



**SCIENCE IP**<sup>®</sup>  
The CAS Search Service

# The Importance of Markush Searching for Chemical Entities

Elaine Cheeseman, Ph.D.  
Fall 2010 ACS Meeting

# Agenda

- What are the patentability requirements of chemical entities?
- Why search the patent and non-patent literature?
- Why is structure searching essential for chemical entities?
- What are the two types of structure-searchable databases for patent and non-patent literature?
- How do exemplified compound databases and Markush databases differ?
- What value does Markush searching provide?



# What are the patentability requirements of chemical entities?

- To be granted a patent, your invention must show
  - utility
  - novelty
  - non-obviousness
- Requirements are the same regardless of invention

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau

(43) International Publication Date  
24 June 2010 (24.06.2010)

(10) International Publication Number  
WO 2010/071853 A1

(51) International Patent Classification:  
C07D 313/04 (2006.01) A61K 31/06 (2006.01)  
A61P 5/00 (2006.01) A61P 9/00 (2006.01)  
A61P 25/00 (2006.01) A61P 37/00 (2006.01)  
A61K 31/47 (2006.01) A61K 31/5377 (2006.01)  
A61P 5/58 (2006.01) A61P 18/00 (2006.01)  
A61P 31/00 (2006.01)

(21) International Application Number:  
PCT/US2009/088865

(22) International Filing Date:  
18 December 2009 (18.12.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
61/203,136 19 December 2008 (19.12.2008) US

(71) Applicant (for all designated States except US): SIRTUIN PHARMACEUTICALS, INC. (US/US), 200 Technology Square, Suite 300, Cambridge, MA 02139 (US).

(72) Inventor(s) and  
Investor(s)/Applicant (for US only): GALLMANN, Christopher (US/US), C/o Sirtin Pharmaceuticals, Inc., 200 Technology Square, Suite 300, Cambridge, MA 02139 (US). DRSCH, Jeremy, S. (US/US), C/o Sirtin Pharmaceuticals, Inc., 200 Technology Square, Suite 300, Cambridge, MA 02139 (US). NG, Pak, Yee (US/US); C/o Sirtin Pharmaceuticals, Inc., 200 Technology Square, Suite 300, Cambridge, MA 02139 (US). PERINI, Robert, B. (US/US), C/o Sirtin Pharmaceuticals, Inc., 200 Technology Square, Suite 300, Cambridge, MA 02139 (US).

(74) Agents: HALSTEAD, David, P. et al.; Ropes & Gray LLP, One International Place, Boston, MA 02110 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AU, AT, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, FI, FR, GB, GR, HU, IE, IT, LI, LU, LV, MT, NL, NO, PL, PT, RO, RU, SE, SI, SK, TR, UA, UK, US), OAPI (BF, BJ, GM, GN, GW, IL, IN, IS, KE, MG, ML, MR, NI, NG, SN, TD, TG).

[Continued on next page]

(54) Title: THIAZOLOPYRIDINE SIRTUIN MODULATING COMPOUNDS

(57) Abstract: Provided herein are novel sirtuin-modulating compounds of Structural Formula (Ia) and methods of use thereof. The sirtuin-modulating compounds may be used for increasing the lifespan of a cell, and treating and/or preventing a wide variety of diseases and disorders including, for example, diseases or disorders related to aging or stress, diabetes, obesity, neurodegenerative diseases, cardiovascular disease, blood clotting disorders, inflammation, cancer, and/or fibrosis as well as diseases or disorders that would benefit from increased mitochondrial activity. Also provided are compositions comprising a sirtuin-modulating compound in combination with another therapeutic agent.

WO 2010/071853 A1

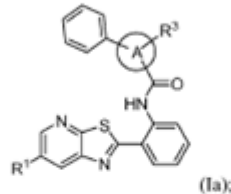
# What do you look for in prior art?

- **Anticipation (lack of novelty)**
  - a reference is said to anticipate an invention if it discloses every element claimed
  - must occur in a single reference
- **Obviousness**
  - usually involves combining more than one reference
  - must be a suggestion or teaching to combine references

WO 2010/071853 PCT/US2009/068865


**CLAIMS:**



1. A compound represented by Structural Formula (Ia):

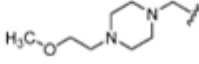
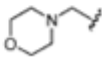
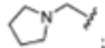


(Ia);

5 or a pharmaceutically acceptable salt thereof, wherein:



ring A is selected from  and , wherein "\*" represents a portion of ring A bound to phenyl, and "•••••" represents a portion of ring A bound to C=O in the compound;

R<sup>1</sup> is selected from hydrogen, ,  and ;

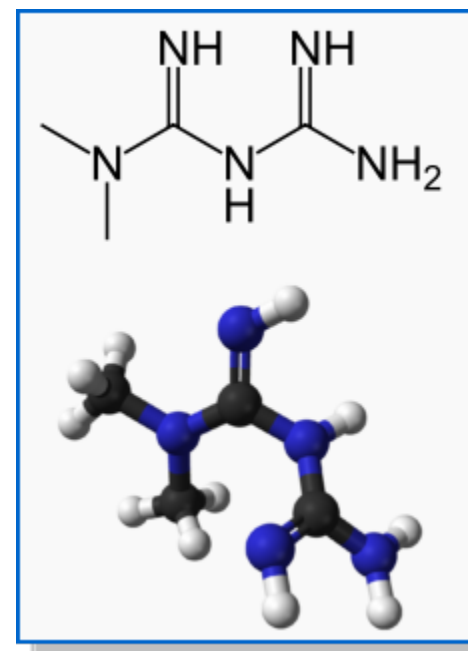
10

# Why search the patent and non-patent literature?

*To gain a comprehensive background (in pharmaceutical treatment of diabetes, for example)*

- Obtain an overall picture of the art in your area of interest
- Become acquainted with the known compounds in your area of interest
- Obtain background information to determine if your invention is patentable
- Explore issues concerning freedom-to-operate, validity, etc.

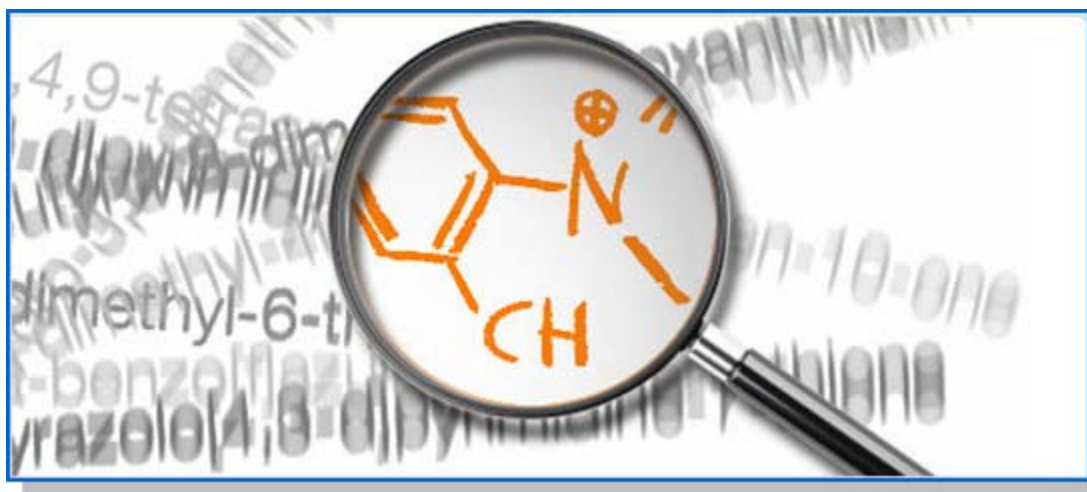
## Metformin



# Why is structure searching essential for chemical entities?

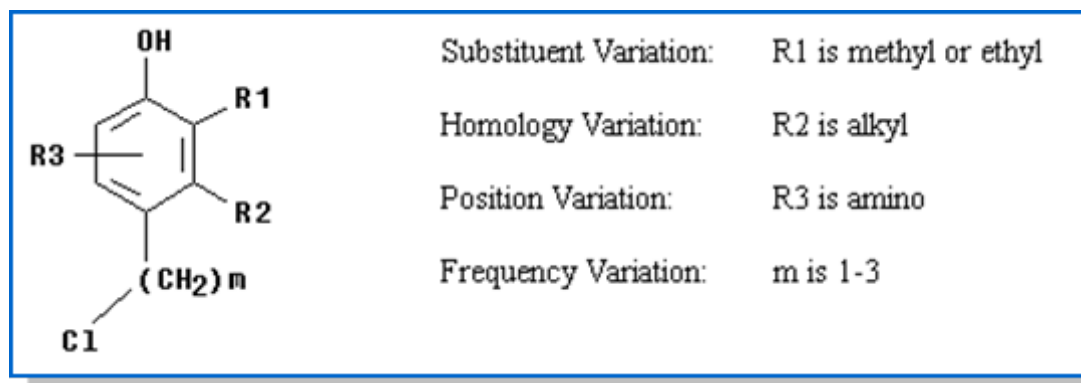
- Identifies the substance of interest precisely
- Useful in locating both known and research compounds
- Best technique to identify chemical entities depicted by structures in references

*Chemical name searching is not precise, but can be useful as a supplement to gain additional references not linked to a structure*



# What are the two types of structure-searchable databases for patent and non-patent literature?

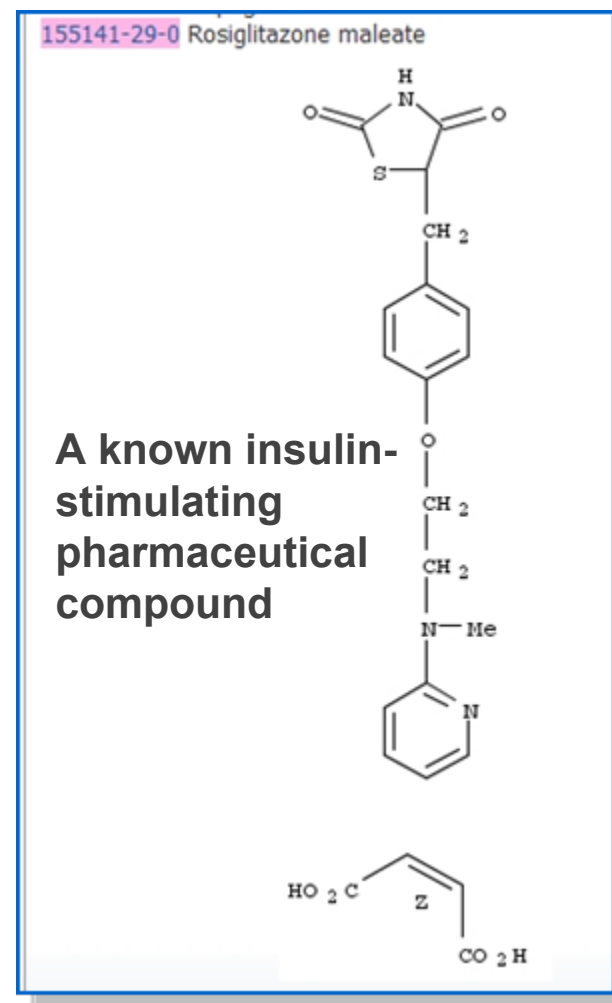
- **Exemplified compound databases**
  - information on specific substances
- **Markush databases**
  - information from Markush structures in patent claims or disclosures
  - R-groups of patent Markush structures are defined to search all possible permutations



# How do exemplified compound databases differ from Markush databases?

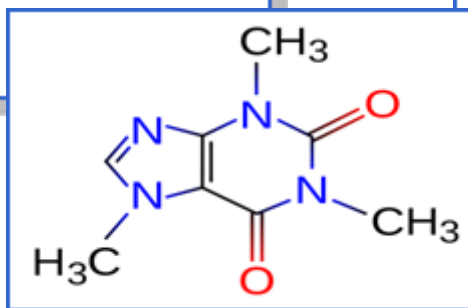
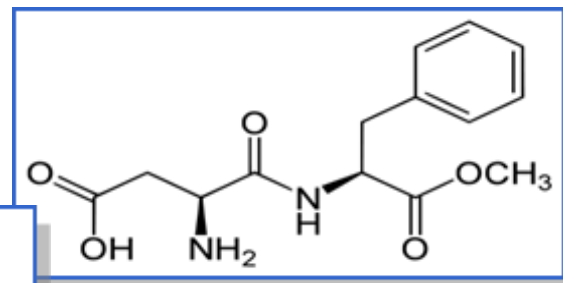
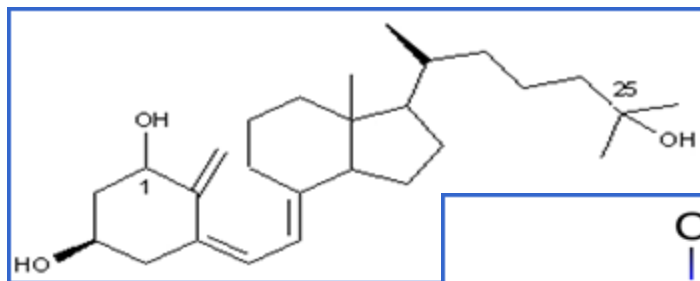
*Search exemplified databases when you need information about specific substances:*

- Are these substances known?
- What are the references to these substances—patents, non-patents, other disclosures?
- What are the uses of these substances?
- Are commercial sources available for these substances?



# Access important exemplified compound databases through STN<sup>®</sup>

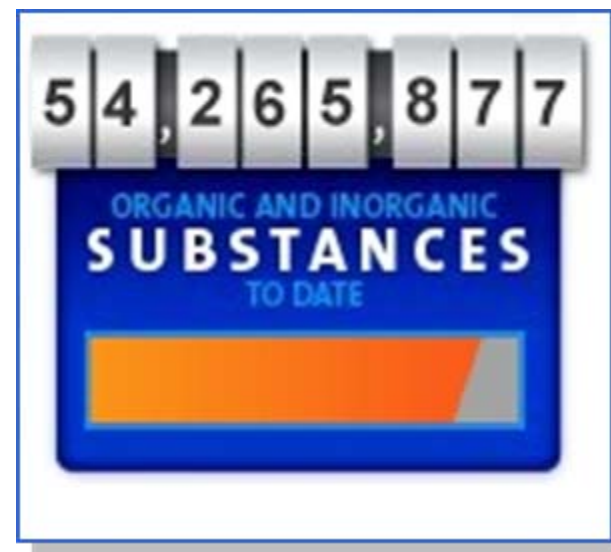
- CAS REGISTRY<sup>SM</sup>/CAplus<sup>SM</sup>
- Derwent World Patents Index<sup>®</sup> (WPINDEX)
- ReaxysFile



# CAS REGISTRY—the world's largest source of chemical substance information (more than 54 million)

*CAS REGISTRY contains a variety of substances:*

- Compounds associated with preparations, property data, etc.
- Compounds that are specifically mentioned in patent claims
- Compounds that play a key role in the journal article or patent
- Reactants and isolated intermediates
- Specific substances from the examples with non-characterizing data



# A variety of useful information is available in a CAS REGISTRY record

```
RN  57-88-5  REGISTRY
ED  Entered STN:  16 Nov 1984
CN  Cholest-5-en-3-ol (3 $\beta$ )- (9CI)  (CA INDEX NAME)
OTHER CA INDEX NAMES:
CN  Cholesterol (8CI)
OTHER NAMES:
CN  (-)-Cholesterol
CN   $\Delta^5$ -Cholesten-3 $\beta$ -ol
CN  3 $\beta$ -Hydroxycholest-5-ene
CN  5:6-Cholesten-3 $\beta$ -ol
CN  Cholest-5-en-3 $\beta$ -ol
CN  Cholesterin
CN  Cholesteryl alcohol
CN  Dythol
CN  Lidinit
CN  Lidinite
CN  NSC 8798
CN  Provitamin D
FS  STEREOSEARCH
DR  732297-95-9, 793670-51-6, 80356-33-8, 209124-38-9, 218965-24-3,
    262418-13-3, 378185-03-6, 676322-57-9
MF  C27 H46 O
CI  COM
...
```

**CAS Registry Number<sup>®</sup>**

**CAS Systematic Name - current**

**CAS Systematic Name - previous**

**Trade names, synonyms**

**Deleted CAS Registry Numbers**

**Molecular formula**

# A variety of useful information is available in a CAS REGISTRY record (cont.)

...

LC STN Files: ADISNEWS, AGRICOLA, ANABSTR, BEILSTEIN\*, BIOBUSINESS, BIOSIS, BIOTECHNO, CA, CABA, CANCERLIT, CAOLD, CAPLUS, CASREACT, CBNB, CEN, CHEMCATS, CHEMINFORMRX, CHEMLIST, CIN, CSChem, CSNB, DDFU, DETHERM\*, DIOGENES, DIPPR\*, DRUGU, EMBASE, GMELIN\*, HODOC\*, HSDB\*, IFICDB, IFIPAT, IFIUDB, IPA, MEDLINE, MRCK\*, MSDS-OHS, NAPRALERT, NIOSHTIC, PDLCOM\*, PIRA, PROMT, RTECS\*, SPECINFO, TOXCENTER, TULSA, ULIDAT, USAN, USPAT2, USPATFULL, VETU, VTB

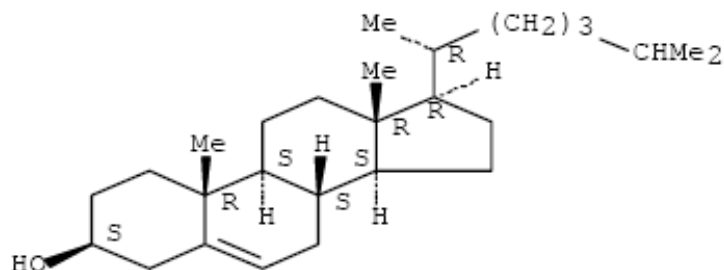
(\*File contains numerically searchable property data)

Other Sources: DSL\*\*, EINECS\*\*, TSCA\*\*

(\*\*Enter CHEMLIST File for up-to-date regulatory information)

Absolute stereochemistry.

**Databases on STN that include this CAS RN**



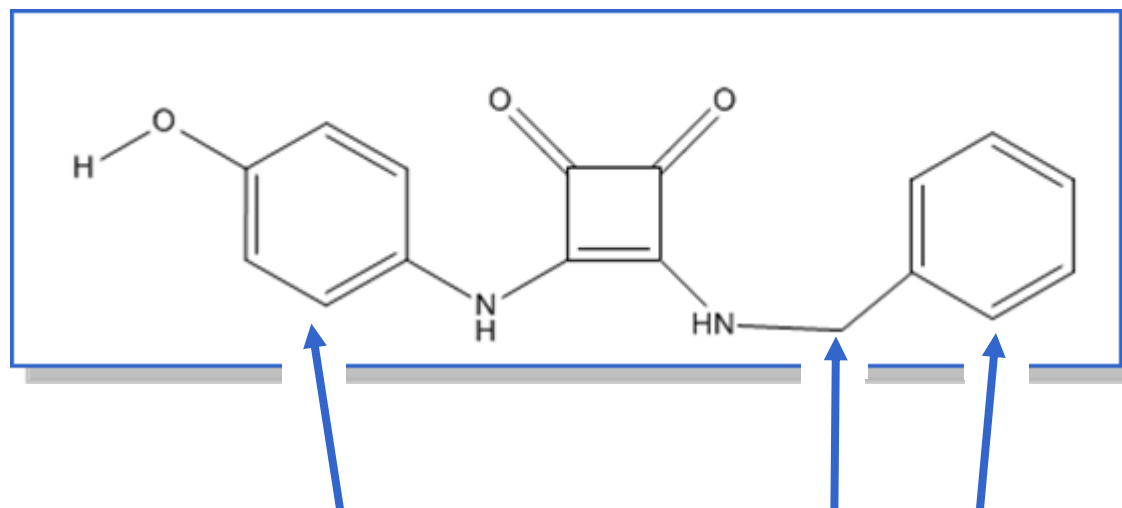
**The substance structure image**

107642 REFERENCES IN FILE CA (1907 TO DATE)  
 9568 REFERENCES TO NON-SPECIFIC DERIVATIVES IN FILE CA  
 107781 REFERENCES IN FILE CAPLUS (1907 TO DATE)

**The number of references in CAS<sup>SM</sup>/Caplus**

# Let's try an example of a structure search in CAS REGISTRY

*Locate patents for squaric acid compounds with these specific substitutions:*

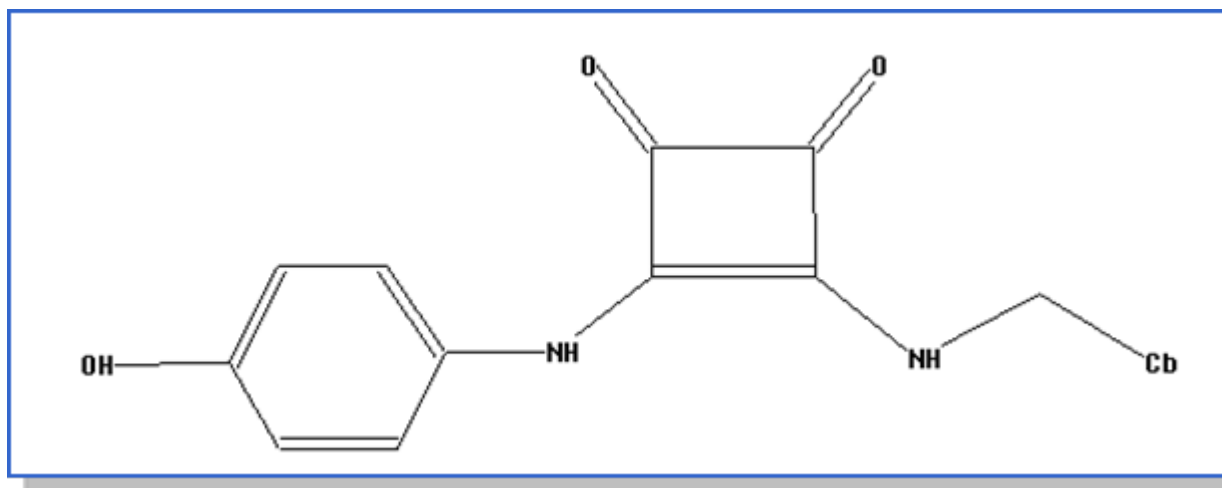


Substitution is allowed on all open sites

# Use STN Express<sup>®</sup> software to create your structure query

Define your query attributes:

- No additional fusion on phenyl or cyclobutene
- Additional substitution is allowed
- Cb = carbocycle: unsaturated, monocyclic



# Search for substances first in exemplified compound database REGISTRY, then in Markush database MARPAT

## Approach:

1. Structure search in CAS REGISTRY
2. Cross over REGISTRY hits to CAplus
3. Markush structure search in MARPAT for patent records

# Reference from CAplus retrieved from the REGISTRY structure search

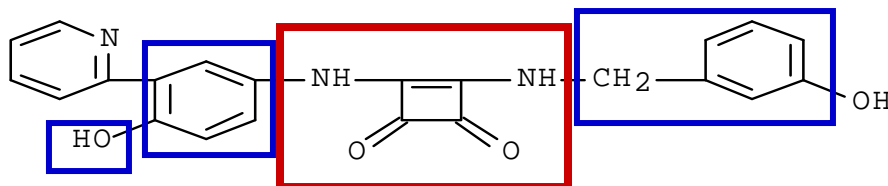
## Reference:

L14 ANSWER 1 OF 3 CAPLUS COPYRIGHT 2010 ACS on STN  
 AN 2007:145730 CAPLUS [Full-text](#)  
 DN 146:229183  
 TI Preparation of N-pyridinylphenyl-3-4-diaminocyclobut-3-ene-1,2-diones as  
 CHK1, CHK2, and/or SGK kinase inhibitors for treating cancer

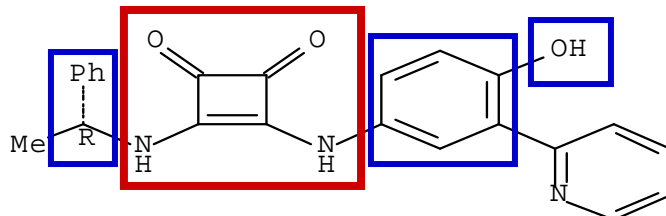
. . .

## Hit structures:

RN 924909-93-3 CAPLUS  
 CN 3-Cyclobutene-1,2-dione, 3-[[[(3-hydroxyphenyl)methyl]amino]-4-[[4-hydroxy-3-(2-pyridinyl)phenyl]amino]- (CA INDEX NAME)



RN 924909-95-5 HCAPLUS  
 CN 3-Cyclobutene-1,2-dione, 3-[[[4-hydroxy-3-(2-pyridinyl)phenyl]amino]-4-[[[(1R)-1-phenylethyl]amino]- (CA INDEX NAME)

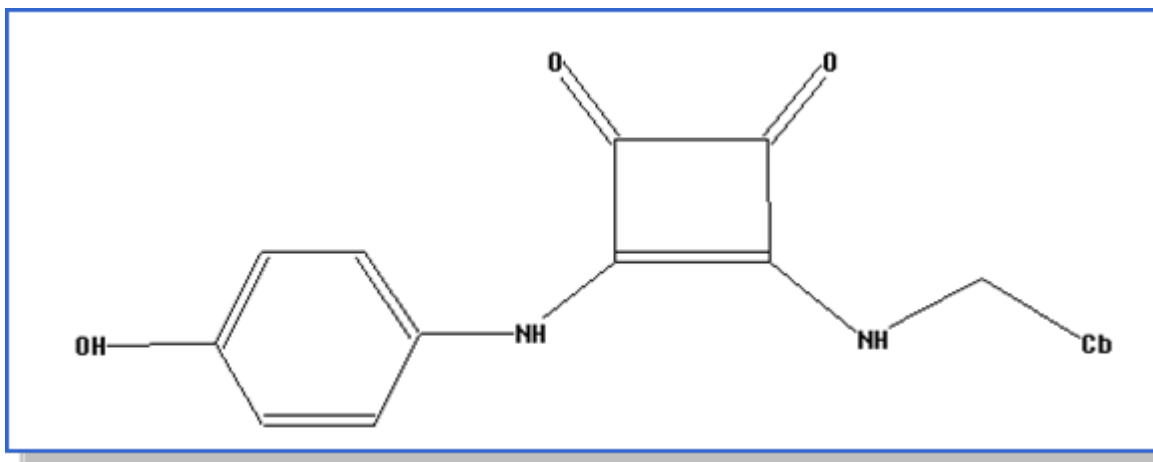


# How do Markush databases differ from exemplified compound databases?

*Search Markush databases to see if a substance is part of a Markush structure patent disclosure*

## A Markush structure search

- Retrieves patent records of interest even if the record does not refer to a specific substance
- Helps determine whether your substances of interest are included in a patent's broad Markush claims



# MARPAT<sup>®</sup> is a Markush database, available only on STN

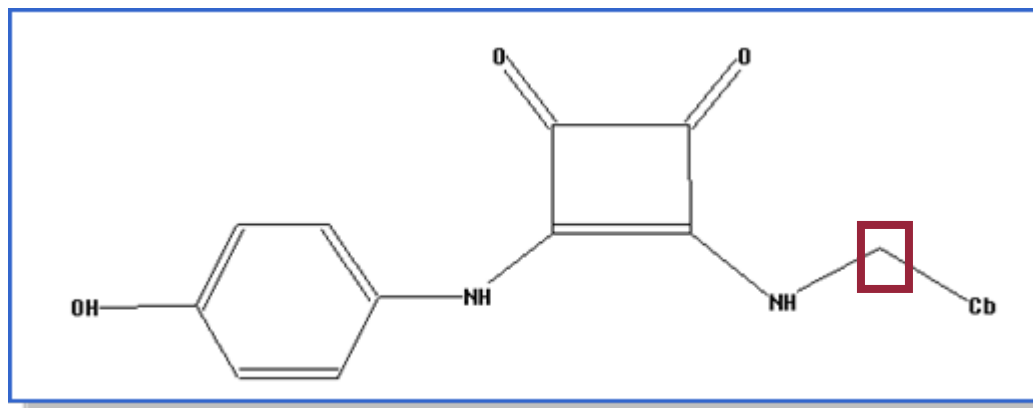
<b><i>Producer</i></b>	<b>CAS</b>
<b>Content</b>	Markush structures covering organic and organometallic compounds in patents
<b>Coverage</b>	1988 to date (also includes 1961-1987 records from INPI)
<b>Approximate numbers</b>	>300,000 patent records >800,000 Markush structures

# Modify your STN Express structure query for Markush searching by adjusting the nodes

Nodes are disclosed in the following ways:

- Specifically: methylene only
- Specifically or generically: methylene or “alkylene”
- Specifically, generically, or with a text descriptor: methylene, “alkylene”, or “linking group”

*MARPAT settings influence the type and amount of hits*



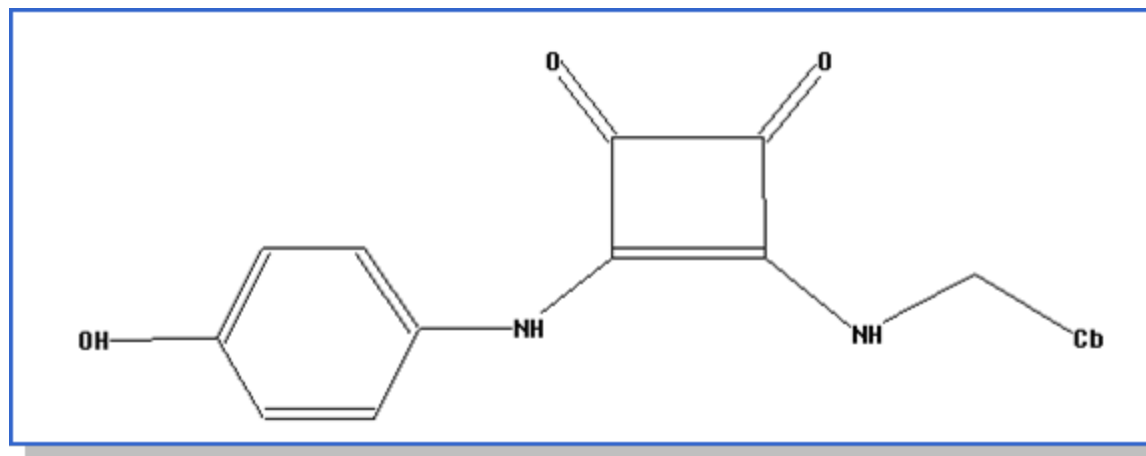
# Tailor your query using MARPAT settings to retrieve substances covered by Markush claims

## Query attributes:

- No additional fusion on phenyl or cyclobutene
- Additional substitution is allowed
- Cb = carbocycle: unsaturated, monocyclic

## MARPAT settings:

- The cyclobutene must be specifically disclosed
- All other nodes can be disclosed specifically or generically



# What value does Markush searching provide?

## 10 additional (unique) patent records

### Approach:

<b>1. Structure search in CAS REGISTRY</b>	<b>• 120 compounds</b>
<b>2. Cross over REGISTRY hits to CAplus</b>	<b>• 3 references (all patent records)</b>
<b>3. Markush structure search in MARPAT for patent records</b>	<b>• 12 patent records</b>
<b>Eliminate duplicates from REGISTRY/CAplus</b>	<b>• (-2 patent records)</b>
<b>MARPAT records unique to MARPAT</b>	<b>• 10 patent records!</b>

# Unique hits retrieved from MARPAT

## Reference:

L14 ANSWER 3 OF 3 MARPAT COPYRIGHT 2010 ACS on STN

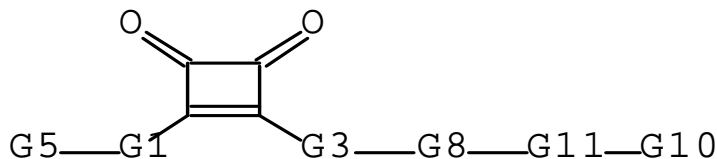
AN 124:175615 MARPAT [Full-text](#)

TI Preparation of substituted N-aryl-1,2-diaminocyclobutene-3,4-dione smooth muscle relaxants

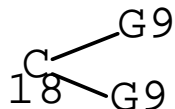
...

## Query hit:

### MSTR 1



G1 = NH  
 G3 = NH  
 G5 = Ph (opt. substd. by (1-2) G6)  
 G6 = OH  
 G8 = 18



G10 = Ph (opt. substd. by (1-2) G6)  
 G11 = (0-6) CH2

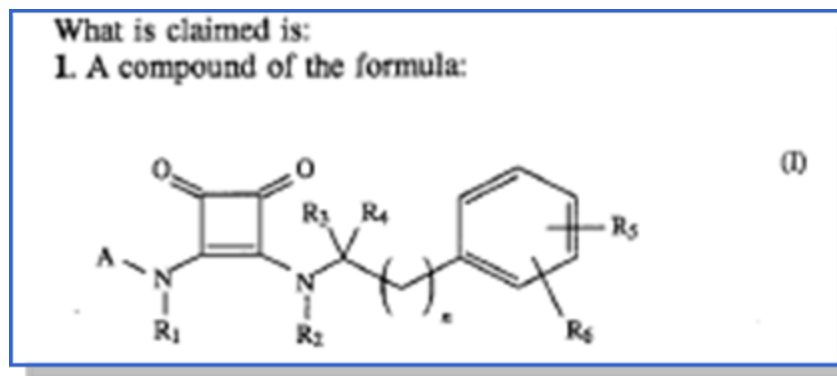
Derivative: or pharmaceutically acceptable salts

Patent location: claim 1

# Summary: the importance of Markush searching for chemical entities

- Locate chemical entities precisely with structure searching
- Begin structure searches in exemplified compound databases for specific compounds
- Search your structure query in MARPAT to supplement your retrieval

*With MARPAT, you can locate additional references not found in exemplified compound databases*



# Put the information retrieval talents of Science IP experts to work for you

- **Tap into the knowledge base of searchers who understand the scientific research process**
  - All have many years of experience searching CAS and STN databases across many scientific areas
  - All have advanced scientific degrees (a number have a Ph.D. in chemistry)
- **Obtain thoughtful and thorough answers to your scientific questions**



**[www.scienceip.org](http://www.scienceip.org)**

**1-866-360-0814 (toll free in N. America)**

**E-mail: [scienceip@cas.org](mailto:scienceip@cas.org)**

***Thank You!***



# Appendix

- **Use the following slides if audience has questions.**

## Results of the CAS REGISTRY and MARPAT searches

### Approach:

- Structure search was run in CAS REGISTRY
- Hits searched in CAplus; limited to patents (L6)
- Structure search was run in MARPAT
- CAplus hits NOTed out (L11)

### Results:

L12      13 DUP REM L6 L11 (0 DUPLICATES REMOVED)  
ANSWERS '1-3' FROM FILE HCAPLUS  
ANSWERS '4-13' FROM FILE MARPAT

# Types of structure-searchable databases

- **Exemplified compound databases:**
  - Search for indexed specific compounds
  - Examples:
    - CAS REGISTRY/CAplus
    - Derwent World Patents Index<sup>®</sup> (WPINDEX)
    - ReaxysFile
  - All of the above databases are on STN.
- **Markush databases:**
  - Search for compounds which fall within a Markush structure in the claims or disclosure.
  - Searches the various permutations of the R-groups.
  - There is no requirement that the compounds were made.
  - Often retrieves patent records which are not retrieved in the exemplified compound databases.
  - Examples:
    - MARPAT
    - MMS (also includes specific compounds)

## CAS REGISTRY/CAplus

- **Producer: Chemical Abstracts Service**
- **Document types: Patents, non-patents, conference proceedings, dissertations, etc.**
- **Content: chemistry and all related areas**
- **Coverage in years: 1907 – present; some earlier records**
- **Additional facts:**
  - CAS Registry/CAplus are full implementation is only on STN
  - CAS Registry/CAplus are the main databases in SciFinder®—
  - A structure search in CAS Registry retrieves CAS Registry Numbers® which can be searched in other databases to get references (patent and non-patent literature).
  - CAplus is the companion reference database to CAS Registry. It includes:
    - enhanced titles and abstracts
    - patent family information for patent records.
    - detailed structure (CAS Registry Numbers) and concept indexing

## MMS (Merged Markush Service)

- **Producer: Thomson-Reuters and INPI for the older information**
- **Host: Questel-Orbit**
- **Document types: patents**
- **Content and coverage in years: pharmaceuticals (1978 – present), agrochemicals (1982 – present), and general chemicals (1982 – present)**
- **Approximate number of compounds: > 2.5 million structure records**
- **Additional facts:**
  - Includes both specific compounds and Markush structures
  - Includes organic compounds, peptides, organometallics, etc.
  - Bibliographic information is retrieved from the World Patent Index (“Derwent”).

# Markush database: MARPAT

- **Producer: Chemical Abstracts Service**
- **Content: Markush structures covering organic and organometallic compounds in patents.**
- **Coverage: 1988 – date (also includes 1961 – 1987 records from INPI)**
- **Approximate numbers: > 300K patent records; > 780K Markush structures**
- **Additional facts:**
  - Records include the Markush indexing, bibliographic information, and abstracts (the latter two are the same as in CAplus).
  - Available only on STN.
  - The same patent record may be retrieved in the CAS REGISTRY/CAplus search (due to the specific compounds) and from the MARPAT search (due to the Markush indexing). Duplicates can be removed.

## MARPAT settings:

- **Query node settings (“match level”): nodes disclosed:**
  - Specifically (“atom”): retrieves only that node or group (narrowest)
    - E.g. : 1,2,4 triazine will only retrieve hits with the 1,2,4 triazine.
  - Specifically or generically (“class”): retrieves that group or a generic version of it
    - E.g. 1,2,4 triazine: will retrieve that triazine, “heterocycle”, “heterocycle” which fits in with the element count (i.e., 1 or more nitrogens), or “ring”
  - Specifically, generically, or with a text descriptor (“any”): (broadest)
    - E.g. **Chloro**-phenyl-1,2,4 triazine: This setting will retrieve chloro-phenyl, “halo”-phenyl, or “optionally substituted”-phenyl based on the terminology in the patent.
- **Settings used depend on the type and amount of hits you want to retrieve.**

## Example: table of results

Table was made using the STN Express Table Tool from the transcript.

Answer	Accession Number	Patent Number	Patent Assignee/Corporate Source	Title
L22-1-OF-10 HCAPLUS	2007:145730 HCAPLUS <a href="#">Full-Text</a>	WO-2007014607 DE-102005035741 AU-2006275160 CA-2616682 EP-1910277 JP-2009502820 US-20080312244	Merck Patent GmbH, Germany	Preparation of N-pyridinylphenyl-3-4-diaminocyclobut-3-ene-1,2-diones as CHK1, CHK2, and/or SGK kinase inhibitors for treating cancer
L22-2-OF-10 HCAPLUS	2006:1066603 HCAPLUS <a href="#">Full-Text</a>	WO-2006105865 DE-102005015254 DE-102005039066 AU-2006231023 CA-2603478 EP-1866288 JP-2008534633 US-20090036449	Merck Patent G.m.b.H., Germany	Preparation of 3-amino-4-[(phenylmethyl)amino]-3-cyclobutene-1,2-diones as CHK1, CHK2 and SGK kinase inhibitors
L22-3-OF-10 HCAPLUS	2006:678244 HCAPLUS <a href="#">Full-Text</a>	WO-2006072354 DE-102005001053 AU-2005324119 CA-2594388 EP-1838662 JP-2008526790 US-20080262096	Merck Patent GmbH, Germany	Preparation of 1,2-dianilino-1-cyclobutene-3,4-diones and related compounds as serine threonine kinase (SGK) inhibitors
L22-4-OF-10 MARPAT	144:350556 MARPAT <a href="#">Full-Text</a>	WO-2006032372 DE-102004046492 AU-2005287667 CA-2580924	Sanofi-Aventis Deutschland G.m.b.H., Germany	Preparation of 4-phenyltetrahydroisoquinolines as transport protein NHE-1 inhibitors

## Appendix: Drawbacks of searching only full-text patent databases on the Internet

- **Not structure searchable.**
  - Often research compounds are only referred to as structures; cannot be retrieved by term searching.
  - Structure searching on chemical structures: retrieves hits not retrieved by term searching
- **Search by terms; need to search all names/chemical names and variations.**
- **Search by codes: generally sufficiently precise.**
- **No intellectual indexing or enhanced title/abstracts.**
- **Often the query options are limited; cannot search with complex queries as can be done on STN.**
- **Often no indication of patent family members of the same invention.**