



SCIENCE IP[®]
The CAS Search Service

CAS Databases as Key Sources of Pharmaceutical Information

Elaine Cheeseman, Ph.D.
Chemistry & the Law Division
Fall 2011 ACS National Meeting

Agenda

- **STN[®] databases and CAS databases**
- **Exemplified compound databases vs. Markush databases**
- **Reaction, sequence and polymer searching**
- **Common search techniques and requests**
- **Example of a pharmaceutical search**
- **Multiple ways to access CAS/STN information**

CAS supports the mission of the ACS

ACS Mission

To advance the broader chemistry enterprise and its practitioners for the benefit of Earth and its people.

CAS Mission

To provide the world's best digital research environment to search, retrieve, analyze and link chemical information.



STN delivers expertly indexed sci-tech and patent content that is superior in quality, scope and currency

Bibliographic databases with detailed structure and concept indexing

- Both patent and non-patent databases
- CAS REGISTRYSM, CAplusSM, Derwent World Patent Index[®], biomedical databases, etc.

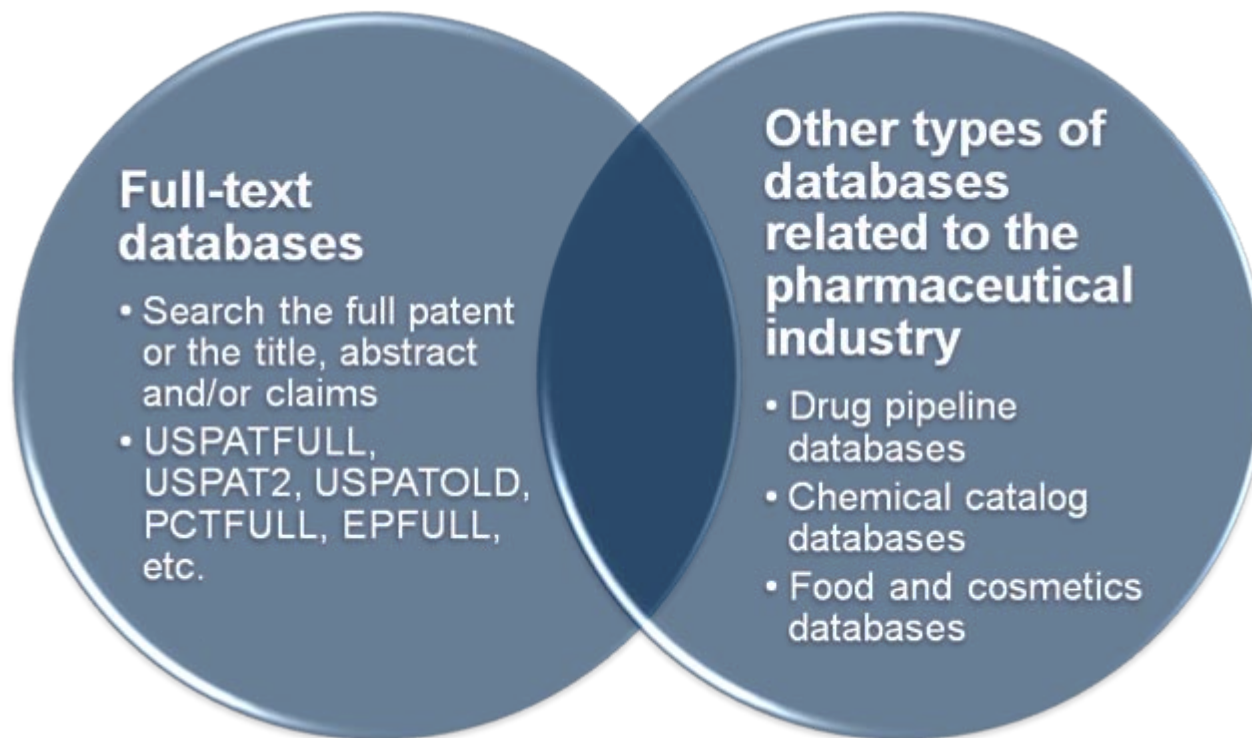
Exemplified compound databases and their reference databases

- Search for specific compounds
- CAS REGISTRY (exemplified compound database) and CAplus (reference database)

Markush database

- Search for Markush structures from patents
- MARPAT[®]

STN delivers expertly indexed sci-tech and patent content that is superior in quality, scope and currency



```
=> d 15 an ti abs
```

```
L1 ANSWER 15 OF 30684 USPATFULL on STN
```

```
AN 2011:192552 USPATFULL
```

```
TI Substituted Imidazopyridinyl-Aminopyridine compounds
```

STN provides access to the premier chemistry collection produced by CAS—CAS REGISTRY and CPlus

Biochemistry	Pharmacology, Plant biochemistry, Toxicology, Immunochemistry, Animal nutrition and Plant nutrition
Organic chemistry	General organic chemistry, Aliphatic compounds, Heterocyclic compounds, Carbohydrates, Terpenes and terpenoids, Steroids, Amino acids, Peptides and Proteins
Macromolecular chemistry	Synthetic high polymers, Synthetic elastomers, natural rubber, Industrial carbohydrates, Surface-active agents and Detergents
Applied chemistry	Industrial inorganic chemicals, Metals and alloys, Pharmaceuticals, Waste treatment and disposal, Fossil fuels and related products
Physical, Inorganic & Analytical chemistry	Nuclear technology, Electric phenomena, Analytical chemistry, Thermodynamics and Thermal properties

CAS REGISTRY is widely recognized as the “gold standard” for chemical substances

- **Types of substances**

- Organic compounds
- Inorganic compounds
- Coordination compounds
- Sequences
- Polymers
- Compounds referred to by designation (e.g., lab code)
- Salts
- Hydrates
- Mixtures

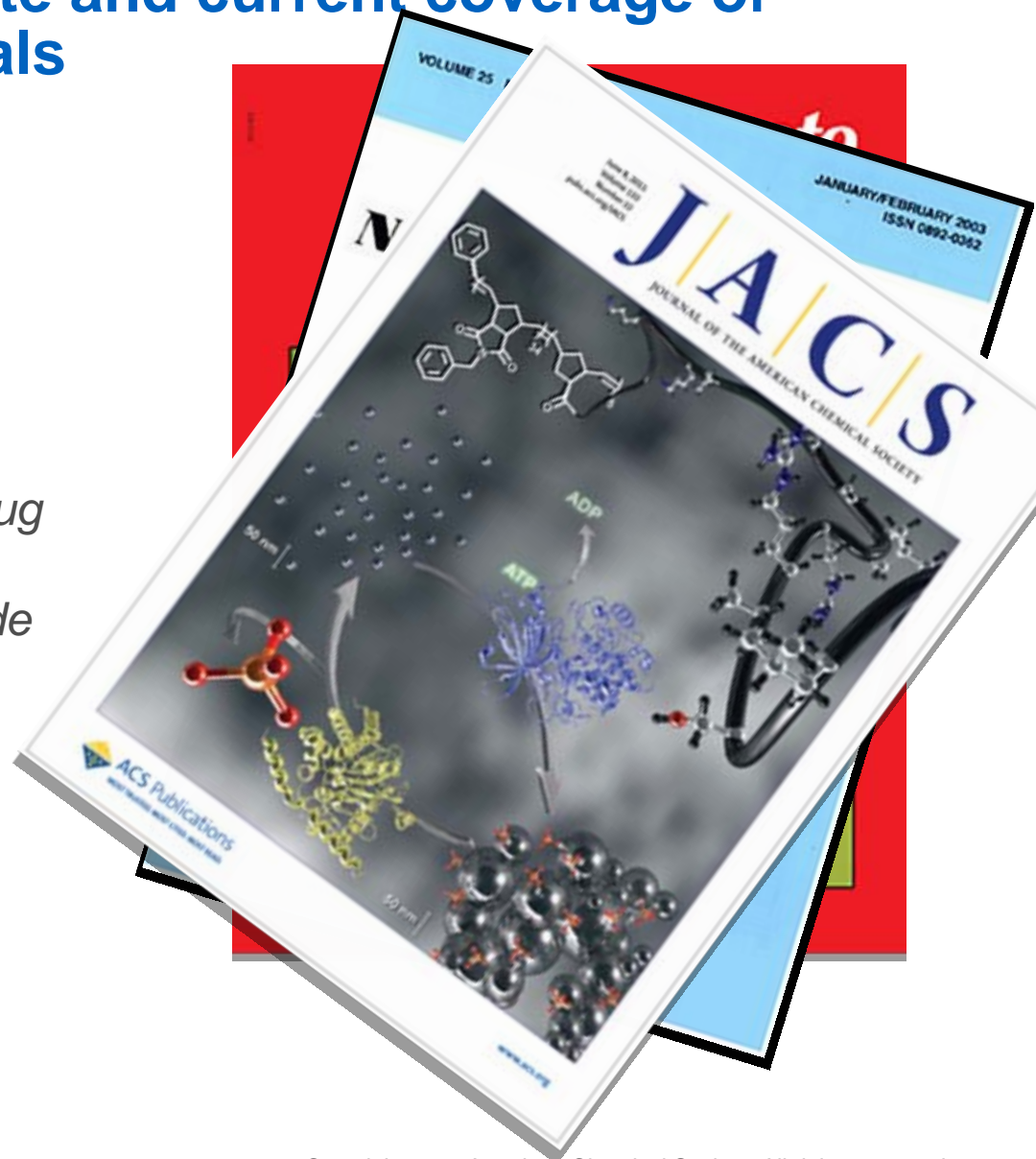
- **Substance sources**

- Patent documents
- Non-patent sources including scientific journals, chemical catalogs and select substance collections on the Web



CPlus provides complete and current coverage of chemically related journals

- More than 10,000 journals
- 1,500 core journals, with bibliographic information and abstracts available within 7 days
 - Core journals include
 - *Biopharmaceutics & Drug Disposition*
 - *Journal de Pharmacie de Belgique*
 - *Neurotoxicology and Teratology*



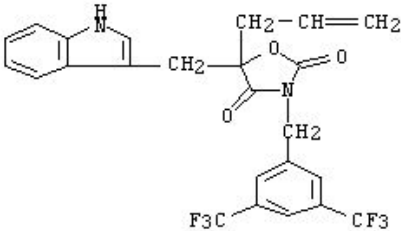
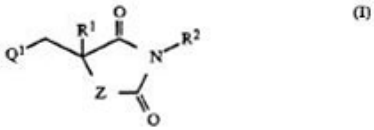
CAS patent coverage is the most comprehensive and timely

- 61 patent authorities covered for chemically related inventions
- 9 major patent offices
 - Bibliographic information available within 2 days of publication
 - Full indexing within 27 days
 - Offices include:



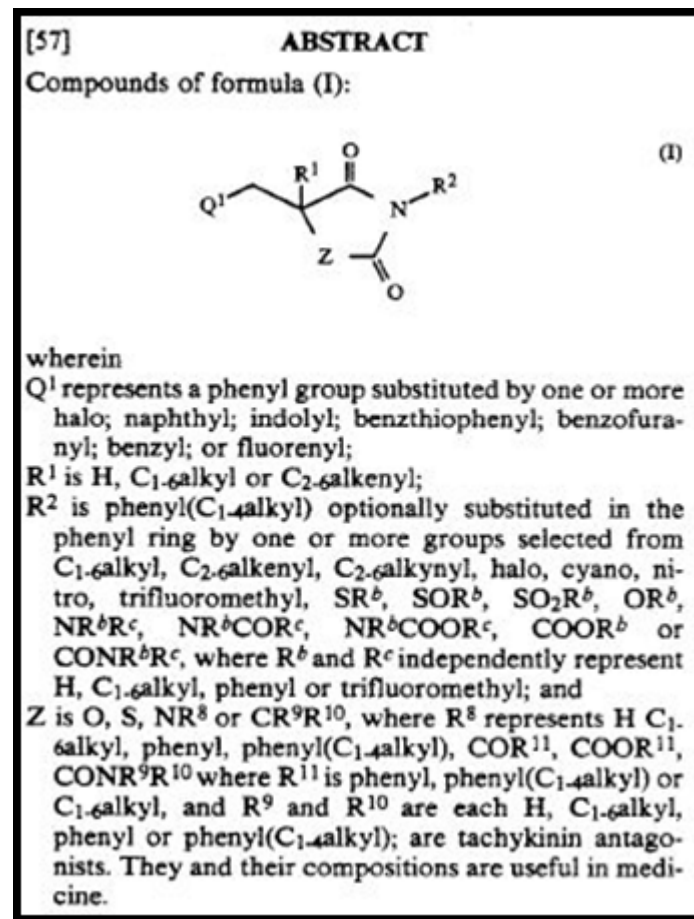
What's the difference?

Exemplified compound databases vs. Markush databases

Type of database	Content	Example	Databases
Exemplified compound databases	Specific compounds (including multi-component species)	REGISTRY COPYRIGHT 2011 ACS on STN 	CAS REGISTRY, WPINDEX [®]
Markush database	Markush structures	[57] ABSTRACT Compounds of formula (I):  wherein Q ¹ represents a phenyl group substituted by one or more halo; naphthyl; indolyl; benzthiophenyl; benzofuran-yl; benzyl; or fluorenyl;	MARPAT [®]

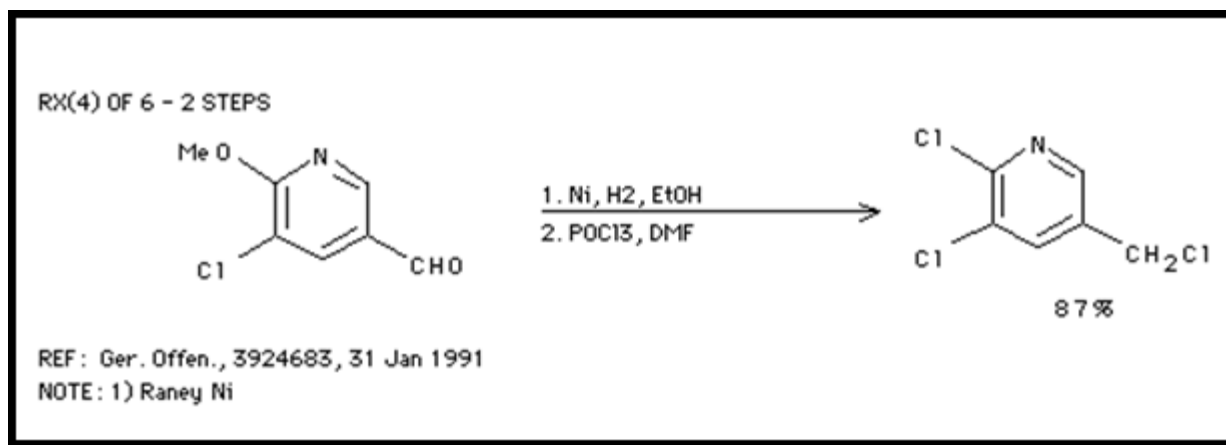
MARPAT searches can retrieve additional sources not found in exemplified compound databases

- Search all permutations of R-groups in Markush structures in patents
 - Compounds do not have to be made or disclosed to be retrieved
- First disclosure of a compound may be embedded in a Markush structure in a patent
- MARPAT settings, narrow to broad:
 - Specifically: e.g. oxazolidine
 - Specifically or generically: indole or “any heterocycle” or “any ring”
 - Specifically, generically or with a text descriptor: allyl, alkyl or optionally substituted



The CAS reaction database provides precise, dependable and timely information on synthetic organic research

- More than 48 million single- and multi-step reactions and synthetic preparations
- Coverage
 - 1840 – present; 30,000 to 50,000 reactions added weekly
 - Sources include journals, patents, dissertations and evaluated reference works
- More than 1.9 million experimental procedures available for searching in SciFinder®



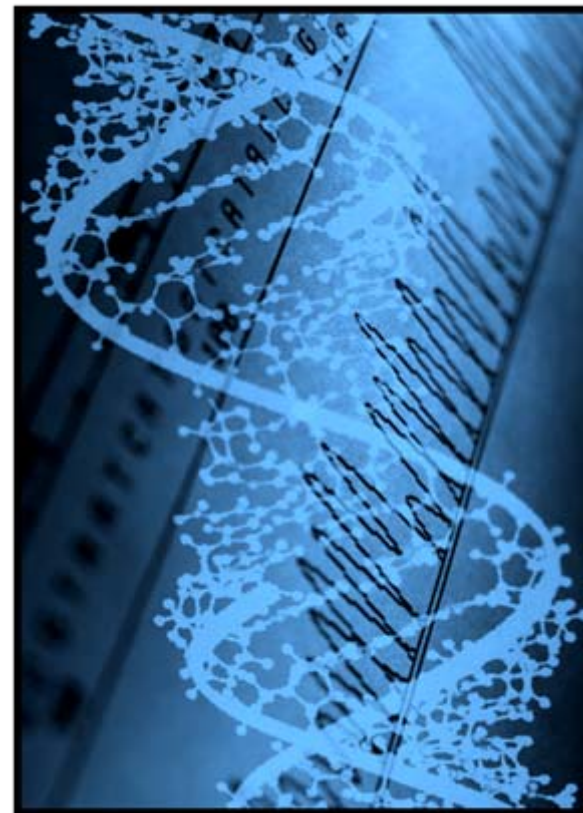
CAS databases are the world's largest and most comprehensive source of polymer information

- **Polymers are often used in pharmaceutical formulations**
 - As excipients to give the desired properties of tablets or capsules
 - In controlled-release drug delivery systems
- **CAS REGISTRY contains >1.3 million polymers**
- **STN also has ample polymer information in its Polymer, Bioscience and Formulation clusters**



STN has multiple sequence databases for comprehensive sequence searching

- Search for sequences using BLAST[®], GETSIM and/or sequence code match
- Use sequence searchable databases on STN:
 - CAS REGISTRY
 - DGENE (Derwent Geneseq)
 - PCTGEN
 - USGENE
- Search by sequence only or sequence and name

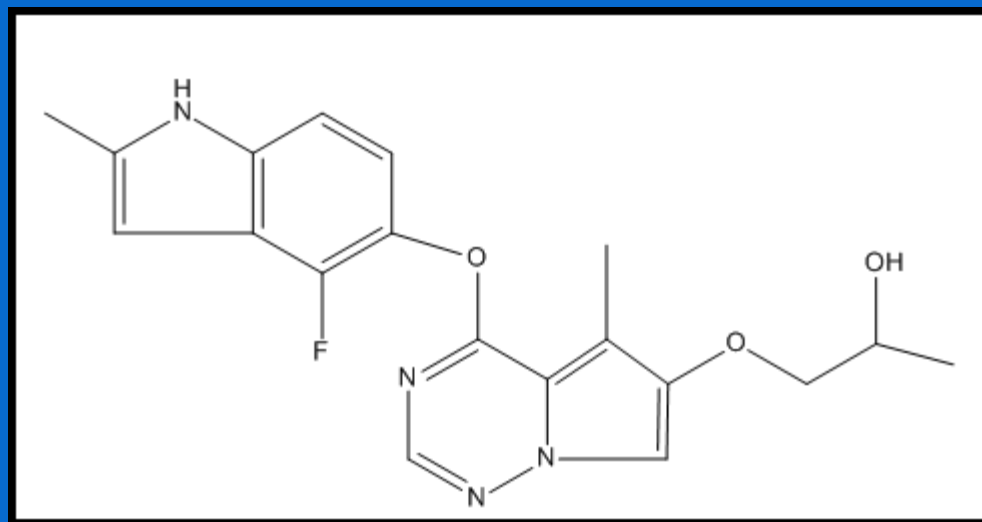


STN searches allow for comprehensive and precision searching

- **Use flexible proximity operators such as (nW), (nA), (L), (S) and (P)**
 - RECOMBINANT(4A)HUMAN(4A)ANGIOTENSIN
- **Upload multiple lines of terms/synonyms or search in multiple databases for comprehensive retrieval**
- **Search in the Basic Index or specific fields (e.g., /TI, /IT, /PN)**
 - ANGIOSTATIN/TI
- **Combine answer sets (L-numbers) to retrieve the answers you need**
 - S L1(5a)L5 AND L10
- **Search in multiple databases for comprehensive retrieval**
 - DUPLICATE REMOVE/IDENTIFY options are available

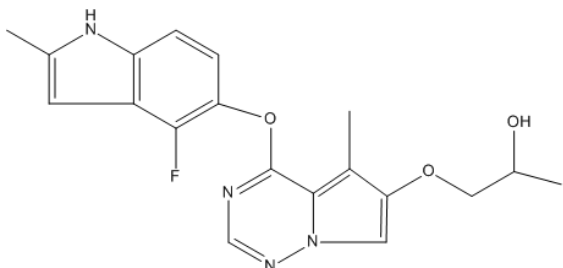
Search Example:

Search for patent references to compounds related to Brivanib (a tyrosine kinase inhibitor for the treatment of cancer).

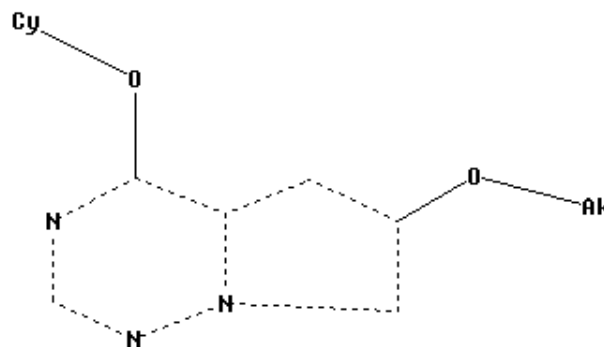


Example: Search by substructure

Brivanib:



Query for Brivanib-like compounds:

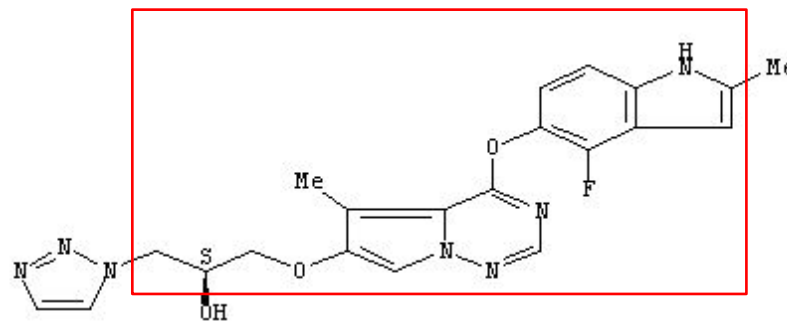
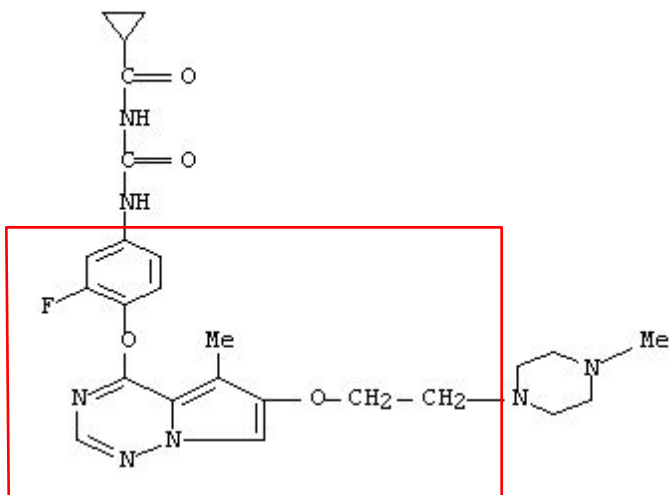


Features of the query

- Cy = any ring system
- Ak = alkyl
- --- = unspecified bonds
- Additional substitution is allowed
- Additional fusion on the fused ring is not allowed
- MARPAT settings:
 - Fused ring must be specifically disclosed
 - All other nodes can be disclosed specifically or generically

Example: Structure search in CAS REGISTRY

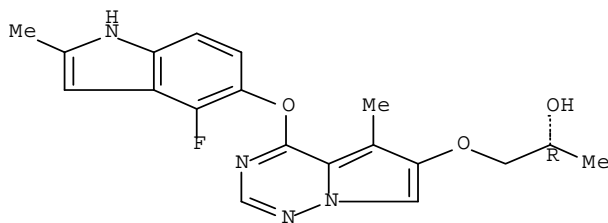
- Upload the structure query
 - The sample substructure search gave 11 compounds in REGISTRY
- Review the sample hits
 - The full substructure search gave 254 compounds in REGISTRY



Example: CAS REGISTRY record for Brivanib

RN 649735-46-6 REGISTRY
ED Entered STN: 14 Feb 2004
CN 2-Propanol, 1-[[4-[(4-fluoro-2-methyl-1H-indol-5-yl)oxy]-5-methylpyrrolo[2,1-f][1,2,4]triazin-6-yl]oxy]-, (2R)- (CA INDEX NAME)
OTHER NAMES:
CN BMS 540215
CN **Brivanib**
FS STEREOSEARCH
MF C19 H19 F N4 O3
SR CA
LC STN Files: CA, CAPLUS, CASREACT, CHEMCATS, EMBASE, IPA, TOXCENTER, USPAT2, USPATFULL

Absolute stereochemistry.



****PROPERTY DATA AVAILABLE IN THE 'PROP' FORMAT****

20 REFERENCES IN FILE CA (1907 TO DATE)

1 REFERENCES TO NON-SPECIFIC DERIVATIVES IN FILE CA

Example: Name and chemical name search strategy

```
L1 que Brivanib?  
L2 que BMS582664 or BMS(w)582664 or BMS582(w)664 or BMS(w)582(w)664  
L3 que BMS540215 or BMS(w)540215 or BMS540(w)215 or BMS(w)540(w)215  
L4 que fluoro(30a)methyl(35a)indol?(30a)(oxy or yloxy)(35a)((methylpyrrolo or  
pyrrolo)(10w)triazin?)(35a)(propanol or propan(2w)ol or propan?(35a)hydroxy  
Or (methylethyl or methyl(2w)ethyl)(35a)alanin? Or AMINOPROPANO? Or  
amino(35a)(propanoic or propanoate))  
L5 que L1 or L2 or L3 or L4  
L6 que L1/ti,ab,clm or L2/ti,ab,clm or L3/ti,ab,clm or L4/ti,ab,clm (optional)
```

Example: Run the search and combine the hits

1. Search by retrieved CAS Registry Numbers®
2. Search by names/chemical names in bibliographic databases and in the patent full-text databases
3. Search in MARPAT
4. Remove duplicates and combine the hits

```
=> DUP REM L18 L26 L38 L59
...
L109          127 DUP REM L18 L26 L38 L59 (0 DUPLICATES REMOVED)
              ANSWERS '1-29' FROM FILE HCAPLUS
              ANSWERS '30-39' FROM FILE MARPAT
              ANSWERS '40-88' FROM FILE USPATFULL
              ANSWERS '89-127' FROM FILE PCTFULL
```

Example: CPlus hit record

L109 ANSWER 17 OF 127 HCAPLUS COPYRIGHT 2011 ACS on STN
 AN 2007:1215848 HCAPLUS Full-text
 DN 147:449087
 TI Process for the preparation of L-alanine
 (1R)-2-[4-(4-fluoro-2-methyl-1H-indol-5-yloxy)-5-
methylpyrrolo[2,1-f][1,2,4]triazin-6-yloxy]-1-methylethyl ester
 IN Crispino, Gerard A.; Hamedi, Mourad; LaPorte, Thomas L.; Thornton, John
 E.; Pesti, Jaan A.; Xu, Zhongmin; Lobben, Paul C.; Leahy, David K.;
 Muslehiddinoglu, Jale; Lai, Chiajen; Spangler, Lori Ann; Discordia, Robert
 P.
 PA Bristol-Myers Squibb Company, USA
 SO U.S. Pat. Appl. Publ., 15 pp.
 CODEN: USXXCO
 DT Patent
 LA English
 FAN.CNT 1

	PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
	-----	----	-----	-----	-----
PI	US 20070249610	A1	20071025	US 2007-737787	20070420
	US 7671199	B2	20100302		
	AU 2007240453	A1	20071101	AU 2007-240453	20070419
	CA 2650059	A1	20071101	CA 2007-2650059	20070419
	...				

Example: Abstract and other codes from CPlus record

AB The invention relates to an improved process for preparing the title compound (brivanib alaninate), a compound which has been shown to be useful for the treatment of certain types of cancer. Intermediate 4-chloro-5-methylpyrrolo[2,1-f][1,2,4]triazin-6-yl pivalate (I) was coupled with 4-fluoro-5-hydroxy-2-methylindole and the product reacted with (R)-propylene oxide and then Cbz-L-alanine. I was prepared from 5-methyl-4-oxo-3,4-dihydropyrrolo[2,1-f][1,2,4]triazine-6-carboxylic acid Et ester by methylation with MeMgCl, continuous oxidation followed by quench reaction, protection with pivalic acid, and chlorination. The continuous process involves better heat transfer and the use of significantly less material in the reactor than a batch process, thus preventing potentially dangerous thermal runaway.

INCL 514243000; 544184000

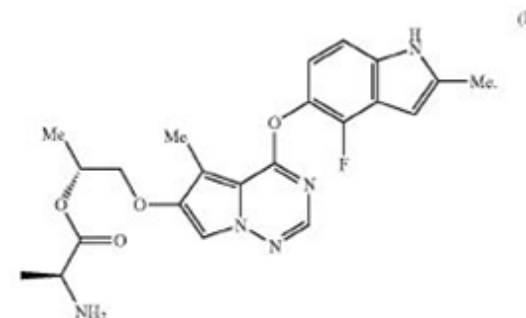
IPCI A61K0031-53 [I,A]; C07D0487-04 [I,A]; C07D0487-04 [I,A];
A61K0031-53 [N,A]; A61P0003-10 [N,A]

IPCR A61K0031-53 [I,A]; C07D0487-04 [I,A]; A61P0003-10 [N,A]

NCL 514/243.000; 544/184.000; 544/183.000

CC 34-2 (Amino Acids, Peptides, and Proteins)
Section cross-reference(s): 1, 28

(57) **ABSTRACT**
The invention relates to an improved process for preparing [(1R,2S)-2-aminopropionic acid 2-[4-(4-fluoro-2-methyl-1H-indol-5-yloxy)-5-methylpyrrolo[2,1-f][1,2,4]triazin-6-yloxy]-1-methylethyl ester of the formula:



Compound I has been shown to be useful for the treatment of certain types of cancer.

Example: Intellectual chemical and concept indexing

```

ST  brivanib alaninate prepn antitumor;
    indolyloxypyrrolotriazinomethylethyl alaninate fluoromethyl
IT  Antitumor agents
    Neoplasm
      (preparation of brivanib alaninate)
IT  25952-53-8
    RL: RGT (Reagent); RACT (Reactant or reagent)
      (claimed; preparation of brivanib alaninate)
IT  649735-63-7P
    RL: PAC (Pharmacological activity); SPN (Synthetic preparation); THU
      (Therapeutic use); BIOL (Biological study); PREP (Preparation); USES
      (Uses)
      (preparation of brivanib alaninate)
IT  1142-20-7  15448-47-2, R-Propylene oxide, reactions  288385-88-6
    RL: RCT (Reactant); RACT (Reactant or reagent)
      (preparation of brivanib alaninate)
IT  427878-70-4P  649735-46-6P  649736-41-4P
    872206-43-4P  872206-45-6P  872206-47-8P  872206-49-0P
    952490-01-6P
  
```

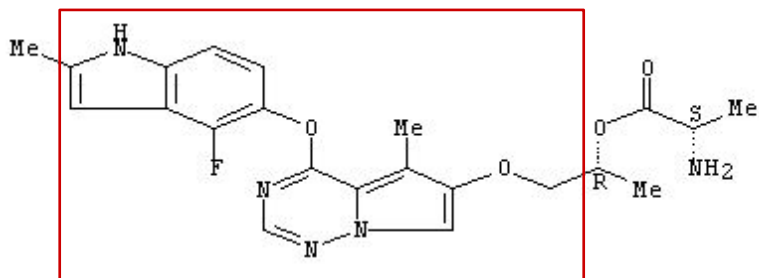
Example: Hit structures

RN 649735-63-7 HCAPLUS

CN L-Alanine, (1R)-2-[[4-[(4-fluoro-2-methyl-1H-indol-5-yl)oxy]-5-methylpyrrolo[2,1-f][1,2,4]triazin-6-yl]oxy]-1-methylethyl ester (CA

INDEX NAME)

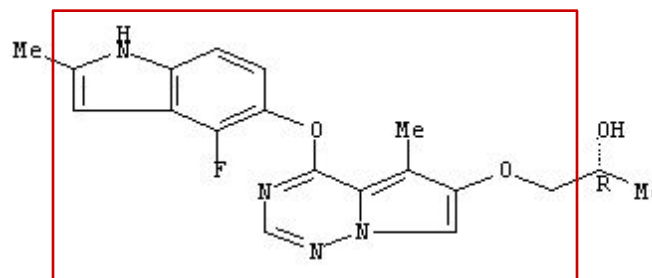
Absolute stereochemistry.



RN 649735-46-6 HCAPLUS

CN 2-Propanol, 1-[[4-[(4-fluoro-2-methyl-1H-indol-5-yl)oxy]-5-methylpyrrolo[2,1-f][1,2,4]triazin-6-yl]oxy]-, (2R)- (CA INDEX NAME)

Absolute stereochemistry.



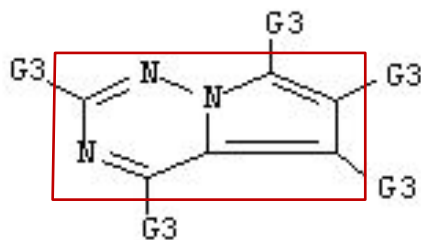
Example: MARPAT hit bibliographic record

```

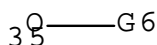
L109 ANSWER 33 OF 127  MARPAT  COPYRIGHT 2011 ACS on STN
AN  143:392509  MARPAT  Full-text
TI  Keratin dyeing compositions and bicyclic 5-6 heteroaromatic dyeing
    compounds with one ring nitrogen junction
IN  Glenn, Robert Wayne, Jr.; Lim, Muill; Gardlik, John Michael; Murphy, Bryan
    Patrick; Rees, Charles
PA  The Procter & Gamble Company, USA
SO  PCT Int. Appl., 91 pp.
    CODEN: PIXXD2
DT  Patent
LA  English
FAN.CNT 1
      PATENT NO.          KIND  DATE          APPLICATION NO.  DATE
      -----          -
PI  WO 2005097052      A1   20051020      WO 2005-US9687  20050324
      W:  AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH,
          CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD,
          GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
...
US 20050229333      A1   20051020      US 2005-61408   20050218
  
```

Example: MARPAT query hit in new assembled format

MSTR 1 Assembled



G3 = 35



G6 = carbon chain <containing 1-10 C> (opt. substd.) /
 carbocycle <containing 3-10 C, non-aromatic> (opt. substd.)

Patent location:

claim 1

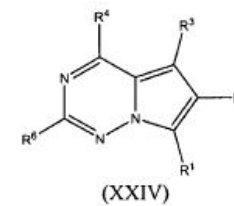
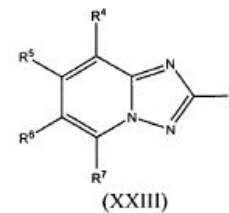
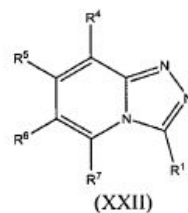
Note:

additional substitution and heteroatom
 interruptions also claimed

CLAIMS:

1. A keratin dyeing composition comprising:
 - (A) a medium suitable for dyeing; and
 - (B) one or more bicyclic 5-6 heteroaromatic keratin dyeing compounds with one ring junction nitrogen and one or two extra heteroatoms according to the following formulas:

...



STN supports all common search types

**Structure search
on specific
compound**

**Structure search
on broad structure
query**

**Structure search
on broad structure
query AND concept**

**Structure, name
and/or chemical
name**

**Structure, name
and/or chemical
name AND a
concept (e.g., an
indication)**

**Multiple concepts:
concept 1 AND
concept 2**

**Sequence search
AND concept**

**Polymer search
AND concept**

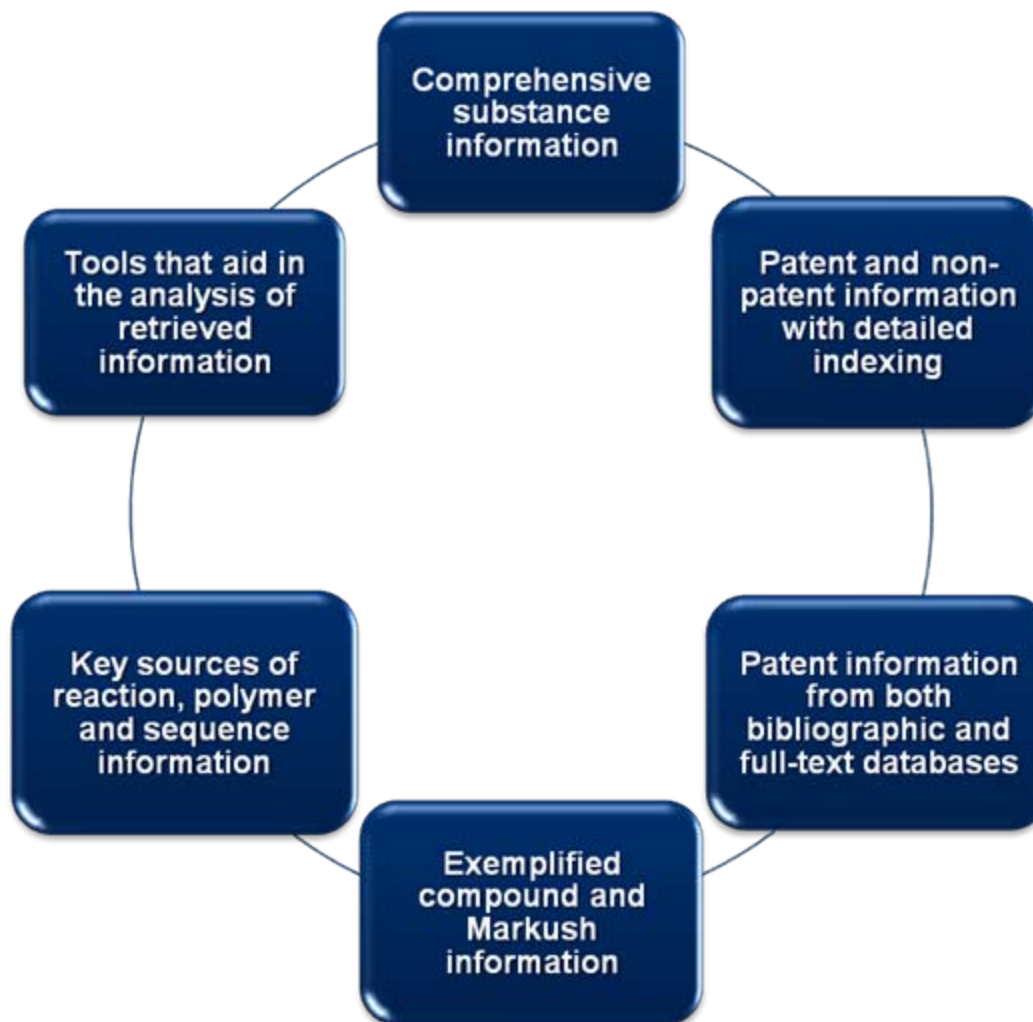
**Concept search in
the claims only (or
full text) of the full-
text patent
databases**

Multiple ways to access CAS/STN information

- **STN Express[®] and STN[®] on the WebSM** – full featured access via command line interface
- **STN Easy[®]** – Web-based access to the STN databases for non-complex searches
- **SciFinder** – access to REGISTRY, CPlus, etc., via intuitive graphical interface
- **Science IP[®]** –
 - CAS' search service
 - Experienced searchers
 - Customized to your needs and budget
 - www.scienceip.org

Summary

STN and CAS databases provide access to:







Thank you!

Elaine Cheeseman, Ph.D.
echeeseman@cas.org

scienceip@cas.org
www.scienceip.org
1-866-360-0814