



SCIENCE IP[®]
The CAS Search Service

Key Sources for Pharmaceutical and Chemical Literature and Patent Searching

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

Agenda

- **Limitations of free web-based patent databases**
- **STN[®] access to key scientific and technical databases**
- **Clusters of STN databases in all areas of science**
- **STN searching for exemplified compounds, Markush structures, reactions, polymers, and sequences**
- **Flexible STN search techniques**
- **Chinese patents well-represented in CAplus[®]**

Limitations of free web-based patent databases

Attribute	Features	Issues
Content	Full-text searching No intellectual indexing	Search only terms in the patent
Types of searching	Terms, patent codes, and date ranges	Limitations on how they can be combined
Queries	Often limits on numbers of characters or terms in a query Often limits on complexity of a query	Often cannot search complex multi-line queries
Search operators	AND, OR, NOT NEAR (within 5 words) – <i>WIPO only</i>	Often no proximity searching Difficult to search precisely
Structure searching	Not structure searchable	Will not retrieve structure-only references

Free patent databases on the web

Database	Content	Truncation	Operators	Proximity	Other
USPTO	Grants: 1976-present Apps: 2001-present 1790-1975: limited search features	Right; \$	AND, OR, ANDNOT	Not available	Grants and applications in separate databases Advanced Search is command language
WIPO (WIPO data) (Patent-scope)	1978-present	Right; *	AND, OR, ANDNOT XOR: one true but not both	NEAR: within 5 words	Query form Cannot OR terms in a single line

Free patent databases on the web (cont.)

Database	Content	Truncation	Operators	Proximity	Other
EPO (EP data) (Espacenet)	EP complete	*=unlimited ?=0 or 1 #=1 character	AND, OR, NOT	Not available	Maximum: 10 terms per field 20 terms per search Query form
Google Patents	US patents and patent applications from the USPTO	Appears to stem words	Advanced Search: “All of the words” “At least one of the words” “Without the words”	Based on Google Patent’s algorithm	Basic search Advanced search (Query form)

Google Scholar “Beta” observations (non-patents)

Attributes	Challenges
<ul style="list-style-type: none">• Full-text search• Basic and advanced search (query form) options• Advanced search – Author, Journal, and Date fields• Stemming appears to be part of the algorithm• Displays:<ul style="list-style-type: none">— Initial list of hits includes keyword in context— Click on a hit to see title, reference, and abstract— Often a charge to see full article	<ul style="list-style-type: none">• Content – journal coverage not given• Indexing – if provided by publisher and/or via automated indexing• Google algorithm:<ul style="list-style-type: none">— Includes relevance ranking— If multiple concepts, can get hits with only one concept— Can result in huge answer sets

PubMed observations

Attributes	Challenges
<ul style="list-style-type: none">• Indexed database• 98% MEDLINE® content• Basic search and advanced search options• Search by “all fields” (default) or specific fields in Advanced search• Operators – AND, OR, NOT; no proximity operator• Search by MeSH terms separately (explode, major descriptors, narrow terms, etc.)• “Automatic Query translation” can include MeSH® terms• Export summary of hits (bibliography), full records, etc.	<ul style="list-style-type: none">• Coverage:<ul style="list-style-type: none">— MEDLINE – 5,400 journals— EMBASE – 7,500 journals• Standalone system – no duplicate removal of hits from other databases

When a comprehensive search is critical, free web-based patent databases are not enough

- **For decisions about patents, such as:**
 - Patentability searches
 - Freedom-to-operate searches
 - Validity searches
- **For decisions about investments in research:**
 - Are there opportunities in this area?
 - Are there better opportunities in other areas?

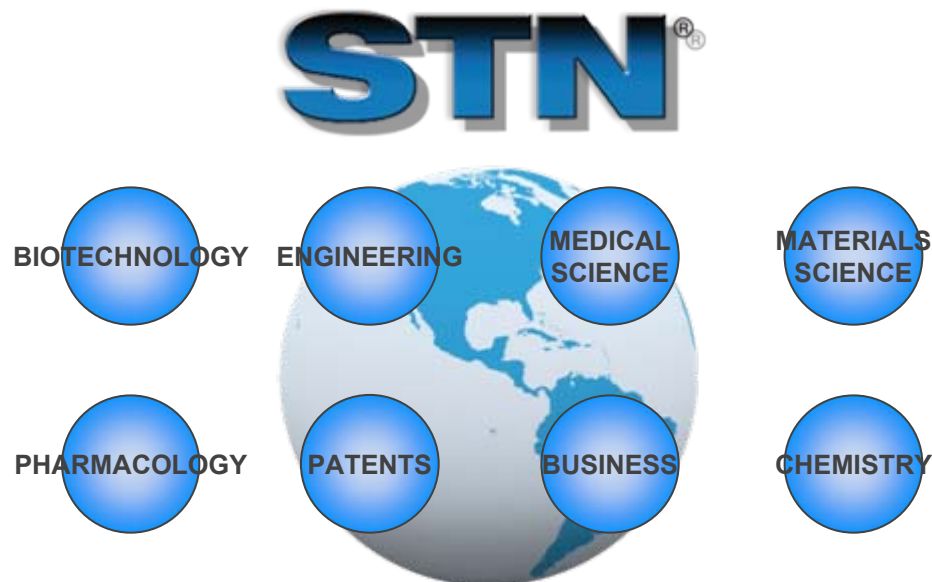


Consider:

- **What are the consequences of missing a key reference?**
- **“An hour spent in the library is worth a month in the laboratory.”**
 - source unknown

STN provides the best sources and techniques for scientific searching

- Chemical structure searching for exemplified compounds and Markush structures
- Any combination of:
 - Biosequence searching
 - Polymer searching
 - Reaction searching
 - Term searching (including complex multi-line strategies)
- Approximately 170 authoritative scientific and technical databases
- Multi-file searching with duplicate removal



STN databases cover all areas of science

Cluster	Databases
Bioscience	MEDLINE, EMBASE, BIOSIS, BIOTECHNO
Medicine	ADISINSIGHT, ADISCTI, ADISNEWS, IPA
Materials	ALUMINUM, MECHENG, SOLIDSTATE, WELDASEARCH
Engineering	COMPENDEX, ENCOMPLIT, ENCOMPAT, INSPEC
Toxicology	TOXCENTER SM , HEALSAFE, RTECS [®] , MEDLINE
Food	FROSTI, FOMAD, FSTA

CAS databases provide complete and timely coverage of all reputable sources

- 10,000 major scientific journals
- 1,500 core journals
 - Bibliographic information and abstracts available within 7 days



- 9 major patent offices
 - Bibliographic information available within 2 days of publication
 - Full indexing within 27 days
 - Offices include
 - US (UPTO)
 - WIPO
 - Europe (EPO)
 - France (INPI)
 - Germany (DMPA)
 - Japan (JPO)
 - Russia (ROSPATENT)
 - UK (UK-IPO)
 - Canada (CIPO)

Search STN for exemplified compounds by structure

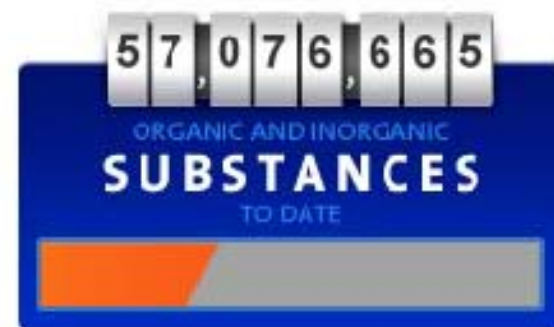
Database	Substances
CAS REGISTRY SM	> 57 million
Derwent World Patents Index – Derwent Chemistry Resource	> 1.5 million

Search steps

1. Create the structure query in STN Express[®]
2. Search for exact compounds or broad substructures
3. Search for references to the retrieved compounds
4. Broaden or narrow the retrieval with keywords, codes, and/or date restrictions

CAS REGISTRY – the world's largest source of chemical substance information

- **Contains a variety of substances:**
 - Organic compounds including pharmaceuticals
 - Inorganics – organometallics, salts, alloys
 - Sequences – proteins, nucleic acids
 - Polymers
- **Supports searching by:**
 - Structure
 - Sequence
 - Chemical names
 - Synonyms
 - Molecular formula
 - And more



CAS REGISTRY records hold a lot of information...

```
L2 ANSWER 1 OF 1 REGISTRY COPYRIGHT 2010 ACS on STN
RN 134523-03-8 REGISTRY
ED Entered STN: 28 Jun 1991
CN 1H-Pyrrole-1-heptanoic acid, 2-(4-fluorophenyl)- $\beta,\delta$ -dihydroxy-
5-(1-methylethyl)-3-phenyl-4-[(phenylamino)carbonyl]-, calcium
salt (2:1), ( $\beta R,\delta R$ )- (CA INDEX NAME)
OTHER CA INDEX NAMES:
CN 1H-Pyrrole-1-heptanoic acid, 2-(4-fluorophenyl)- $\beta,\delta$ -dihydroxy-
5-(1-methylethyl)-3-phenyl-4-[(phenylamino)carbonyl]-, calcium
salt (2:1), [R-(R*,R*)]-
OTHER NAMES:
CN Atocor
CN Atorvastatin calcium
CN Atorvastatin hemicalcium
CN Atorvastatin hemicalcium salt
CN CI 981
CN Lipitor
CN Rotacor
```

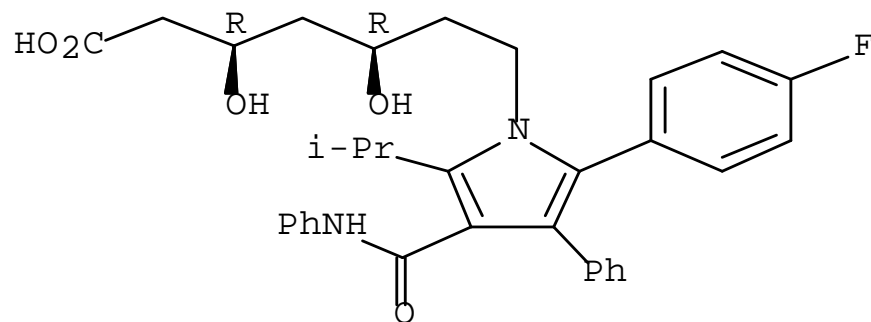
...

...including the molecular formula and databases associated with a CAS Registry Number®

```
FS  STEREOSEARCH
DR  1108202-55-6, 334757-04-9
MF  C33 H35 F N2 O5 . 1/2 Ca
CI  COM
SR  CA
...
LC  STN Files:  ADISINSIGHT, ADISNEWS, AGRICOLA, ANABSTR,
              REAXYSFILE*, BIOSIS, BIOTECHNO, CA, CAPLUS, CASREACT,
              CHEMCATS, CIN, CSCHEM, DDFU, DRUGU, EMBASE, HSDB*,
              IMSPATENTS, IMSRESEARCH, IPA, MEDLINE, MRCK*, PATDPASPC,
              PIRA, PROMT, PROUSDDR, PS, RTECS*, TOXCENTER,
              USAN, USPAT2, USPATFULL
              (*File contains numerically searchable property data)
CRN (134523-00-5)
```

The structure and number of references in the CA/CAPLUS databases are available

Absolute stereochemistry.



● 1/2 Ca

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

743 REFERENCES IN FILE CA (1907 TO DATE)

11 REFERENCES TO NON-SPECIFIC DERIVATIVES IN FILE CA

754 REFERENCES IN FILE CAPLUS (1907 TO DATE)

CAplus records for patents include an enhanced title, inventors, and patent assignees

L6 ANSWER 27 OF 166 HCA COPYRIGHT 2010 ACS on STN
AN 150:83896 HCA Full-text
ED Entered STN: 22 Jan 2009
TI Solid dosage forms containing an HMG-CoA reductase inhibitor
such as atorvastatin or fluvastatin in combination with
colloidal clay such as attapulgite, and method for the
preparation thereof
IN Karavas, Evangelos; Koutris, Efthimios; Ioannidou, Elli;
Samara, Vicky; Bikiaris, Dimitrios
PA Parmatheen S.A., Greece
SO PCT Int. Appl., 20pp.
CODEN: PIXXD2
DT Patent
LA English
CC 63-6 (Pharmaceuticals)

...

Patent family members with patent numbers, application numbers, dates, and kind codes are provided

FAN.CNT 1

	PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
	-----	----	-----	-----	-----
PI	WO 2009000286	A1	20081231	WO 2007-EP5568	20070625
	W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY,				
	...				
	AU 2007355452	A1	20081231	AU 2007-355452	20070625
				WO 2007-EP5568	A 20070625
	CA 2691956	A1	20081231	CA 2007-2691956	20070625
				WO 2007-EP5568	W 20070625
	EP 2170294	A1	20100407	EP 2007-764809	20070625
	...				

Enhanced abstracts and supplementary terms help for comprehensive searching

- AB The present invention relates to the formulation of solid dosage forms comprising a therapeutically effective amount of an HMG-CoA reductase inhibitor, and especially atorvastatin or fluvastatin or salts thereof, in combination with colloidal clay such as attapulgite, and a process for the preparation thereof by direct compression. Thus, 80 mg tablets of atorvastatin comprised (in wt%): atorvastatin Ca 6.89, microcelac 100 64.61, starch 1500 25.00, attapulgite 3.00, Mg stearate 0.50.
- ST atorvastatin fluvastatin HMG CoA reductase inhibitor
attapulgite clay solid

Detailed indexing is intellectually added

IT Pharmaceutical excipients
 (bindings; solid dosage forms containing an HMG-CoA reductase inhibitor such as atorvastatin or fluvastatin in combination with colloidal clay such as attapulgate, and method for preparation thereof)

IT Clays
 RL: THU (Therapeutic use); BIOL (Biological study); USES (Uses)
 (colloidal; solid dosage forms containing an HMG-CoA reductase inhibitor such as atorvastatin or fluvastatin in combination with colloidal clay such as attapulgate, and method for preparation thereof)

IT Pharmaceutical excipients
 (diluents; solid dosage forms containing an HMG-CoA reductase inhibitor such as atorvastatin or fluvastatin in combination with colloidal clay such as attapulgate, and method for preparation thereof)

...

CAS Registry Numbers are part of concept indexing

IT 9028-35-7

RL: BSU (Biological study, unclassified); BIOL (Biological study)

(inhibitors; solid dosage forms containing an HMG-CoA reductase inhibitor such as atorvastatin or fluvastatin in combination with colloidal clay such as attapulgite, and method for preparation thereof)

IT 557-04-0, Magnesium Stearate 9005-25-8, Starch, biological studies 12174-11-7, Attapulgite 93957-54-1, Fluvastatin 134523-00-5, Atorvastatin 134523-03-8, Atorvastatin Calcium 198828-48-7, Microcelac 100

RL: THU (Therapeutic use); BIOL (Biological study); USES (Uses) (solid dosage forms containing an HMG-CoA reductase inhibitor such as atorvastatin or fluvastatin in combination with colloidal clay such as attapulgite, and method for preparation thereof)

Cited references are included for displaying and searching

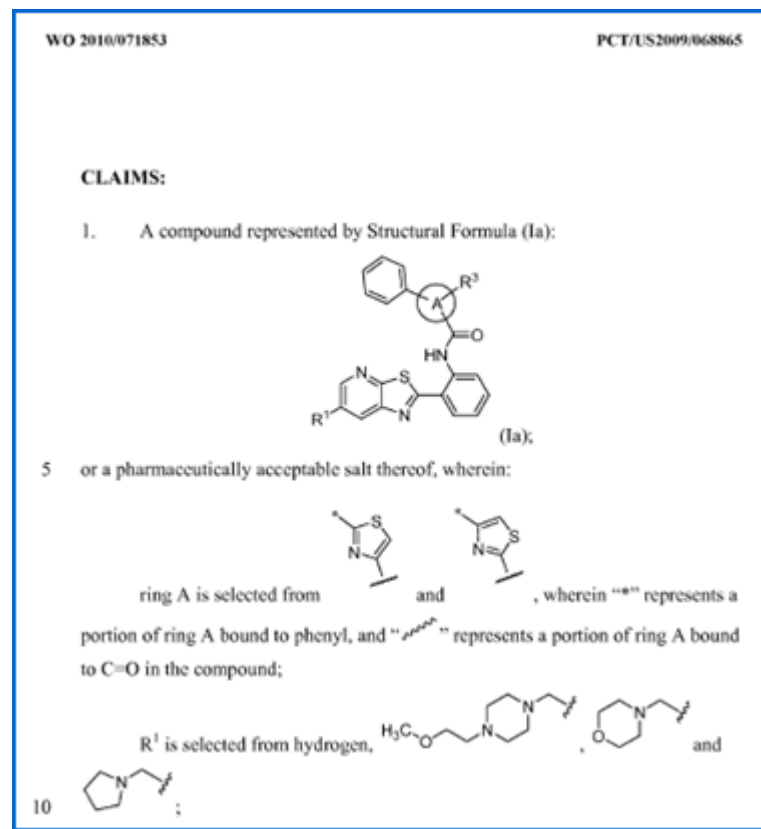
RE.CNT 5 THERE ARE 5 CITED REFERENCES AVAILABLE FOR THIS RECORD

RE CITED REFERENCES

- (1) Astra Ab; WO 9815290 A 1998 HCA
- (2) Kerc Janez; US 20040072894 A1 2004 HCA
- (3) Michael, F; WO 2004110431 A 2004 HCA
- (4) Pharmathen S A; WO 2007031801 A 2007 HCA
- (5) Ratiopharm Gmbh; DE 10316087 A1 2004 HCA

The MARPAT database covers Markush structures

- Structure searchable database with intellectual indexing based on Markush structures in claims and disclosure
- Retrieve additional patent references of interest for patentability, freedom-to-operate, etc.
- Three levels of settings provide specificity as needed
- Supports ANDed retrieval with terms, date restrictions, etc. in HCAplus



The MARPAT bibliography is similar to the CAplus bibliography

```

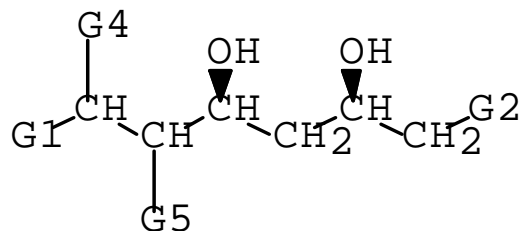
L6 ANSWER 10 OF 48 MARPAT COPYRIGHT 2011 ACS on STN
AN 147:522015 MARPAT Full-text
TI Novel process for statins and its pharmaceutically acceptable salts
   thereof
IN Reddy, Manne Satyanarayana; Rajan, Srinivasan Thirumalai; Reddy,
   Maramreddy Sahadeva
PA Satyanarayana Reddy, Manne, India; Thirumalai Rajan, Srinivasan; Sahadeva
   Reddy, Maramreddy
SO PCT Int. Appl., 114 pp.
   CODEN: PIXXD2
DT Patent
LA English
FAN.CNT 1

```

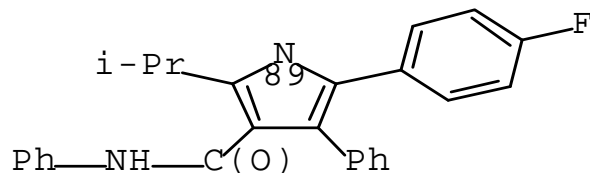
	PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
	-----	----	-----	-----	-----
PI	WO 2007125547	A2	20071108	WO 2007-IN172	20070430
	WO 2007125547	A9	20071221		
	WO 2007125547	A3	20080403 ...		

The QHIT format gives the parts of the Markush representation that caused the retrieval

MSTR 1



G1 = 89



G2 = CO₂H

Patent location:

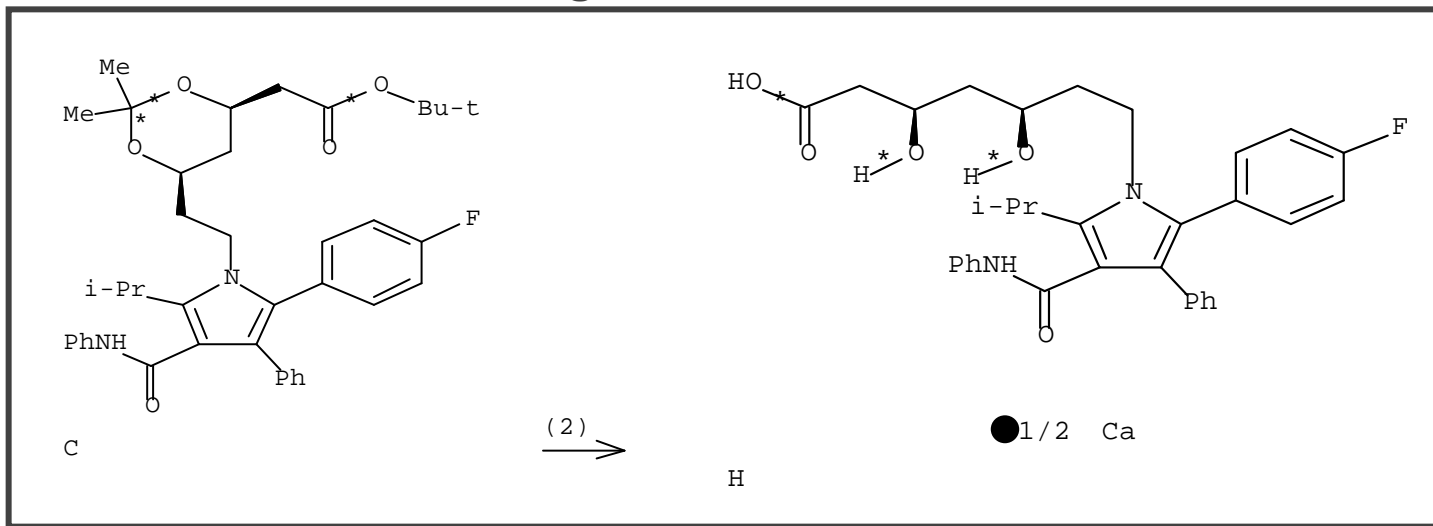
claim 18

Note:

and pharmaceutically acceptable salts

Reaction searching is available on STN

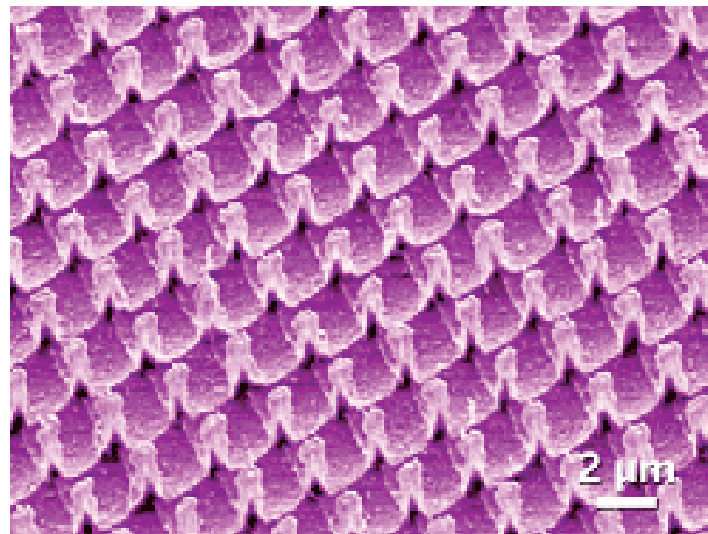
A → B reaction searching



Database	Content
CASREACT	Chemical Abstracts reactions, plus additional content added; journals and patents; 1840-present
ChemInformRX	Journals; 1991-present
Derwent Journal of Synthetic Methods	Journals; 1975-6/2003

CAS databases are the best sources for polymer searching

- **World's largest, most current source of polymer information**
- **REGISTRY contains more than 1.3 million polymers**
- **Polymers types include**
 - Homopolymers
 - Copolymers
 - Graft or Block polymers
- **Most REGISTRY polymer records are cited in CPlus; many have references in other STN databases**
- **REGISTRY is a source of additional names for use in STN databases that do not contain CAS Registry numbers**



pubs.acs.org/subscribe/journals/cen/89/i02/html/8902scic3.html

STN provides key sources for sequence searching

- **CAS REGISTRY/CAplus**
 - > 62 million peptide and nucleic acid sequences
 - 1907-present
- **Derwent GeneSeq (DGENE/Derwent World Patent Index)**
 - > 26 million peptide and nucleic acid sequences
 - 1981-present
- **USGENE**
 - All available peptide and nucleic acid sequences from the USPTO
 - 1981-present
- **PCTGEN**
 - Peptide and nucleic acid sequences submitted electronically to WIPO
 - 2001-present

Sequence Algorithms: BLAST[®], FASTA, Sequence Code Match

STN also provides flexible search techniques

- Command-line interface for precision searching
- Operators – AND, OR, NOT, (nW), (nA), (L), (S), (P)
- Search in specific fields or in basic index
- Search multiple synonyms
- Search by structure or sequence and terms, codes, etc.
- Multi-file searching
- Multiple analysis tools available

Flupirtine:

```
L1  que FLUPIRTIN? Or KATADOLON? Or
    TRANCOPAL or effirma or elb245 or
    elb(w)245
L2  que D9998 or D(w)9998 or D9(w)998 or
    D(w)9(w)998
L3  que W2964? or W(w)2964? or W2(w)(964
    or 964M) or W(w)2(w)(964 or 964M)
L4  que (fluorophenyl or
    fluoro(4a)phenyl)(15a)((methyl or
    me)(15a)amino or methylamino)
L5  que (fluorobenzyl or
    fluoro(4a)benzyl)(15a)amino or
    fluorobenzylamino
L6  que pyrid?(15a)(carbam? Or ylcarbam?)
    Or pyridinecarbam?
L7  que (L4 or L5)(15a)L6(15a)(ethyl or
    et)
L8  que L1 or L2 or L3 or L7
```

In STN, Chinese patent information is well represented and timely

- **Chinese patent coverage in CAplus includes**
 - A Unexamined Patent Application – 1985-present
 - U Granted Utility Model Patent – April 7, 2010-present
 - Y Granted Utility Model Patent – 2006-April 6, 2010
 - World Traditional Medicine – 50,000 patent records, many Chinese
- **Translated patent bibliographic records appear in CAplus approx. 5 days after their publication with SIPO, which is 2-5 weeks faster than with any other provider**

“China's patent office is now the world's leading producer of patent invention applications in chemistry.”

<http://www.cas.org/newsevents/releases/chinesepatents112309.html>



CAS databases are accessed through the following products and services

- STN Express – client software, free to STN users
- STN[®] on the WebSM – web-based access
- SciFinder[®] – web version now available
- STN Easy[®] – simplified STN, web-based access
- Science IP[®] – the CAS Search Service

The logo for STN Express, featuring the text "STN Express" in a bold, yellow, italicized sans-serif font with a slight drop shadow.The logo for STN on the web, featuring the text "STN on the web" in a purple and black font. The "STN" is in a large, bold, purple font, and "on the web" is in a smaller, black font. The text is set against a background of a white grid pattern that recedes into the distance.The logo for SciFinder, featuring a blue and purple diamond-shaped graphic with a white wave-like pattern inside. Below the graphic, the text "SciFinder" is written in a large, black, serif font with a registered trademark symbol.The logo for STN Easy, featuring the text "STN Easy" in a bold, white, italicized sans-serif font with a red and black glow effect.The logo for Science IP, featuring a gold sunburst graphic above the text "SCIENCE IP" in a blue, sans-serif font. Below this, the text "The CAS Search Service" is written in a smaller, blue, sans-serif font.

Summary

- Free web-based databases have their limitations
- STN offers key value-added scientific and technical databases
- STN database clusters cover all areas of science
- CAS indexing enables more comprehensive retrieval
- STN supports search for exemplified compounds, Markush structures, reactions, polymers, and sequences
- Flexible STN search techniques enable precision searching for comprehensive retrieval
- Large increase in Chinese patenting





Thank you!

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

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