Abstract

Title: "The Complete Markush Structure Search: Mission Impossible...?"

Abstract: The author will present the techniques required to bring together Markush search results generated from a combination of Derwent Fragmentation Codes and CAS Registry/MARPAT structure searching. In addition, special focus will be given to the techniques required to bring search results from the Merged Markush Service (MMS) on Questel.Orbit, over to the STN platform, for family sorting and de-duplication with STN-based CAS and Derwent Fragmentation Code searches. Using a few examples, the author will compare the content of the results obtained using the three systems, and hopefully inspire lively debate on both the search techniques employed and the philosophy of indexing from the the three database producers: Derwent, CAS and INPI.

Robert Austin
Regional Sales Manager USA
FIZ Karlsruhe
Hermann-von-Helmholtz-Platz 1
76344 Eggenstein-Leopoldshafen
Germany
Phone: +49 7247 808 263
Fax: +49 7247 808 131
Email: robert.austin@fiz-karlsruhe.de
www.fiz-karlsruhe.de

Biography: Robert Austin is the U.S. representative for FIZ Karlsruhe, the European partner of the Scientific and Technical Information Network (STN). After graduating with a bachelors degree in Applied Chemistry in 1991, he joined Derwent Information Ltd in the UK, as a Pharmaceutical Patent Indexer. After 9 years as an Indexer, Customer Trainer and, ultimately, Product Manager for Derwent World Patents Index, he moved on to work with FIZ Karlsruhe in January 2001.
Acknowledgements

Permission to reproduce search examples required for this paper, and also for unlimited access time to the various databases required to prepare them, was kindly provided by the following organisations. Many thanks to them all for this. In addition I made thorough use of the MMS training manual, available for download from the Millenium Information Services web site. So I have listed them here too. I would also like to thank Andrew Berks and Edlyn Simmons for suggesting some meaningful example structures to search.

All opinions/comments about search results are my own, and do not necessarily express the opinion of my employer nor of any of the other organisations or people named here.

Robert Austin
Tuesday, 23. October 2001

Chemical Abstracts Service (CAS)
2540 Olentangy River Road
PO Box 3012
Columbus, Ohio 43210
USA
Tel: +1 800 848 6533 / +1 614 447 3698
Fax: +1 614 447 3798
Email: help@cas.org
www.cas.org

Derwent Information Americas
1725 Duke Street
Suite 250
Alexandria, VA 22314
USA
Tel: +1 800 DERWENT / +1 703 706 4220
Fax: +1 703 838 5240 / +1 800 457 0850
Email: custserv@derwentus.com
www.derwent.com

Institut National de la Propriété Industrielle (INPI)
26 bis rue de Saint Pétersbourg
75800 Paris Cedex 08
France
Tel: +33 1 53 04 57 60
Fax: +33 1 42 94 01 16
www.inpi.fr
Introduction

Where did the term Markush come from?

In 1923 Dr Eugene A Markush filed a patent application in the United States concerning a method of preparing pyrazoline dyes that could be used for wool or silk:

![Patent Image]

Claim 1 of this application read:

```
Claims:
1. The process for the manufacture of dyes which comprises coupling with a halogen-substituted pyrazolone, a diazotized unsubstituted material selected from the group consisting of aniline, homologues of aniline and halogen substitution products of aniline.
```

The claim was challenged as being too unspecific. On appeal the US Commissioner of Patents ruled on the propriety of such claims. The patent was granted in 1924 as US 1,506,316.

What “Markush” means today

Today the term Markush denotes a substance or substituent, agent, reactant or other material that is described as being from a group consisting of certain specified materials. The specified materials can be an element, a chemical structure, a functional group, a class of chemical structures (such as alkyl or aryl), a class of functional groups (such as esters), etc. The value of Markush structures in patents is that a number of different chemical compounds can be described in a single patent claim. Markush structures are allowed in patents to protect an invention of related compounds without requiring the inventor to prepare and test each and every possible compound. A typical example of a modern Markush claim is shown on page 12.

There are three generally available systems\(^1\) to search Markush structure prior-art:

- Derwent World Patents Index® Chemical Fragmentation Codes
- INPI Merged Markush Service (produced in association with Derwent Information Ltd)
- Chemical Abstracts Service MARPAT®

---

\(^1\) A useful overview of the history of the development of Markush structure indexing can be found at: [www.lib.uchicago.edu/cinf/218nm/paper42/index.html](http://www.lib.uchicago.edu/cinf/218nm/paper42/index.html) (Edlyn Simmons, ACS National Meeting, August 1999).
The aim of this paper

Comprehensiveness is critical for patent searching, and in order to complete a thorough graphic-based patent prior-art search for a given chemical, class of chemicals, or fragment of substructure, all available online resources should be utilised. The Scientific and Technical Information Network (STN®) has powerful tools for this kind of searching, including the Chemical Abstracts Service (CAS) Registry File and the MARPAT database. However, STN does not offer one other major file available to the searcher for chemical structure prior-art: the Merged Markush Service (MMS). Currently this service is only available on Questel. Orbit. It is the intention of the author to explain both the basics of how to use MMS, and also how to transfer subsequent MMS results to the STN platform, for de-duplication and sorting into a comprehensive search result.
The Databases

_registry / CAPLUS_

The REGISTRY file, produced by the Chemical Abstracts Service (CAS), contains 33 million specific substance references from Journal and patent literature from 1957 to date. Each has a unique CAS Registry Number® and the fully functional version of REGISTRY is unique to STN. The CAPLUS file provides CAS bibliographic references for substance records via their Registry Number, and comprises over 18 million bibliographic references, including 3.3 million patents, dating back to 1907. Patent coverage is for 33 world-wide authorities. Approximately 2.5 million CAPLUS patent records (i.e. roughly 80%) are indexed with Registry numbers. In addition to CAPLUS, a large number of other bibliographic databases, e.g. USPATFULL, provide substance indexing using Registry Numbers. Together REGISTRY and CAPLUS represent the world’s most comprehensive source of chemical substance information.

MARPAT

The MARPAT database, also produced by the Chemical Abstracts Service (CAS) and available only on STN, contains representations of Markush structures that appear in patents published since January 1st 1988. Patent country coverage since that date is the same as for the CAPLUS1 file. Whilst specific compound references in patents are generally indexed in the REGISTRY file (above), importantly MARPAT also covers substances specifically disclosed by Markush formulae, but which are not unambiguously drawn out or named in the patent. MARPAT comprises 457,000 Markush compounds from 161,000 patent records. Unlike REGISTRY, MARPAT records are bibliographic (patent) based rather than substance based. Patent information is displayable but only searchable in CAPLUS. The link between the two files is via MARPAT Accession Number. CAS also provide work-in-progress indexing in MARPATPREV.

Merged Markush Service

The Merged Markush Service (MMS), jointly produced by INPI (the French Patent Office) and Derwent Information Ltd, contains Markush structures that appear in patents dating back to 1979 for pharmaceuticals, and 1987 for other chemical subjects. Derwent World Patents Index (Derwent WPI) and INPI Pharmsearch databases provide the bibliographic references via MMS Compound Numbers (CN). MMS and INPI Pharmsearch are only available on Questel.Orbit. MMS patent country coverage is for selected Derwent WPI authorities from 1987 (see page 16) and as the 4 authorities of INPI Pharmsearch (US, WO, EP, FR) from 1979-1986. Specific compounds disclosed in patents are typically represented as options within corresponding Markush compound indexing in the MMS backfile (i.e. unlike MARPAT/REGISTRY above). However, the file also incorporates all 200,000 specific substance references which form the Derwent Chemistry Resource (DCR) file on STN (1999-date). MMS comprises 976,000 substance records corresponding to 537,000 patent records in Derwent WPI, and to 104,000 patent records in INPI Pharmsearch. MMS is currently being indexed forward in time by Derwent, and backwards in time by INPI. The backfile indexing target is 1978 for all INPI authorities and subject coverage. Pre-1987 backfile bibliographic indexing is only available in INPI Pharmsearch on Questel.Orbit.

1 Except Russian Patents, for which coverage is included from January 10th 2000.
Derwent WPI: Chemical Fragmentation Codes

The Derwent World Patents Index® (Derwent WPI) is the most comprehensive value-added patent database in the world, comprising 10.8 million patent records. Chemical Fragmentation Codes are a non-graphical, code-based system of chemical substance indexing spanning the entire backfile of Derwent WPI to 1963. They are provided as a text searchable index of the Derwent WPI bibliographic file on several platforms, including STN. The system incorporates all data1 indexed by Derwent in MMS on Questel.Orbit (1987-date) and in the Derwent Chemistry Resource (DCR) on STN (1999-date). Subject coverage is from 1963 for pharmaceuticals, 1965 for agrochemicals and 1970 for other areas of chemistry. Country coverage is for selected Derwent WPI authorities (see page 16). From 1992 onwards the majority of Chemical Fragmentation Codes have been automatically generated from the graphical structures contained in MMS. The codes represent the oldest source of Markush substance indexing from patents and are available for 1.1 million Derwent WPI records. Access is restricted to Derwent Subscribers.

Coverage summary

This is a simple graphical representation of the comparative backfile coverage of each of the services provided by Derwent, INPI and Chemical Abstracts Service. More detailed tables are given on pages 15 and 16.

---

1 Except Steroid structures from 1999 onwards.
## Content Summary

<table>
<thead>
<tr>
<th>Feature</th>
<th>REGISTRY</th>
<th>MARPAT</th>
<th>MMS</th>
<th>Derwent WPI Fragment Codes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1987- (other chem)</td>
<td>1965- (agrochem)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1963- (pharma)</td>
<td>1970- (other chem)</td>
</tr>
<tr>
<td>Specific Substances</td>
<td>33 million¹</td>
<td>N/A</td>
<td>202,000</td>
<td>N/A</td>
</tr>
<tr>
<td>Markush Compounds</td>
<td>N/A</td>
<td>457,000</td>
<td>774,000</td>
<td>N/A</td>
</tr>
<tr>
<td>Patent records</td>
<td>2.5 million (in CA)</td>
<td>161,000</td>
<td>537,000 (in WPI)</td>
<td>1.1 million²</td>
</tr>
<tr>
<td>Compounds Covered</td>
<td>Specific; all classes</td>
<td>Markush; all classes except: alloys, metal oxides, inorganic salts, intermetallic, polymers</td>
<td>Specific and markush; all classes except: alloys, non-pharma /agrochem polymers</td>
<td>Specific and markush; all classes except: alloys, non-pharma /agrochem polymers, Steriods (from 1999)</td>
</tr>
<tr>
<td>Access points</td>
<td>STN structure Searching; graphic upload; text search</td>
<td>STN structure Searching; graphic upload; text search</td>
<td>DARC Structure Searching</td>
<td>Code-based structure Searching; full bib. search; MC, IPC</td>
</tr>
<tr>
<td>Bibliographic information</td>
<td>Displayable; Searchable in CAPLUS</td>
<td>Displayable; Searchable in CAPLUS</td>
<td>Displayable (INPI only): search &amp; display in DWPI or Pharmsearch</td>
<td>Searchable &amp; displayable</td>
</tr>
<tr>
<td>Work in progress</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Sample search</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Access</td>
<td>Open</td>
<td>Open</td>
<td>Open</td>
<td>Subscription only</td>
</tr>
<tr>
<td>SDI (charge)</td>
<td>Yes ($25)</td>
<td>Yes ($25)</td>
<td>No</td>
<td>Yes ($16)</td>
</tr>
<tr>
<td>Structure Search charge</td>
<td>$133</td>
<td>$94</td>
<td>$58</td>
<td>FREE</td>
</tr>
<tr>
<td>Availability</td>
<td>STN</td>
<td>STN</td>
<td>Questel.Orbit</td>
<td>STN, Questel.Orbit, Dialog</td>
</tr>
</tbody>
</table>

¹ This number includes biosequences, alloys, polymers and multi-component substances; not all substances will have references to patents.

² Fragmentation Coding also covers some non-chemical subjects, e.g. drug delivery devices.
# Country Coverage Comparison

<table>
<thead>
<tr>
<th>Authority</th>
<th>REGISTRY / CAPLUS</th>
<th>MARPAT</th>
<th>MMS</th>
<th>Derwent WPI Frag. Code</th>
<th>Derwent WPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>AR</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1975 (only)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1983</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1983</td>
</tr>
<tr>
<td>Austria</td>
<td>AT</td>
<td>1957</td>
<td>1988</td>
<td>1993</td>
<td>1975</td>
</tr>
<tr>
<td>Belgium</td>
<td>BE</td>
<td>1957</td>
<td>1988</td>
<td>1987</td>
<td>1963</td>
</tr>
<tr>
<td>Canada</td>
<td>CA</td>
<td>1957</td>
<td>1988</td>
<td>1987</td>
<td>1963</td>
</tr>
<tr>
<td>China</td>
<td>CN</td>
<td>1986</td>
<td>1988</td>
<td>-</td>
<td>1985</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>CZ</td>
<td>1993 *</td>
<td>1988 *</td>
<td>-</td>
<td>1994</td>
</tr>
<tr>
<td>Denmark</td>
<td>DK</td>
<td>1957 *</td>
<td>1988 *</td>
<td>-</td>
<td>1974</td>
</tr>
<tr>
<td>Finland</td>
<td>FI</td>
<td>1960 *</td>
<td>1988 *</td>
<td>-</td>
<td>1974</td>
</tr>
<tr>
<td>France</td>
<td>FR</td>
<td>1957</td>
<td>1988</td>
<td>1979</td>
<td>1963</td>
</tr>
<tr>
<td>Hungary</td>
<td>HU</td>
<td>1957-1958; 1988</td>
<td>-</td>
<td>-</td>
<td>1975</td>
</tr>
<tr>
<td>India</td>
<td>IN</td>
<td>1957</td>
<td>1988</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ireland</td>
<td>IE</td>
<td>-</td>
<td>-</td>
<td>1995</td>
<td>1963-1969;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1963-1994</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1994</td>
</tr>
<tr>
<td>Italy</td>
<td>IT</td>
<td>1957</td>
<td>1988</td>
<td>-</td>
<td>1977</td>
</tr>
<tr>
<td>Latvia</td>
<td>LV</td>
<td>1996 *</td>
<td>1996 *</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Lithuania</td>
<td>LT</td>
<td>1996 *</td>
<td>1996 *</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>LU</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1984</td>
</tr>
<tr>
<td>Mexico</td>
<td>MX</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1997</td>
</tr>
<tr>
<td>Netherlands</td>
<td>NL</td>
<td>1957</td>
<td>1988</td>
<td>1987</td>
<td>1963</td>
</tr>
<tr>
<td>New Zealand</td>
<td>NZ</td>
<td>-</td>
<td>-</td>
<td>1992</td>
<td>1992</td>
</tr>
<tr>
<td>Norway</td>
<td>NO</td>
<td>1957 *</td>
<td>1988</td>
<td>-</td>
<td>1974</td>
</tr>
<tr>
<td>Philippines</td>
<td>PH</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1992</td>
</tr>
<tr>
<td>Poland</td>
<td>PL</td>
<td>1960 *</td>
<td>1988 *</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Portugal</td>
<td>PT</td>
<td>-</td>
<td>-</td>
<td>1987 (only)</td>
<td>1983-1987</td>
</tr>
<tr>
<td>Singapore</td>
<td>SG</td>
<td>-</td>
<td>-</td>
<td>1996</td>
<td>1996</td>
</tr>
<tr>
<td>Slovakia</td>
<td>SK</td>
<td>1996 *</td>
<td>1996 *</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Switzerland</td>
<td>CH</td>
<td>1957</td>
<td>1988</td>
<td>1987</td>
<td>1963</td>
</tr>
<tr>
<td>Taiwan</td>
<td>TW</td>
<td>-</td>
<td>-</td>
<td>1993-1998</td>
<td>1993</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>GB</td>
<td>1957</td>
<td>1988</td>
<td>1987</td>
<td>1963</td>
</tr>
</tbody>
</table>

---

1 CAS Registry Indexing begins in 1957. CAPLUS patent coverage dates back to 1907. An asterisk ‘*’ indicates that only patents issued to residents in the granting country are abstracted/indexed.

2 MMS currently begins in 1979 with pharmaceutical coverage only; all chemistry is covered from 1987.

3 This is the specific coverage of DWPI Fragmentation Code Indexing (i.e. from a subset of Derwent “Major Authorities”).

4 This is the general coverage of DWPI, including all title-only “Minor Authorities”.

---

16
Structure Query Preparation Software

STN Express 6.0

*STN Express®* is online communications software produced by the *Chemical Abstracts Service* (CAS) and is primarily designed for searching STN databases. However, the software can also be used to search other command line based systems, including *Questel.Orbit* and *Dialog*. Amongst its many features STN Express includes a structure query preparation module, which can be used to prepare queries for all STN databases which support *Messenger* structure searching, e.g. REGISTRY/MARPAT. The structure query module can also be used to convert graphical queries into Chemical Fragmentation Code strategies, exclusively for searching Derwent WPI on STN. Version 6.0 is the latest version of STN Express.

**note**

It is also possible to search for chemical structures references via *STN on the Web* ([stnweb.cas.org](http://stnweb.cas.org) or [stnweb.fiz-karlsruhe.de](http://stnweb.fiz-karlsruhe.de)), including using Derwent Chemical Fragmentation Code strategies.

Derwent Markush TOPFRAG 3.1

*Markush TOPFRAG* (MTF) is an online search aid produced by Derwent Information Ltd. MTF can be used by searchers to convert chemical structure queries into (a) Chemical Fragmentation Code strategies for searching Derwent WPI on either *STN, Questel.Orbit* or *Dialog*; or (b) Markush DARC format text queries for uploading to MMS on *Questel.Orbit*. The TOPFRAG module - which is used to create the Fragmentation Code strategies - is additionally provided by Derwent and CAS as a part of STN Express structure query preparation software (see above). Version 3.1 is the latest version of MTF.

For further information

*STN International*  
[www.cas.org/stn.html](http://www.cas.org/stn.html)  
OR  
[www.stn-international.de](http://www.stn-international.de)

*Chemical Abstracts Service*  
[www.cas.org](http://www.cas.org)

*FIZ Karlsruhe*  
[www.fiz-karlsruhe.de](http://www.fiz-karlsruhe.de)

*Derwent Information Ltd*  
[www.derwent.com](http://www.derwent.com)
Search Example

Objectives

- Conduct a Merged Markush Service (MMS) structure search via STN Express
- Extract relevant search results from MMS, and transfer them to the Derwent World Patents Index® on STN (files WPINDEX/WPIDS/WPIX).
- Complement MMS results with answers retrieved using an STN Express Derwent Chemical Fragmentation Code search query in WPIDS/WPIX
- Sort and de-duplicate the MMS/Derwent results, with answers retrieved using corresponding search queries in CAS Registry and MARPAT databases on STN, to end up with a single, comprehensive answer set.

**note**

This example does not attempt to show the full diversity of structure search options which are available in MMS, Derwent Chemical Fragmentation Codes, CAS Registry or MARPAT, just the principles of how to bring all these resources together for a comprehensive search result.

Search Question: What is the patent prior-art for the drug diazepam (valium): either when described directly, or when unambiguously defined within a Markush chemical formula?

Diazepam:
Search Strategy
To find patents which specifically describe the drug diazepam

Step 1  MMS structure search for diazepam
Step 2  Extract Derwent WPI compound numbers
Step 3  Crossover MMS results to PHARM
Step 4  Extract patent numbers from PHARM
Step 5  Transfer MMS results to WPIDS
Step 6  WPIDS chemical code search for diazepam
Step 7  CASLINK structure search for diazepam
Step 8  Combine, de-duplicate and FSORT all results

MMS

Prepare the query offline using Derwent Markush TOPFRAG (MTF):

1. Select New Structure from the Query Menu; select Markush format
2. Draw the query; select **Utilities & Generate Codes**; save the text query

---

**Logon to MMS via STN Express (see page 75):**

---

**Welcome to Markush DARC / Bienvenue sur Markush DARC**

**English Version**

---

**** BASE MMS - 28/09/01 ****
972630 Compounds - Last On: 9999-ISE03

**Selected Segment(s):**

- BACKF
- MPHARM
- WPIM
- FRONIF
Enter QT (Query Text) at the main -ST- prompt:

-ST- (BA, CN, QT, QG, RF, PE, AA, SB, BL, BL, ID, INFO) ? QT

*** QT ***

**** SELECTED DATA BASE : MMS ****

-QU- (CN, CA, GM, GL, GR, BO, AT, FS, AP, VP, ATTR, VE) ?

Upload the MTF Markush DARC text query:

1. Activate the Command Window

2. Open the MTF text query into the Command window

3. Upload the text query in from the Command window:
Check that the structure has uploaded correctly with the VE (verify) command:

Type FI (Finish) when you have verified the query:

You are currently defining GROUP 0
Type FI (Finish) again to exit the QU (Query) mode:

```
-QU- (CN,CA,GM,GI,GR,BO,AT,FS,AP,VP,ATTR,VE) ? FI
OTHER SPECIFICATIONS (Y/N) ? N
FILE SEGMENTS (Y/N) ? N
```

Save (SV) the structure query (QU) for future use:

```
-ST- (BA,CN,QT,QG,RF,RE,AA,SB,BL,BI,GD,INFO) ? SV
QU / DN,CN,PN / ? QU
NAME (8CHAR.) ? diazepam
```

```
1 diazepam QU 1 MARKUSH DARC 22/09/01 1
```

Type RE (Retrieve Candidates) to begin the first stage of the structure search:

```
-RE- (BA,CN,QT,QG,RF,RE,AA,SB,BL,BI,GD,INFO) ? RE
*** RE ***
```

```
R1 - RE / MMS - : 23249 ANSWER(S)
```

Continue the structure search with AA (atom by atom):

```
-AA- (BA,CN,QT,QG,RF,RE,AA,SB,BL,BI,GD,INFO) ? AA
*** AA ***
```

```
17 ANSWER(S) FOR 2637 CANDIDATES
24 ANSWER(S) FOR 5057 CANDIDATES
```

```
- AA - NUMBER OF ANSWER(S) : 25
- FILE RX - NUMBER OF CANDIDATE(S) : 0
- CANDIDATES REMAINING TO BE PROCESSED : 15959
```

Continue the search online:

```
CONTINUE AA (A), BATCH SEARCH (B), CANCEL (C), POWER-BATCH (P) ? A
*** AA *** CONTINUED
```

```
26 ANSWER(S) FOR 9695 CANDIDATES
27 ANSWER(S) FOR 12269 CANDIDATES
30 ANSWER(S) FOR 14855 CANDIDATES
30 ANSWER(S) FOR 17185 CANDIDATES
```

MMS offers a useful offline Batch facility. This is particularly helpful with more complex queries (page 60). This particular example is continued online (option A).
The Complete Markush Structure Search: Mission Impossible?

32 ANSWER(S) FOR 19598 CANDIDATES
33 ANSWER(S) FOR 22423 CANDIDATES

R2 - AA / MMS : 34 ANSWER(S)
R3 - RX / MMS : 0 ANSWER(S)

Review the search history with HI R (History R-answer sets):

** LIST OF VALID ANSWER FILES **
R 1 RE / MMS 23249 ANSWER(S) CURRENT R0
R 2 AA / MMS 34 ANSWER(S) CURRENT AA

Save (SV) the compound number (CN) based answer set:

MARKUSH/DARC 2 / 34 CN : 0018-19801 MMS

Review answers graphically with VI FO (Visualise Focus):

VI FO just displays the Markush subgroups (Gx) where the hit atoms are located.
VI FO highlights hit atoms with a box.
The Complete Markush Structure Search: Mission Impossible?

**MARKUSH/DARC** 2/34 **CN:** 0018-19801 **MMS**

- **FG:** 0
- **GM:** 4

---

**MARKUSH/DARC** 14/34 **CN:** R16158 **MMS**

- **GM:** 0

---

It is possible to select the record to be displayed, e.g. 14 (of 34)

**SEGMENTS:** ABEVY2:G0

/DIAZEPAM SALTS GENERAL/

---

**Type FI (Finish) to leave the VI (Visualise) mode:**

? **FI**

- **ST:** (BA, CN, QT, QG, RF, PE, AA, SB, BL, BL, GD, INFO)?
Display a list of the compound numbers with LI (List) command:

<table>
<thead>
<tr>
<th>No.</th>
<th>Compound Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CN = RA3H0Q</td>
</tr>
<tr>
<td>2</td>
<td>CN = 0018–19801</td>
</tr>
<tr>
<td>3</td>
<td>CN = 9903–H8301</td>
</tr>
<tr>
<td>4</td>
<td>CN = 9807–26403</td>
</tr>
<tr>
<td>5</td>
<td>CN = 9645–39703</td>
</tr>
<tr>
<td>6</td>
<td>CN = 9637–00802</td>
</tr>
<tr>
<td>7</td>
<td>CN = 9637–00801</td>
</tr>
<tr>
<td>8</td>
<td>CN = 9418–44103</td>
</tr>
<tr>
<td>9</td>
<td>CN = 9238–14101</td>
</tr>
<tr>
<td>10</td>
<td>CN = 9008–31802</td>
</tr>
<tr>
<td>11</td>
<td>CN = 9002–12202</td>
</tr>
<tr>
<td>12</td>
<td>CN = 8940–09710</td>
</tr>
<tr>
<td>13</td>
<td>CN = 8914–02901</td>
</tr>
<tr>
<td>14</td>
<td>CN = R1.6158</td>
</tr>
<tr>
<td>15</td>
<td>CN = 8821–12501</td>
</tr>
<tr>
<td>16</td>
<td>CN = R01255</td>
</tr>
<tr>
<td>17</td>
<td>CN = 96105966–03</td>
</tr>
<tr>
<td>18</td>
<td>CN = 95110993–02</td>
</tr>
<tr>
<td>19</td>
<td>CN = R70037640–01</td>
</tr>
<tr>
<td>20</td>
<td>CN = 66074206–07</td>
</tr>
<tr>
<td>21</td>
<td>CN = 86040114–01</td>
</tr>
<tr>
<td>22</td>
<td>CN = 88060463–01</td>
</tr>
<tr>
<td>23</td>
<td>CN = 90020299–01</td>
</tr>
<tr>
<td>24</td>
<td>CN = 61100839–01</td>
</tr>
<tr>
<td>25</td>
<td>CN = 89120432–02</td>
</tr>
<tr>
<td>26</td>
<td>CN = 90010072–02</td>
</tr>
<tr>
<td>27</td>
<td>CN = 89030183–01</td>
</tr>
<tr>
<td>28</td>
<td>CN = R86060272–02</td>
</tr>
<tr>
<td>29</td>
<td>CN = 79055233–01</td>
</tr>
<tr>
<td>30</td>
<td>CN = 80105158–01</td>
</tr>
<tr>
<td>31</td>
<td>CN = 81035060–01</td>
</tr>
<tr>
<td>32</td>
<td>CN = R81085339–07</td>
</tr>
<tr>
<td>33</td>
<td>CN = R81085339–06</td>
</tr>
<tr>
<td>34</td>
<td>CN = 85010433–09</td>
</tr>
</tbody>
</table>

**Derwent WPI compound numbers:**
- RXXXXX  Specific Compounds
- XXXX–XXXXX  Markush Compounds

**Note:** these numbers are searchable in the STN WPINDEX file (DCN field).

**INPI Pharmsearch compound numbers:**
- RXXXXXXXX–XX  Specific Compounds
- XXXXXXXX–XX  Markush Compounds

**Note:** these numbers are only searchable in the INPI Pharmsearch file on Questel.Orbit.
PHARMSEARCH

In order to transfer Pharmsearch answers from MMS to STN, it is necessary to identify the corresponding bibliographic records in Pharmsearch, and extract a list of patent numbers to use as search terms for locating WPINDEX records on STN. The Questel.Orbit Memsort (MEMS) command can be used to extract the patent numbers. MEMS is equivalent to the STN ANALYZE command.

From MMS, enter BI (Bibliographic) to access Questel.Orbit:

```
- ST- (BA,CN,QT,QG,RF,RE,AA,SB,BL,GD,INFO) ? BI

* SEE YOU LATER *
******

(C) QUESTEL 1994 QUESTEL. ORBIT (TM) 1998 28/09/01 17*47*35
...
..FILE / ..INFO / ..GUIDE
```

Access the Pharmsearch database (File PHARM):

```
? FT PHARM

Selected file: PHARM

Period covered: weeks 7901-9952 (US), 7902-9952 (EP), 7902-9952 (FR)
Backfile indexing continues as usual but no further indexing concerning the Frontfile will take place in 2001. A replacement service is being designed to allow any further searching.
```

---note---

Regular updates to INPI Pharmsearch are to the Backfile of the database. The frontfile is not currently being updated (see above). All pre-1987 MMS bibliographic references are found in Pharmsearch on Questel.Orbit (i.e. not in Derwent WPI).

---Helpful HINT---

To return from Questel.Orbit to MMS the command is:

```
? ST MASC
```
Crossover compound numbers (CN) from MMS using the *MD command:

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>R70037640-01/CN</td>
</tr>
<tr>
<td>1</td>
<td>R81085339-06/CN</td>
</tr>
<tr>
<td>1</td>
<td>R81085339-07/CN</td>
</tr>
<tr>
<td>135</td>
<td>R86060272-02/CN</td>
</tr>
<tr>
<td>1</td>
<td>61100839-01/CN</td>
</tr>
<tr>
<td>1</td>
<td>66074206-07/CN</td>
</tr>
<tr>
<td>1</td>
<td>79055233-01/CN</td>
</tr>
<tr>
<td>1</td>
<td>80105158-01/CN</td>
</tr>
<tr>
<td>1</td>
<td>81035060-01/CN</td>
</tr>
<tr>
<td>1</td>
<td>85010433-09/CN</td>
</tr>
<tr>
<td>1</td>
<td>86040114-01/CN</td>
</tr>
<tr>
<td>1</td>
<td>88060463-01/CN</td>
</tr>
<tr>
<td>1</td>
<td>89030183-01/CN</td>
</tr>
<tr>
<td>1</td>
<td>89120432-02/CN</td>
</tr>
<tr>
<td>1</td>
<td>90010072-02/CN</td>
</tr>
<tr>
<td>1</td>
<td>90020299-01/CN</td>
</tr>
<tr>
<td>1</td>
<td>95110993-02/CN</td>
</tr>
<tr>
<td>1</td>
<td>96105966-03/CN</td>
</tr>
</tbody>
</table>

** SS 1: Results 151

INPI Pharmsearch compound numbers:
- RRRRRRRRRR-XX Specific Compounds
- RRRRRRRRRR-XX Markush Compounds

Helpful HINT

Like other patent files on Questel.Orbit, PHARM has a standardised patent number field (XPN), designed for cross-file searching. The numbers in this field are in standardised Derwent format, and as such can be transferred directly to Derwent WPI on STN.

Extract a list of standardised basic patent numbers (XPN RK 1) using MEMS:

<table>
<thead>
<tr>
<th>Search statement 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>? MEMS SS 1 SET DIAZEPAM STORE /XPN RK 1</td>
</tr>
</tbody>
</table>

Memsort (MEMS) command:
- SS 1: Search Statement 1
- SET: Use whole answer set
- STORE: Keep the patent number list
- DIAZEPAM: Customised name for the list
- XPN: Patent Number(s)
- RK 1: The first patent number only

18 INPI Compound Numbers
Retrieve 151 patent records
Display the list of patent numbers using the LI MEMS (List Memsort) command:

Search statement 2

? LI MEMS DIAZEPAM NONSTOP

Memory is of type MEMSORT (statistical analysis)

<table>
<thead>
<tr>
<th>#</th>
<th>FREQ</th>
<th>TERM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>DE4207922</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>DE4222826</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>EP---1924</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>EP---11745</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>EP---21337</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>EP---22771</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>EP---25546</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>EP---31561</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>EP---35132</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>EP---69307</td>
</tr>
<tr>
<td>145</td>
<td>1</td>
<td>WO9615815</td>
</tr>
<tr>
<td>146</td>
<td>1</td>
<td>WO9630393</td>
</tr>
<tr>
<td>147</td>
<td>1</td>
<td>WO9637232</td>
</tr>
<tr>
<td>148</td>
<td>1</td>
<td>WO9640052</td>
</tr>
<tr>
<td>149</td>
<td>1</td>
<td>WO9640053</td>
</tr>
<tr>
<td>150</td>
<td>1</td>
<td>WO9719112</td>
</tr>
<tr>
<td>151</td>
<td>1</td>
<td>WO9802148</td>
</tr>
</tbody>
</table>

Logoff from Questel.Orbit:

Search statement 2

? ST

Session finished: 28 SEP 2001 Time 11:27:41

Your session will be retained for 2 hours.

QUESTEL.ORBIT thanks you. Hope to hear from you again soon.

List Memsort (LI MEMS) command:
DIAZEPAM Customised name for the list
NONSTOP Display complete list

Logoff with the ST command. Questel.Orbit automatically retains your search for two hours. This useful feature means that you are able to return and refine your search, if needed, after searching on STN.
Turn MMS results into STN search scripts

The list of MMS-Derwent WPI compound numbers on page 27, and the list of MMS-Pharmsearch patent numbers on page 30, can be turned into simple search scripts for searching in the WPINDEX file on STN. One method of doing this using straightforward *Microsoft Excel* functions is shown below.

### MMS compound numbers

1. Copy the list of MMS Derwent WPI compound numbers (page 27)

2. Paste the compound numbers into a Excel Spread sheet; use standard “text to columns” feature to produce a simple list of CNs
3. Use the CONCATENATE function to turn each compound number into a simple search statement: e.g. CONCATENATE(“S ”; A1;“/DCN”).

A simple list of compound number (DCN) search statements is produced:

- S RA3H0Q/DCN
- S 0018-19801/DCN
- S 9903-HB001/DCN
- S 9807-26403/DCN
- S 9645-39703/DCN
- S 9637-00802/DCN
- S 9637-00801/DCN
- S 9418-44103/DCN
- S 9238-14101/DCN
- S 9008-31802/DCN
- S 9002-12202/DCN
- S 8940-09710/DCN
- S 8914-02901/DCN
- S R16158/DCN
- S 8821-12501/DCN
- S R01255/DCN

Helpful HINT

On STN you are limited to 999 L-numbered sets per search session. If you have more than 500 or so CNs to crossover a slightly modified technique is advised: simply create search statements which comprise two or three CNs per line, e.g.

S (8914-02901 OR R16158 OR 8821-12501)/DCN
MMS-Pharmsearch patent numbers

1. Copy the list of MMS-Pharmsearch Patent Numbers (page 30)

2. Paste the patent numbers into an Excel spreadsheet; use standard “text to columns” feature to produce a simple list of PNs
3. Use the CONCATENATE function to turn each patent number into a simple search statement: e.g. CONCATENATE(“S ”; A1;“/PN”).

A simple list of patent number search statements is produced:

<table>
<thead>
<tr>
<th>Patent Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>S DE4207922/PN</td>
</tr>
<tr>
<td>S DE4222826/PN</td>
</tr>
<tr>
<td>S EP—1924/PN</td>
</tr>
<tr>
<td>S EP—11745/PN</td>
</tr>
<tr>
<td>S EP—21337/PN</td>
</tr>
<tr>
<td>S EP—22771/PN</td>
</tr>
<tr>
<td>S EP—25546/PN</td>
</tr>
<tr>
<td>S EP—69307/PN</td>
</tr>
<tr>
<td>S EP—91964/PN</td>
</tr>
<tr>
<td>S EP—98113/PN</td>
</tr>
<tr>
<td>S EP—105065/PN</td>
</tr>
<tr>
<td>S EP—106335/PN</td>
</tr>
<tr>
<td>S EP—117531/PN</td>
</tr>
<tr>
<td>S EP—120248/PN</td>
</tr>
</tbody>
</table>
Recall the Search Question…

What is the patent prior-art for the drug diazepam (valium): either when described directly, or when unambiguously defined within a Markush chemical formula?

Diazepam:

Search Strategy
To find patents which specifically describe the drug diazepam

Step 1  MMS structure search for diazepam
Step 2  Extract Derwent WPI compound numbers
Step 3  Crossover MMS results to PHARM
Step 4  Extract patent numbers from PHARM
Step 5  Transfer MMS results to WPIDS
Step 6  WPIDS chemical code search for diazepam
Step 7  CASLINK structure search for diazepam
Step 8  Combine, de-duplicate and FSORT all results
WPINDEX

Having retrieved and formatted relevant search terms from MMS these can be uploaded to the WPINDEX (WPIDS) file on STN via the Express Command Window.

Derwent WPI compound numbers from MMS

Access Derwent WPI (e.g. subscriber file WPIDS):

=> FILE WPIDS

FILE ’WPIDS’ ENTERED AT 20:32:57 ON 28 SEP 2001
COPYRIGHT (C) 2001 DERWENT INFORMATION LTD

Copy & paste the compound numbers (DCN) to the Command Window and upload:

=> S RA3H0Q/DCN
L1 1 RA3H0Q/DCN

=> S 0018–19801/DCN
L2 1 0018–19801/DCN

=> S 9903–HBJ01/DCN
L3 1 9903–HBJ01/DCN

=> S 9807–26403/DCN
L4 1 9807–26403/DCN

=> S 9645–39703/DCN
L5 1 9645–39703/DCN

=> S 9637–00802/DCN
L6 1 9637–00802/DCN

=> S 9637–00801/DCN
L7 1 9637–00801/DCN

=> S 9418–44103/DCN
L8 1 9418–44103/DCN

=> S 9238–14101/DCN
L9 1 9238–14101/DCN

=> S 9008–31802/DCN
L10 1 9008–31802/DCN
The Derwent Compound Number (DCN) field on STN also has a simple Thesaurus function, incorporating the names and numbers of 20,000 of the most common compounds from 1987 to date.

Helpful HINT

```
=> E DIAZEPAM+ALL/DCN
E1 0 --> DIAZEPAM/DCN
E2 230 USE R01255/DCN
******* END***
```
Pharmsearch Patent Numbers

Copy & paste the patent numbers to the Command Window and upload:

- S DE4207922/PN
- S FR-839/PN
- S WO9802148/PN

Combine into one answer set and delete unnecessary L-numbers:

- S L2-L152
- DEL L2-L152 Y

L-Numbers reassigned. Enter 'DIS HIST' to see current assignment.

Helpful HINT

L-Numbers are re-assigned when SET RENUMBER is ON. Add PERM to make this the default:

- SET RENUMBER ON PERM

SET COMMAND COMPLETED
Derwent WPI Chemical Fragmentation Codes

STN Express incorporates an offline module for generating Derwent Chemical Fragmentation Codes from a graphical structure query. The codes can then be searched by Derwent Subscribers in files WPIDS or WPIX. STN is the only host to provide this facility within its online communication software.

**note**
The Merged Markush Service (MMS) is an open access database available to all online searchers. However, subscription to Derwent *Chemical Patents Index (CPI)* is required to use Chemical Fragmentation Codes.

**Prepare the fragmentation code query offline:**

1. Click on the structure query preparation button

![Structure Query Preparation Button](image1)

2. Open a new WPI format window

![WPI Format Window](image2)

3. Prepare and Save the query:

![Fragmentation Code Preparation](image3)
4. Click on the Generate WPI button

5. Open the WPI-format structure file; select the search fields (SUBS)

*note* Each online search field - or Subsection (SUBS) - relates to a broad chemical subject area. Searching in fields M0, M2, M3 and M4 covers all areas of chemistry relevant to our structure.
Logon to STN and access file WPIDS (or WPIDS):

FILE WPIDS ENTERED AT 20:22:23 ON 28 SEP 2001
COPYRIGHT (C) 2001 DERWENT INFORMATION LTD

Run the Chemical Fragmentation Code script file:

```plaintext
=> S (D780(P)G100(P)H602(P)J521(P)M412(P)M531) /M0,M2,M3,M4
L3 1170 (D780(P)G100(P)H602(P)J521(P)M412(P)M531) /M0,M2,M3,M4
=> S L3(P)M113(P)M511(P)M520) /M2,M3,M4
L4 910 L3(P)M113(P)M511(P)M520) /M2,M3,M4
=> S L4(P)M281(P)M320(P)M320(P)M270 OR M273) /M2,M3,M4
L5 721 L4(P)M211(P)M320(P)M320(P)M270 OR M273) /M2,M3,M4
=> S L5(P)01829/RIN
L6 603 L5(P)01829/RIN
=> S L6(P)D014(P)D022(P)G010(P)H641(P)"L941") /M2,M3,M4
L7 370 L6(P)D014(P)D022(P)G010(P)H641(P)"L941") /M2,M3,M4
=> S (L3(P)M900/M0) OR (L4(P)M901/M2,M3,M4) OR (L6(P)M902/M2,M3,M4)
L8 410 (L3(P)M900/M0) OR (L4(P)M901/M2,M3,M4) OR (L6(P)M902/M2,M3,M4)
=> S L8 OR L7
L9 780 L8 OR L7
=> S L9(NOTP)(H1 OR H3 OR H4 OR H6 OR H8 OR H9 OR J0 OR J1 OR J3 OR J2) /M2,M3,M4
L10 741 L9(NOTP)(H1 OR H3 OR H4 OR H6 OR H8 OR H9 OR J0 OR J1 OR J3 OR J2) /M2,M3,M4
=> S L10(NOTP)(J3 OR J4 OR J6 OR J8 OR J9 OR J1 OR K1 OR K2 OR K3 OR K4 OR K5 OR K6) /M2,M3,M4
L11 741 L10(NOTP)(J3 OR J4 OR J6 OR J8 OR J9 OR J1 OR K1 OR K2 OR K3 OR K4 OR K5 OR K6) /M2,M3,M4
```
The Complete Markush Structure Search: Mission Impossible?

⇒ S L11 (NOTP) (K7 OR K8 OR K9 OR "L1" OR "L2" OR "L3" OR "L4" OR "L5" OR "L6") / M2, M3, M4
L12 741 L11 (NOTP) (K7 OR K8 OR K9 OR "L1" OR "L2" OR "L3" OR "L4"
"L5" OR "L6") / M2, M3, M4
⇒ S L12 (NOTP) ("L7" OR "L8") / M2, M3, M4
L13 741 L12 (NOTP) ("L7" OR "L8") / M2, M3, M4

Delete the unnecessary L-numbers:

⇒ DEL L3-L12 Y
L-numbers reassigned. Enter 'DIS HIST' to see current assignment

⇒ D HIS

FILE 'WPIDS' ENTERED AT 20:22:23 ON 28 SEP 2001
L1 241 L***-L***
L2 149 L***-L***
L3 741 SEA L*** (NOTP) ("L7" OR "L8") / M2, M3, M4

Isolate fragmentation code answers which pre-date MMS-DWPI results (1963-1986):

⇒ L3 (NOTP) M904/M2, M3, M4
L4 487 L3 (NOTP) M904/M2, M3, M4

M904 is a fragmentation control code which means MMS indexing is also available. The M904 compounds are therefore NOTed out to give just the pre-MMS answers (i.e. 1963-1986).

Bring INPI and Derwent answers together

⇒ S L1 OR L2 OR L4
L5 791 L1 OR L2 OR L4

L1 MMS-DWPI: 241 records
L2 MMS-Pharmsearch: 149 records
L3 DWPI-Frag Code: 741 records
**CASLINK**

The Chemical Abstracts Service (CAS) provides the following structure files of interest for this search example: REGISTRY, MARPAT and MARPATPREV. In addition to providing the databases individually, CAS also links them together - along with the CAPLUS bibliographic file - to form the specialised CASLINK database cluster. Using CASLINK simplifies the structure search process by allowing the three structure databases to be interrogated simultaneously and de-duplicating the results.

*Prepare the structure query:*

1. Click on the structure query preparation button

![Structure Query Preparation Button](image)

2. Prepare and Save the query

![Structure Query](image)

*Notes (see page 12):*

1. STN Express default MARPAT match level for Cl and CH₃ must be changed to atom to be the same as the MTF default.
2. Unfilled valencies must be defined as H to be the same as the MTF default. (Alternatively the closed CSS search option can be used).

---

**Logon to STN and access CASLINK:**

```bash
=> FILE CASLINK

FILE 'REGISTRY' ENTERED AT 20:22:23 ON 28 SEP 2001
COPYRIGHT (C) 2001 American Chemical Society (ACS)

FILE 'MARPAT' ENTERED AT 20:22:23 ON 28 SEP 2001
COPYRIGHT (C) 2001 American Chemical Society (ACS)

FILE 'MARPATPREV' ENTERED AT 20:22:23 ON 28 SEP 2001
COPYRIGHT (C) 2001 American Chemical Society (ACS)
```
Upload the structure query:

Upload C:\Programm\stnexp\Queries\diazepam stn.str

L6 STRUCTURE UPLOADED

=> D
L6 HAS NO ANSWERS
L6 STR

Structure attributes must be viewed using STN Express query preparation.
Test the query with a SAMPLE search:

⇒ S L6 SAM

S L6 SSS SAM FILE=REGISTRY
SAMPLE SEARCH INITIATED 20:23:07 FILE 'REGISTRY'
SAMPLE SCREEN SEARCH COMPLETED - 117 TO ITERATE

100.0% PROCESSED 117 ITERATIONS 10 ANSWERS
SEARCH TIME: 00.00.01

FULL FILE PROJECTIONS: ONLINE **COMPLETE**
BATCH **COMPLETE**
PROJECTED ITERATIONS: 1692 TO 2988
PROJECTED ANSWERS: 11 TO 389

L7 10 SEA SSS SAM L1
1 FILES SEARCHED...

S L6 SSS SAM FILE=MARPAT
SAMPLE SEARCH INITIATED 20:23:08 FILE 'MARPAT'
SAMPLE SCREEN SEARCH COMPLETED - 26 TO ITERATE

100.0% PROCESSED 26 ITERATIONS 1 ANSWERS
SEARCH TIME: 00.00.02

FULL FILE PROJECTIONS: ONLINE **COMPLETE**
BATCH **COMPLETE**
PROJECTED ITERATIONS: 214 TO 826
PROJECTED ANSWERS: 1 TO 80

L8 1 SEA SSS SAM L1
1 FILES SEARCHED...

Review effectiveness of the search in REGISTRY:

⇒ D L7 SCAN

L7 10 ANSWERS REGISTRY COPYRIGHT 2001 ACS
IN 2H-1,4-Benzodiazepin-2-one, 7-chloro-1,3-dihydro-1-methyl-5-phenyl-,
mixt. with 2-[2-[4-[(4-chlorophenyl)phenylmethyl]-1-
piperazinyl]ethoxy]ethanol (9CI)
MF C21 H27 CI N2 O2 . C16 H13 CI N2 O
CI MS

Ph

Cl

O

Me

CM 1

CM 2

45
Review effectiveness of the search from MARPAT:

L8  1 ANSWERS  MARPAT  COPYRIGHT 2001 ACS
IC  ICM  C07D403-06
    ICS  G01N033-53; C07K016-00
ICA  C12Q001-28
ICI  C07D403-06, C07D243-24, C07D207-416
CC  28-21 (Heterocyclic Compounds (More Than One Hetero Atom))
    Section cross-reference(s): 15
TI  Preparation of benzodiazepine protein conjugates.
    ST  benzodiazepine protein conjugate prep.; immunoassay benzodiazepine
    IT  Immunoassay
        (immunoassay of benzodiazepines using antibodies prepd. by using
         benzodiazepine protein conjugates)
IT  604-75-1, Oxazepam  846-49-1, Lorazepam  846-50-4, Temazepam
    1812-30-2, Bromazepam  3703-53-5, Temazepam glucuronide  6801-81-6,
    Oxazepam glucuronide  28981-97-7, Alprazolam  30896-57-2,
    4-Hydroxyalprazolam  32781-79-6, Lorazepam glucuronide  34084-50-9
    65686-11-5, 4-Hydroxytriazolam
RL: ANT (Analyte); ANST (Analytical study)
    (immunoassay of benzodiazepines using antibodies prepd. by using
     benzodiazepine protein conjugates)

ALL ANSWERS HAVE BEEN SCANNED

note  MMS does not offer equivalent features to the SCAN format or SAMPLE
      search options. Both of features are free-of-charge.
Conduct a full file search:

⇒ S L6 FUL

S L6 SSS FUL FILE=REGISTRY
FULL SEARCH INITIATED 20:23:37 FILE 'REGISTRY'
FULL SCREEN SEARCH COMPLETED - 2194 TO ITERATE

100.0% PROCESSED 2194 ITERATIONS 149 ANSWERS
SEARCH TIME: 00.00.01

L9 149 SEA SSS FUL L6
1 FILES SEARCHED...

S L9 SSS FUL FILE=MARPAT
FULL SEARCH INITIATED 20:23:38 FILE 'MARPAT'
FULL SCREEN SEARCH COMPLETED - 424 TO ITERATE

100.0% PROCESSED 424 ITERATIONS 4 ANSWERS
SEARCH TIME: 00.00.06

L10 4 SEA SSS FUL L6
1 FILES SEARCHED...

S L10 SSS FUL FILE=MARPATPREV
FULL SEARCH INITIATED 20:23:46 FILE 'MARPATPREV'
FULL SCREEN SEARCH COMPLETED - 2 TO ITERATE

100.0% PROCESSED 2 ITERATIONS 0 ANSWERS
SEARCH TIME: 00.00.01

L11 0 SEA SSS FUL L6
1 FILES SEARCHED...

S L9 FILE=CAPLUS
L12 11073 FILE CAPLUS
1 FILES SEARCHED...

SET DUPORDER FILE
SET COMMAND COMPLETED

DUP REM L11 L10 L12
L11 HAS NO ANSWERS
PROCESSING COMPLETED FOR L11
PROCESSING COMPLETED FOR L10
PROCESSING COMPLETED FOR L12

L13 11069 DUP REM L11 L10 L12 (8 DUPLICATES REMOVED)
ANSWERS '1-4' FROM FILE MARPAT
ANSWERS '5-11069' FROM FILE CAPLUS

CASLINK search results:
L9 Registry substance records (149)
L10 Marpat patent records (4)
L11 Marpat Previews patent records (0)
L12 CAPLUS bibliographic records (11,073)
L13 CAPLUS/MARPAT deduplicated results (11,069)
**Limit the CASLINK search to patent references:**

| L12 AND P/DT FILE=CAPLUS | L14 562 FILE CAPLUS
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 FILES SEARCHED...</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>L10 AND P/DT FILE=CAPLUS</th>
<th>L15 4 FILE CAPLUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 FILES SEARCHED...</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>L11 AND P/DT FILE=CAPLUS</th>
<th>L16 0 FILE CAPLUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 FILES SEARCHED...</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>L15 AND L10 FILE=MARPAT</th>
<th>L17 4 FILE MARPAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 FILES SEARCHED...</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>L16 AND L11 FILE=MARPATPREV</th>
<th>L18 0 FILE MARPATPREV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 FILES SEARCHED...</td>
<td></td>
</tr>
</tbody>
</table>

DUP REM L18 L17 L14
L18 HAS NO ANSWERS
PROCESSING COMPLETED FOR L18
PROCESSING COMPLETED FOR L17
PROCESSING COMPLETED FOR L14

<table>
<thead>
<tr>
<th>L19 562 DUP REM L18 L17 L14 (4 DUPLICATES REMOVED)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANSWERS '1-4' FROM FILE MARPAT</td>
</tr>
<tr>
<td>ANSWERS '5-562' FROM FILE CAPLUS</td>
</tr>
</tbody>
</table>

The CASLINK search has retrieved **562 patent records** (L19).
Combining CAPLUS and WPINDEX answers

Remove duplicates:

DUP REM L5 L19

FILE 'WPIDS' ENTERED AT 21:00:04 ON 28 SEP 2001
COPYRIGHT (C) 2001 DERWENT INFORMATION LTD

FILE 'CAPLUS' ENTERED AT 21:00:04 ON 28 SEP 2001
COPYRIGHT (C) 2001 AMERICAN CHEMICAL SOCIETY (ACS)

FILE 'MARPAT' ENTERED AT 21:00:04 ON 28 SEP 2001
COPYRIGHT (C) 2001 American Chemical Society (ACS)

L20 1132 DUP REM L5 L19 (221 DUPLICATES REMOVED)
ANSWERS '1-791' FROM FILE WPIDS
ANSWERS '792-1131' FROM FILE CAPLUS
ANSWER '1132' FROM FILE MARPAT

DUP REM de-duplicates the results in the order of file preference specified: WPIDS (L5) over CAS (L19).

note

DUP REM is based on the basic patent. CAPLUS and WPINDEX do not necessarily have the same basic patent number for a given family. It is therefore important to use FSORT to determine the actual number of patent families (below).

Determine the number of patent families with FSORT:

FSORT L20

SET SMARTSELECT ON
SET COMMAND COMPLETED

SET HIGHLIGHTING OFF
SET COMMAND COMPLETED

SEL L20 1- PN,APPS
L21 SEL L20 1- PN APPS : 9178 TERMS

'I21 ' DELETED
L21 1132 FS0 L20

156 Multi-record Families
Family 1
.
.
Family 155
752 Individual Records
0 Non-patent Records

From 1132 records FSORT provides: 908 patent families, i.e. 156 + 752.

FSORT identifies and groups patent families - including any duplicate records - across the databases. Within the groups the records are placed in file preference order, i.e. WPIDS first.
Review the answers one record per family:

Using the PFAM display it is possible to display just one member of each family.

CAPLUS multi record families where no WPIDS record was retrieved.

CAPLUS single record families where no WPIDS record was retrieved.

A uniquely retrieved MARPAT record.
A review of the search history:

=> D HIS

(FILE 'HOME' ENTERED AT 20:22:04 ON 28 SEP 2001)

FILE 'WPIDS' ENTERED AT 20:22:23 ON 28 SEP 2001

<table>
<thead>
<tr>
<th>L</th>
<th>S</th>
<th>L***-L***</th>
<th>MMS DWPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>241</td>
<td>S L***-L***</td>
<td>MMS Pharmsearch (+1 BSM)</td>
</tr>
<tr>
<td>L2</td>
<td>149</td>
<td>S L***-L***</td>
<td>DWPI Frag Codes, 1963-</td>
</tr>
<tr>
<td>L3</td>
<td>741</td>
<td>SEA L*** (NOTP) (&quot;L7&quot; OR &quot;L8&quot;)/M2,M3,M4</td>
<td>DWPI Frag Codes, 1963-86</td>
</tr>
<tr>
<td>L4</td>
<td>487</td>
<td>SEA L3 (NOTP)/M504/M2,M3,M4</td>
<td>Total DWPI/Pharmsearch</td>
</tr>
<tr>
<td>L5</td>
<td>791</td>
<td>L1 OR L2 OR L4</td>
<td></td>
</tr>
</tbody>
</table>

FILE 'REGISTRY, MARPAT, MARPATPREV, CAPLUS' ENTERED AT 20:32:57 ON 28 SEP 2001

<table>
<thead>
<tr>
<th>L</th>
<th>S</th>
<th>L***-L***</th>
<th>MMS DWPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>L6</td>
<td></td>
<td>STRUCTURE UPLOADED</td>
<td></td>
</tr>
<tr>
<td>L7</td>
<td>10</td>
<td>S L6 SSS SAM FILE=REGISTRY</td>
<td></td>
</tr>
<tr>
<td>L8</td>
<td>1</td>
<td>S L7 SSS SAM FILE=MARPAT</td>
<td></td>
</tr>
<tr>
<td>L9</td>
<td>149</td>
<td>S L6 SSS FILE=REGISTRY</td>
<td></td>
</tr>
<tr>
<td>L10</td>
<td>4</td>
<td>S L9 SSS FILE=MARPAT</td>
<td></td>
</tr>
<tr>
<td>L11</td>
<td>0</td>
<td>S L10 SSS FILE=MARPATPREV</td>
<td></td>
</tr>
<tr>
<td>L12</td>
<td>11073</td>
<td>S L9 FILE=CAPLUS</td>
<td></td>
</tr>
<tr>
<td>L13</td>
<td>11069</td>
<td>DUP REM L11 L10 L12 (8 DUPLICATES REMOVED)</td>
<td>CASLINK, total</td>
</tr>
<tr>
<td>L14</td>
<td>562</td>
<td>S L12 AND P/D FILE=CAPLUS</td>
<td></td>
</tr>
<tr>
<td>L15</td>
<td>4</td>
<td>S L10 AND P/D FILE=CAPLUS</td>
<td></td>
</tr>
<tr>
<td>L16</td>
<td>0</td>
<td>S L11 AND P/D FILE=CAPLUS</td>
<td></td>
</tr>
<tr>
<td>L17</td>
<td>4</td>
<td>S L15 AND L5 FILE=MARPAT</td>
<td></td>
</tr>
<tr>
<td>L18</td>
<td>0</td>
<td>S L16 AND L6 FILE=MARPATPREV</td>
<td></td>
</tr>
</tbody>
</table>
| L19 | 562 | DUP REM L18 L17 L14 (4 DUPLICATES REMOVED) | CASLINK, patents only

FILE 'WPIDS, CAPLUS, MARPAT' ENTERED AT 21:00:04 ON 28 SEP 2001

<table>
<thead>
<tr>
<th>L</th>
<th>S</th>
<th>L***-L***</th>
<th>MMS DWPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>L20</td>
<td>1132</td>
<td>DUP REM L5 L19 (221 DUPLICATES REMOVED)</td>
<td>CAPLUS/WPIDS deduplicated</td>
</tr>
<tr>
<td>L21</td>
<td>1132</td>
<td>FSO L26</td>
<td>CAPLUS/WPIDS FamilySorted</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SET SMARTSELECT ON</td>
<td>(908 patent families)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SET HIGHLIGHTING OFF</td>
<td></td>
</tr>
</tbody>
</table>

note

The Derwent Registry Number (DRN) field provides additional indexing for 2000 compounds from 1981 to date, and may often yield extra search results. The field has a simple Thesaurus.

== E DIAZEPAM/ALL/DRN
E1 0 --> DIAZEPAM/DRN
E2 341 USE 1255/DRN
******** END***
Identifying unique results

If needed it is possible to continue the search to identify the comparative volume of unique answers that each database producer has brought to the final answer set.

Patent families can be spread over several records, and not necessarily over the same number of records in CAPLUS compared to WPINDEX. Therefore simple record number counts for each file is not a precise method for assessing the number of unique answers from CA and Derwent/INPI. Using the DUP REM command gives a good impression of unique results from each service, but is only based upon the basic patent and therefore does not produce 100% de-duplication (see note on page 49).

One technique to make the comparison of the results more meaningful, is to attempt to merge the answer sets into one of either WPINDEX or CAPLUS, and count comparative records there. This is achieved by crossing over patent numbers from one file to the other using the TRANSFER (or SELECT) commands. In the continuation of the search example, WPINDEX (WPIDS) has been chosen as the target file to help explain this approach.

| note | The following steps are not actually necessary to complete the search. They are, however, of interest to the theme of this paper. This is why they are included here. |

Search Example continued....
The following search strategy continues directly on from the previous search example (see search history display on page 51).

**Search Strategy**
To identify unique answers from each service

**Step 1**  Transfer CAPLUS PNs to WPIDS

**Step 2**  Identify CAPLUS records not in WPIDS

**Step 3**  Use a series of Boolean statements to identify unique answers provided by each service as represented by WPIDS record counts
Transfer Patent Numbers from CAPLUS to WPIDS

Access Derwent WPI (file WPIDS):

=> FIL WPIDS

FILE 'WPIDS' ENTERED AT 21:19:15 ON 28 SEP 2001
COPYRIGHT (C) 2001 DERWENT INFORMATION LTD

SET AUDIT ON to monitor the accuracy of the cross-file searches:

=> SET AUDIT ON
SET COMMAND COMPLETED

Transfer basic patent numbers (PN.B) from CAPLUS (L19) to WPIDS:

=> TRA L19 PN.B 1-
L22 TRANSFER L19 1- PN.B : 562 TERMS
L23 508 L22
L24 QUE TERMS FROM L22 WITH NO HITS: 40 TERMS

=> D L24 1-
L24 QUE TERMS FROM L28 WITH NO HITS: 40 TERMS

TERM # TERMS
------ -------------------------
1 AT324343/PN

The 40 missing PNs (L24) comprise:
AT (2), AU (1), CN (2), CS (9), DD (1), DE (1), EP (1), ES (4), HU (2), IN (1), PL (7), RO (2), US (2), WO (2), ZA (1)

note There might be equivalents to some of these 40 missing PNs present in WPIDS. These can be located, e.g. via their Priority Number (PRN).

Return to CAPLUS and retrieve those patents not found in WPIDS (L24):

=> FIL CAPLUS

FILE 'CAPLUS' ENTERED AT 21:40:12 ON 28 SEP 2001
COPYRIGHT (C) 2001 AMERICAN CHEMICAL SOCIETY (ACS)

=> S L24
125 40 L24
Extract Priority Numbers from CAPLUS to find extra records in WPIDS:

\[
=> 	ext{FIL WPIDS}
\]

FILE 'WPIDS ENTERED AT 21:42:47 ON 28 SEP 2001
COPYRIGHT (C) 2001 DERWENT INFORMATION LTD

\[
=> 	ext{TRA L25 PRN 1-}
\]

L26 TRANSFER L25 1- PRN : 53 TERMS
L27 14 L26
L28 QUE TERMS FROM L26 WITH NO HITS: 26 TERMS

14 WPIDS records are retrieved (L27) via PRN.

Combine the PN.B answer set (L23) with the extra PRN answer set (L27):

\[
=> \text{L23 OR L27}
\]

L29 522 L23 OR L27

L23 (508) + L27 (14) = L29 (522)

Identify CAPLUS records not in WPIDS

How many of the “40 missing PNs” CAPLUS records (L25) correspond directly to the 14 additional WPIDS records obtained via PRN (L27)? In other words, how many of these 40 records are only found in CAPLUS?

Extract PRNs from WPIDS (L27) and search in CAPLUS:

\[
=> 	ext{FIL CAPLUS}
\]

FILE 'CAPLUS' ENTERED AT 21:48:12 ON 28 SEP 2001
COPYRIGHT (C) 2001 AMERICAN CHEMICAL SOCIETY (ACS)

\[
=> \text{TRA L27 PRN 1-}
\]

L30 TRANSFER L27 1- PRN : 35 TERMS
L31 14 L29
L32 QUE TERMS FROM L30 WITH NO HITS: 4 TERMS

AND the result (L30) with the “40 missing PNs” CAPLUS answer set (L25):

\[
=> \text{L25 AND L31}
\]

L33 6 L25 AND L31

\[
=> \text{L25 NOT L30}
\]

L34 34 L25 NOT L31

Conclusion: only 6 of the 40 CAPLUS records were found in WPIDS via PRN (L33). This would appear to leave 34 without directly corresponding records in WPIDS (L33). I.e. 34 out of 562 answers (L19) are not present in WPIDS at all.
Unique answers based on WPIs record counts

**A brief review of the search history:**

| L1  | 241 | MMS (Derwent via DCN) |
| L2  | 149 | MMS (Pharmsearch via PN) (+1 BSM patent, see page 31) |
| L3  | 741 | Derwent Chemical Fragmentation Code, 1963- |
| L4  | 487 | Derwent Chemical Fragmentation Code, 1963-86 |
| L5  | 791 | Total Derwent + Pharmsearch |
| L13 | 11069 | CASLINK, total |
| L19 | 562 | CASLINK, patents only |
| L20 | 1132 | CASLINK (L19) + WPIs (L5) (221 records de-duplicated) |
| L21 | 1132 | CASLINK + WPIs (L20) FamilySorted (908 patent families) |
| L23 | 508 | WPIs from CASLINK (L19) via PN.B |
| L25 | 40 | CASLINK (L19) PN.B not in WPIs |
| L27 | 14 | WPIs from CASLINK (L25) via PRN |
| L29 | 522 | WPIs from CASLINK via PN.B (L19) & PRN (L25) |
| L33 | 6 | CASLINK (L25) in WPIs via PRN |
| L34 | 34 | CASLINK (L25) not in WPIs |

Simple Boolean search statements can be used to further analyse the unique records derived from CA, Derwent and INPI resources – as represented by Derwent WPI (WPIs) records. It is important to remember when reviewing the following figures that 34 CAS and 1 Pharmsearch (French BSM) patent results were not found in WPIs.

**Example Boolean Search statements in WPIs:**

```
=> D HIS L35-

FILE 'WPIDS' ENTERED AT 21:53:39 ON 28 SEP 2001

L35  997 L5 OR L29  Total CAS*, Derwent & INPI
L36  475 L5 NOT L29  Derwent + INPI not CAS
L37  206 L29 NOT L5  CAS* not Derwent + INPI
L38  344 L1 OR L2  MMS (Derwent + INPI)
L39  195 L1 NOT L2  MMS (Derwent) not MMS (INPI)
L40  103 L2 NOT L1  MMS (INPI) not MMS (Derwent)
L41  335 L29 NOT L38  CAS* not MMS
L42  157 L38 NOT L29  MMS not CAS
L43  113 L1 NOT L29  MMS (Derwent) not CAS
L44  54 L2 NOT L29  MMS (INPI) not CAS
L45  153 L41 RAN=(-1979)  Pre-1980 CAS not MMS
L46  182 L41 RAN=(1980-)  1980-date CAS not MMS
L47  334 L4 NOT L29  Derwent Fragment Code (1963-86) not CAS
L48  240 L29 RAN=(-1986)  Pre-1987 CAS
L49  87 L48 NOT L4  Pre-1987 CAS not Derwent Fragment Code (1963-86)

(* note L34 above for CAS)
```
Examples of unique answers

Below are some selected examples demonstrating the unique value that each service has brought to achieving the final result.

Chemical Abstracts Service

**CAS records retrieved not in Derwent WPI (28/9/01):**

| AN   | 2001:581687 CAPLUS |
| TI   | Continuous method for preparing pharmaceutical granules |
| PA   | Rhodia Chimie, Fr. |
| PI   | WO 2001056549 A1 20010809 WO 2001-FR225 20010124 |
| **Text from WO 2001056549 (Claim 10):** |
| ...les psychotropes (trimipramine, amineptine, chlorpromazine et derives des phenothiazines, diazepam, lorazepam, is nitrazepam, mebroamate, .... |

| AN   | 1981:145357 CAPLUS |
| TI   | Pharmaceutical composition with improved absorption properties |
| PI   | AT 360163 B 19801229 AT 1978-4020 19780602 |

**CAS unique versus Merged Markush Service (MMS)**

| AN   | 2001-091624 WPIDS |
| TI   | New method for assaying a drug candidate with a biosensor having one or more surface-bound biomolecules where the absorption, distribution, metabolism, or excretion data may predict the behavior and fate of the drug candidate. |
| PA   | (BIAC-N) BIACORE AB |
| PI   | WO 2000079268 A2 20001228 (200110)* EN 78p G01N033-48 |
| **Text from WO 2000079268 (disclosure):** |
| ...on the same sensor surface, in duplicate, on the day of HSA immobilization (shown) and in duplicate a week later (not shown). Diazepam and warfarin were used as marker compounds.... |

| AN   | 2001-258613 WPIDS |
| TI   | Medicine comprising traditional Chinese medicine and western medicine is used for treating cough and asthma diseases. |
| PA   | (LUSS-I) LU S |
| PI   | CN 1277866 A 20001227 (200127)* A61K035-78 |

**CAS unique versus Derwent Fragmentation Code (1963-1986):**

| AN   | 1980-87365C WPIDS |
| TI   | Diazepam oily drug used as tranquilliser - composed of diazepam and fatty acid tri glyceride to increase absorption velocity. |
| PA   | (SUMO) SUMITOMO CHEM CO LTD |
| PI   | JP 55136219 A 19801023 (198049)* |
**Merged Markush Service (Derwent)**

**Derwent unique versus INPI:**

<table>
<thead>
<tr>
<th>AN</th>
<th>1996-363309 [37] WPIDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TI</td>
<td>Benzodiazepin protein conjugates which are immunogens – for prodn. of antibodies against benzodiazepin(s) which can be used in immunoassays to detect benzodiazepin(s).</td>
</tr>
<tr>
<td>PA</td>
<td>(BOEF) BOEHRINGER MANNHEIM GMBH; (HOFF) ROCHE DIAGNOSTICS GMBH</td>
</tr>
<tr>
<td>PI</td>
<td>DE 19503320 A1 19960808 (199637)* 14p C07D403-06</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AN</th>
<th>1992-310718 [38] WPIDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TI</td>
<td>Crystallisation of 1,4-benzodiazepine with high quality – by using cyclohexane as crystallisation solvent.</td>
</tr>
<tr>
<td>PA</td>
<td>(SUMO) SUMITOMO CHEM CO LTD</td>
</tr>
<tr>
<td>PI</td>
<td>JP 04210681 A 19920731 (199238)* 4p C07D243-24</td>
</tr>
</tbody>
</table>

**Derwent unique versus CAS:**

<table>
<thead>
<tr>
<th>AN</th>
<th>2000-352522 [31] WPIDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TI</td>
<td>Composition for formation of powder for preparation of injection solution comprises poorly water-soluble drug, cyclodextrin and water soluble organic solvent.</td>
</tr>
<tr>
<td>PA</td>
<td>(TAKE) TAKEDA CHEM IND LTD</td>
</tr>
<tr>
<td>PI</td>
<td>EP 1004318 A2 20000531 (200031)* EN 12p A61K047-48</td>
</tr>
</tbody>
</table>

Text from EP 1004318 (Claim 5):

...the slightly water-soluble drug is selected from diazepam, lorazepam, oxazepam, griseofulvin, lankacidins, ...

<table>
<thead>
<tr>
<th>AN</th>
<th>1992-310718 [38] WPIDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TI</td>
<td>Crystallisation of 1,4-benzodiazepine with high quality – by using cyclohexane as crystallisation solvent.</td>
</tr>
<tr>
<td>PA</td>
<td>(SUMO) SUMITOMO CHEM CO LTD</td>
</tr>
<tr>
<td>PI</td>
<td>JP 04210681 A 19920731 (199238)* 4p C07D243-24</td>
</tr>
</tbody>
</table>

**Merged Markush Service (INPI)**

**INPI Pharmsearch record not retrieved in Derwent WPI (or CAPLUS):**

PHARM - (C) INPI
AN - 61100839
PN - FR839M M 19611030 [FR——839]
PA - F. HOFFMANN-LA ROCHE & Cie (Socit anonyme)
EAB - 5-phenyl-3H-1,4-benzodiazepin-(1H)-2-one derivatives ....
### INPI unique versus CAS:

<table>
<thead>
<tr>
<th>AN</th>
<th>1986-095329</th>
<th>WPIDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TI</td>
<td>New use of bi cyclic diazepine derivs. - as inhibitors of platelet activating factor.</td>
<td></td>
</tr>
<tr>
<td>PA</td>
<td>BOEHRINGER INGELHEIM</td>
<td></td>
</tr>
<tr>
<td>PI</td>
<td>EP 176928</td>
<td>A 19860409</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AN</th>
<th>1980-80636C</th>
<th>WPIDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TI</td>
<td>Sub-lingual and buccal admin. of benzodiaepine(s) - for rapid absorption to affect the CNS.</td>
<td></td>
</tr>
<tr>
<td>IN</td>
<td>PORTER, W R</td>
<td></td>
</tr>
<tr>
<td>PA</td>
<td>AMERICAN HOME PROD CORP</td>
<td></td>
</tr>
<tr>
<td>PI</td>
<td>US 4229447</td>
<td>A 19801021</td>
</tr>
</tbody>
</table>

### INPI unique versus Derwent:

<table>
<thead>
<tr>
<th>AN</th>
<th>1998-002241</th>
<th>WPIDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TI</td>
<td>Use of antipyretic and benzodiazepine derivative stabilised by benzoic acid, salt or derivative - to treat e.g. infant convulsions, arthrosis and other painful disorders.</td>
<td></td>
</tr>
<tr>
<td>PA</td>
<td>LAB CRINEX SA</td>
<td></td>
</tr>
<tr>
<td>PI</td>
<td>FR 2747923</td>
<td>A1 19971031</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AN</th>
<th>1985-013506</th>
<th>WPIDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TI</td>
<td>Compositions for percutaneous administration - contg. e.g. a benzodiazepine in a mixture of higher mono alcohol and a urea, lactam, amide, thio glycerol or lactic acid cpd.</td>
<td></td>
</tr>
<tr>
<td>PA</td>
<td>LILLY &amp; CO ELLI; NITTO ELECTRIC IND CO</td>
<td></td>
</tr>
<tr>
<td>PI</td>
<td>DE 3424057</td>
<td>A 19850110</td>
</tr>
</tbody>
</table>

### Derwent Fragmentation Codes (1963-1986)

#### Derwent unique versus CAS:

<table>
<thead>
<tr>
<th>AN</th>
<th>1986-335074</th>
<th>WPIDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TI</td>
<td>Drug-contg. porous medical gels - produced without heating or using crosslinking agent to avoid damaging drug.</td>
<td></td>
</tr>
<tr>
<td>PA</td>
<td>NITTO ELECTRIC IND CO</td>
<td></td>
</tr>
<tr>
<td>PI</td>
<td>JP 61249920</td>
<td>A 19861070</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AN</th>
<th>1966-41591F</th>
<th>WPIDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TI</td>
<td>Process for 2 3-dihydro-1 4-benzodiazepine-2-one derivs anti-convulsant - ts muscle relaxants sedatives.</td>
<td></td>
</tr>
<tr>
<td>PA</td>
<td>HOFFMANN LA ROCHE &amp; CO AG F</td>
<td></td>
</tr>
<tr>
<td>PI</td>
<td>NL 6908658</td>
<td>A 196800*</td>
</tr>
</tbody>
</table>
More search examples

Objectives

The basic steps in the search example outlined on pages 18 to 58 were as follows.

- Conduct a Merged Markush Service (MMS) structure search via STN Express
- Extract relevant search results from MMS, and transfer them to the Derwent World Patents Index® on STN (files WPINDEX/WPIDS/WPIX).
- Complement MMS results with answers retrieved using an STN Express Derwent Chemical Fragmentation Code search query in WPIDS/WPIX
- Sort and de-duplicate the MMS/Derwent results, with answers retrieved using corresponding search queries in CAS Registry and MARPAT databases on STN, to end up with a single, comprehensive answer set.
- Transfer answers from CASLINK to WPIDS/WPIX to further assess unique results.

A further four searches were carried out following this technique giving five in total:

- **Diazepam**: a search for specific references to diazepam.
- **Carbapenem**: a search for carbapenem derivatives based on a common substructure.
- **Tyrosine Kinase Inhibitor**: a search for a class of tyrosine kinase inhibitors with a common substructure.
- **Sartan**: a more complex search for Sartan derivatives, based on a common substructure, but including a generic heterocycle substituent.
- **Nonanoxybenzenesulphonate (NOBS)**: a non-pharmaceutical example for which a full substructure search is not required. Answers relating specifically to the three isomers of NOBS is required. This features a variable point of attachment.

The results of all these searches, tabulated to follow the precise search order of the first example *Diazepam*, is given in the table on page 70.
Carbapenem

Search Question: Find all patent references to carbapenem derivatives based on a specific substructure.

STN Express query for CASLINK

Notes:
1. For CASLINK all undefined positions are assumed to be free for substitution.
2. The MARPAT default Match Level for chain atoms is Class (i.e., to retrieve Alkyl as well as C). For this query this is set to Atom to find an exact match only (i.e., one carbon)

Markush TOPFRAG query for MMS

Notes:
1. For MMS queries the number of substitutable positions (Free Sites) must be specified.
2. The default Translation is Exact for all real atoms in MMS. This is equivalent to MARPAT Atom Match Level (as set above).

See page 21 for how to upload this query via STN Express.
The Complete Markush Structure Search: Mission Impossible?

STN Express Query for Derwent Chemical Fragmentation Codes

**Authors advice:**
1. Follow instructions on page 39.
2. Prepare the query as if it were not a substructure query (i.e. do not specify free sites).
3. Make suitable manual modifications to the output text (see below).

Use *Generate WPI* function of *Express* to create the command file (page 39):

```plaintext
=> S (D690(P)H401(P)H481(P)J111(P)J521(P)M331(P)M412(P))/M0,M2,M3,M4 \>_line1
=> S _line1(P)(M511(P)M520(P)M530(P)M540(P))/M2,M3,M4 \>_line2
=> S _line2(P)(M280(P)M321(P)M340(P)M342(P)M391(P)(M370 OR M373))/M2,M3,M4 \>_line3
=> S _line3(P)41252/RIN \>_line4
=> S _line4(P)(D011(P)D013(P)J011(P)"L941")/M2,M3,M4 \>_line5
=> S (_line1(P)M900(M0)) OR (_line2(P)M901(M2,M3,M4)) OR (_line4(P)M902(M2,M3,M4)) \>_line6
=> S _line6 OR _line5 \>_line7
=> S _line7(NOTP)(H1 OR H2 OR H3 OR H5 OR H6 OR H7 OR H9 OR J2 OR J3 OR J4)/M2,M3,M4 \>_line8
=> S _line8(NOTP)(J6 OR J9 OR K1 OR K2 OR K3 OR K4 OR K5 OR K6 OR K7 OR K8)/M2,M3,M4 \>_line9
=> S _line9(NOTP)(K9 OR "L1" OR "L2" OR "L3" OR "L4" OR "L5" OR "L6" OR "L7")/M2,M3,M4 \>_line10
=> S _line10(NOTP)("L8" OR M1)/M2,M3,M4 \>_line11
```

To manually modify the automatic output:
1. Remove Negation codes (H1, etc)
2. Include higher priority Basic Group Code(s), e.g. M411.
3. Remove No Carbon Chain codes (M320, M280)
4. Add higher ring count code options (M5:)
5. Add further ring substitution options (D01:)
6. Add higher functional group counts, e.g. J212, J112 etc.

Manually modified command file for substructure search:

```plaintext
=> S (D690(P)J111 OR J211) OR J521(P) OR M411 OR M412) /M0,M2,M3,M4 \>_line1
=> S _line1(P)(M511 OR M512 OR M513)/M2,M3,M4 \>_line2
=> S _line2(P)(M321 OR M322 OR M323)/M2,M3,M4 \>_line3
=> S _line3(P)41252/RIN \>_line4
=> S _line4(P)(D011 OR D013) OR D019 OR D014) "L941")/M2,M3,M4 \>_line5
=> S (_line1(P)M900(M0)) OR (_line2(P)M901(M2,M3,M4)) OR (_line4(P)M902(M2,M3,M4)) \>_line6
=> S _line6 OR _line5 \>_line7
```

**note** Search results are given in the table on page 70.
Tyrosine Kinase Inhibitor

Search Question: Find all patent references to a class of tyrosine kinase inhibitors with a specific substructure.

STN Express query for CASLINK

Notes:
1. For CASLINK all undefined positions are assumed to be free for substitution.

Markush TOPFRAG query for MMS

Notes:
1. For MMS queries the number of substitutable positions (Free Sites) must be specified.

See page 21 for how to upload this query via STN Express.
STN Express Query for Derwent Chemical Fragmentation Codes

Authors advice:
1. Copy & paste the Standard Format drawing (above) into the WPI format window and edit if necessary.
2. Follow instructions on page 39.
3. Prepare the query as if it were not a substructure query (i.e. do not specify free sites)
4. Make manual modifications to the output text (see below).

Use Generate WPI function of Express to create the command file (page 39):

```plaintext
=>S (D621(P)G100(P)H121(P)M412(P)M531)/M0,M2,M3,M4 _>line1
=>S _line1(P)(M123(P)M143(P)M511(P)M520)/M2,M3,M4 _>line2
=>S _line2(P)(M280(P)M320)/M2,M3,M4 _>line3
=>S _line3(P)D011(P)G010(P)H102)/M2,M3,M4 _>line4
=>S (_line1(P)M900/M0) OR (_line2(P)M901/M2,M3,M4) OR (_line3(P)M902/M2,M3,M4) _>line5
=>S _line5 OR _line4 _>line6
=>S _line6(NOTP)(H2 OR H3 OR H4 OR H5 OR H6 OR H7 OR H8 OR H9 OR J0 OR J1)/M2,M3,M4 _>line7
=>S _line7(NOTP)(J2 OR J3 OR J4 OR J5 OR J6 OR J9 OR K0)/M2,M3,M4 _>line8
```

To manually modify the automatic output:
1. Remove Negation codes (H1, etc)
2. Include higher priority Basic Group Code(s), e.g. M411.
3. Remove No Carbon Chain codes (M320, M280)
4. Add higher ring count code options (M5:)
5. Add further ring substitution options (D01:)
6. Add higher functional group counts, e.g. H122, etc.
7. Can alcohols also be esters? Amines be amides? etc

Manually modified command file for substructure search:

```plaintext
=>S (D621(P)H12! OR J31! OR J321) (P) (M412 or M411) (P) (M531 OR M532 OR M533) )/M0,M2 _>line1
=>S _line1(P)(M123(P)M143(P)M511 OR M512 OR M513))/M2 _>line2
=>S _line2 _>line3
=>S _line3(P)((D011 OR D013 OR D014 OR D015) (P)G1!!(P) (H102 OR H103 OR J31! OR J321))/M2 _>line4
=>S (_line1(P)M900/M0) OR (_line2(P)M901/M2) OR (_line3(P)M902/M2) _>line5
=>S _line5 OR _line4 _>line6
```

note Search results are given in the table on page 70.
Sartan

Search Question: Find all patent references for Sartan derivatives, based on a common substructure, which includes an unspecified heterocycle substituent.

STN Express query for CASLINK

Notes:
1. For CASLINK all undefined positions are assumed to be free for substitution.
2. For generic groups, e.g. Hy (heterocycle), the default Express Match Level is atom. Change this to Class so that both Hy and specific heterocycles will be retrieved.

Markush TOPFRAG query for MMS

Notes:
1. For MMS there are three different generic terms for heterocycle: HEA Heteroaromatic, HEF Heterocycle Fused, HET monocyclic heterocycle (non-aromatic). A Gx group is used to include them all as fragments.
2. For generic groups, e.g. HEA, HEF, HET, the default MTF Translation is Narrow. Both generic and specific heterocycles will be retrieved.
3. For MMS queries the number of substitutable positions (Free Sites) must be specified: this includes generic groups such as HEA, HET and HEF.
4. Unspecified bonds have been used in the tetrazole ring. This avoids any problems with bond normalisation.
STN Express Query for Derwent Chemical Fragmentation Codes

Authors advice:
1. Create a query similar to the MMS query (see above).
2. Prepare the query as if were not a substructure query (i.e. do not specify free sites)
3. Follow instructions on page 39.
4. Make manual modifications to the output text (see below).

Use Generate WPI function of Express to create the command file (page 39):

```plaintext
=>S (F570 (P)G100 (P)M532)/M0,M2,M3,M4 \>_line1
=>S _line1 (P) (M111 (P)M113 (P)M132)/M2,M3,M4 \>_line2
=>S _line2 (P) (M280 (P)M311 (P)M321 (P)M342)/M2,M3,M4 \>_line3
=>S _line3 (P)00061/RIN \>_line4
=>S _line4 (P) (F015 (P)G011 (P)G013)/M2,M3,M4 \>_line5
=>S (_line1 (P)M900/M0) OR (_line2 (P)M901/M2,M3,M4) OR (_line4 (P)M902/M2,M3,M4) \>_line6
=>S _line6 OR _line5 \>_line7
```

See page 21 for how to upload this query via STN Express.
The Complete Markush Structure Search: Mission Impossible?

To manually modify the automatic output:
1. Remove Negation codes (H1, etc)
2. Include higher priority Basic Group Code(s), e.g. M411.
3. Remove No Carbon Chain codes (M320, M280)
4. Add higher ring count code options (M5:)
5. Add further ring substitution options (D01:)
6. Add further ring linkage code options (M1:)

Manually modified command file for substructure search:

Helpful HINT
Do you want to know how to manipulate Fragmentation Code strategies? There are no short cuts to understanding how to do this. The best option is to get in touch with Derwent Information and ask for a training seminar...!

Visit: www.derwent.com/training.html

note
Search results are given in the table on page 70.
Nonanoyloxybenzenesulfonate (NOBS)

**Search Question:** Find all patent references for the three isomers of the bleach additive nonanoyloxybenzenesulfonate (NOBS), including Alkanoyl answers which cover nonanoyl.

**STN Express query for CASLINK**

Notes:
1. For this search the closed substructure (CSS) search option is to be used. This emulates an MMS default: all undefined positions are assumed to be hydrogen.
2. For chain atoms the default MARPAT Match Level is Class. This is exactly what is required here: specific nonanoyl answers and MARPAT alkyl groups which include nonanoyl within their definition.

**Markush TOPFRAG query for MMS**

Notes:
1. As this is not a substructure search no free sites for substitution are required (see CSS above).
2. The atoms of the alkyl chain must be labelled with Broad Translation to retrieve both specific and generic Alkyl (superatom CHK) indexing.
STN Express Query for Derwent Chemical Fragmentation Codes

Authors advice:
1. Copy & paste the Standard Format drawing (above) into the WPI format window and edit if necessary.
2. Follow instructions on page 39.
3. Make a minor manual modification to the output text (see below).

Use Generate WPI function of Express to create the command file (page 39):

```
=>S  (G100(P)J241(P)K431(P)M414(P)M5311)/M0,M2,M3,M4 \_line1
=>S  _line1(P)(M222(P)M231(P)M281(P)M320(P) (M260 OR M262))/M2,M3,M4 \_line2
=>S  _line2(P)(J011(P) (G011 OR G012 OR G013))/M2,M3,M4 \_line3
=>S  (_line1(P)M900/M0) OR (_line1(P)M901/M2,M3,M4) OR (_line2(P)M902/M2,M3,M4) \_line4
=>S  _line4 OR _line3 \_line5
=>S  _line5( NOTP) (H1 OR H2 OR H3 OR H4 OR H5 OR H6 OR H7 OR H8 OR H9 OR J1)/M2,M3,M4 \_line6
=>S  _line6( NOTP) (J3 OR J4 OR J5 OR J6 OR J9 OR K1 OR K2 OR K3 OR K5 OR K6)/M2,M3,M4 \_line7
=>S  _line7( NOTP) (K7 OR K8 OR K9 OR "L1" OR "L2" OR "L3" OR "L4" OR "L5" OR "L6")/M2,M3,M4 \_line8
=>S  _line8( NOTP) ("L7" OR "L8" OR "L9" OR M1)/M2,M3,M4 \_line9
```

(Continued...)
To manually modify the automatic output:
1. OR the general alkyl chain code M220 (7 or more carbons) with the specific code M222 (8 carbons), as generated by STN Express.

Manually modified command file for substructure search:

```plaintext
=>S (G100(P)J241(P)K431(P)M14(P)M531)/M0,M2,M3,M4)\_line1
=>S \_line1(P)((M222 OR M220)(P)M231(P)M281(P)M320(P)(M260 OR M262))\_line2
=>S \_line2(P)(J011(P)(G011 OR G012 OR G013))/M2,M3,M4)\_line3
=>S (_line1(P)M900/M0) OR (_line1(P)M901/M2,M3,M4) OR (_line2(P)M902/M2,M3,M4)\_line4
=>S \_line4 OR \_line3\_line5
=>S \_line5(NOTP)(H1 OR H2 OR H3 OR H4 OR H5 OR H6 OR H7 OR H8 OR H9 OR J1)/M2,M3,M4)\_line6
=>S \_line6(NOTP)(J3 OR J4 OR J5 OR J6 OR J9 OR K1 OR K2 OR K3 OR K5 OR K6)/M2,M3,M4)\_line7
=>S \_line7(NOTP)(K7 OR K8 OR K9 OR "L1" OR "L2" OR "L3" OR "L4" OR "L5" OR "L6")/M2,M3,M4)\_line8
=>S \_line8(NOTP)("L7" OR "L8" OR "L9" OR M1)/M2,M3,M4)\_line9
```

**note** Search results are given in the table on page 70.

**note** The general CAS Registry Number for NOBS, 101482-85-3 (where the 1,2-, 1,3- or 1,4 isomer is not specified), is not retrieved in this structure search. In order to yield the results on page 70, this RN was added manually to the search and is included in the CASLINK result.

**Helpful HINT** Do you want to know how to manipulate Fragmentation Code strategies? There are no short cuts to understanding how to do this. The best option is to get in touch with Derwent Information and ask for a training seminar...!

Visit: [www.derwent.com/training.html](http://www.derwent.com/training.html)
Search Results

The results for all searches are tabulated to follow the order of the first example *Diazepam* (see search history display on page 55)

<table>
<thead>
<tr>
<th>Result description</th>
<th>Lx</th>
<th>Diazepam</th>
<th>Carbapenem</th>
<th>Tyrosine</th>
<th>Sartan</th>
<th>NOBS 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MMS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total CNs</td>
<td>34</td>
<td>1704</td>
<td>495</td>
<td>1250</td>
<td>258</td>
<td></td>
</tr>
<tr>
<td>Derwent WPI CNs</td>
<td>16</td>
<td>919</td>
<td>326</td>
<td>712</td>
<td>256</td>
<td></td>
</tr>
<tr>
<td>INPI Pharmsearch CNs</td>
<td>18</td>
<td>785</td>
<td>169</td>
<td>538</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td><strong>INPI PHARMSEARCH</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INPI Pharmsearch PNs</td>
<td>151</td>
<td>497</td>
<td>154</td>
<td>329</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td><strong>DERWENT WPI (STN)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Records from Derwent CNs</td>
<td>L1</td>
<td>241</td>
<td>386</td>
<td>204</td>
<td>396</td>
<td>261</td>
</tr>
<tr>
<td>Records from INPI PNs</td>
<td>L2</td>
<td>149</td>
<td>481</td>
<td>146</td>
<td>310</td>
<td>2</td>
</tr>
<tr>
<td>INPI PNs not in DWPI</td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Frag. Code (1963-date)</td>
<td>L3</td>
<td>741</td>
<td>826</td>
<td>2108</td>
<td>2249</td>
<td>1091</td>
</tr>
<tr>
<td>Total Derwent + INPI</td>
<td>L5</td>
<td>791</td>
<td>981</td>
<td>1460</td>
<td>2023</td>
<td>894</td>
</tr>
<tr>
<td><strong>CASLINK</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CASLINK (dups, removed)</td>
<td>L13</td>
<td>11069 (-4)</td>
<td>4838 (-338)</td>
<td>1328 (-51)</td>
<td>2450 (-297)</td>
<td>265 (-43)</td>
</tr>
<tr>
<td>CASLINK (Patents)</td>
<td>L19</td>
<td>562</td>
<td>891</td>
<td>464</td>
<td>546</td>
<td>257</td>
</tr>
<tr>
<td><strong>COMBINED RESULTS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (duplicates removed)</td>
<td>L20</td>
<td>1132 (-221)</td>
<td>1257 (-615)</td>
<td>1794 (-130)</td>
<td>2157 (-412)</td>
<td>1011 (-140)</td>
</tr>
<tr>
<td>FSORT patent families</td>
<td>L21</td>
<td>908</td>
<td>1024</td>
<td>1499</td>
<td>1955</td>
<td>958</td>
</tr>
<tr>
<td>MMS &amp; CAS (dup. rem.) 3</td>
<td></td>
<td>-</td>
<td>-</td>
<td>667 (-104)</td>
<td>623 (-410)</td>
<td>-</td>
</tr>
<tr>
<td>FSORT (MMS/CAS)</td>
<td></td>
<td>-</td>
<td>-</td>
<td>585</td>
<td>540</td>
<td>-</td>
</tr>
<tr>
<td><strong>DERWENT WPI (STN)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAS DWPI records (via PNs)</td>
<td>L29</td>
<td>522</td>
<td>843</td>
<td>426</td>
<td>525</td>
<td>245</td>
</tr>
<tr>
<td>CAS Patents not in DWPI</td>
<td>L34</td>
<td>34</td>
<td>11</td>
<td>13</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Total Derwent/INPI/CAS</td>
<td>L35</td>
<td>997</td>
<td>1163</td>
<td>1692</td>
<td>2122</td>
<td>995</td>
</tr>
<tr>
<td>Derwent/INPI not CAS</td>
<td>L36</td>
<td>475</td>
<td>320</td>
<td>1266</td>
<td>1597</td>
<td>750</td>
</tr>
<tr>
<td>CAS not Derwent/INPI</td>
<td>L37</td>
<td>205</td>
<td>182</td>
<td>232</td>
<td>99</td>
<td>101</td>
</tr>
<tr>
<td>Total MMS in DWPI</td>
<td>L38</td>
<td>344</td>
<td>710</td>
<td>307</td>
<td>487</td>
<td>263</td>
</tr>
<tr>
<td>MMS Derwent not MMS INPI</td>
<td>L39</td>
<td>195</td>
<td>229</td>
<td>161</td>
<td>177</td>
<td>261</td>
</tr>
<tr>
<td>MMS INPI not MMS Derwent</td>
<td>L40</td>
<td>103</td>
<td>324</td>
<td>103</td>
<td>91</td>
<td>2</td>
</tr>
<tr>
<td>CAS not MMS</td>
<td>L41</td>
<td>335</td>
<td>299</td>
<td>291</td>
<td>100</td>
<td>109</td>
</tr>
<tr>
<td><strong>MMS not CAS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMS Derwent not CAS</td>
<td>L42</td>
<td>157</td>
<td>166</td>
<td>172</td>
<td>62</td>
<td>127</td>
</tr>
<tr>
<td>MMS INPI not CAS</td>
<td>L43</td>
<td>113</td>
<td>58</td>
<td>120</td>
<td>50</td>
<td>125</td>
</tr>
<tr>
<td>Pre-1980 CAS not MMS</td>
<td>L44</td>
<td>54</td>
<td>132</td>
<td>70</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>1980-date CAS not MMS</td>
<td>L45</td>
<td>153</td>
<td>42</td>
<td>13</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Fragmentation Code not CAS</td>
<td>L46</td>
<td>182</td>
<td>257</td>
<td>159</td>
<td>100</td>
<td>109</td>
</tr>
<tr>
<td>Frag Code (63-86) not CAS</td>
<td>L47</td>
<td>334</td>
<td>176</td>
<td>1102</td>
<td>1535</td>
<td>625</td>
</tr>
<tr>
<td>CAS not Frag Code (63-86)</td>
<td>L49</td>
<td>87</td>
<td>65</td>
<td>21</td>
<td>0</td>
<td>9</td>
</tr>
</tbody>
</table>

1 All searches were all conducted on 02/10/01, except *Diazepam* conducted on 28/09/01.
2 The general CAS Registry Number for Nonanoyloxybensenesulfonate (NOBS) 101482-85-3, where the 1,2-, 1,3- or 1,4 isomer is not specified, is not retrieved in this structure search. In order to yield the CASLINK results given here the general RN was added manually to the search. This is also a Closed Substructure Search (CSS). See page 67.
3 In these two cases the Fragmentation Code substructure search clearly retrieves a large number of incorrect answers (as well as the correct ones). Thus a total featuring just MMS and CAS is a useful reference point.
Search Results

Reorganising the results, including adding the missing PN (L34) to all the relevant WPIDS totals, an abbreviated Excel Sheet is produced (below). This is then used to create Excel Charts.

<table>
<thead>
<tr>
<th>Result description</th>
<th>Diazepam</th>
<th>Carbapenem</th>
<th>Tyrosine</th>
<th>Sartan</th>
<th>NOBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAS Total</td>
<td>556</td>
<td>854</td>
<td>439</td>
<td>529</td>
<td>246</td>
</tr>
<tr>
<td>CAS not Derwent/INPI</td>
<td>239</td>
<td>193</td>
<td>245</td>
<td>103</td>
<td>102</td>
</tr>
<tr>
<td>CAS not Frag Code (63-86)</td>
<td>87</td>
<td>65</td>
<td>21</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>CAS not MMS</td>
<td>369</td>
<td>310</td>
<td>304</td>
<td>104</td>
<td>110</td>
</tr>
<tr>
<td>CAS Pre-1980 not MMS</td>
<td>154</td>
<td>42</td>
<td>13</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CAS 1980-date not MMS</td>
<td>216</td>
<td>268</td>
<td>172</td>
<td>104</td>
<td>110</td>
</tr>
<tr>
<td>MMS Total</td>
<td>345</td>
<td>710</td>
<td>307</td>
<td>487</td>
<td>263</td>
</tr>
<tr>
<td>MMS not CAS</td>
<td>158</td>
<td>166</td>
<td>72</td>
<td>62</td>
<td>127</td>
</tr>
<tr>
<td>MMS Derwent</td>
<td>241</td>
<td>386</td>
<td>204</td>
<td>396</td>
<td>261</td>
</tr>
<tr>
<td>MMS Derwent not CAS</td>
<td>113</td>
<td>58</td>
<td>120</td>
<td>50</td>
<td>125</td>
</tr>
<tr>
<td>MMS Derwent not MMS INPI</td>
<td>195</td>
<td>229</td>
<td>161</td>
<td>177</td>
<td>261</td>
</tr>
<tr>
<td>MMS INPI</td>
<td>150</td>
<td>481</td>
<td>146</td>
<td>310</td>
<td>2</td>
</tr>
<tr>
<td>MMS INPI not CAS</td>
<td>55</td>
<td>132</td>
<td>70</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>MMS INPI not MMS Derwent</td>
<td>104</td>
<td>324</td>
<td>103</td>
<td>91</td>
<td>2</td>
</tr>
<tr>
<td>Total Frag Code (1963-1986)</td>
<td>487</td>
<td>429</td>
<td>1198</td>
<td>1540</td>
<td>693</td>
</tr>
<tr>
<td>Frag Code (63-86) not CAS</td>
<td>334</td>
<td>176</td>
<td>1102</td>
<td>1535</td>
<td>625</td>
</tr>
<tr>
<td>Derwent/INPI Total</td>
<td>792</td>
<td>981</td>
<td>1460</td>
<td>2023</td>
<td>894</td>
</tr>
<tr>
<td>Derwent/INPI not CAS</td>
<td>476</td>
<td>320</td>
<td>1266</td>
<td>1597</td>
<td>750</td>
</tr>
<tr>
<td>Total Derwent/INPI/CAS</td>
<td>1032</td>
<td>1174</td>
<td>1705</td>
<td>2126</td>
<td>996</td>
</tr>
<tr>
<td>Total Derwent/INPI/CAS FSORT</td>
<td>909</td>
<td>1024</td>
<td>1499</td>
<td>1955</td>
<td>958</td>
</tr>
</tbody>
</table>

This table of results is converted into the chart below.
Additional search steps were carried out for Tyrosine Kinase Inhibitor and Sartan queries (see page 70). This was to combine, de-duplicate and FSORT CAS answers with MMS answers. The additional data is highlighted with in a circle (below).

For completeness the remaining three sets of search results are separated from Tyrosine Kinase Inhibitor and Sartan (above) and charted below.
A note on Fragmentation Codes

It would appear from the results tables that the Derwent Chemical Fragmentation code produces a large volume of incorrect answers. This is especially the case when the query does not contain a suitable Derwent Ring Index Number (RIN) to help with precision.

Unlike STN and MMS structure searching, Chemical Fragmentation Codes are not a graphic based system of unambiguously connected nodes and bonds. There is an inherent “looseness” in retrieval which, for certain types of broad chemical structure search question, can be a useful feature. Despite the obvious noise in the answer sets (for certain queries), it is important to understand that the Derwent Chemical Fragmentation Code will retrieve highly relevant records which are unretrievable via straightforward CAS Registry Number searching. This is due to the nature of Markush patent claims, which are only covered fully by Derwent Chemical Fragmentation Code indexing prior to 1987.

For example, in the NOBS search (page 67) the following relevant patent was retrieved via Fragmentation Code, but was not found via Registry Number. This is not an indexing error by CAS, it is simply because NOBS was not specifically disclosed in the patent so the number is not assigned by CA. NOBS appears only as a “prophetic” substance defined within a Markush structure.

AN 1983-820415 [46] WPIDS
DNC C1983-112603

TI Bleaching compns. effective over wide temp. ranges - contg. per oxygen
bleaching cpd. and specific bleach activator.
DC E25 E12 F09
IN SPADINI, G L; YAT-MING CHUNG, S; CHUNG, S Y
PA (PROC) PROCTER & GAMBLE CO
CYC 19
PI,B US 4412934 A 19831101 (198346)* 11p
AB US 4412934 A UPAB: 19980617

Bleaching compn. contains: (a) a peroxygen bleaching cpd. (I) (to give H2O2 in aq. soln.); and
(b) a bleach activator of formula RCO-L (II) (R is 5-18C alkyl with longest linear alkyl chain
from CO contg. 6-10C inclusions; and L is a leaving gp. the conjugate acid of which has pKa 6-
13.) The mol. ratio H2O2 yielded by (I):(II) is above ca 1.5.

Compns. have mol. ratio H2O2 yielded by (I):(II) pref. at least 2, and contain 0.1-95%
pref. 1-60% by wt. (I) (esp. 1-20% when compns. are also detergent compns.); 0.1-60% pref. 0.5-
40% (II) (esp. 0.5-20% when detergent compns.). Opt. compns. may also contain 0-50% pref. 1-30%
esp. 10-25% of a detergent surfactant (US 3664961 and 3919678 etc.), 0-70% pref. 10-60% esp. 20-
60% degreency builder, Suds suppressing agents, etc. Prepd. (I) are Na carbonate peroxyhydrate,
Na pyrophosphate peroxyhydrate urea peroxyhydrate, Na2O2 etc., esp. Na perborate.4H2O, partic.
Na perborate.H2O.

The compns. are extremely effective in surface bleaching of textiles, and are esp. useful
in removing 'dingy soils' from textiles. The compns. are effective over a wide temp. range.

Text from US 4412934 (Claim 13):

13. The composition of claim 12 wherein R is a linear alkyl chain
containing from about 5 to about 9 carbon atoms.
Coverage summary

This is a simple graphical representation of the comparative backfile coverage of each of the services provided by Derwent, INPI and Chemical Abstracts Service. More detailed tables are given on pages 15 and 16.

Conclusions

The five search examples show that whilst both CAS and Derwent/INPI bring common results in a given chemical structure based search strategy, there will almost always be unique results from each service. This is largely due to country coverage, timeliness and backfile coverage differences, but a certain amount of the hard-to-assess “human-factor” is probably present within the intellectual indexing. It is certainly clear from these results that information professionals interested in complete Markush search retrieval should make use of all the tools available to them, where ever they may be located.
Connecting to MMS via STN Express

Setup STN Express to connect to MMS

Set up a new connection option in the “Setup” part of STN Express.

Click on “New” to create a new Setup for connecting to MMS

Give your new Set up a name, e.g. “MMS”; choose “Questel” as the Host Name; enter in your Login ID (optional) and Password (optional)

---

1 The standard logon method is via the Internet (Winsock). For connection using other routes please contact your STN or Questel.Orbit representative for expert advice.
Click on Host settings; type in “M” in the Choice of Service box; click on OK

Save the new Setup

Set MMS online options to work with STN Express

Logon via STN Express using the new MMS Setup (described above)
The Complete Markush Structure Search: Mission Impossible?

Welcome to Markush DARC / Bienvenue sur Markush DARC

**** Last selected data base : MMS ****
**** Base MMS - 21/09/01 ****

972630 compounds - Last CN : 9999-NSD03

Selected Segment(s):
BACKF
MPARM
WPIM
FRONTF

-ST- (BA, CN, QT, QG, RF, PE, AA, SE, BL, BI, GD, INFO) ?

At the main -ST- prompt enter OP (options):

-ST- (BA, CN, QT, QG, RF, PE, AA, SE, BL, BI, GD, INFO) ? OP

Answer ALL the questions as prompted:

1. Specify your language preference and if you wish to see a cost estimation:

   Français (1)/English (2) ? 2
   Display cost of the session : Y/N ? Y

2. Choose your Terminal/Emulation type: enter “S” for STN Express

   ** TERMINAL INFORMATION **
   0. Non-graphics terminal  ---> 0
   1. Tektronix 4010, 4014  ---> 1
   2. Tektronix 4020  ---> 2
   3. Tektronix 41XX, 42XX  ---> 3
   4. VT 24X, 330, 34X (Tektron Mode)  ---> 4
   5. VT 640 (or PCL/IV EMULATOR)  ---> 5
   6. HENLETT-PACKARD 2647A, 2648  ---> 6
   ** EMULATORS **
   7. IMAGINATION  ---> 7
   8. VERSATERM-PRO (Macintosh)  ---> 8
   9. EMUTEK 5+, 7+  ---> 9
   10. EMUTEK LEVEL 2  ---> 10
   11. INFOLOG  ---> 11
   12. ZSTEM 240  ---> 12
   S. STN EXPRESS  ---> S

   Please enter your terminal code ---> S
3. Select the transmission rate and format for MMS bibliographic images

<table>
<thead>
<tr>
<th>TRANSMISSION RATE (BAUD)</th>
<th>NO IMAGE (0)</th>
<th>GIF FORMAT (G)</th>
<th>TIFF FORMAT (T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>300</td>
<td>T</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Helpful HINT** Set the transmission rate to **300**. In reality this is not the transmission rate, but answering the question in this way is a necessary “trick” to allow STN Express to communicate correctly with MMS.