873 A Based on WO 2023215220 A
PRAI US 2023-481045P 20230123
US 2022-337344P 20220502

AN.S DCR-6815203
MF C20 H14 F2 N2 O6 S
STR



AN.M 2302-05802
SDM B : Pharmaceuticals, Agrochemicals; V : Simple organic compounds; Z : Salts
STR



CAS
A division of the American Chemical Society

Document deduplication

Please understand what is going on

- This example will deduplicate the same invention/similar patent family records
  - Patent family records, not structures!!!
- Caution: Patent databases' records may have different, **relevant** information
  - i.e., Different patent family members
- This technique will remove a record if any patent number from the first database's records matches any patent numbers in the records from the second database

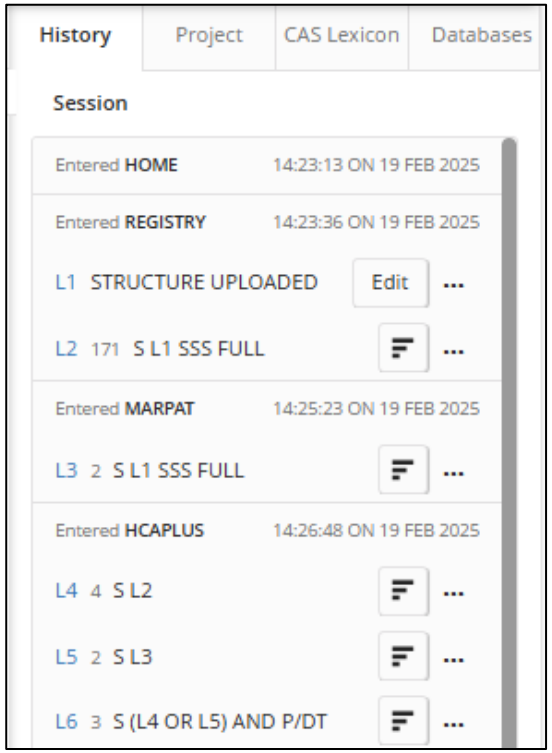
CAS
A division of the American Chemical Society

Document deduplication

TRANSFER command

TRANSFER patent numbers from the first database

NOT the corresponding set from the second database set



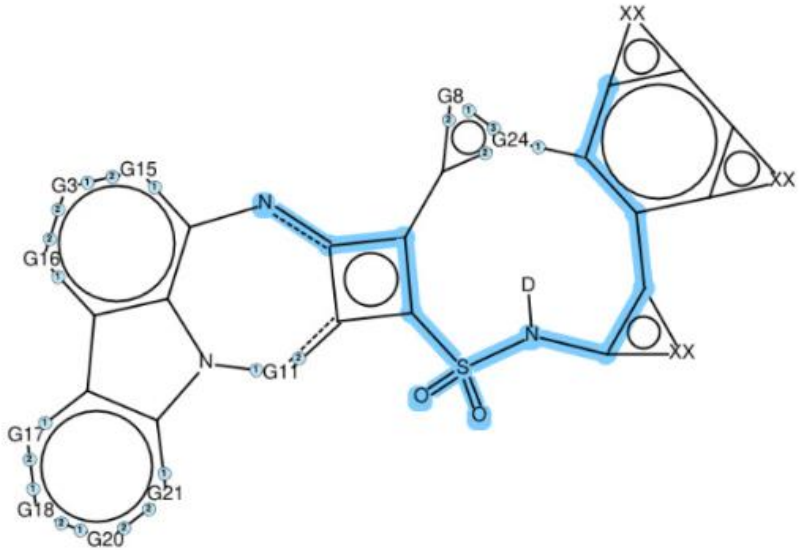

```
=> TRA L6 PN 1-

L12          TRANSFER L6 1- PN :      22 TERMS
L13           3 L12
L14             QUE  TERMS FROM L12 WITH NO HITS:     1 TERM

=> S L11 NOT L13

L15           7 L11 NOT L13

=> D BIB HITSTR AHITSTR
```


CAS
A division of the American Chemical Society

Unique DWPI record

L15 ANSWER 1 OF 7 WPINDEX COPYRIGHT 2025 CLARIVATE on STN
AN 2023-652468 [2023066] WPINDEX Full-text
CR 2023-64747V; 2023-64815X; 2023-64888U; 2023-64889R; 2023-64945A; 2023-649642; 2023-64995J; 2023-651142; 2023-65238C; 2023-65242W; 2023-652443; 2023-68023M; 2024-D2099V
TI New nitrogen-containing cyclic compound used as organic layer of organic light emitting device for consumer product e.g. flat panel display, curved display, computer monitor, medical monitor, television, rollable display, and digital camera
DC E11; L03; T01; U11; U12; W03; W04; W05
IN FLEETHAM T; SULEYMANOVA A
PA (UVDI-C) UNIVERSAL DISPLAY CORP
CYC 2
PI KR 2023091831 A 20230623 (2023066)* KO 101[2]
US 20230286989 A1 20230914 (2023077) EN
ADT KR 2023091831 A KR 2022-177552 20221216; US 20230286989 A1 Provisional US 2021-265495P 20211216; US 20230286989 A1 Provisional US 2022-363047P 20220415; US 20230286989 A1 Provisional US 2022-363068P 20220415; US

AN.M 2292-80901
SDM E : General Chemicals; V : Simple organic compounds
STR



CAS
A division of the American Chemical Society

Repeat strategy in Reaxys databases

No unique records were found compared to HCAplus search

```
=> FILE REAXYSFILESUB; S L1 SSS FULL
```

```
L16          191 SEA SSS FUL L1

=> FILE REAXYSFILEBIB; S L16 AND P/DT
```

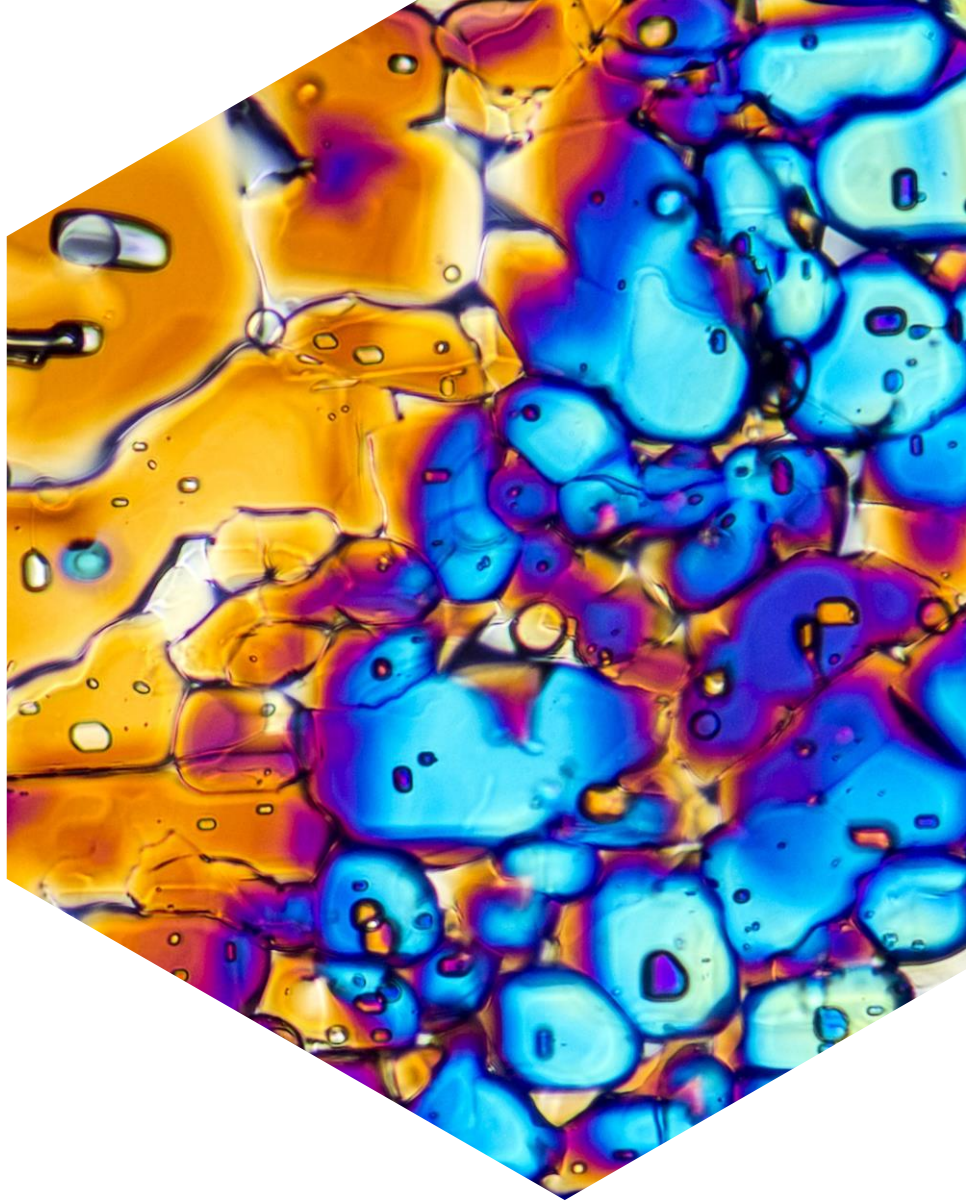
```
            12 L16
      45265362 P/DT
L17         12 L16 AND P/DT

=> TRA L6 PN 1-

L18         TRANSFER L6 1- PN :      22 TERMS
L19         24 L18
L20            QUE  TERMS FROM L18 WITH NO HITS:     3 TERMS

=> S L17 NOT L19

L21          0 L17 NOT L19
```


CAS
A division of the American Chemical Society

Agenda

Strategies for complex chemical structures

- How to search for macrocyclic compounds with variable ring sizes across multiple databases
- Deduplicate patent results
- When to include stereochemistry in Derwent databases
- How to search for radicals or ions
- Restricting further substitution on VPA rings

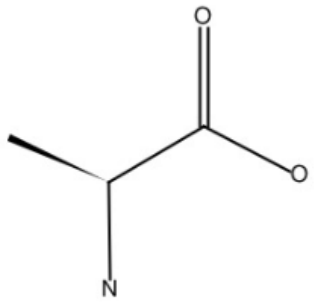


CAS
A division of the American Chemical Society

Stereochemistry in DCR

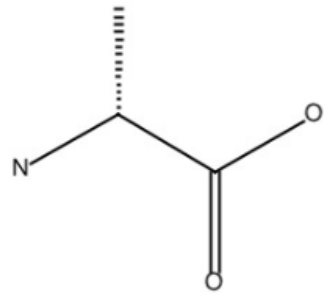
Will include stereochemistry whenever a specific stereoisomer is claimed

L1 ANSWER 1 OF 3 DCR COPYRIGHT 2025 CLARIVATE on STN.
AN DCR-8757 DCR
DCSE 8189-2-0-0
CN.P L-ALANINE
CN.S 2-Amino-propionic acid
SY (S)-ALANINE; ABUFENE; ALA-OH; ALANINA; **ALANINE**; ALANINE, L-; ALANINE-L; L-(S)-ALANINE; L-ALANINE



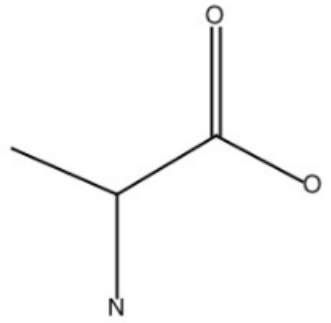
CMT L-stereoisomer
MF C3 H7 N O2
ED Entered STN: 26 Apr 1999
Last updated on STN: 16 Nov 2015
Update DWPI Cross Ref.: 20 Feb 2025

L1 ANSWER 2 OF 3 DCR COPYRIGHT 2025 CLARIVATE on STN.
AN DCR-8214 DCR
DCSE 8189-1-0-0
CN.P D-ALANINE
CN.S (R)-2-Amino-propionic acid
SY **ALANINE**; ALANINE, D-; ALANINE-D; D-ALANINE; DEXTROALANINE



CMT D-stereoisomer
MF C3 H7 N O2
ED Entered STN: 3 Nov 1999
Last updated on STN: 17 Jul 2024
Update DWPI Cross Ref.: 4 Feb 2025

L1 ANSWER 3 OF 3 DCR COPYRIGHT 2025 CLARIVATE on STN.
AN DCR-8189 DCR
DCSE 8189-0-0-0
CN.P **ALANINE**
SY **ALANINE**



CMT Unspecified stereoisomer
MF C3 H7 N O2
ED Entered STN: 26 Apr 1999
Last updated on STN: 20 Jun 2022
Update DWPI Cross Ref.: 20 Feb 2025

CAS
A division of the American Chemical Society

DCR indexing in DWPI

Unique hits for each stereoisomer

```
=> S DCR-8757/IT

L2          3151 DCR-8757/IT

=> S DCR-8214/IT

L3          1908 DCR-8214/IT

=> S DCR-8189/IT

L4          7945 DCR-8189/IT

=> S L2 OR L3 OR L4

L5          9619 L2 OR L3 OR L4
```

DCR-8757 = L- Alanine

DCR-8214 = D-Alanine

DCR-8189 = Alanine

Note: Be sure to include all stereoisomers for comprehensive searches.

CAS
A division of the American Chemical Society

Structure searching in DCR and DWPIM

All structures are considered flat, even if they show stereochemistry

Drawing a structure flat, or with stereochemistry, will retrieve the same records

CAS
A division of the American Chemical Society

Structure searching in DCR and DWPIM

The flat version of Alanine

File **DCR**

```
Uploading structure file: ANALINE
```


```
Node Attributes
Chain Nodes : 1  2  3  4  5  6
Bond Attributes
Chain Bonds : 1-2  1-3  1-4  2-5  2-6
Exact Bonds : 1-2  1-3
Exact/Normalized Bonds : 1-4  2-5  2-6
Markush Attributes
Match Level (CLASS) : 1  2  3  4  5  6
Element Count Level (LIMITED) : 1  2  3  4  5  6


L6      STRUCTURE UPLOADED
```

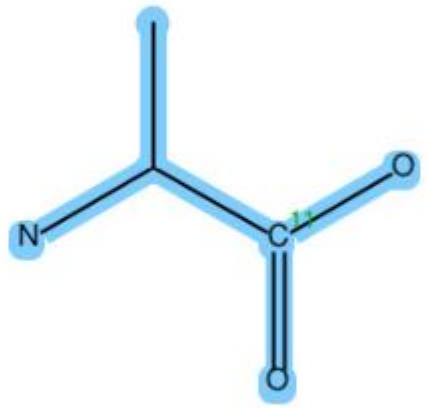
```
=> S L6 EXA FULL

FULL SEARCH INITIATED 12:10:15
FULL SCREEN SEARCH COMPLETED -        0 TO ITERATE

  0.0% PROCESSED    6024235 ITERATIONS                              16 ANSWERS
SEARCH TIME: 00.00.05

L7           16 SEA EXA FUL L6

=> S L1 AND L7

L8            3 L1 AND L7

=> S L7 NOT L1

L9           13 L7 NOT L1

=> D 1-2
```


CAS
A division of the American Chemical Society

Structure searching in DCR and DWPIM

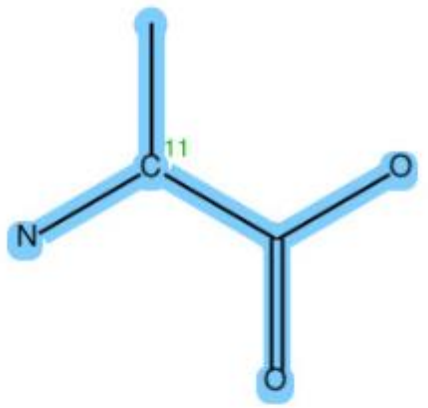
Should these also be included in a search for Alanine?

L9 ANSWER 1 OF 13 DCR COPYRIGHT 2025 CLARIVATE on STN.
AN DCR-6133476 DCR
DCSE 8189-0-0-12



CMT Isotope labelled
MF C3 H7 N O2
ED Entered STN: 25 Jan 2023
Last updated on STN: 25 Jan 2023
Update DWPI Cross Ref.: 25 Jan 2023

L9 ANSWER 2 OF 13 DCR COPYRIGHT 2025 CLARIVATE on STN.
AN DCR-6133475 DCR
DCSE 8189-0-0-11



CMT Isotope labelled
MF C3 H7 N O2
ED Entered STN: 25 Jan 2023
Last updated on STN: 25 Jan 2023
Update DWPI Cross Ref.: 25 Jan 2023

CAS
A division of the American Chemical Society

Structure searching in DCR and DWPIM

The D version of Alanine

```
Uploading structure file: D-ANALINE
```


```
Node Attributes
Chain Nodes : 1  2  3  4  5  6
Bond Attributes
Chain Bonds : 1-2  1-3  1-4  2-5  2-6
Exact Bonds : 1-2  1-3
Exact/Normalized Bonds : 1-4  2-5  2-6
Markush Attributes
Match Level (CLASS) : 1  2  3  4  5  6
Element Count Level (LIMITED) : 1  2  3  4  5  6


L10     STRUCTURE UPLOADED
```

```
=> S L10 EXA FULL

FULL SEARCH INITIATED 12:21:03
FULL SCREEN SEARCH COMPLETED -        0 TO ITERATE

  0.0% PROCESSED    6024235 ITERATIONS                          16 ANSWERS
SEARCH TIME: 00.00.05

L11          16 SEA EXA FUL L10

=> S L1 AND L11

L12           3 L1 AND L11

=> S L11 AND L7

L13          16 L11 AND L7
```


CAS
A division of the American Chemical Society

Stereochemistry considerations

If you must …

If you wish to be comprehensive, structure searching should be considered

- Isotopes, stereochemistry, etc.

Searching for a substance record with stereochemistry requires some finesse

- Part of a chemical name (/CN vs. /CNS)
- Frag code (for subscribers)

CAS
A division of the American Chemical Society

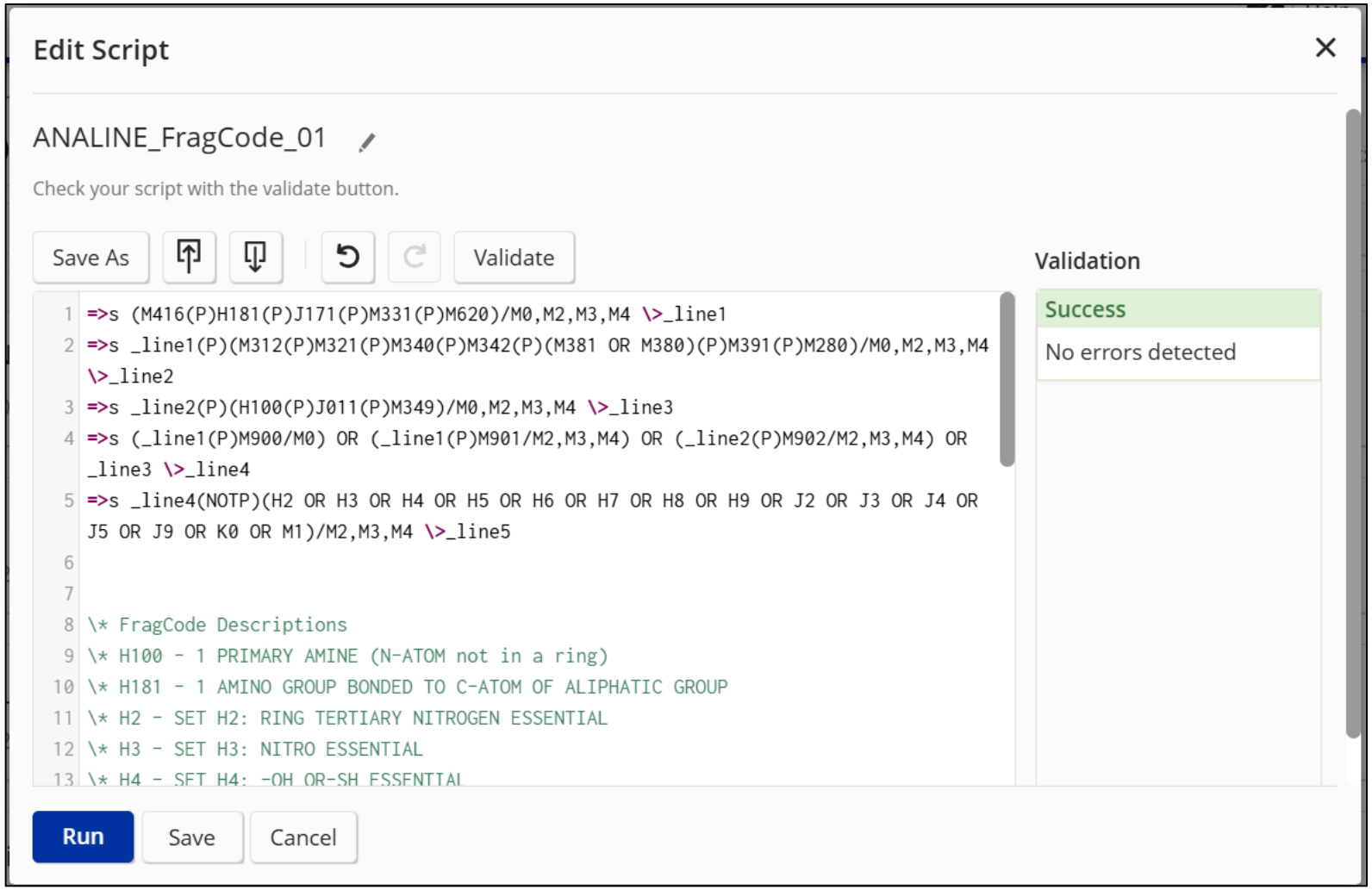
Chemical name vs. chemical name segment

```
=> S (D (W) ALANINE)/CNS

        278627 D/CNS
           206 ALANINE/CNS
L15         13 (D (W) ALANINE)/CNS

=> D KWIC 1-4


L15    ANSWER 1 OF 13 DCR  COPYRIGHT 2025 CLARIVATE on STN.
CN.P   D-ALANYL-D-ALANINE CARBOXYPEPTIDASE, FRACTION B
SY     D-ALANYL-D-ALANINE CARBOXYPEPTIDASE, FRACTION B;
       D-ALANYL-D-ALANINE-CARBOXYPEPTIDASE-FRACTION-B; EC-3.40.710.10

L15    ANSWER 2 OF 13 DCR  COPYRIGHT 2025 CLARIVATE on STN.
CN.P   D-ALANYL-D-ALANINE-CARBOXYPEPTIDASE
SY     D ALANYL D ALANINE CARBOXYPEPTIDASE;
       D-ALANYL-D-ALANINE-CARBOXYPEPTIDASE; E.C-3.4.16.4

L15    ANSWER 3 OF 13 DCR  COPYRIGHT 2025 CLARIVATE on STN.
CN.P   N-METHYL-D-ALANINE
SY     N-METHYL-D-ALANINE

L15    ANSWER 4 OF 13 DCR  COPYRIGHT 2025 CLARIVATE on STN.
SY     (D-MEALA)3-(ETVAL)4-CSA; (D-MEALA)3-(ETVAL)4-CYCLOSPORIN-A;
       (D-MEALA)3-(ETVAL)4-CYCLOSPORINE-A;
       8-(N-METHYL-D-ALA),9-(N-ETHYL-VAL)CYCLOSPORINE;
       8-(N-METHYL-D-ALANINE),9-(N-ETHYL-VALINE)CYCLOSPORINE; ALISPORIVIR
```

```
=> D KWIC


L17    ANSWER 1 OF 1 DCR  COPYRIGHT 2025 CLARIVATE on STN.
CN.P   D-ALANINE
SY     ALANINE; ALANINE, D-; ALANINE-D; D-ALANINE; DEXTROALANINE
```

Derwent fragmentation code strategy

Derwent auto-generated script for Alanine without M800 code



CAS
A division of the American Chemical Society

Alanine script run in WPIX

No M800 code

```
=> s (M416(P)H181(P)J171(P)M331(P)M620)/M0,M2,M3,M4

          4297 M416/M0
        560648 M416/M2
        746822 M416/M3
          2590 M416/M4
         13261 H181/M0
        551301 H181/M2
```

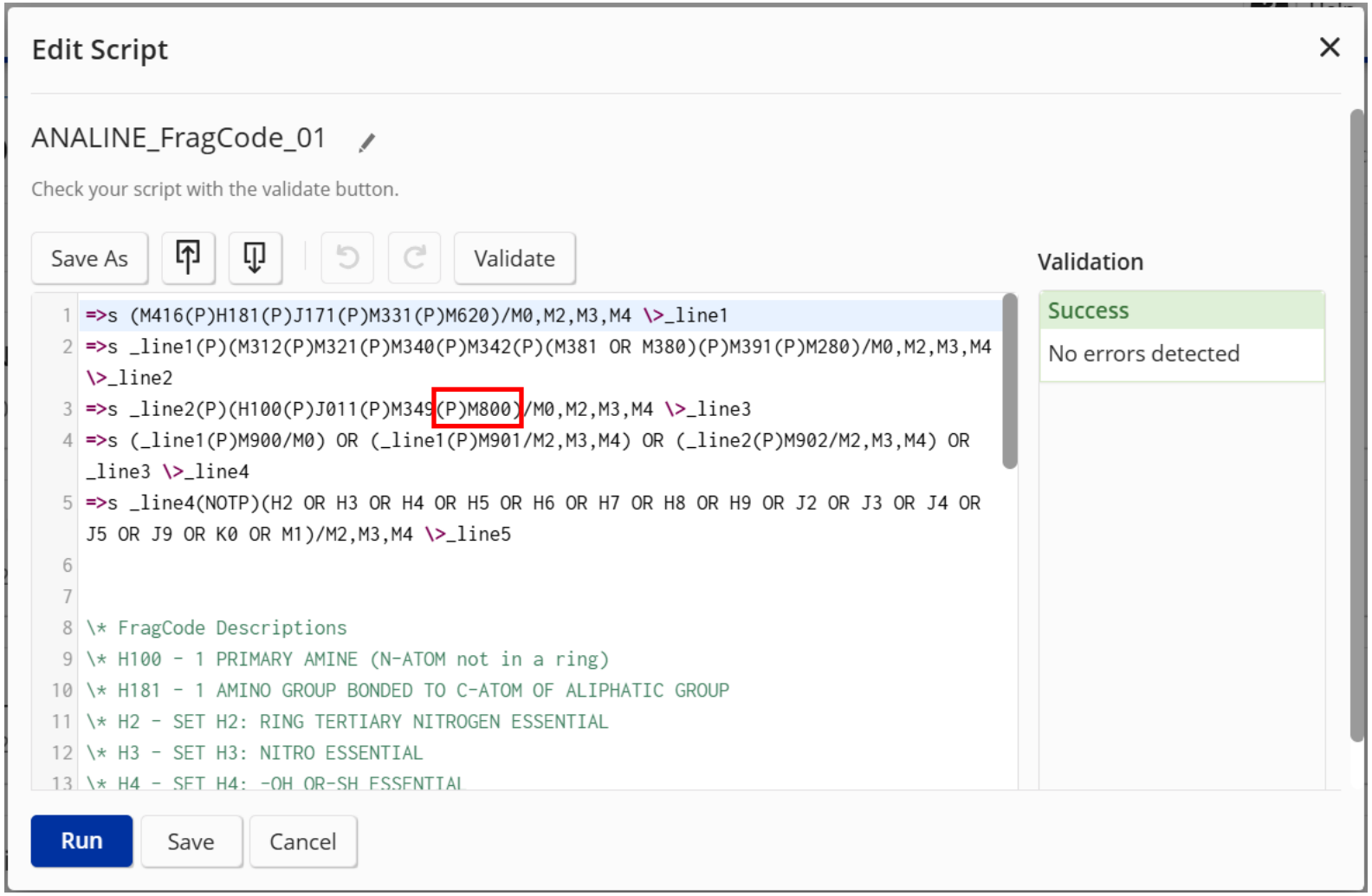
. . .

```
        549448 M1/M2
        244837 M1/M3
         74582 M1/M4
L22      30653 L21(NOTP)(H2 OR H3 OR H4 OR H5 OR H6 OR H7 OR H8 OR H9 OR J2 OR
               J3 OR J4 OR J5 OR J9 OR K0 OR M1)/M2,M3,M4
```


CAS
A division of the American Chemical Society

Derwent fragmentation code strategy

Derwent auto-generated script for Alanine with M800 code manually added



CAS
A division of the American Chemical Society

Alanine script run in WPIX

Including M800 fragmentation code

```
=> s (M416(P)H181(P)J171(P)M331(P)M620)/M0,M2,M3,M4

          4297 M416/M0
        560648 M416/M2
        746822 M416/M3
          2590 M416/M4
         13261 H181/M0
        551301 H181/M2
```

. . .

```
        549448 M1/M2
        244837 M1/M3
         74582 M1/M4
L27       3660 L26(NOTP)(H2 OR H3 OR H4 OR H5 OR H6 OR H7 OR H8 OR H9 OR J2 OR
               J3 OR J4 OR J5 OR J9 OR K0 OR M1)/M2,M3,M4
```


CAS
A division of the American Chemical Society

Stereochemistry searching in Derwent databases

Structure searching captures all records with stereochemistry, whether the structure was drawn flat or with stereochemistry

When stereochemistry is a must for the search strategy –

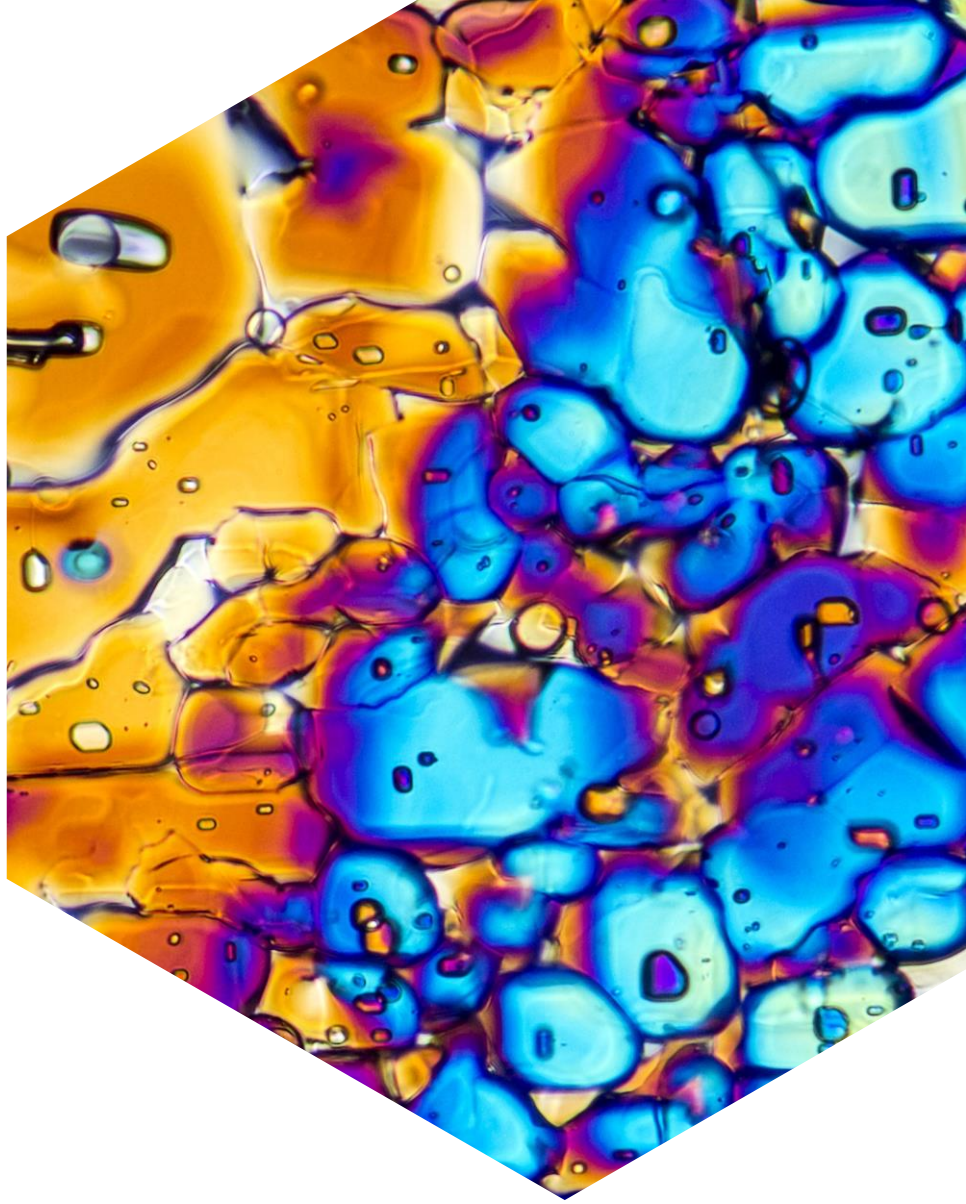
- Consider chemical name/chemical name segment searching
- M800 fragmentation code for Derwent subscribers

CAS
A division of the American Chemical Society

Agenda

Strategies for complex chemical structures

- How to search for macrocyclic compounds with variable ring sizes across multiple databases
- Deduplicate patent results
- When to include stereochemistry in Derwent databases
- How to search for radicals or ions
- Restricting further substitution on VPA rings

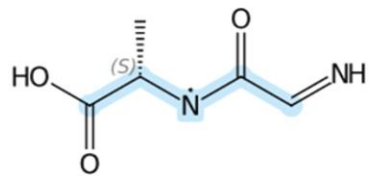


CAS
A division of the American Chemical Society

Unusual charges or radicals

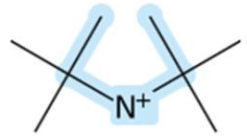
How to specifically search for these compounds

RN 1260679-31-9 REGISTRY
ED Entered STN: 27 Jan 2011
CN **Amidogen, [(1S)-1-carboxyethyl](2-iminoacetyl)-** (CA INDEX NAME)
OTHER CA INDEX NAMES:
CN **[(1S)-1-Carboxyethyl](2-iminoacetyl)amidogen**
FS STEREOSEARCH
MF C5 H7 N2 O3
SR CA
LC STN Files: CA, CAPLUS

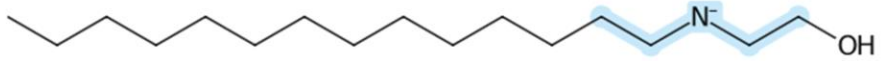


Absolute stereochemistry shown

RN 2228844-87-7 REGISTRY
ED Entered STN: 28 Jun 2018
CN **Aminylium, bis(1,1-dimethylethyl)-** (CA INDEX NAME)
OTHER CA INDEX NAMES:
CN **Bis(1,1-dimethylethyl)aminylium**
MF C8 H18 N
SR CA
LC STN Files: CA, CAPLUS



RN 2918190-33-5 REGISTRY
ED Entered STN: 29 Mar 2023
CN **Ethanol, 2-(tetradecylamino)-, ion(1-)** (CA INDEX NAME)
MF C16 H34 N O
SR CA
LC STN Files: CA, CAPLUS



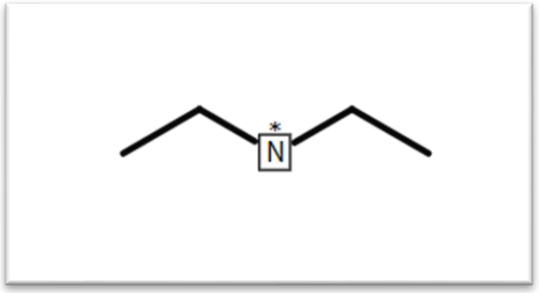
Structure search used for these results

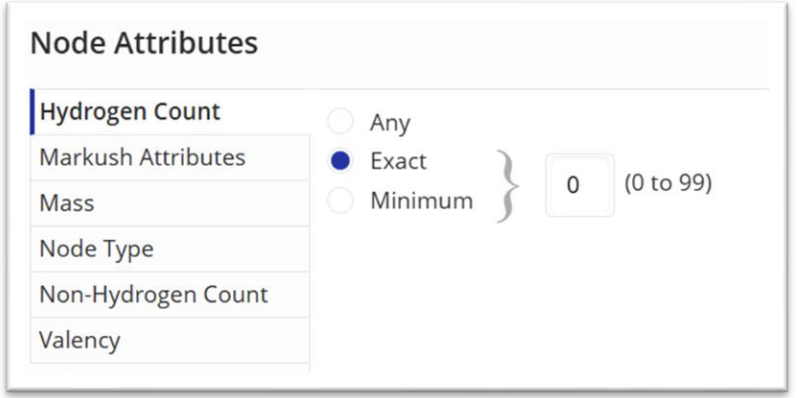
Important to define bonds as exact not exact/normalized

Bonds from N-C should be Exact to avoid finding normal amides in a substructure search

Nitrogen is locked against further substitution

Nitrogen has a H-attachment of Exact zero to avoid getting normal secondary amines





This retrieves radicals and any compound with a nitrogen charge

CAS
A division of the American Chemical Society

Looking for specific variations

How to search for a presence of a radical

For positively or negatively charged compounds, use one of the icons in the structure editor to apply these

Radicals do not have a charge, but they would have an unusual valance on the nitrogen

Valence can be defined as an attribute to the nitrogen node

A amidogen node would have a valence of 2

Applying this would retrieve 112 compounds

Of these, 100 were named Amidogen. Could also see “free radical form” in the name

CAS
A division of the American Chemical Society

Charges and radicals in other files

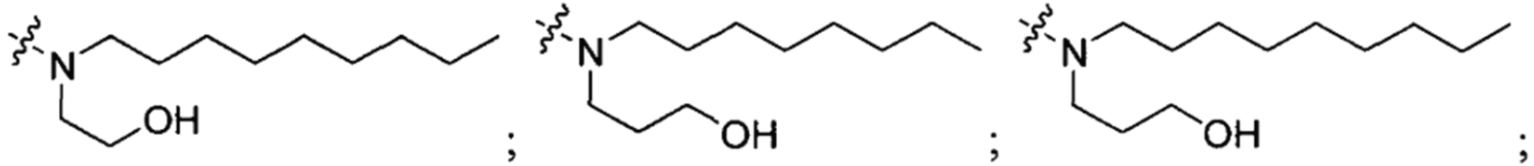
DCR and ReaxysfileSub

Registry produced 211 compounds, including 112 radicals, 44 positively charged and 49 negatively charged compounds

DCR produced 10 compounds, all being radicals

ReaxysfileSub produced 4,320 charged and radicals, but the majority are from incorrect interpretation of algorithmic curation of journals and patents, for example in WO2019036008:

30. The compound of any one of claims 1-29, wherein R^3 has one of the following structures:

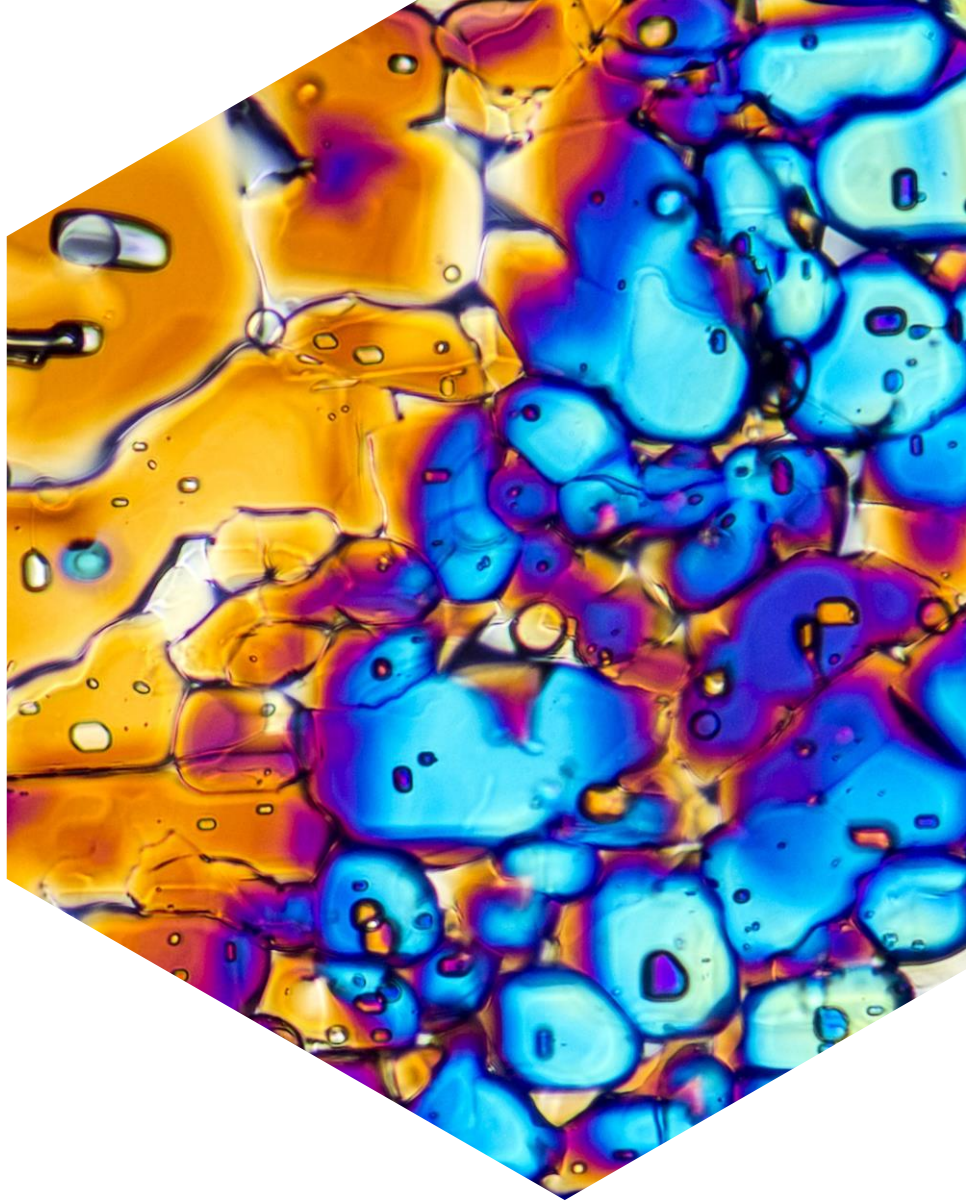


CAS
A division of the American Chemical Society

Agenda

Strategies for complex chemical structures

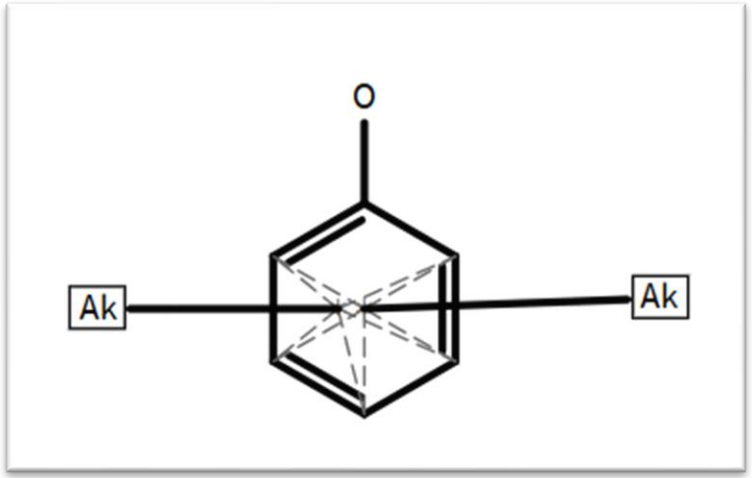
- How to search for macrocyclic compounds with variable ring sizes across multiple databases
- Deduplicate patent results
- When to include stereochemistry in Derwent databases
- How to search for radicals or ions
- Restricting further substitution on VPA rings



CAS
A division of the American Chemical Society

Restrict further substitution on Ring

Would be interested in finding dialkylphenols where the oxygen may have any further substituent, but no further substitution on the phenyl ring



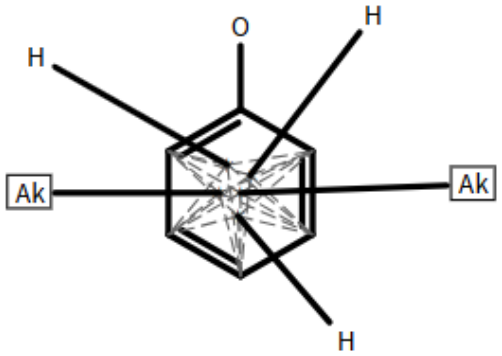
Searched as such, this query retrieves 1.7 mil compounds in Registry

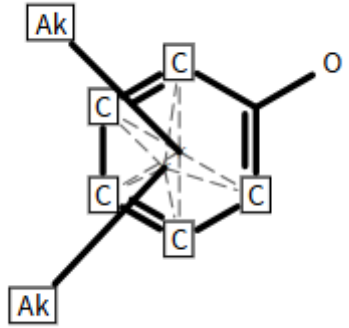
CAS
A division of the American Chemical Society

What doesn’t work

Cannot combine with an atom lock or add more hydrogens

Adding more VPA hydrogen nodes or locking VPA locations doesn’t produce the correct results in Registry:





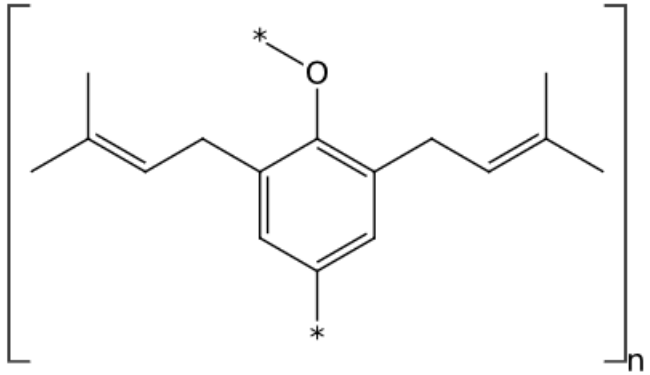
Structure on the right gives good results in DCR and ReaxysfileSub

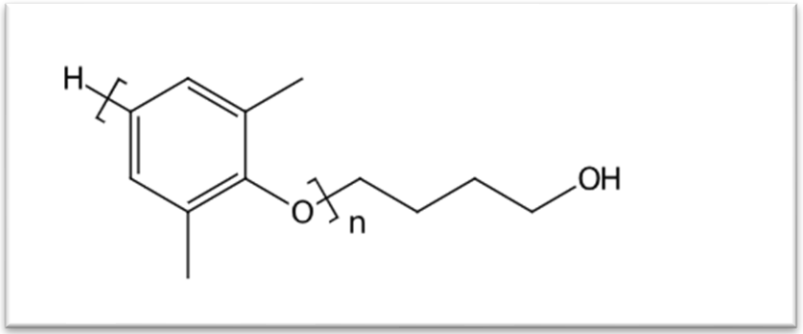
CAS
A division of the American Chemical Society

Run a Closed Substructure Search

Open the Oxygen with non-H attachment Minimum 1

The closed substructure search still brings in some unexpected results:





Where a polymer repeating group creates the 4th attachment (left)

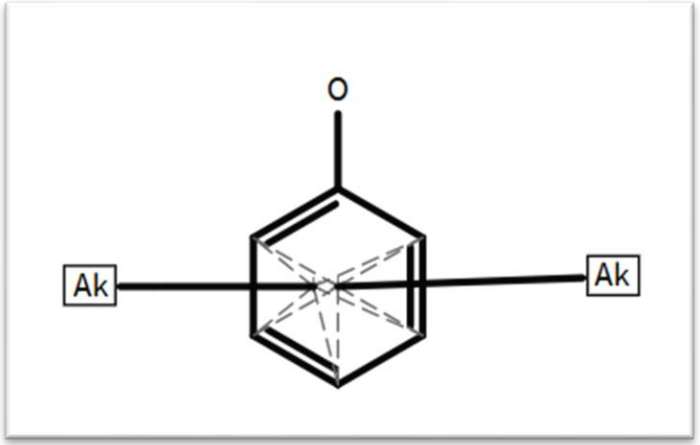
It does include a compound with a hard-coded Hydrogen (right)

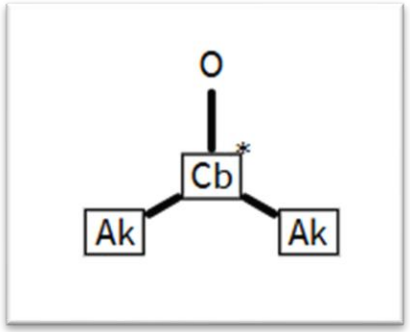
CAS
A division of the American Chemical Society

Some options to restrict further substitution

Combine with a generic representation of this structure

Create a second L-number where the phenyl ring is reduced to a generic Cb group with 6 carbon atoms, unsaturated and monocyclic, which can then be locked against further substitution






```
619009 S L5 AND L4 SSS FUL
```


CAS
A division of the American Chemical Society

Prepared but not intermediate products

Linking and Not-linking roles

```
      FILE 'REGISTRY' ENTERED AT 16:50:13 ON 26 FEB 2025
L3         619077 S L1 AND L2 SSS FUL
L4         619070 S L3/COM
L5          92035 S L4 AND CAPLUS/LC

      FILE 'CAPLUS' ENTERED AT 16:51:23 ON 26 FEB 2025
L6          29671 S L5/PREP
L7          26735 S L5/PREP(NOTL)RACT/RL
```


CAS
A division of the American Chemical Society

PREP not-linked to RACT

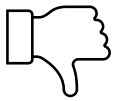
Find docs where an RN is linked to PREP, but not also to RACT

TI Phenol compound, poly(phenylene ether), curable composition, cured product, and electronic component

PI **JP 2025006572** A 20250117

IT **3025233-87-5 P**

RL: IMF (Industrial manufacture); PRP (Properties); TEM (Technical or engineered material use); **PREP (Preparation)**; USES (Uses)
(phenol compd., poly(phenylene ether), curable compn., cured product, and electronic component)



IT **3025233-83-1** P **3025233-84-2 P** **3068200-73-4** P **3068200-75-6** P

RL: IMF (Industrial manufacture); RCT (Reactant); **PREP (Preparation)**; **RACT (Reactant or reagent)**
(phenol compd., poly(phenylene ether), curable compn., cured product, and electronic component)

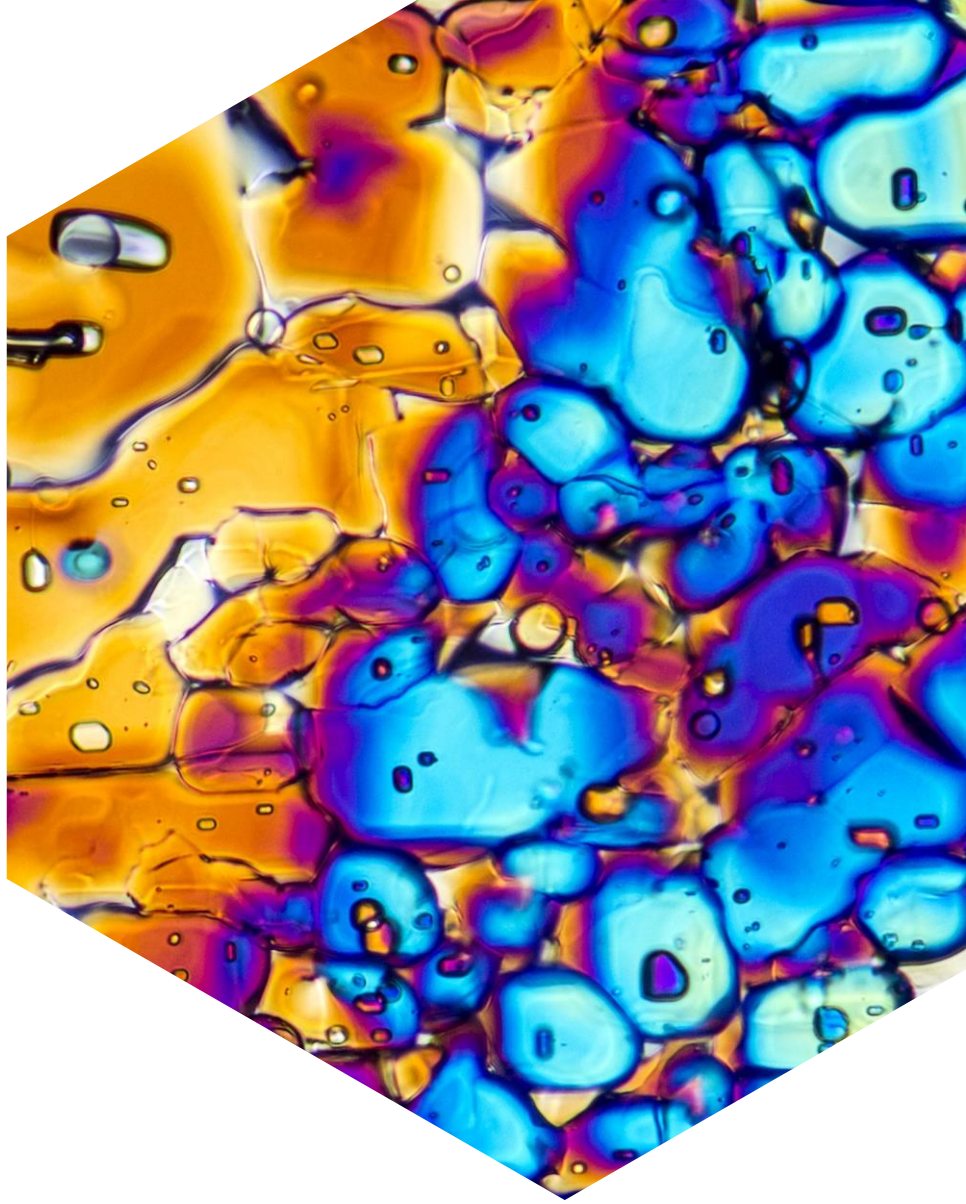
```
     FILE 'CAPLUS' ENTERED AT 16:51:23 ON 26 FEB 2025
L6         29671 S L5/PREP
L7         26735 S L5/PREP(NOTL)RACT/RL
```


CAS
A division of the American Chemical Society

Agenda

Strategies for complex chemical structures

- How to search for macrocyclic compounds with variable ring sizes across multiple databases
- Deduplicate patent results
- When to include stereochemistry in Derwent databases
- How to search for radicals or ions
- Restricting further substitution on VPA rings



CAS
A division of the American Chemical Society

CAS accelerates breakthroughs

Paul Peters
Director, Customer Success Specialist
ppeters@acs-i.org

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

CAS
A division of the American Chemical Society