



CAS STNext[®] E-SEMINAR

Searching Ring System Data on CAS STNext[®]

Paul Peters, ACS International / CAS

© 2023 American Chemical Society. All rights reserved.

 **FIZ Karlsruhe**
Leibniz Institute for Information Infrastructure

CAS 
A division of the
American Chemical Society

Agenda

Analyzing Ring System Data

Search Fields

Combining Ring Search Fields with Structure Search

Use cases for using Ring System Data

Rings in REGISTRY

For substances with up to 252 non-hydrogen atoms

Any structure in REGISTRY with a ring contains Ring System Data.

- Over 207 million substances have Ring System Data (RSD)
- Over 77.8 million substances have polycyclic rings
- Only 78.5 mil substances have a structure without a ring (NO RSD/FA)

Some Ring System Data may be displayed for a substance.

Other Ring System Data is searchable only.

Often, broad structure claims are best searched with RSD in combination with structure query.

Challenges in Searching Rings

The name of a ring system is not always clear.

New substances entering REGISTRY may not yet have a name.

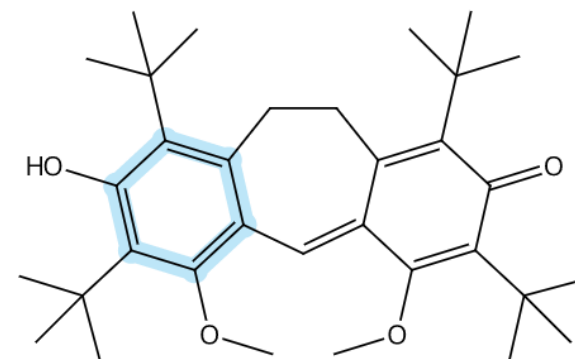
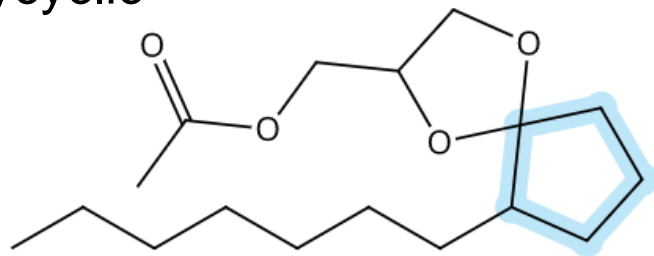
Structure search may be too broad to run within system limits.

The claimed ring system may be too generic to be comprehensively covered in a structure search query.

Terminology of rings

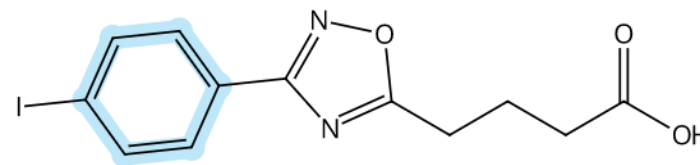
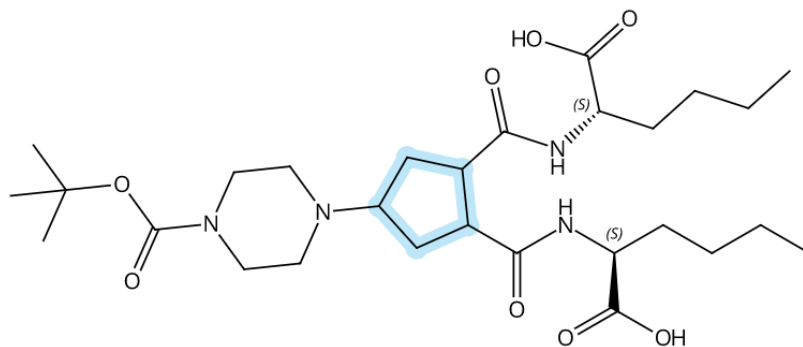
Fused Ring”: Shares at least one ring node with another ring.

AKA: embedded or polycyclic



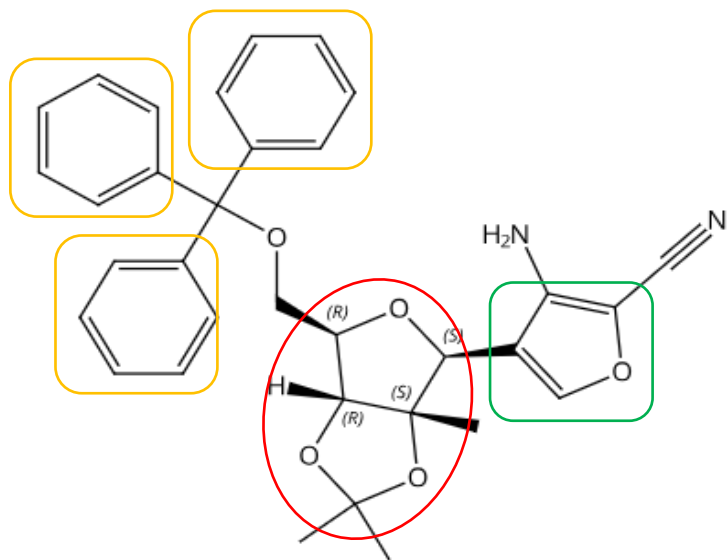
Unfused Ring”: Does not share a ring node with another ring.

AKA: isolated or monocyclic



Displaying Ring System Data

Substance CAS RN 872201-71-3



This substance contains five Ring Systems

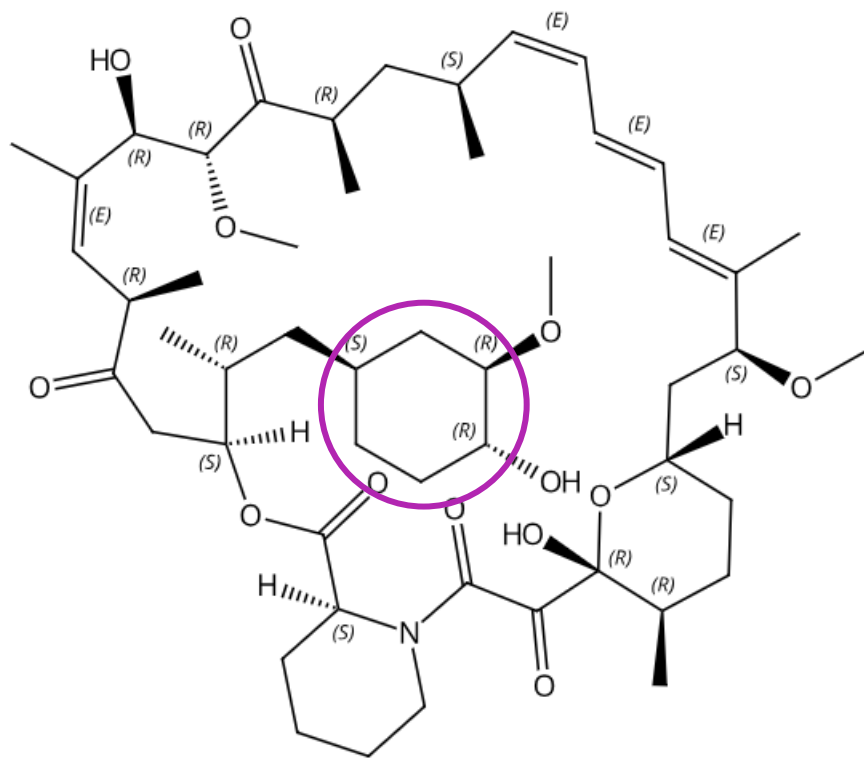
Ring System Data

Elemental Analysis	Elemental Sequence	Size of the Rings	Ring System Formula	Ring Identifier	RID Occurrence
EA	ES	SZ	RF	RID	Count

C40	OC4	5	C40	16.138.5	1
C6	C6	6	C6	46.150.18	3
C302-C40	OCOC2-OC4	5-5	C503	180.182.1	1

The ring data for Rapamycin

CAS RN is 53123-88-9



Absolute stereochemistry shown
E/Z labels describe double bond geometry

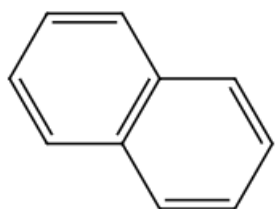
Ring System Data

Elemental Analysis	Elemental Sequence	Size of the Rings	Ring System Formula	Ring Identifier	RID Occurrence
EA	ES	SZ	RF	RID	Count
C6	C6	6	C6	46.150.1	1
C5N-C5O-	NC5-OC5-	6-6-29	C33N02	22445.1.1	1
C26N02	NC20C21OC3				

This substance has two Ring System with the larger one containing three rings.

Finding the smallest rings in the ring system

Sometimes leads to unexpected choices



Elemental Analysis	Elemental Sequence	Size of the Rings	Ring System Formula	Ring Identifier	RID Occurrence Count
EA	ES	SZ	RF	RID	Count
C6-C6	C6-C6	6-6	C10	591.49.57	1

Nothing strange here. Naphthalene is built as two 6-membered rings fused together

Finding the smallest rings in the ring system

Sometimes leads to unexpected choices



Elemental Analysis	Elemental Sequence	Size of the Rings	Ring System Formula	Ring Identifier	RID Occurrence
EA	ES	SZ	RF	RID	Count
C5-C5	C5-C5	5-5	C7	103.10.1	1

There is no 6-membered ring in the bicycloheptane ring system.
The smallest rings are two 5-membered rings

Finding the smallest rings in the ring system

Sometimes leads to unexpected choices



Elemental Analysis	Elemental Sequence	Size of the Rings	Ring System Formula	Ring Identifier	RID Occurrence
EA	ES	SZ	RF	RID	Count
C2Fe-C2Fe-	FeC2-FeC2-	3-3-3-3-3-3-	C10Fe	14785.1.1	1
C2Fe-C2Fe-	FeC2-FeC2-	3-3-3-3			
C2Fe-C2Fe-	FeC2-FeC2-				
C2Fe-C2Fe-	FeC2-FeC2-				
C2Fe-C2Fe	FeC2-FeC2				

What?! There is no 5-membered ring in these cyclopentadienyl-based complexes? Indeed, these are built by 10 3-membered rings.

Naphthalene is not a 10-membered ring system with a bridge, is it?

Why search on ring system data?

Considering using ring system data for your searches

Refining molecular formula or chemical name searches.

Segmenting large answer sets based on types of ring systems.

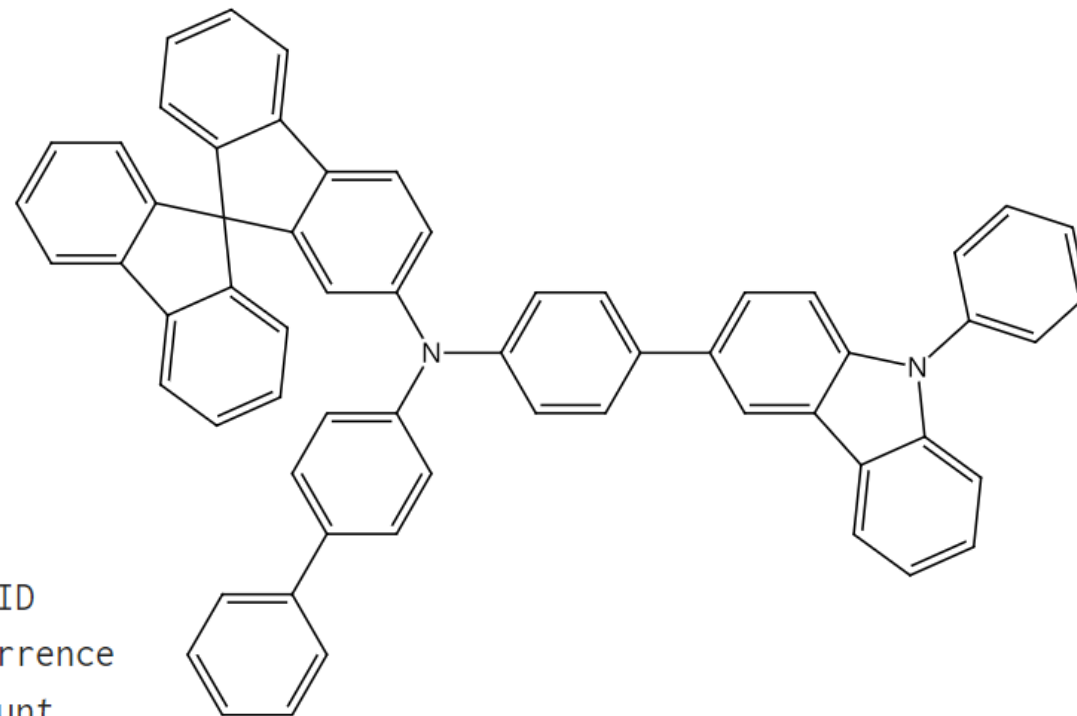
Removing unwanted ring systems from broad searches.

Creating a subset when structure queries exceed system limits.

Understanding the search fields in RSD

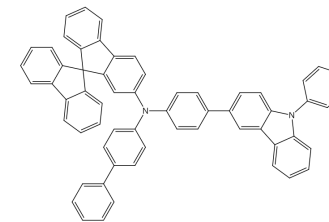
OLED compound by LG Chem

For reference on
following slides



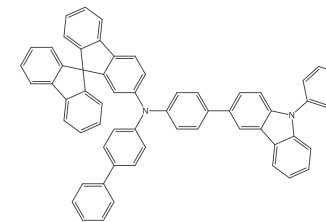
Elemental Analysis EA	Elemental Sequence ES	Size of the Rings SZ	Ring System Formula RF	Ring Identifier RID	RID Occurrence Count
C6	C6	6	C6	46.150.18	4
C4N-C6-C6	NC4-C6-C6	5-6-6	C12N	1839.22.20	1
C5-C5-C6-C6-	C5-C5-C6-C6-	5-5-6-6-6-6	C25	9841.9.1	1
C6-C6	C6-C6				

RSD: Counting rings and ring systems



To specify...	Field Code	Examples
Total number of rings in a component	CNR	13/CNR 10-14/CNR
Total number of ring systems in an entire substance	NRS	6/NRS 3-6/NRS
Total number of ring systems in a single component	CNRS	6/CNRS 3-6/CNRS
Number of rings in a ring system	NRRS	3/NRRS 3-10/NRRS

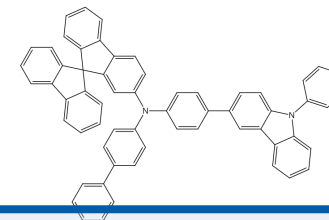
RSD: Ring sizes and counts



To specify...	Field Code	Examples
Size of (smallest) ring in a ring system, and number of occurrences of that ring in the ring system	SZS	5/SZS 2 5/SZS 1-10 5/SZS
Sizes of all (smallest) rings in a ring system, and number of occurrences of that system in a component	SZ	5-6-6/SZ 1 5-6-6/SZ 1-3 5-6-6/SZ

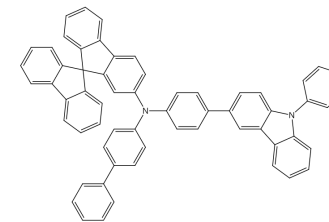
Values for occurrences are optional

RSD: Formulas and elements in rings



To specify...	Field Code	Example
Ring system formula, and occurrences of that system formula in a component	RF	C12N/RF 2 C12N/RF
Ring system formula (without element counts), and occurrences of that system formula in a component	RELF	C N/RELF 3 C N/RELF
One element in a ring system, and occurrences of that element in the ring system	REL	C/REL 10 C/REL
Number of unique elements in a ring system	RELC	2/RELC
Number of atoms in a ring system	RATC	13/RATC

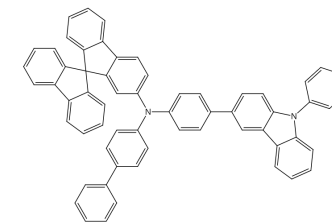
RSD: Formulas and elements in rings



To specify...	Field Code	Example
Formulas for all the rings in a ring system, occurrences of that system formula in a component	EA	C4N-C6-C6/EA 3 C4N-C6-C6/EA
Formula for a ring in a ring system, occurrences of that ring in the ring system	EAS	C4N/EAS 2 C4N/EAS
Elemental sequence for all the rings in a ring system, occurrences of that system formula in a component	ES	NC4-C6-C6/ES 3 NC4-C6-C6/ES
Elemental sequence for a ring in a ring system, occurrences of that ring in the ring system	ESS	NC4/ESS 2 NC4/ESS

RSD: Proximity operators

Combining RSD fields applied to same component or system



OPERATOR	PROXIMITY
(P)	Same single-component substance or same component of a multi-component substance
(NOTP)	Not in the same component
(S)	In the same ring system
(NOTS)	Not in the same ring system

1. RSD is not always conclusive

Keeps various options open for elements in the ring

Ring system of two 6-membered rings

One ring has NCNC3 as the sequence

Second ring has 0-2 N in the ring

Only C and optionally N in that ring

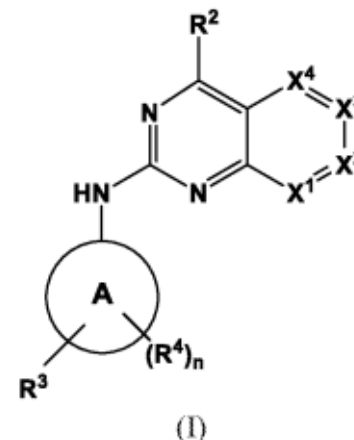
What are the element sequences for the second ring?

WO 2023/138412

PCT/CN2023/071020

CLAIMS

1. A compound of Formula (I)



or a pharmaceutically acceptable salt, solvate, stereoisomer, or isotopic variant thereof, wherein each of X^1 , X^2 , X^3 , and X^4 is independently N or CR^A , in which R^A is H or R^1 , with the proviso that no more than two of X^1 , X^2 , X^3 , and X^4 are N;

ring A is a 5- or 6-membered aromatic or heteroaromatic ring which contains 0, 1, or 2 heteroatoms selected from the group consisting of N, O, and S, optionally further fused to one or two cyclic rings independently selected from -cycloalkyl, -heterocyclyl, -aryl, -heteroaryl ring;

Developing the search strategy

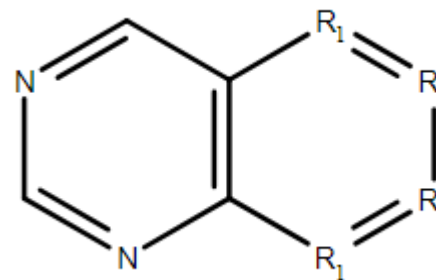
Translate the known information to respective search fields

```
FILE 'REGISTRY' ENTERED
L3      20765955 S 6-6/SZ (S) 10/RATC
L4      2841932  S L3 (S) NCNC3/ESS
L5      2725361  S L4 (S) 6-8 C/REL (S) 2-4 N/REL (S) C N/RELF
```

Consider stopping here and using L5 as the basis for our structure search, or add additional ring sequence information for the second ring:

```
L6      2722462  L5 (S) (2 NCNC3/ESS OR NCNC3/ESS(S)C6/ESS OR
        NCNC3/ESS(S)NC5/ESS OR NCNC3/ESS(S)NC2NC2/ESS
        OR NCNC3/ESS(S)N2C4/ESS)
```

```
L7      STRUCTURE UPLOADED
L8      50 S L7 SSS SAM
L9      1061861 S L7 SSS FUL SUB=L6
L10     1062485 S L7 SSS FUL
```

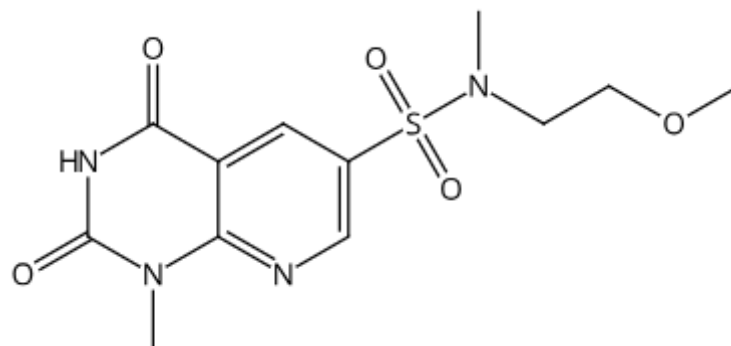
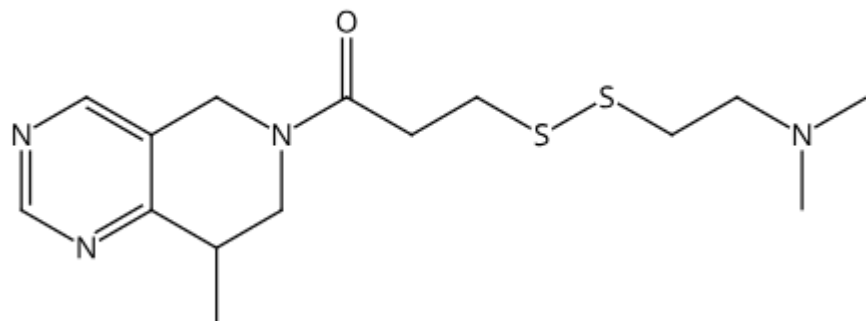
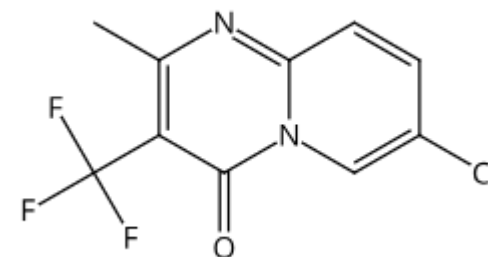
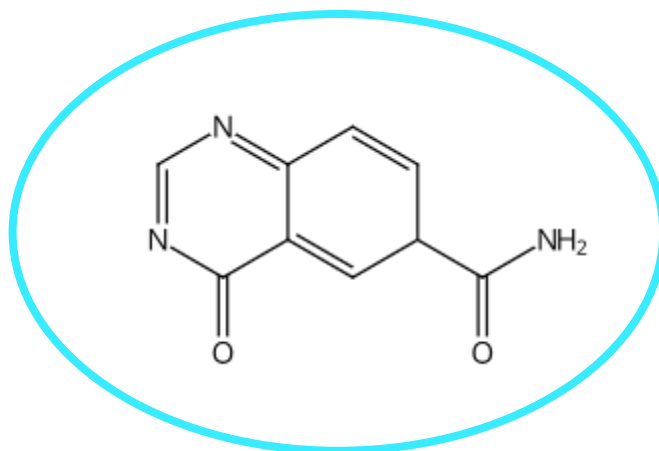
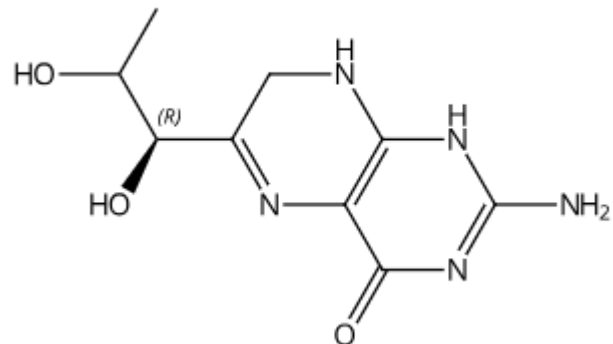


R1=C/N

What are still some ring systems in L6 that do not match our structure query?

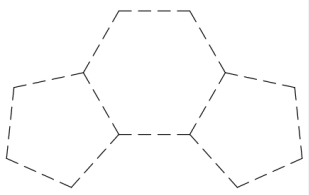
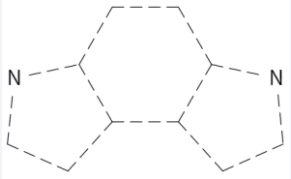
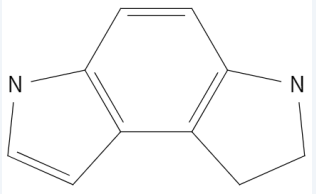
L11 1646033 S L6 NOT L9

L12 93091 S L11 AND 1/NRS



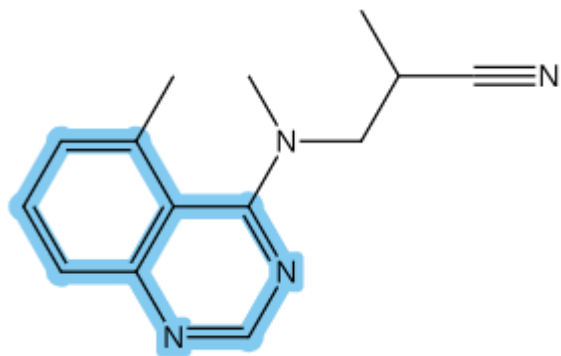
Ring Identifier (RID)

Defines shape – atoms – bonds in a ring system

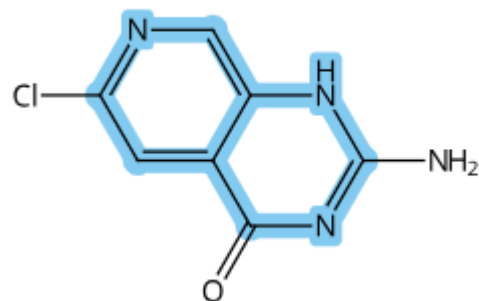
Retrieval	Portion of Ring Identifier	Notes
Skeleton 	1310/RID (leftmost part of the number up to the first period)	<ul style="list-style-type: none">• Any atom may be present• Any type of bonding is allowed• Additional fusion or bridging not allowed
Skeleton and atoms 	1310.20/RID (leftmost part of the number up to the second period)	<ul style="list-style-type: none">• Any type of bonding may be present• Additional fusion or bridging not allowed
Skeleton, atoms, and bonds 	1310.20.3/RID (entire 3-part number)	<ul style="list-style-type: none">• Additional fusion or bridging not allowed

RIDs for our nitrogen heterocycles

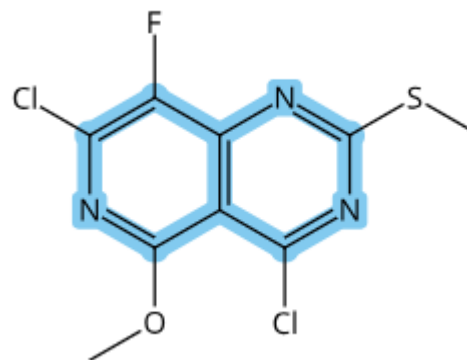
Results from our previous search strategy



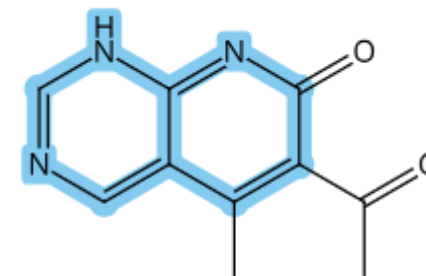
RID: 591.100.47



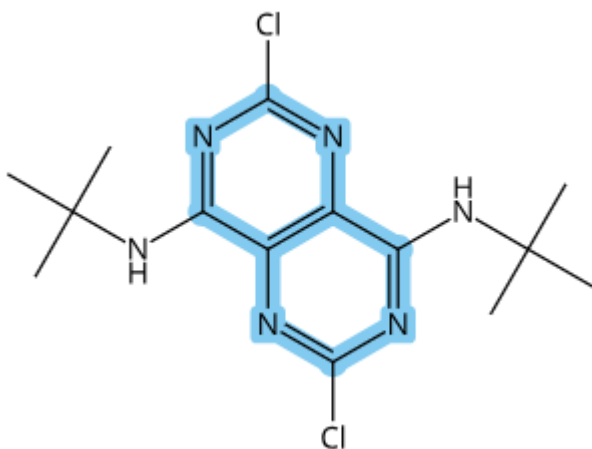
RID: 591.109.6



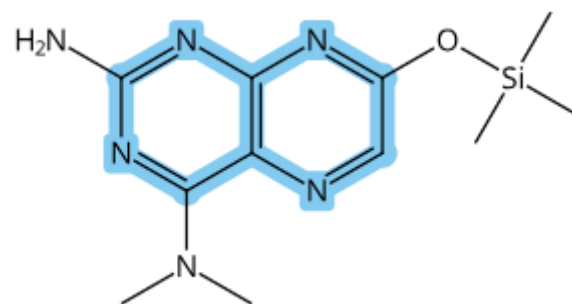
RID: 591.113.10



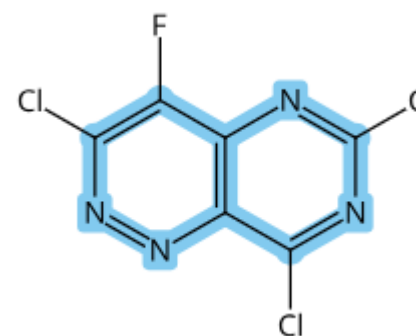
RID: 591.304.27



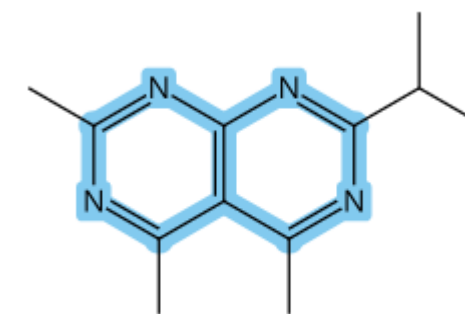
RID: 591.343.3



RID: 591.385.57



RID: 591.338.2



RID: 591.310.9

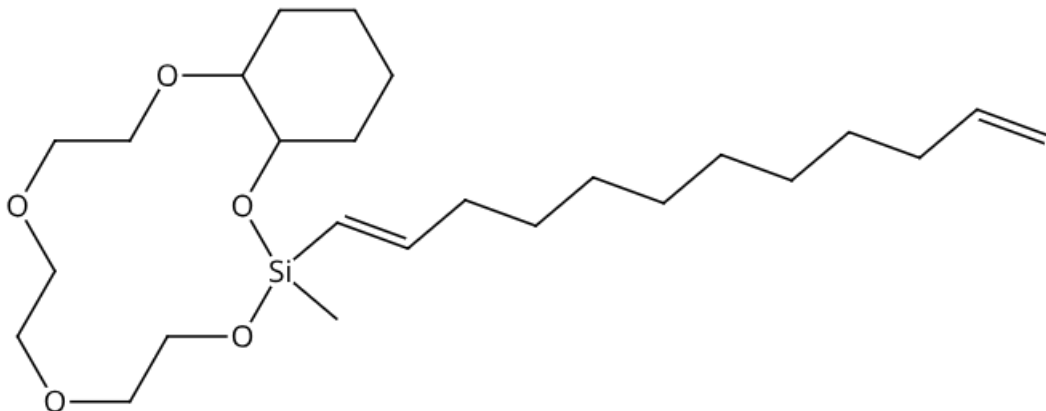
2. Patent claims with broadly described rings

US20230257403 claims silacrown rings very broadly

1. A silacrown ether having eleven to about twenty-seven ring atoms of which one of the ring atoms is silicon, wherein the silacrown ether contains one alkyl group on the silicon atom and at least one exocyclic substituent on the silicon ring atom and/or on a carbon ring atom, and wherein the at

RSD for one example substance

Helpful to see what fields may be useful to include in our strategy



Ring System Data

Elemental Analysis EA	Elemental Sequence ES	Size of the Rings SZ	Ring System Formula RF	Ring Identifier RID	RID Occurrence Count
C6-C8O5Si	C6- OSiOC2OC2OC2O C2	6-14	C12O5Si	5410.197.1	1

The ESS field describing the elemental sequence of the smallest ring starts at the lowest alphabetical heteroatom and finds the shortest path to the next heteroatom

What are the components of the ring search?

Verify which search fields would be relevant

11-27 ring atoms

1 silicon atom in the ring

Ring only consist of Si, O, C atoms

Si atom directly linked to two oxygen atoms

Crown structure of C-C-O repetition

All characteristics must be found in one ring system

What are the components of the ring search?

Verify which search fields would be relevant

11-27 ring atoms

11-27/RACT

1 silicon atom in the ring

1 SI/REL

Ring only consist of Si, O, C atoms

C O SI/RELF

Si atom directly linked to two oxygen atoms

OSiO?/ESS

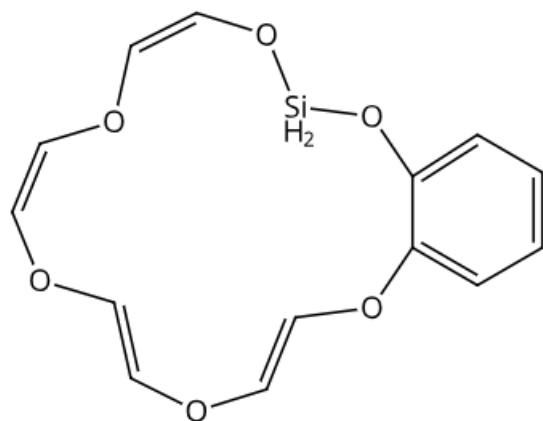
Crown structure of C-C-O repetition

OSiOC2OC2O?/ESS

All characteristics must be found in one ring system (S) proximity

Running the search on STNext

```
FILE 'REGISTRY' ENTERED
L3      16858932 S 11-27/RATC
L4      80356 S L3 (S) 1 SI/REL
L5      9630 S L4 (S) C O SI/RELF
L6      3952 S L5 (S) OSIO?/ESS
L7      47 S L5 (S) OSIOC2OC2O?/ESS
```

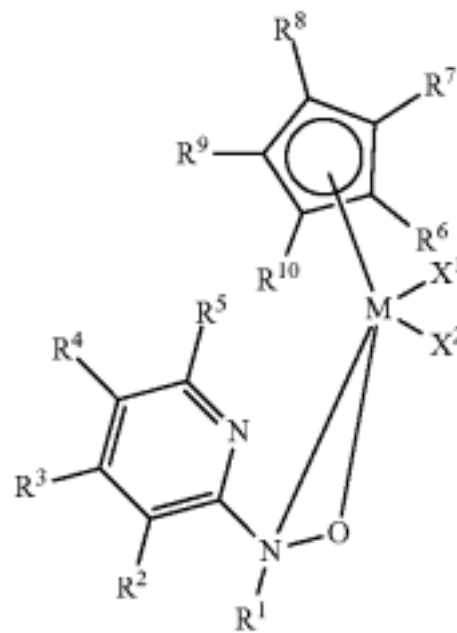
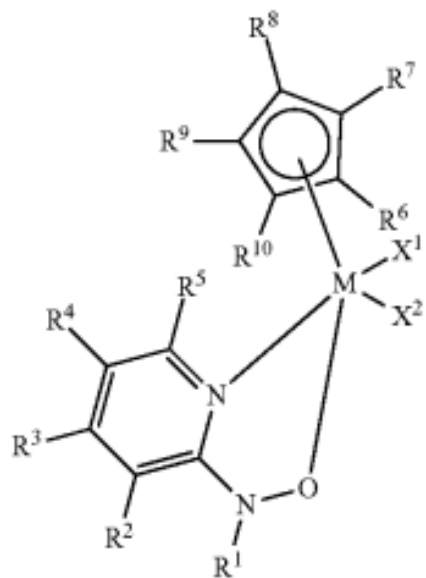


3. Rings with multiple options

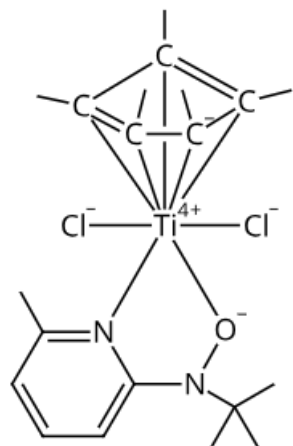
US20220127394 claims metal complexes in two forms

What is claimed is:

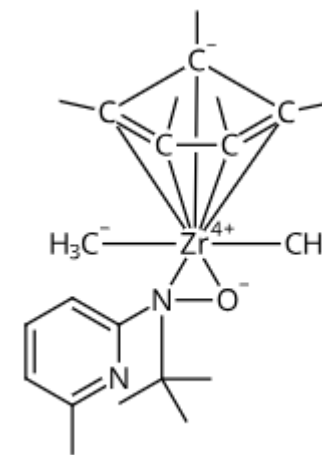
1. A catalyst compound represented by Formula I(a) or I(b):



RSD for two of the exemplified substances



Elemental Analysis EA	Elemental Sequence ES	Size of the Rings SZ	Ring System Formula RF	Ring Identifier RID
C2Ti-C2Ti-	TiC2-TiC2-	3-3-3-3-3-5-	C10N20Ti	1981.162.1
C2Ti-C2Ti-	TiC2-TiC2-	6		
C2Ti-CN20Ti-	TiC2-NOTiNC-			
C5N	NC5			

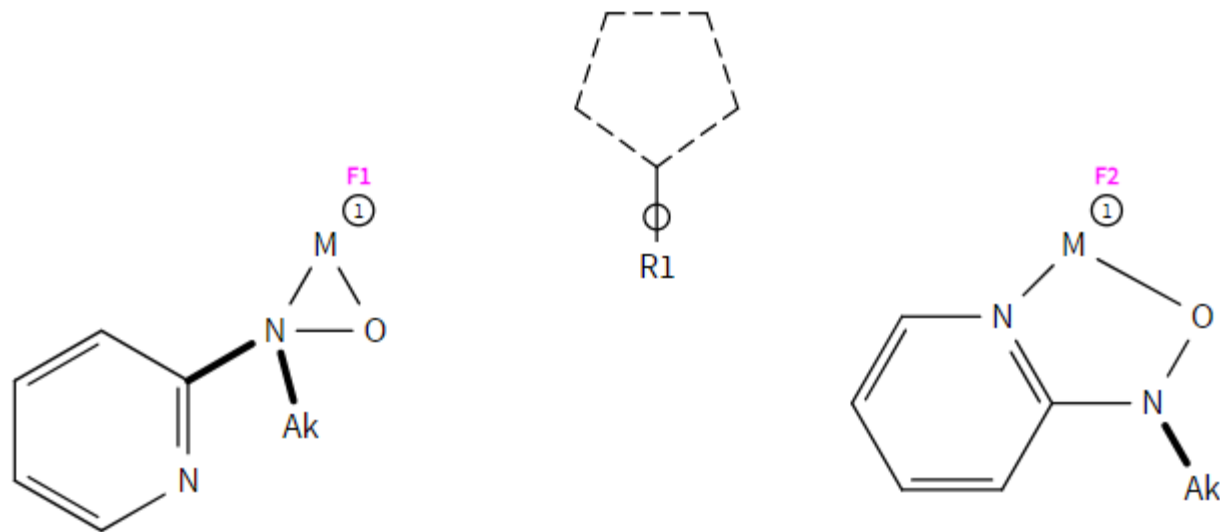


Elemental Analysis EA	Elemental Sequence ES	Size of the Rings SZ	Ring System Formula RF	Ring Identifier RID
C5N	NC5	6	C5N	46.156.30
C2Zr-C2Zr-	ZrC2-ZrC2-	3-3-3-3-3-3	C5N0Zr	112.280.1
C2Zr-C2Zr-	ZrC2-ZrC2-			
C2Zr-NOZr	ZrC2-NOZr			

Could we just search these two RIDs?

Compare results to structure based search

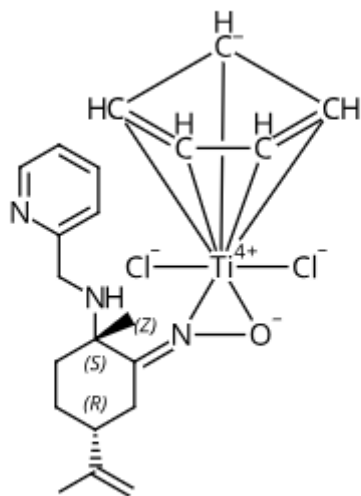
L1 19 S 1981/RID(S)M/REL(S)10 C/REL(S)2 N/REL(S)1 O/REL(S)14/RATC
L2 18 S 46.156.30/RID(L) (112/RID(S)M/REL(S)5 C/REL(S)1 N/REL(S)1 O/REL(S)8/RATC)
L3 37 S L1 OR L2



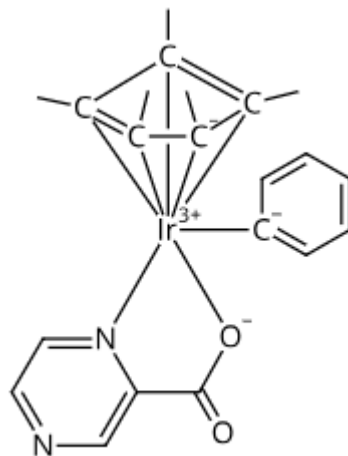
L4 STRUCTURE UPLOADED
L5 32 S L4 SSS FUL
L6 32 S L3 AND L5
L7 5 S L3 NOT L5

What are the 5 extra RSD hits we retrieved?

There were two main structure variations found in L7



• HCl



These are mentioned in 4 prior publications potentially relevant as inventive step

AN 2007:229702 CAPLUS Full-text

DN 146:441913

TI Mono and dinuclear iridium, rhodium and ruthenium complexes containing chelating carboxylato pyrazine ligands: Synthesis, molecular structure and electrochemistry

AU Govindaswamy, Padavattan; Therrien, Bruno; Suess-Fink, Georg; Stepnicka, Petr; Ludvik, Jiri

CS Institut de Chimie, Universite de Neuchatel, Neuchatel, CH-2009, Switz.

SO Journal of Organometallic Chemistry (2007), 692(8), 1661-1671

Analyze by RSD

May be helpful to get an overview of possibilities

Analyze

37 answers available

1-37 ↓ Ascending, → Ascending Switch Axes [Download]

Select up to 2 fields to analyze

- Author / Inventor
- Corporate Source / Patent Assignee
- Publication Year
- Controlled Terms
- Patent Country
- ^ Cooperative Patent Classification
- ^ International Patent Classification
- ∨ Custom Analyze
 - sz
 - rid

	3-3-3-3-3-...	3-3-3-3-3-3	6
112.180.1		10	10
112.280.1		8	8
1981.100.1	1		
1981.161.1	8		6
1981.162.1	8		6
1981.99.1	2		1
46.150.1		2	2
46.150.18	13	12	25
46.156.30		18	18

Analyze Cancel

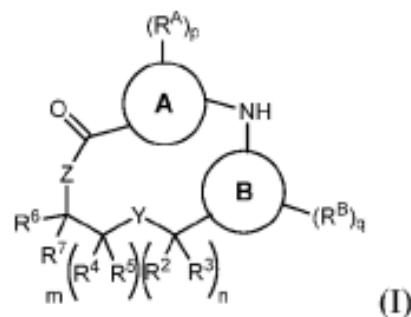
Loading Data: 100% complete

Searching macrocyclic rings

Additional fusion allowed but no formation of bridges

Example taken from WO2023178235

1. A compound represented by Formula I:



or a pharmaceutically acceptable salt and/or a stereoisomer thereof; wherein

Ring A is a 6,5 or 6,6 fused bicyclic heteroaryl containing two, three, or four ring nitrogens;

Ring B is a 6,5 or 6,6 fused bicyclic heteroaryl containing one, two or three ring nitrogens;

m is 1 or 2; Y is selected from the group consisting of O, N(R^a), S(O)_w, CH₂ and a bond;

n is 0 or 1; Z is selected from the group consisting of N(R^a) and O;

Effectively creating an 11-19-membered macrocycle that has 2 times a 5,6 or 6,6-membered ring fused into the ring with 4-10 nitrogens; 0-2 oxygens; 0-1 sulfurs

Developing the search strategy

Converting known elements and ring sizes to RSD search fields

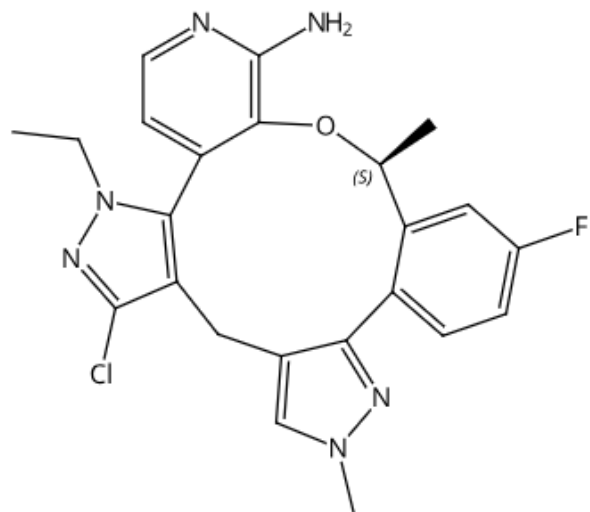
```
FILE 'REGISTRY' ENTERED
L1      258474 S 11-19/SZS(S) ((2 5/SZS(S)2 6/SZS) OR (1 5/SZS(S)3 6/SZS) OR 4 6/SZS)
L2      42901 S L1(S) 4-10 N/REL
L3      31669 S L2(NOTS)M/REL
L4      28823 S L3(NOTS) >2 O/REL
L5      28299 S L4(NOTS) >1 S/REL
L6      19277 S L5(S)5/NRRS
```

It is not possible to search for 0-2 O/REL since the zero value is not searchable. Alternatively, we exclude using (NOTS) rings with more than 2 oxygen atoms, and more than 1 sulfur atom in the ring

Display one of the results

=> D STR RSD

L6 ANSWER 1 OF 19277 REGISTRY COPYRIGHT 2023 ACS on STN

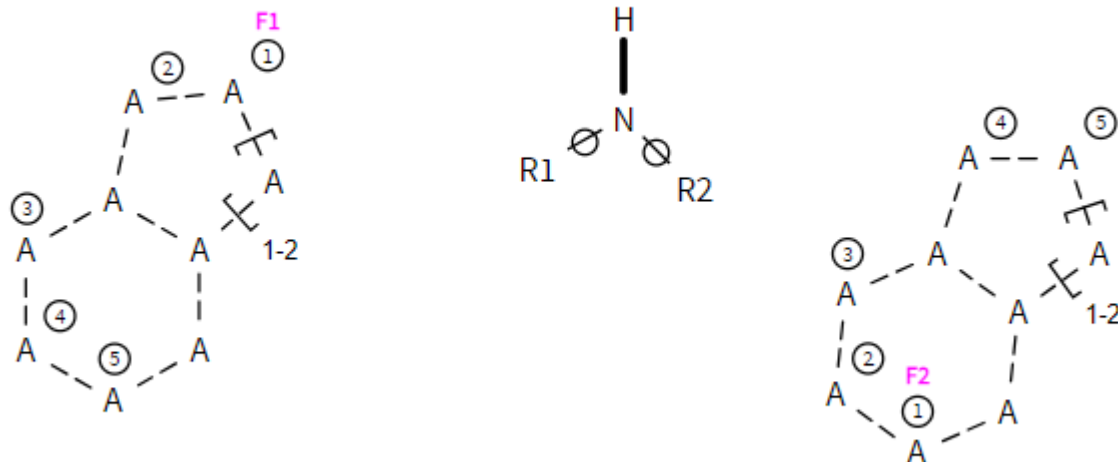


Absolute stereochemistry shown

There is not an option to force rings A and B to be the fused bicyclic 5-6 or 6-6 rings, with the NH group separating them, other than running a substructure search within the results in L6

Creating the structure query

The ring size is either 5-6 or 6-6 for both rings connecting to N-H



Create multiple attachment points for R1 and R2 (except the repeating group)
Since it is unknown which atoms in these rings are C, N, O, S, use variable A
Make sure that the bonds from R1-N and N-R2 are rings bonds!

Run the structure search as a subset of L6

Structure L7 is too broad to be run by itself

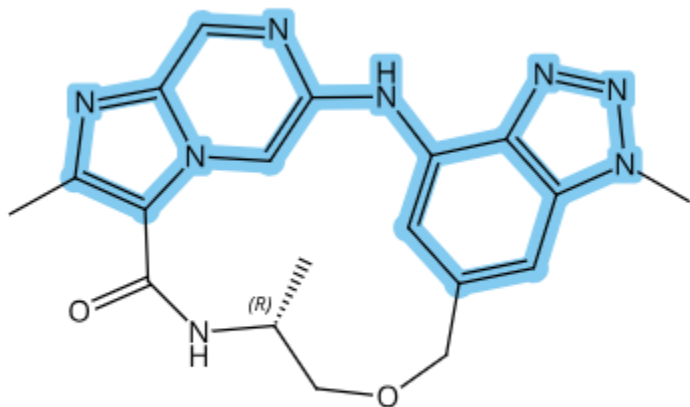
```
L1          258474 S 11-19/SZS (S) ((2 5/SZS (S) 2 6/SZS) OR (1 5/SZS (S) 3 6/SZS)
OR 4 6/SZS)
L2          42901 S L1 (S) 4-10 N/REL
L3          31669 S L2 (NOTS) M/REL
L4          28823 S L3 (NOTS) >2 O/REL
L5          28299 S L4 (NOTS) >1 S/REL
L6          19277 S L5 (S) 5/NRRS

L7          STRUCTURE UPLOADED
L8          15 S L7 SSS FUL SUB=L6

FILE 'CAPLUS' ENTERED
L9          2 S L8
```

Displaying two answers

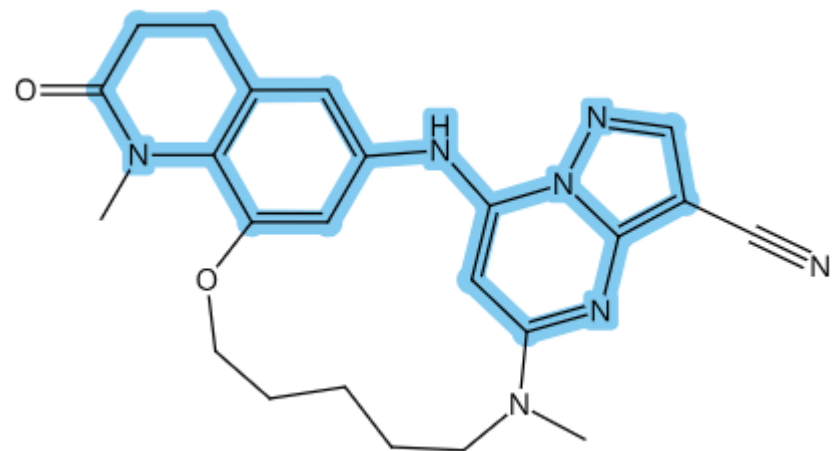
These are the most recent and oldest structures of L8



Absolute stereochemistry shown

RN 2987998-75-2 in WO2023178235

Elemental Analysis	Elemental Sequence	Size of the Rings	Ring System Formula	Ring Identifier	RID	Count
EA	ES	SZ	RF	RID		
C2N3-C3N2-	N3C2-NCNC2-	5-5-6-6-14	C16N8O	208217.21.1	1	
C4N2-C6-	NC2NC2-C6-					
C10N3O	NC2NC2NC2OC4					



RN 2390553-28-1 in J Med Chem (2019), p.9418

Elemental Analysis	Elemental Sequence	Size of the Rings	Ring System Formula	Ring Identifier	RID	Count
EA	ES	SZ	RF	RID		
C3N2-C4N2-	N2C3-NCNC3-	5-6-6-6-14	C20N6O	201064.1.1	1	
C5N-C6-C11N2O	NC5-C6-					
	NC3NC3OC5					

Summary of Ring System Data

Search Ring System Data to refine broad substance searches.

Create a subset when structure queries exceed system limits.

Analyze and search RIDs for easy templates.

Searching RSD is cost effective and faster than structure search.

See [HELP SRING](#); [HELP SFIELDS](#); [HELP SEARCH SUBSET](#)

Between problems
and progress
are connections
that matter



Paul Peters

Director, Customer Success Specialists
ppeters@acs-i.org

CONTACT

CAS

help@cas.org
cas.org

FIZ Karlsruhe

EMEAhelp@cas.org
stn-international.de