
The Complete Markush Structure Search: Mission Impossible?

PIUG North East Workshop

October 16th 2001

© Fachinformationszentrum Karlsruhe, October 2001

Robert Austin
Fachinformationszentrum (FIZ) Karlsruhe
Hermann von Helmholtz Platz 1
76344 Eggenstein-Leopoldshafen
Germany

Tel: +49 7247 808 555

Fax: +49 7247 808 131

Email: helpdesk@fiz-karlsruhe.de

Web: www.fiz-karlsruhe.de

Contents

ABSTRACT	7
ACKNOWLEDGEMENTS	9
INTRODUCTION	11
WHERE DID THE TERM MARKUSH COME FROM?	11
WHAT “MARKUSH” MEANS TODAY.....	11
THE AIM OF THIS PAPER	12
THE DATABASES	13
REGISTRY / CAPLUS	13
MARPAT	13
MERGED MARKUSH SERVICE.....	13
DERWENT WPI: CHEMICAL FRAGMENTATION CODES	14
COVERAGE SUMMARY	14
CONTENT SUMMARY	15
COUNTRY COVERAGE COMPARISON	16
STRUCTURE QUERY PREPARATION SOFTWARE	17
STN EXPRESS 6.0	17
DERWENT MARKUSH TOPFRAG 3.1	17
FOR FURTHER INFORMATION	17
SEARCH EXAMPLE	18
OBJECTIVES	18
MMS.....	19
PHARMSEARCH.....	28
TURN MMS RESULTS INTO STN SEARCH SCRIPTS	31
WPINDEX.....	36
CASLINK.....	43
IDENTIFYING UNIQUE RESULTS	52
SEARCH EXAMPLE CONTINUED.....	52
EXAMPLES OF UNIQUE ANSWERS.....	56
MORE SEARCH EXAMPLES	59
OBJECTIVES	59
CARBAPENEM	60
TYROSINE KINASE INHIBITOR	62
SARTAN.....	64
NONANOXYLOXYBENZENESULFONATE (NOBS)	67
SEARCH RESULTS	70
A NOTE ON FRAGMENTATION CODES	71
COVERAGE SUMMARY	74
CONCLUSIONS	74
CONNECTING TO MMS VIA STN EXPRESS	75
SETUP STN EXPRESS TO CONNECT TO MMS	75
SET MMS ONLINE OPTIONS TO WORK WITH STN EXPRESS	76

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 to. 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



Millenium Information Services

215 12th Street, SE
Washington, DC 20005-1427
United States
Tel: +1 202 543 5120
Fax: +1 800 949 5120
www.milleniuminfo.com



Questel.Orbit, Inc

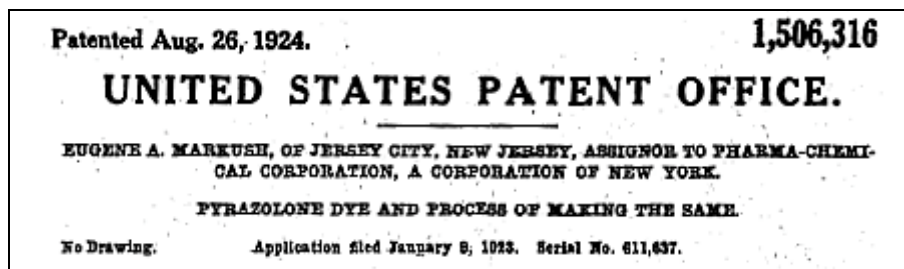
8000 Westpark Drive
McLean, VA 22102 - U.S.A.
Tel: +1 703 442 0900 / +1 800 456 7248
Facsimile: +1 703 893 5632
Email: help@questel.orbit.com
www.questel.orbit.com



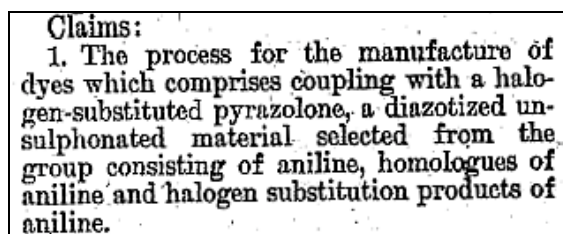
Introduction

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 pyrazolone dyes that could be used for wool or silk:



Claim 1 of this application read:



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 systems¹ 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*®

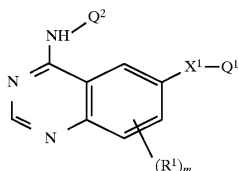
¹ A useful overview of the history of the development of Markush structure indexing can be found at: www.lib.uchicago.edu/cinf/218nm/paper42/index.html (Edlyn Simmons, ACS National Meeting, August 1999).

A typical example of a modern Markush patent claim

US 5866572 (Claim 1):

What we claim is:

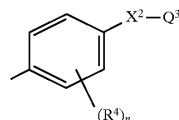
1. A quinazoline derivative of the formula I



wherein X¹ is a direct link;

wherein Q¹ is a 5-membered heteroaryl moiety containing one heteroatom selected from oxygen and sulphur, which heterocyclic moiety is a single ring or is fused to a benzo ring, and Q¹ optionally bears up to 3 substituents selected from halogeno, hydroxy, amino, trifluoromethoxy, trifluoromethyl, cyano, nitro, carboxy, carbamoyl, (1-4C) alkoxy, (1-4C) alkyl, (1-4C) alkoxy, (2-4C) alkenyloxy, (2-4C) alkynyl, (1-3C) alkenedioxy, (1-4C) alkylamino, di-[(1-4C)alkyl]amino, pyrrolidin-1-yl, piperidino, morpholino, piperazin-1-yl, 4-(1-4C) alkylpiperazin-1-yl, (2-4C) alkanoylamino, N-(1-4C) alkylcarbamoyl, N,N-di-[(1-4C)alkyl]carbamoyl, amino-(1-4C)alkyl, (1-4C)alkylamino-(1-4C)alkyl, di-[(1-4C)alkyl]amino-(1-4C)alkyl, pyrrolidin-1-yl-(1-4C)alkyl, piperidino-(1-4C)alkyl, morpholino-(1-4C)alkyl, piperazin-1-yl-(1-4C)alkyl, 4-(1-4C)alkylpiperazin-1-yl-(1-4C)alkyl, halogeno-(2-4C)alkoxy, hydroxy-(2-4C)alkoxy, (1-4C)alkoxy-(2-4C)alkoxy, amino-(2-4C)alkoxy, (1-4C)alkylamino-(2-4C)alkoxy, di-[(1-4C)alkyl]amino-(2-4C)alkoxy, pyrrolidin-1-yl-(2-4C)alkoxy, piperidino-(2-4C)alkoxy, morpholino-(2-4C)alkoxy, piperazin-1-yl-(2-4C)alkoxy, 4-(1-4C)alkylpiperazin-1-yl-(2-4C)alkoxy, (1-4C)alkylthio-(2-4C)alkoxy, (1-4C)alkylsulphonyl-(2-4C)alkoxy, (1-4C)alkylsulphonyl-(2-4C)alkoxy, halogeno-(2-4C)alkylamino, hydroxy-(2-4C)alkylamino, (1-4C)alkoxy-(2-4C)alkylamino, amino-(2-4C)alkylamino, (1-4C)alkylamino-(2-4C)alkylamino, di-[(1-4C)alkyl]amino-(2-4C)alkylamino, pyrrolidin-1-yl-(2-4C)alkylamino, piperidino-(2-4C)alkylamino, morpholino-(2-4C)alkylamino, piperazin-1-yl-(2-4C)alkylamino, 4-(1-4C)alkylpiperazin-1-yl-(2-4C)alkylamino, N-(1-4C)alkyl-halogeno-(2-4C)alkylamino, N-(1-4C)alkyl-hydroxy-(2-4C)alkylamino, N-(1-4C)alkyl-(1-4C)alkoxy-(2-4C)alkylamino, halogeno-(2-4C)alkanoylamino, hydroxy-(2-4C)alkanoylamino, (1-4C)alkoxy-(2-4C)alkanoylamino, (3-4C)alkenoylamino, (3-4C)alkynoylamino, amino-(2-4C)alkanoylamino, (1-4C)alkylamino-(2-4C)alkanoylamino, di-[(1-4C)alkyl]amino-

(2-4C)alkanoylamino, pyrrolidin-1-yl-(2-4C)alkanoylamino, piperidino-(2-4C)alkanoylamino, morpholino-(2-4C)alkanoylamino, piperazin-1-yl-(2-4C)alkanoylamino and 4-(1-4C)alkylpiperazin-1-yl-(2-4C)alkanoylamino, and wherein any of the above-mentioned substituents comprising a CH₂ (methylene) group which is not attached to a halogeno, SO or SO₂ group or to a N, O or S atom optionally bears on said CH₂ group a substituent selected from hydroxy, amino, (1-4C)alkoxy, (1-4C)alkylamino and di-[(1-4C)alkyl]amino; wherein m is 1 or 2 and each R¹ is independently hydrogen, halogeno, trifluoromethyl, hydroxy, amino, nitro, cyano, carboxy, carbamoyl, (1-4C)alkoxy, (1-4C)alkyl, (1-4C)alkoxy, (1-4C)alkylamino, di-[(1-4C)alkyl]amino, (2-4C)alkanoylamino, N-(1-4C)alkylcarbamoyl or N,N-di-[(1-4C)alkyl]carbamoyl; and wherein Q² is phenyl optionally bearing up to 3 substituents selected from halogeno, trifluoromethyl, cyano, hydroxy, amino, nitro, carboxy, carbamoyl, (1-4C)alkoxy, (1-4C)alkyl, (1-4C)alkoxy, (1-4C)alkylamino, di-[(1-4C)alkyl]amino, (2-4C)alkanoylamino, N-(1-4C)alkylcarbamoyl and N,N-di-(1-4C)alkylcarbamoyl, or Q² is a group of the formula II



wherein X² is a group of the formula CO, C(R³)₂, CH(OR³), C(R³)₂-C(R³)₂, C(R³)=C(R³), C≡C, CH(CN), O, S, SO, SO₂, N(R³), CON(R³), SO₂N(R³), N(R³)CO, N(R³)SO₂, OC(R³)₂, SC(R³)₂, C(R³)₂O or C(R³)₂S wherein each R³ is independently hydrogen or (1-4C)alkyl, Q³ is phenyl or naphthyl or a 5- or 6-membered heteroaryl moiety containing up to 3 heteroatoms selected from oxygen, nitrogen and sulphur, which heteroaryl moiety is a single ring or is fused to a benzo ring, and wherein said phenyl or naphthyl group or heteroaryl moiety optionally bears up to 3 substituents selected from halogeno, trifluoromethyl, cyano, hydroxy, amino, nitro, carboxy, carbamoyl, (1-4C)alkoxy, (1-4C)alkyl, (1-4C)alkoxy, (1-4C)alkylamino, di-[(1-4C)alkyl]amino, (2-4C)alkanoylamino, N-(1-4C)alkylcarbamoyl and N,N-di-[(1-4C)alkyl]carbamoyl, n is 1, 2 or 3 and each R⁴ is independently hydrogen, halogeno, trifluoromethyl, cyano, hydroxy, amino, nitro, (1-4C)alkyl, (1-4C)alkoxy, (1-4C)alkylamino, di-[(1-4C)alkyl]amino or (2-4C)alkanoylamino; or a pharmaceutically-acceptable salt thereof.

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*.

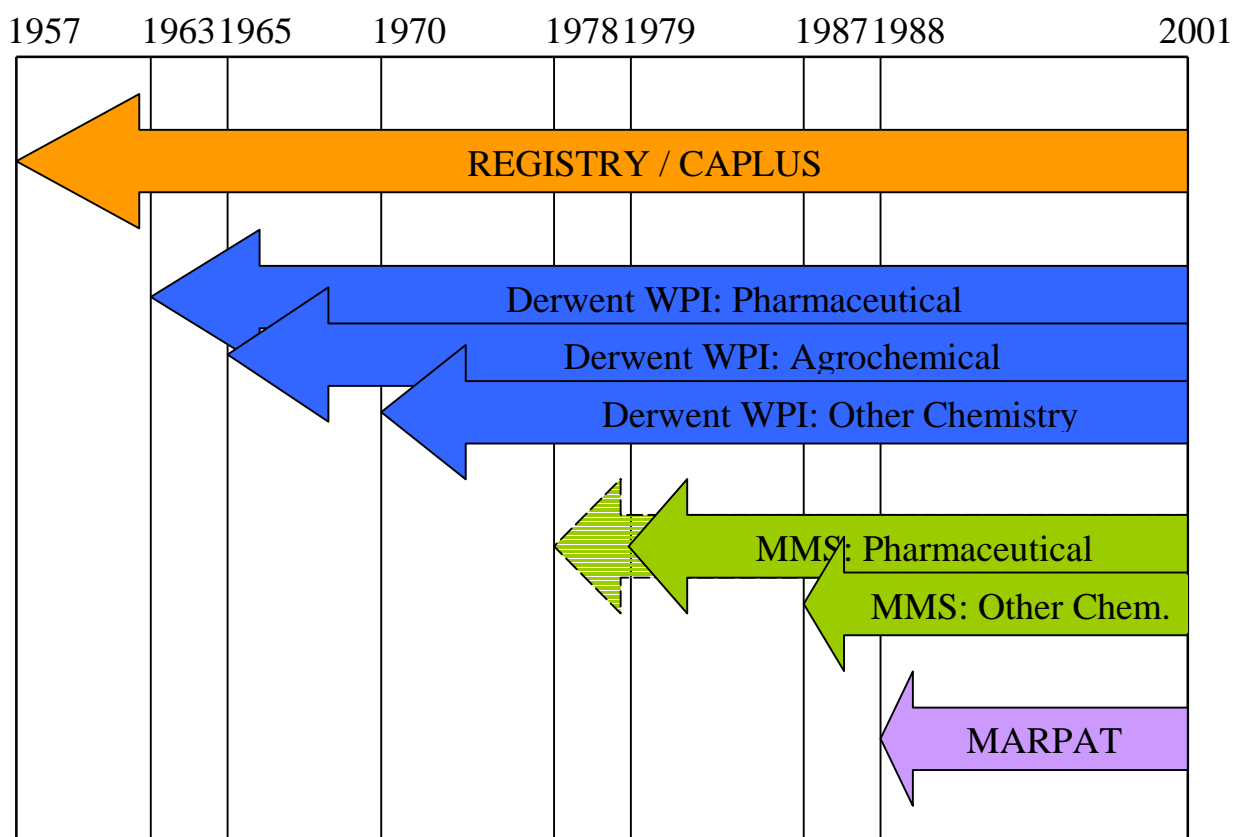
¹ Except Russian Patents, for which coverage is included from January 10th 2000.

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

Feature	REGISTRY	MARPAT	MMS	Derwent WPI Fragment Codes
Source of Information	33 Patent Authorities; Journals, Conferences	33 Patent Authorities (See page 16)	19 Patent Authorities (See page 16)	19 Patent Authorities (See page 16)
Time period	1957-	1988-	1979- (pharma) 1987- (other chem)	1963- (pharma) 1965- (agrochem) 1970- (other chem)
Specific Substances	33 million ¹	N/A	202,000	N/A
Markush Compounds	N/A	457,000	774,000	N/A
Patent records	2.5 million (in CA)	161,000	537,000 (in WPI) 104,000 (in Pharm)	1.1 million ²
Compounds Covered	Specific; all classes	Markush; all classes except: alloys, metal oxides, inorganic salts, intermetallic, polymers	Specific and markush; all classes except: alloys, non-pharma /agrochem polymers	Specific and markush; all classes except: alloys, non-pharma /agrochem polymers, Steriods (from 1999)
Access points	STN structure Searching; graphic upload; text search	STN structure Searching; graphic upload; text search	DARC Structure Searching	Code-based structure Searching; full bib. search; MC, IPC
Bibliographic information	Displayable; Searchable in CAPLUS	Displayable; Searchable in CAPLUS	Displayable (INPI only); search & display in DWPI or Pharmsearch	Searchable & displayable
Work in progress	Yes	Yes	No	No
Sample search	Yes	Yes	No	No
Access	Open	Open	Open	Subscription only
SDI (charge)	Yes (\$25)	Yes (\$25)	No	Yes (\$16)
Structure Search charge	\$133	\$94	\$58	FREE
Availability	STN	STN	Questel.Orbit	STN, Questel.Orbit, Dialog

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

Authority		REGISTRY / CAPLUS ¹	MARPAT	MMS ²	Derwent WPI Frag. Code ³	Derwent WPI ⁴
Argentina	AR	-	-	-	-	1975 (only)
Australia	AU	1957	1988	1987	1963-1969; 1983	1963-1969; 1983
Austria	AT	1957	1988	1993	1993	1975
Belgium	BE	1957	1988	1987	1963	1963
Brazil	BR	1976	1988	1987-1998	1987-1998	1976
Canada	CA	1957	1988	1987	1963	1963
China	CN	1986	1988	-	-	1985
Czech Republic	CZ	1993 *	1988 *	-	-	1994
Czechoslovakia	CS	1957-1992	1988-1992	-	-	1975-1992
Denmark	DK	1957 *	1988 *	-	-	1974
European	EP	1979	1988	1979	1978	1978
Finland	FI	1960 *	1988 *	-	-	1974
France	FR	1957	1988	1979	1963	1963
Germany (East)	DD	1959-1994	1988-1994	1987-1994	1963-1994	1963-1994
Germany	DE	1959	1988	1987	1963	1963
Hungary	HU	1957-1958; 1963	1988	-	-	1975
India	IN	1957	1988	-	-	-
Int. Tech. Disclosure	TP	-	-	1987-1993	1984-1993	1984-1993
Ireland	IE	-	-	1995	1963-1969; 1995	1963-1969; 1995
Israel	IL	1958	1988	1987	1983	1975
Italy	IT	1957	1988	-	-	1977
Japan	JP	1957	1988	1987	1963	1963
Korea (South)	KR	1995-1997	1995-1997	-	-	1986
Latvia	LV	1996 *	1996 *	-	-	-
Lithuania	LT	1996 *	1996 *	-	-	-
Luxembourg	LU	-	-	-	-	1984
Mexico	MX	-	-	-	-	1997
Netherlands	NL	1957	1988	1987	1963	1963
New Zealand	NZ	-	-	1992	1992	1992
Norway	NO	1957 *	1988	-	-	1974
PCT (World)	WO	1978	1988	1979	1978	1978
Philippines	PH	-	-	-	-	1992
Poland	PL	1960 *	1988 *	-	-	-
Portugal	PT	-	-	1987 (only)	1983-1987	1974
Research Disclosure	RD	1974	1988	1987-2001	1978-2001	1978-2001
Romania	RO	1967	1988	1987-1997	1983-1997	1975
Russian Federation	RU	1993	2000	1994-2000	1994-2000	1994
Singapore	SG	-	-	1996	1996	1996
Slovakia	SK	1996 *	1996 *	-	-	1993
South Africa	ZA	1968	1988	1987	1963	1963
Soviet Union	SU	1958-1993	1988-1993	1987-1993	1963-1993	1963-1993
Spain	ES	1957 *	1988 *	1987-1992	1987-1992	1983
Sweden	SE	1957	1988	1987	1974	1974
Switzerland	CH	1957	1988	1987	1963	1963
Taiwan	TW	-	-	1993-1998	1993-1998	1993
United Kingdom	GB	1957	1988	1987	1963	1963
United States	US	1957	1988	1979	1963	1963

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”.

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 or 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

OR

www.stn-international.de

Chemical Abstracts Service

www.cas.org

FIZ Karlsruhe

www.fiz-karlsruhe.de

Derwent Information Ltd

www.derwent.com

Search Example

Objectives

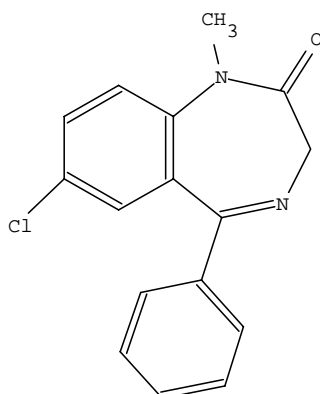
- Conduct a *Merged Markush Service* (MMS) structure search via *STN Express*
- Extract relevant search results from MMS, and transfer them to the *Derwent World Patents Index*® on STN (files WPINDEX/WPIDS/WPIX).
- Complement MMS results with answers retrieved using an STN Express Derwent Chemical Fragmentation 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

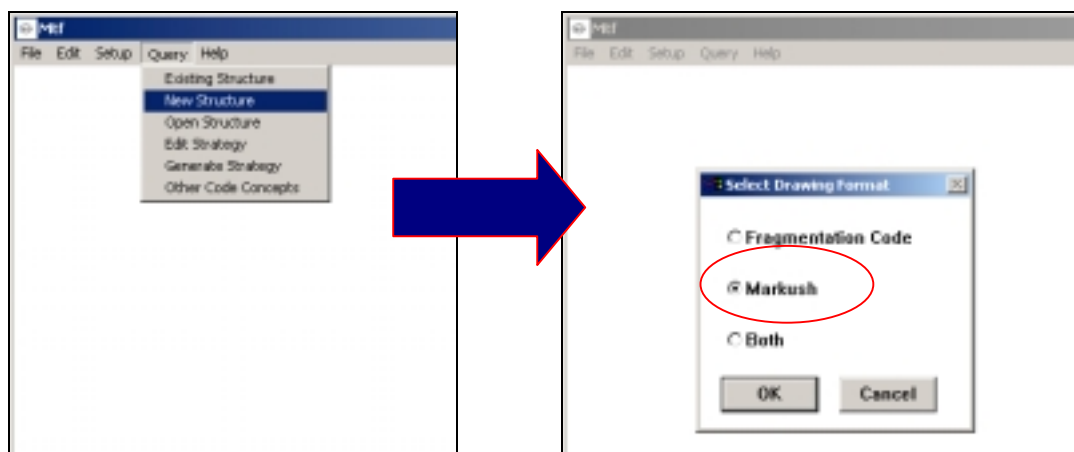
To find patents which specifically describe the drug diazepam

- 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

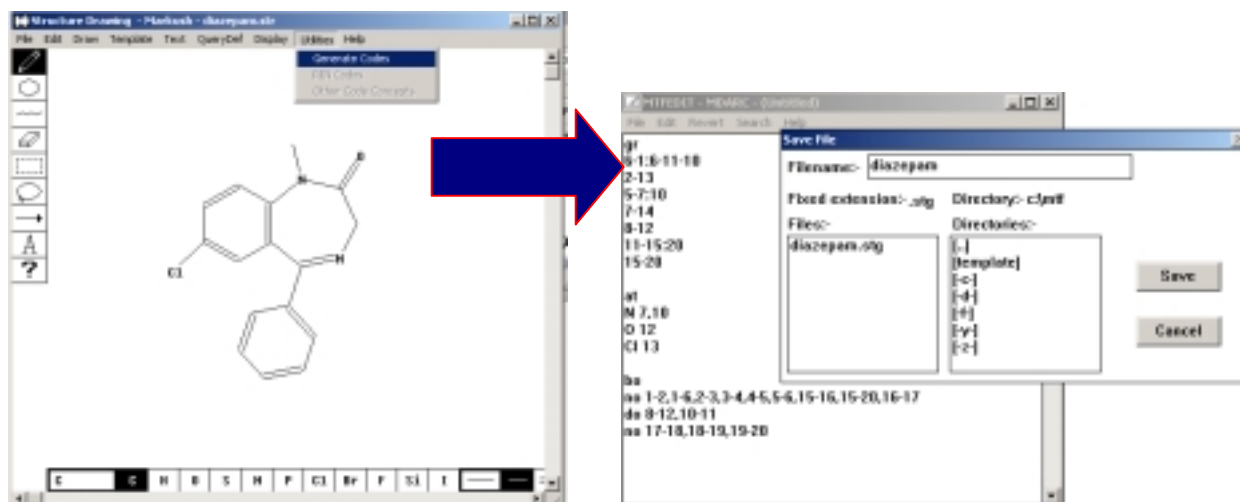
Prepare the query offline using Derwent Markush TOPFRAG (MTF):

1. Select *New Structure* from the *Query* Menu; select *Markush* format

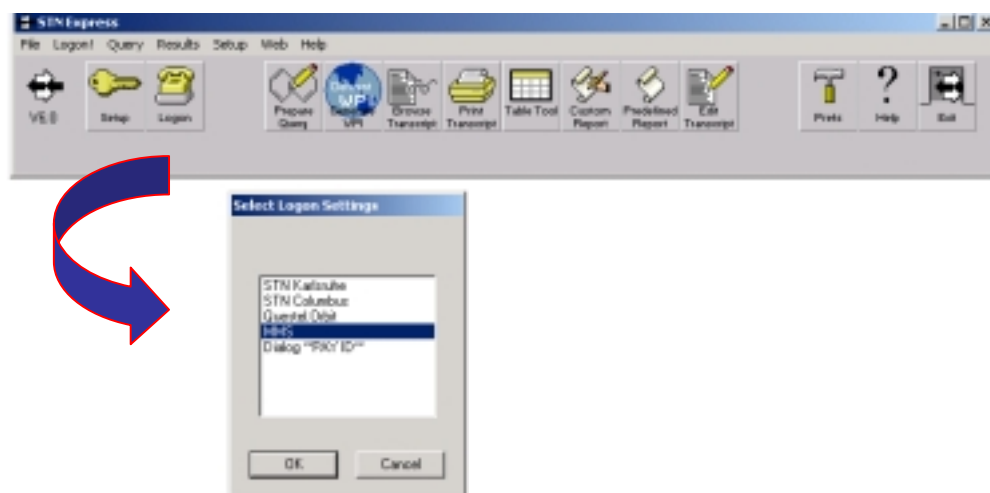


The Complete Markush Structure Search: Mission Impossible?

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 CN : 9999-NSE03

SELECTED SEGMENT(S):

BACKF
MPHARM
WPIM
FRONTF

The Complete Markush Structure Search: Mission Impossible?

Enter QT (Query Text) at the main -ST- prompt:

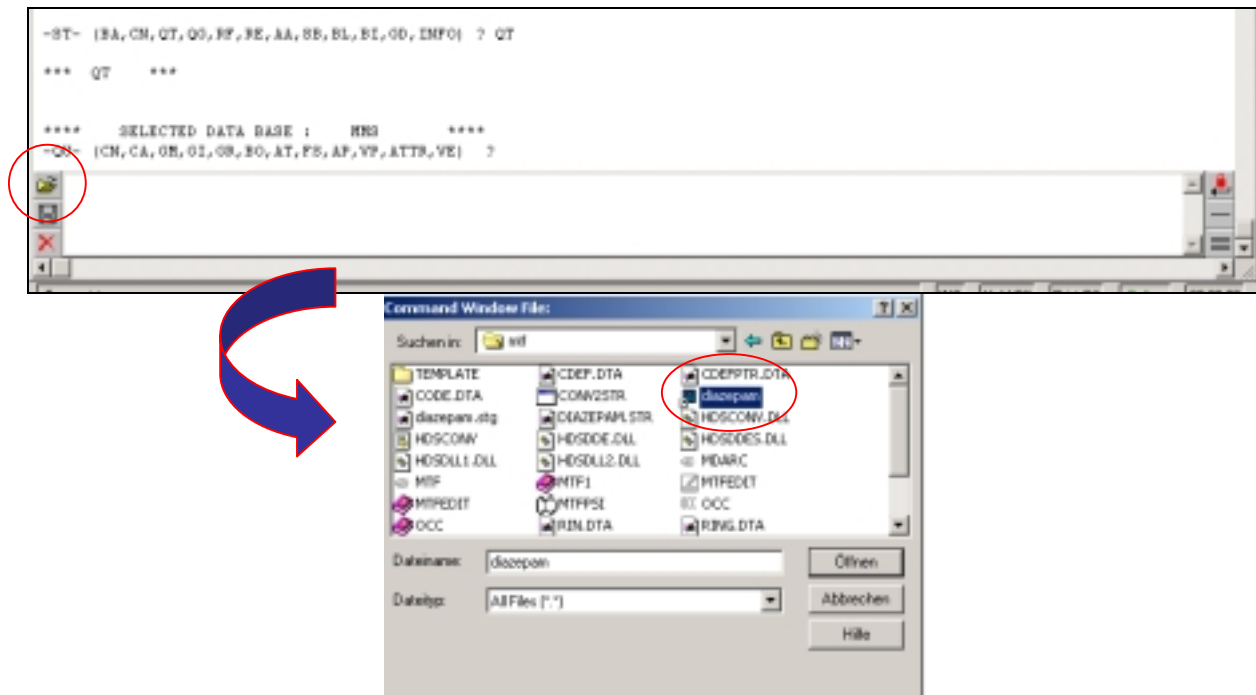
```
-ST- (BA,CN,QT,QG,RF,RE,AA,SB,BL,BI,GD,INFO) ? QT
*** QT ***
**** SELECTED DATA BASE : MMS ****
-QU- (CN,CA,GM,GI,GR,BO,AT,FS,AP,VP,ATTR,VE) ?
```

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*:

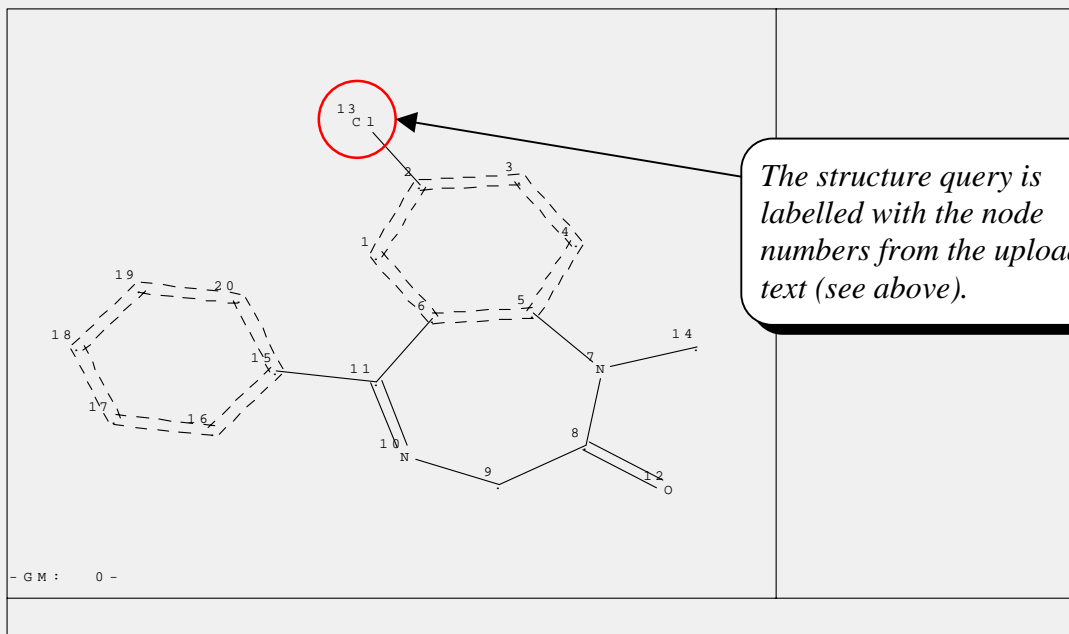


The Complete Markush Structure Search: Mission Impossible?

```
-QU- (CN,CA,GM,GI,GR,BO,AT,FS,AP,VP,ATTR,VE) ? gr
*GRAPH
? 6-1:6-11-10
? 2-13
? 5-7:10
? 7-14
? 8-12
? 11-15:20
? 15-20
?
-QU- (CN,CA,GM,GI,GR,BO,AT,FS,AP,VP,ATTR,VE) ? at
*ATOMS
? N 7,10
? O 12
? Cl 13
?
-QU- (CN,CA,GM,GI,GR,BO,AT,FS,AP,VP,ATTR,VE) ? bo
*BONDS
? no 1-2,1-6,2-3,3-4,4-5,5-6,15-16,15-20,16-17
? do 8-12,10-11
? no 17-18,18-19,19-20
?
```

Check that the structure has uploaded correctly with the VE (verify) command:

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



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

```
? FI
You are currently defining GROUP 0
```

The Complete Markush Structure Search: Mission Impossible?

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
```

Help
ful
HINT

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
```

	NAME	TYPE	NUMBER	ORIGIN	DATE	SPACE	DATABASE
1	DIAZEPAM	QU	1	MARKUSH DARC	22/09/01	1	

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

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

```
*** RE ***
```

```
* SEARCH IN PROGRESS *
```

```
* SEARCH STILL IN PROGRESS *
```

```
* SEARCH STILL IN PROGRESS *
```

```
* SEARCH STILL IN PROGRESS *
```

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

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

```
-ST- (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):

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

** 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:

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

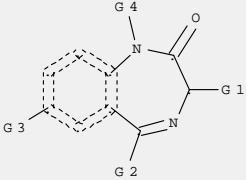
QU / DN,CN,PN / ? CN

NAME (8CHAR.) ? DIAZEPAM

	NAME	TYPE	NUMBER	ORIGIN	DATE	SPACE	DATABASE
1	DIAZEPAM	CN	34	MARKUSH DARC	22/09/01	1	MMS

Review answers graphically with VI FO (Visualise Focus):

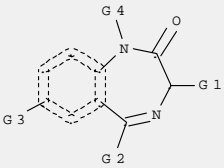
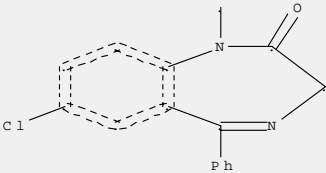
-ST- (BA,CN,QT,QG,RF,RE,AA,SB,BL,BI,GD,INFO) ? VI FO 2

MARKUSH/DARC	2 / 34	CN : 0018-19801	MMS
-GM: 0 / 4 -			
			
SEGMENTS : BEV		G0 (G3, G2, G4)	

VI FO just displays the Markush subgroups (Gx) where the hit atoms are located.

? <ENTER>

The Complete Markush Structure Search: Mission Impossible?

MARKUSH/DARC	2 / 34	CN : 0018-19801	MMS
-FG: 0-	-GM: 4/ 4-		
		<p>H <input type="checkbox"/> C</p>	
G0 (G3 , G2 , G4)			
<p>? 14</p> <div style="border: 2px solid black; border-radius: 15px; padding: 5px; display: inline-block;"> <i>It is possible to select the record to be displayed, e.g. 14 (of 34)</i> </div>			
MARKUSH/DARC	14 / 34	CN : R16158	MMS
-GM: 0/ 0-			
			
<p>SEGMENTS : ABEVYZ1 G0 / DIAZEPAM SALTS GENERAL/ .</p>			

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

? **FI**

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

Display a list of the compound numbers with LI (List) command:

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

```
R2 - AA / MMS : 34 ANSWER(S)
```

? 1-34

```
1 CN = RA3HQQ
2 CN = 0018-19801
3 CN = 9903-HBJ01
4 CN = 9807-26403
5 CN = 9645-39703
6 CN = 9637-00802
7 CN = 9637-00801
8 CN = 9418-44103
9 CN = 9238-14101
10 CN = 9008-31802
11 CN = 9002-12202
12 CN = 8940-09710
13 CN = 8914-02901
14 CN = R16158
15 CN = 8821-12501
16 CN = R01255
17 CN = 96105966-03
18 CN = 95110993-02
19 CN = R70037640-01
20 CN = 66074206-07
21 CN = 86040114-01
22 CN = 88060463-01
23 CN = 90020299-01
24 CN = 61100839-01
25 CN = 89120432-02
26 CN = 90010072-02
27 CN = 89030183-01
28 CN = R86060272-02
29 CN = 79055233-01
30 CN = 80105158-01
31 CN = 81035060-01
32 CN = R81085339-07
33 CN = R81085339-06
34 CN = 85010433-09
```

Derwent WPI compound numbers:

- RXXXXXX Specific Compounds
- XXXX-XXXXX Markush Compounds

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

INPI Pharmsearch compound numbers:

- RXXXXXXXXX-XX Specific Compounds
- XXXXXXXXXXX-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,QI,QG,RF,RE,AA,SB,EL,BI,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):

```
? FI PHARM
Selected file: PHARM
Period covered: weeks 7901-9952 (US), 7902-9952 (EP), 7902-9952 (FR)
9236-9952 (DE and GB), 7902-9952 (PCT). BSM: full collection.
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 MDARC
```

Crossover compound numbers (CN) from MMS using the *MD command:

```
Search statement 1
?*MD DIAZEPAM/CN

Frequency Term
1 R70037640-01/CN
1 R81085339-06/CN
1 R81085339-07/CN
135 R86060272-02/CN
1 61100839-01/CN
1 66074206-07/CN
1 79055233-01/CN
1 80105158-01/CN
1 81035060-01/CN
1 85010433-09/CN
1 86040114-01/CN
1 88060463-01/CN
1 89030183-01/CN
1 89120432-02/CN
1 90010072-02/CN
1 90020299-01/CN
1 95110993-02/CN
1 96105966-03/CN

** SS 1: Results 151
```

INPI Pharmsearch compound numbers:

- RXXXXXXXX-XX Specific Compounds
- XXXXXXXX-XX Markush Compounds

**18 INPI Compound Numbers
Retrieve 151 patent records**

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:

```
Search statement 2
? MEMS SS 1 SET DIAZEPAM STORE /XPN RK 1

Total number of terms extracted: 151
Number of terms now in DIAZEPAM : 151
```

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

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)

#	FREQ	TERM
1	1	DE4207922
2	1	DE4222826
3	1	EP---1924
4	1	EP--11745
5	1	EP--21337
6	1	EP--22771
7	1	EP--25546
8	1	EP--31561
9	1	EP--35132
10	1	EP--69307
	.	.
	.	.
	.	.
145	1	WO9615815
146	1	WO9630393
147	1	WO9637232
148	1	WO9640052
149	1	WO9640053
150	1	WO9719112
151	1	WO9802148

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

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.

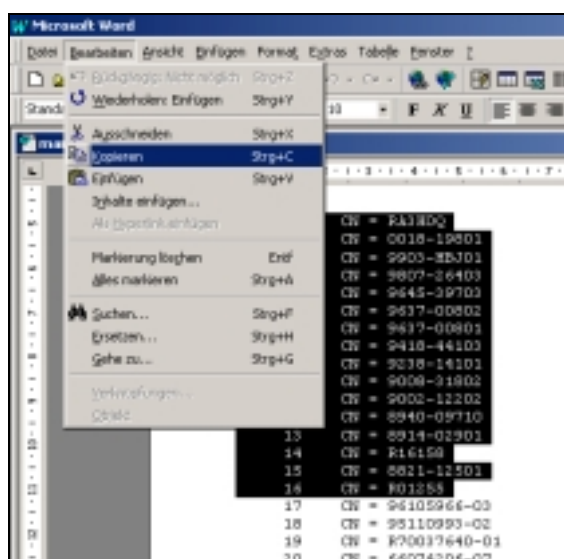
Logoff with the ST command. Questel.Orbit automatically retains your search for two hours. This useful feature means that you are able to return and refine your search, if needed, after searching on STN.

Turn MMS results into STN search scripts

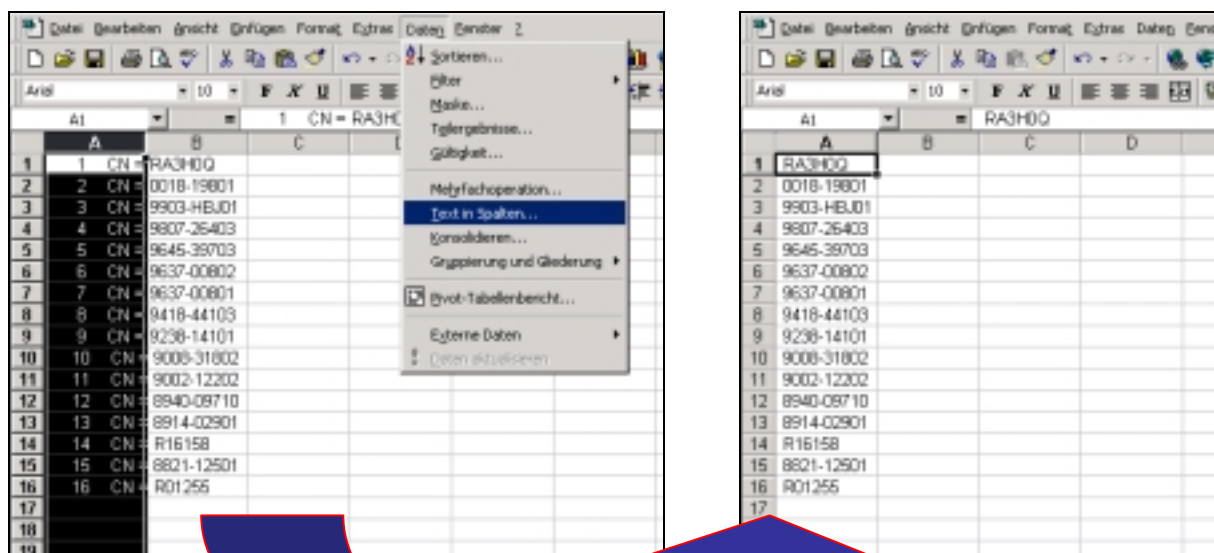
The list of MMS-Derwent WPI compound numbers on page 27, and the list of MMS-Pharmsearch patent numbers on page 30, can be turned into simple search scripts for searching in the WPINDEX file on STN. One method of doing this using straightforward *Microsoft Excel* functions is shown below.

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



The Complete Markush Structure Search: Mission Impossible?

- Use the CONCATENATE function to turn each compound number into a simple search statement: e.g. CONCATENATE("S "; A1;"/DCN").

	A	B	C	D	E
1	RA3HQ	S RA3HQ/DCN			
2	0018-19801	S 0018-19801/DCN			
3	9903-HBJ01	S 9903-HBJ01/DCN			
4	9807-26403	S 9807-26403/DCN			
5	9645-39703	S 9645-39703/DCN			
6	9637-00802	S 9637-00802/DCN			
7	9637-00801	S 9637-00801/DCN			
8	9418-44103	S 9418-44103/DCN			
9	9238-14101	S 9238-14101/DCN			
10	9008-31802	S 9008-31802/DCN			
11	9002-12202	S 9002-12202/DCN			
12	8940-09710	S 8940-09710/DCN			
13	8914-02901	S 8914-02901/DCN			
14	R16158	S R16158/DCN			
15	8821-12501	S 8821-12501/DCN			
16	R01255	S R01255/DCN			
17		S /DCN			
18		S /DCN			
19		S /DCN			

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

```
S RA3HQ/DCN
S 0018-19801/DCN
S 9903-HBJ01/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
```

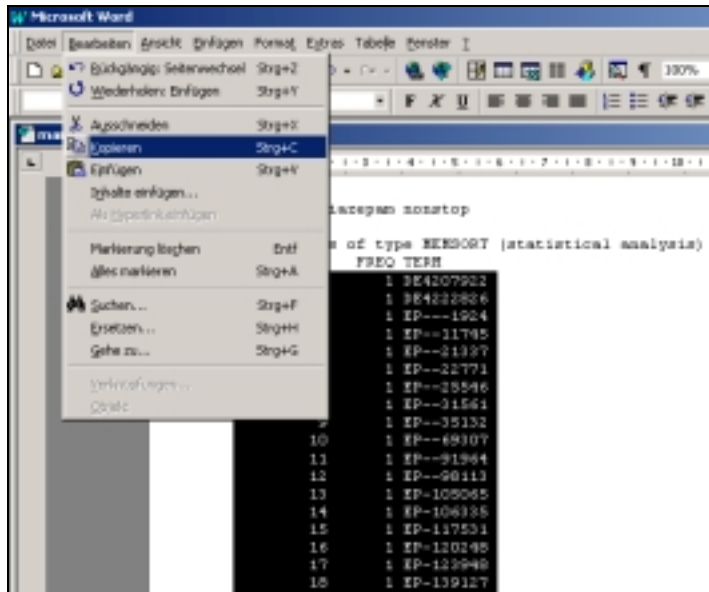
Helpful HINT

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.

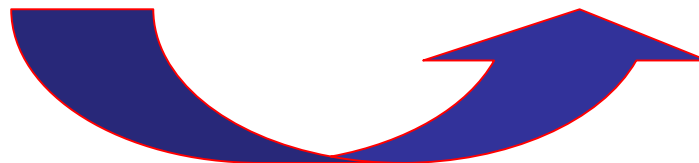
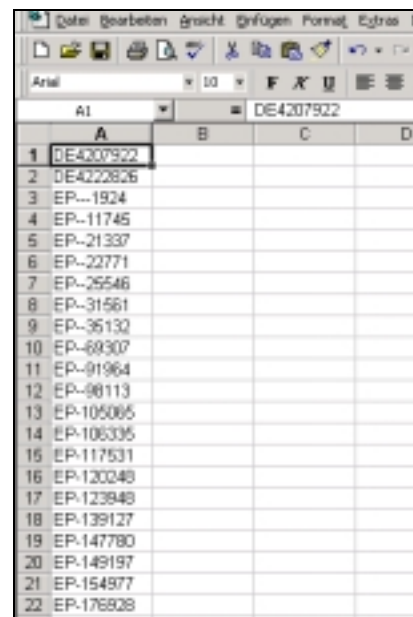
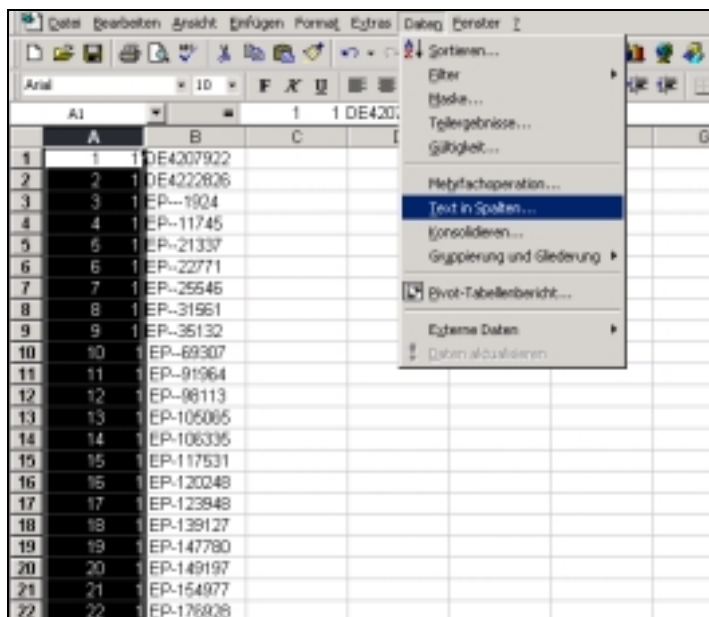
```
S (8914-02901 OR R16158 OR 8821-12501)/DCN
```

MMS-Pharmsearch patent numbers

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

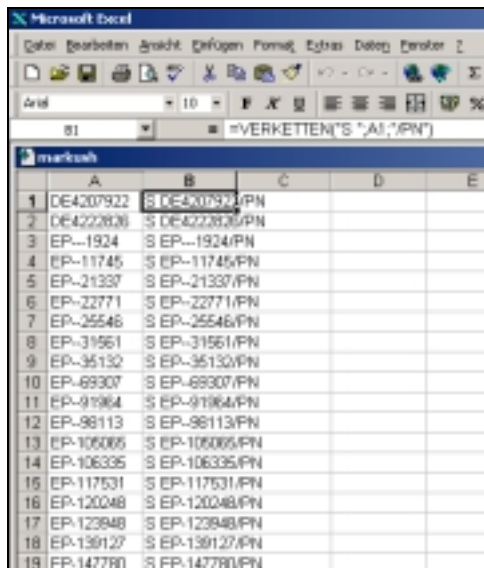


2. Paste the patent numbers into a Excel Spread sheet; use standard “text to columns” feature to produce a simple list of PNs



The Complete Markush Structure Search: Mission Impossible?

3. Use the CONCATENATE function to turn each patent number into a simple search statement: e.g. CONCATENATE("S "; A1;" /PN").



	A	B	C	D	E
1	DE4207922	S DE4207922/PN			
2	DE4222826	S DE4222826/PN			
3	EP--1924	S EP--1924/PN			
4	EP--11745	S EP--11745/PN			
5	EP--21337	S EP--21337/PN			
6	EP--22771	S EP--22771/PN			
7	EP--25546	S EP--25546/PN			
8	EP--31561	S EP--31561/PN			
9	EP--35132	S EP--35132/PN			
10	EP--69307	S EP--69307/PN			
11	EP--91964	S EP--91964/PN			
12	EP--98113	S EP--98113/PN			
13	EP-105065	S EP-105065/PN			
14	EP-106335	S EP-106335/PN			
15	EP-117531	S EP-117531/PN			
16	EP-120248	S EP-120248/PN			
17	EP-123948	S EP-123948/PN			
18	EP-139127	S EP-139127/PN			
19	EP-147780	S EP-147780/PN			

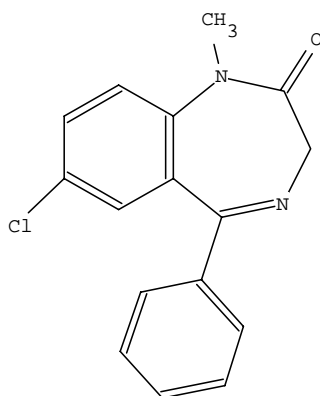
A simple list of patent number search statements is produced:

```
S DE4207922/PN
S DE4222826/PN
S EP---1924/PN
S EP--11745/PN
S EP--21337/PN
S EP--22771/PN
S EP--25546/PN
S EP--31561/PN
S EP--35132/PN
S EP--69307/PN
S EP--91964/PN
S EP--98113/PN
S EP-105065/PN
S EP-106335/PN
S EP-117531/PN
S EP-120248/PN
.
.
.
```

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

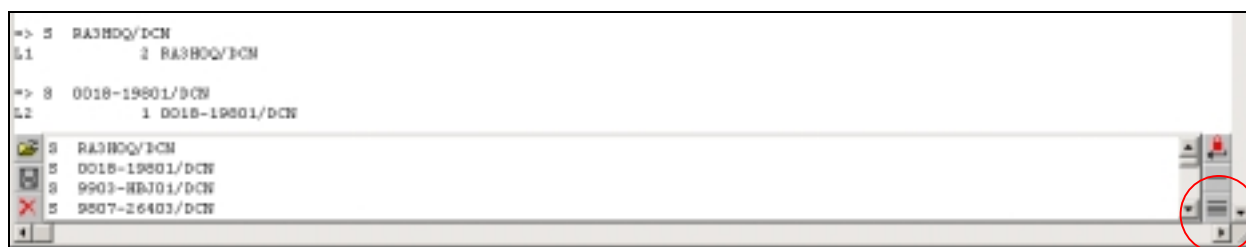
Access Derwent WPI (e.g. subscriber file WPIDS):

=> 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 RA3HQ/DCN
L1      2 RA3HQ/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 Complete Markush Structure Search: Mission Impossible?

```
=> S 9002-12202/DCN  
L11      1 9002-12202/DCN
```

```
=> S 8940-09710/DCN  
L12      0 8940-09710/DCN
```

```
=> S 8914-02901/DCN  
L13      1 8914-02901/DCN
```

```
=> S R16158/DCN  
L14      54 R16158/DCN
```

```
=> S 8821-12501/DCN  
L15      1 8821-12501/DCN
```

```
=> S R01255/DCN  
L16      230 R01255/DCN
```

Note: one compound number is not retrieved. This is an error in Derwent WPI (and has already been reported to Derwent!)

This is the Derwent WPI Specific Compound Number for diazepam.

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

```
=> S L1-L16
```

```
L17      241 (L1 OR L2 OR L3 OR L4 OR L5 OR L6 OR L7 OR L8 OR L9 OR L10 OR  
          L11 OR L12 OR L13 OR L14 OR L15 OR L16)
```

```
=> DEL L1-L16 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 S L***-L***
```

Helpful HINT

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.

```
=> 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
L2 1 DE4207922/PN
.
.

=> S FR----839/PN
L54 0 FR----839/PN
      (FR839/PN)
.
.

=> S WO9802148/PN
L152 1 WO9802148/PN
```

In this example one patent is not retrieved. This is a 1961 French BSM patent, which pre-dates WPINDEX coverage (1963). The patent is also not available in CAPLUS.

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

```
=> S L2-L152

L153 149 (L2 OR L3 OR L4 OR L5 OR L6 OR L7 OR L8 OR L9 OR L10 OR ....)

=> DEL L2-L152 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 S L***-L***
L2 149 S L***-L***
```

*L1 MMS-DWPI: 241 records
L2 MMS-Pharmsearch: 149 records*

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

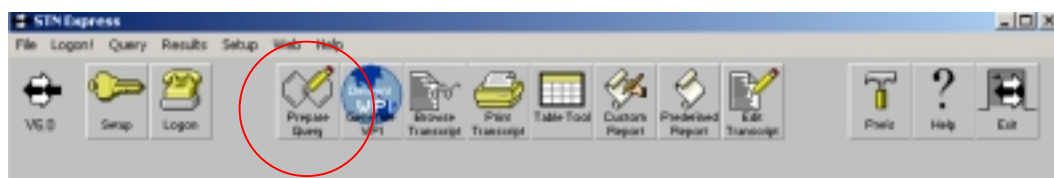
STN Express incorporates an offline module for generating Derwent Chemical Fragmentation Codes from a graphical structure query. The codes can then be searched by Derwent Subscribers in files WPIDS or WPIX. STN is the only host to provide this facility within its online communication software.

note

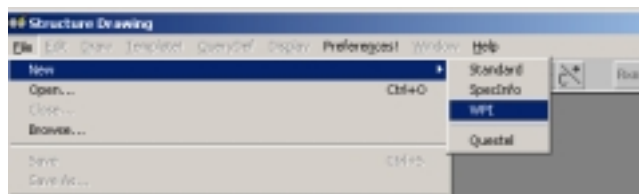
The Merged Markush Service (MMS) is an open access database available to all online searchers. However, subscription to Derwent *Chemical Patents Index (CPI)* is required to use Chemical Fragmentation Codes.

Prepare the fragmentation code query offline:

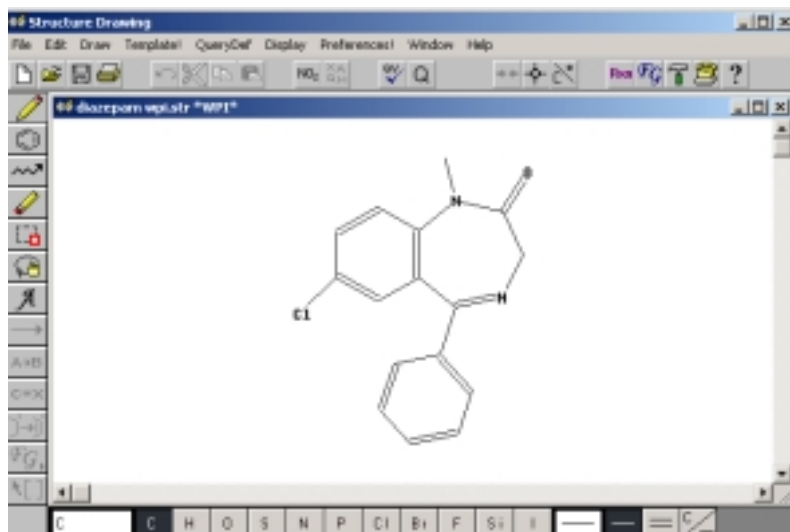
1. Click on the structure query preparation button



2. Open a new WPI format window

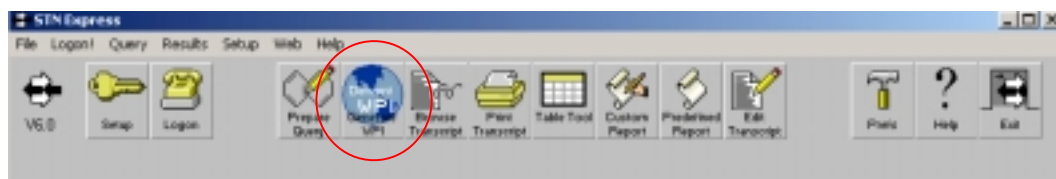


3. Prepare and Save the query:

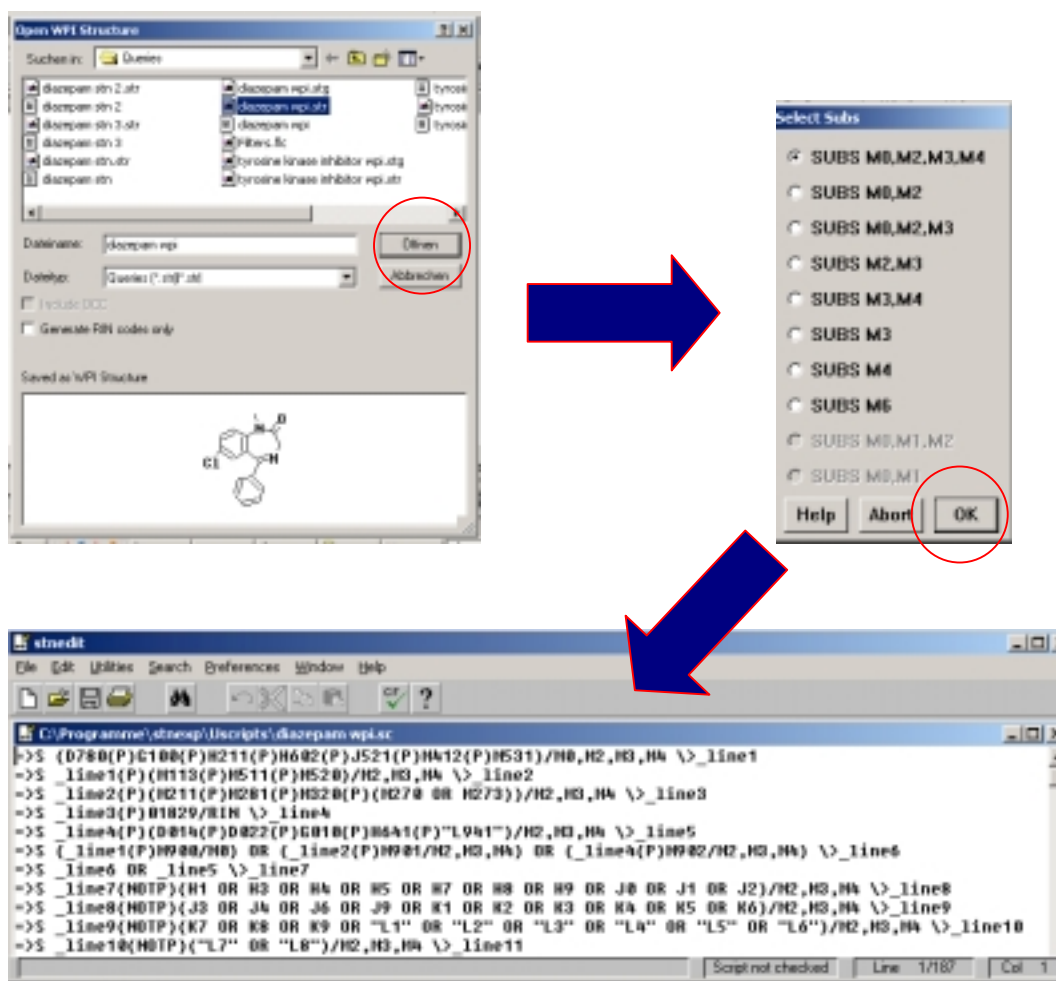


The Complete Markush Structure Search: Mission Impossible?

- Click on the *Generate WPI* button



- 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.

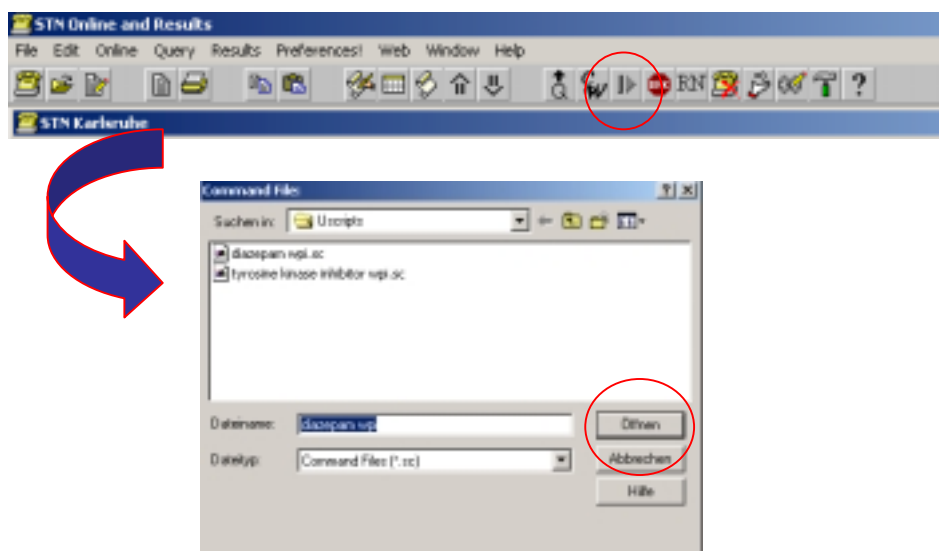
The Complete Markush Structure Search: Mission Impossible?

Logon to STN and access file WPIDS (or WPIDS):

=> **FILE 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:



```
=> S (D780(P)G100(P)H211(P)H602(P)J521(P)M412(P)M531)/M0,M2,M3,M4
L3      1170 (D780(P)G100(P)H211(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)(M211(P)M281(P)M320(P)(M270 OR M273))/M2,M3,M4
L5      721 L4(P)(M211(P)M281(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 H5 OR H7 OR H8 OR H9 OR J0 OR J1 OR J2)/M2,M3,M4
L10     741 L9(NOTP)(H1 OR H3 OR H4 OR H5 OR H7 OR H8 OR H9 OR J0 OR J1
        OR J2)/M2,M3,M4

=> S L10(NOTP)(J3 OR J4 OR J6 OR J9 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 J9 OR K1 OR K2 OR K3 OR K4 OR K5
        OR K6)/M2,M3,M4
```

The Complete Markush Structure Search: Mission Impossible?

```
=> S L11(NOIP)(K7 OR K8 OR K9 OR "L1" OR "L2" OR "L3" OR "L4" OR "L5" OR "L6")/M2,M3,M4
L12      741 L11(NOIP)(K7 OR K8 OR K9 OR "L1" OR "L2" OR "L3" OR "L4" OR
          "L5" OR "L6")/M2,M3,M4

=> S L12(NOIP)("L7" OR "L8")/M2,M3,M4
L13      741 L12(NOIP)("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***-I***
L2      149 L***-I***
L3      741 SEA L***(NOIP)("L7" OR "L8")/M2,M
```

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

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

```
=> L3 (NOIP) M904/M2,M3,M4
```

```
L4      487 L3 (NOIP) 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
```

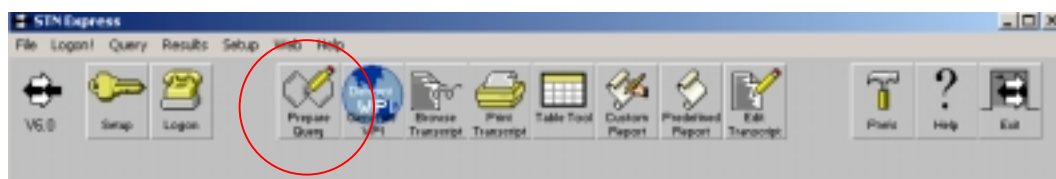
```
L5      791 L1 OR L2 OR L4
```

CASLINK

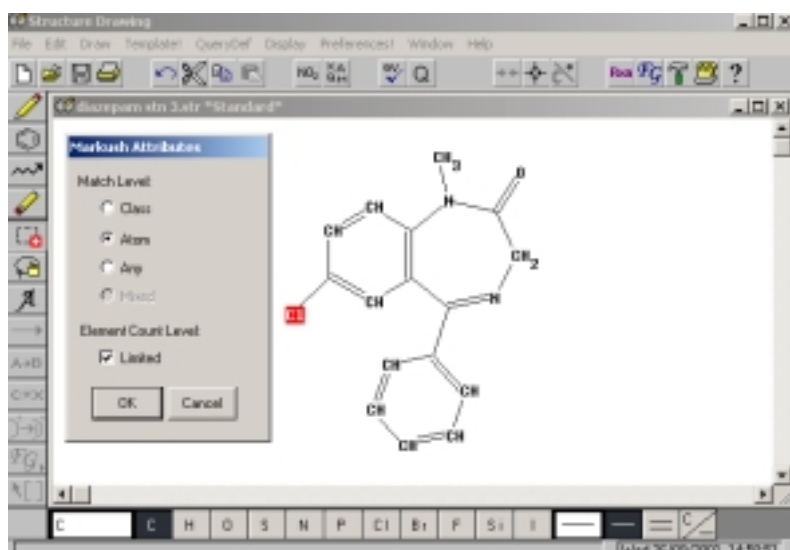
The Chemical Abstracts Service (CAS) provides the following structure files of interest for this search example: REGISTRY, MARPAT and MARPATPREV. In addition to providing the databases individually, CAS also links them together - along with the CAPLUS bibliographic file - to form the specialised CASLINK 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



2. Prepare and Save the query



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:

=> **FILE CASLINK**

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

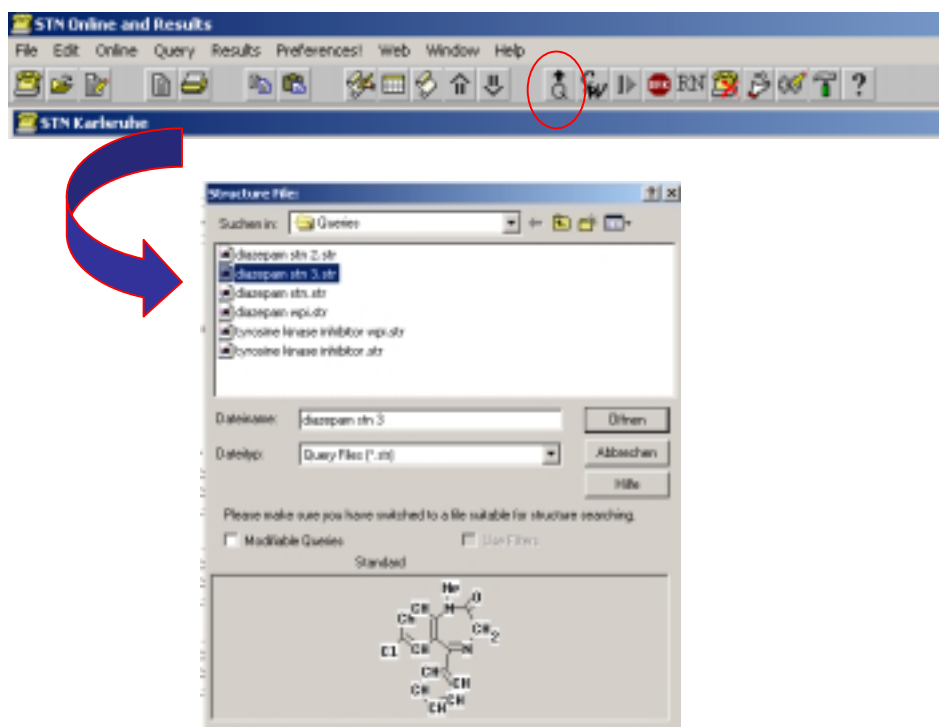
FILE 'MARPAT' ENTERED AT 20:22:23 ON 28 SEP 2001
COPYRIGHT (C) 2001 American Chemical Society (ACS)

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

The Complete Markush Structure Search: Mission Impossible?

FILE 'CAPLUS' ENTERED AT 20:22:23 ON 28 SEP 2001
COPYRIGHT (C) 2001 AMERICAN CHEMICAL SOCIETY (ACS)

Upload the structure query:



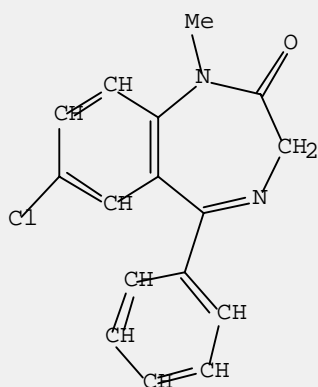
=>
Uploading C:\Programme\stnexp\Queries\diazepam.stn.str

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:

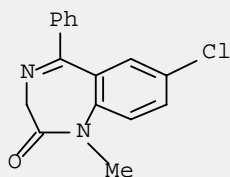
⇒ D L7 SCAN

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-

MF C21 H27 Cl N2 O2 . C16 H13 Cl N2 O

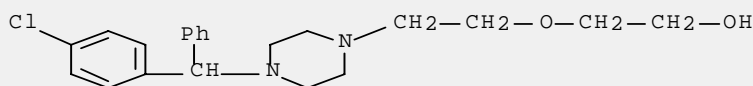
CI MXS

CM 1



CM 2

The Complete Markush Structure Search: Mission Impossible?



HOW MANY MORE ANSWERS DO YOU WISH TO SCAN? (1):0

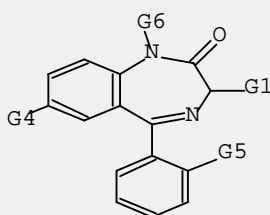
Review effectiveness of the search from MARPAT:

⇒ D L8 SCAN

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 prepn; 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)

...

MSTR 1



G4 = Cl
G6 = Me
MPL: claim 1

ALL ANSWERS HAVE BEEN SCANNED

Compare ease-of-interpretation of results using MARPAT SCAN with MMS VI FO on page 24.

note

MMS does not offer equivalent features to the SCAN format or SAMPLE search options. Both of features are free-of-charge.

The Complete Markush Structure Search: Mission Impossible?

Conduct a full file search:

=> S L6 FUL

S L6 SSS FUL FILE=REGLSIRY

FULL SEARCH INITIATED 20:23:37 FILE 'REGLSIRY'
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 DUORDER 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:

⇒ S L13 AND P/DT

S L12 AND P/DT FILE=CAPLUS
L14 562 FILE CAPLUS
1 FILES SEARCHED...

S L10 AND P/DT FILE=CAPLUS
L15 4 FILE CAPLUS
1 FILES SEARCHED...

S L11 AND P/DT FILE=CAPLUS
L16 0 FILE CAPLUS
1 FILES SEARCHED...

S L15 AND L10 FILE=MARPAT
L17 4 FILE MARPAT
1 FILES SEARCHED...

S L16 AND L11 FILE=MARPATPREV
L18 0 FILE MARPATPREV
1 FILES SEARCHED...

DUP REM L18 L17 L14

L18 HAS NO ANSWERS

PROCESSING COMPLETED FOR L18

PROCESSING COMPLETED FOR L17

PROCESSING COMPLETED FOR L14

L19 562 DUP REM L18 L17 L14 (4 DUPLICATES REMOVED)

ANSWERS '1-4' FROM FILE MARPAT

ANSWERS '5-562' FROM FILE CAPLUS

*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

'L21 ' DELETED
L21 1132 FSO L20

156 Multi-record Families	Answers 1-380
Family 1	Answers 1-2
.	
.	
Family 155	Answers 377-378
752 Individual Records	Answers 381-1132
0 Non-patent Records	

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.

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

SET SMARTSELECT OFF
SET COMMAND COMPLETED

The Complete Markush Structure Search: Mission Impossible?

SET HIGHLIGHTING DEF
SET COMMAND COMPLETED

Review the answers one record per family:

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

=> D PFAM=1- 1 AN TI

L21 ANSWER 1 OF 1132 WPIDS COPYRIGHT 2001 DERWENT INFORMATION LTD **FAMILY1**
AN 2001-202821 [20] WPIDS
TI Treating generalized anxiety disorder by administering low dose of cyclobenzaprine.

L21 ANSWER 3 OF 1132 WPIDS COPYRIGHT 2001 DERWENT INFORMATION LTD **FAMILY2**
AN 2001-474811 [51] WPIDS
TI Sustained-release composition in tablet form comprises a medicament and a blend of ethylcellulose, at least one other polymer and a medicament.

...

L21 ANSWER 377 OF 1132 CAPLUS COPYRIGHT 2001 ACS **FAMILY 155**
AN 1973:537208 CAPLUS
DN 79:137208
TI Anticonvulsive, muscle-relaxant, and sedative benzodiazepines

L21 ANSWER 379 OF 1132 CAPLUS COPYRIGHT 2001 ACS **FAMILY 156**
AN 1972:501691 CAPLUS
DN 77:101691
TI Benzodiazepine derivatives

CAPLUS multi record families where no WPIDS record was retrieved.

L21 ANSWER 381 OF 1132 WPIDS COPYRIGHT 2001 DERWENT INFORMATION LTD
AN 2001-488631 [53] WPIDS
TI Combination formulation used for treating proliferative diseases, especially tumor diseases comprises abl-, PDGF-R- and/or Kit receptor tyrosine kinase inhibitor and organic compound binding to alpha1-acidic glycoprotein.

L21 ANSWER 382 OF 1132 WPIDS COPYRIGHT 2001 DERWENT INFORMATION LTD
AN 2001-417787 [44] WPIDS
TI Use of levetiracetam alone or as a potentiator of compound(s) inducing neural inhibition mediated by gamma aminobutyric acid-A receptors, especially valproate, for treating bipolar disorders, mania, migraine and chronic or neuropathic pain.

...

L21 ANSWER 926 OF 1132 CAPLUS COPYRIGHT 2001 ACS
AN 2001:597840 CAPLUS
DN 135:157707
TI Preparation of carrier particles for drug delivery

L21 ANSWER 927 OF 1132 CAPLUS COPYRIGHT 2001 ACS
AN 2001:581687 CAPLUS
DN 135:157684
TI Continuous method for preparing pharmaceutical granules

CAPLUS single record families where no WPIDS record was retrieved.

L21 ANSWER 1132 OF 1132 MARPAT COPYRIGHT 2001 ACS
AN 130:296702 MARPAT
TI Preparation of benzodiazepinones as protein tyrosine kinase inhibitors

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

L1	241 S L***-L***	MMS DWPI
L2	149 S L***-L***	MMS Pharmsearch (+1 BSM)
L3	741 SEA L***(NOIP) ("L7" OR "L8")/M2,M3,M4	DWPI Frag Codes, 1963-
L4	487 SEA L3(NOIP)M904/M2,M3,M4	DWPI Frag Codes, 1963-86
L5	791 L1 OR L2 OR L4	Total DWPI/Pharmsearch

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

L6	STRUCTURE UPLOADED	
L7	10 S L6 SSS SAM FILE=REGISTRY	
L8	1 S L7 SSS SAM FILE=MARPAT	
L9	149 S L6 SSS FUL FILE=REGISTRY	
L10	4 S L9 SSS FUL FILE=MARPAT	
L11	0 S L10 SSS FUL FILE=MARPATPREV	
L12	11073 S L9 FILE=CAPLUS SET DUPORDER FILE	
L13	11069 DUP REM L11 L10 L12 (8 DUPLICATES REMOVED)	CASLINK, total
L14	562 S L12 AND P/DT FILE=CAPLUS	
L15	4 S L10 AND P/DT FILE=CAPLUS	
L16	0 S L11 AND P/DT FILE=CAPLUS	
L17	4 S L15 AND L5 FILE=MARPAT	
L18	0 S L16 AND L6 FILE=MARPATPREV	
L19	562 DUP REM L18 L17 L14 (4 DUPLICATES REMOVED) SET SMARTSELECT ON SET HIGHLIGHTING OFF	CASLINK, patents only

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

L20	1132 DUP REM L5 L19 (221 DUPLICATES REMOVED) SET SMARTSELECT ON SET HIGHLIGHTING OFF	CAPLUS/WPIDS deduplicated
L21	1132 FSO L26 SET SMARTSELECT OFF SET HIGHLIGHTING DEF SET SMARTSELECT OFF SET HIGHLIGHTING DEF	CAPLUS/WPIDS FamilySorted (908 patent families)

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 to 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/EN
.
.
40 ZA6803041/EN
***** END OF L24 *****

***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**

L25 40 L24

Extract Priority Numbers from CAPLUS to find extra records in WPIDS:

⇒ **FIL WPIDS**

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

⇒ **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):

⇒ **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:

⇒ **FIL CAPLUS**

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

⇒ **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):

⇒ **L25 AND L31**

L33 6 L25 AND L31

⇒ **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 WPIDS 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) + WPIDS (L5) (221 records de-duplicated)
L21	1132	CASLINK + WPIDS (L20) FamilySorted (908 patent families)
L23	508	WPIDS from CASLINK (L19) via PN.B
L25	40	CASLINK (L19) PN.B not in WPIDS
L27	14	WPIDS from CASLINK (L25) via PRN
L29	522	WPIDS from CASLINK via PN.B (L19) & PRN (L25)
L33	6	CASLINK (L25) in WPIDS via PRN
L34	34	CASLINK (L25) not in WPIDS

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 (WPIDS) 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 WPIDS.

Example Boolean Search statements in WPIDS:

```

=> 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, meprobanate,

AN 1981:145357 CAPLUS
TI Pharmaceutical composition with improved absorption properties
PA Kali-Chemie Phama G.m.b.H., Fed. Rep. Ger.
PI AT 360163 B 19801229 AT 1978-4020 19780602

CAS unique versus Merged Markush Service (MMS)

AN 2001-091624 [10] 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 [27] WPIDS
TI Medicine comprising traditional Chinese medicine and western medicine is used for treating cough and asthma diseases.
PA (LIJSS-I) LIJ S
PI CN 1277866 A 20001227 (200127)* A61K035-78

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

AN 1980-87365C [49] 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)*

AN 1966-36636F [00] WPIDS
TI Benzodiazepine derivs - useful as muscle relaxants, anti-convulsants, tranquillisers.
PA (RICT) RICHTER GEDEON VEGYESZETI GYAR
PI NL 6812907 A 19690313 (196800)*

Merged Markush Service (Derwent)

Derwent unique versus INPI:

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

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

Derwent unique versus CAS:

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

Text from EP 1004318 (Claim 5):

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

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

Merged Markush Service (INPI)

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

PHARM - (C) INPI
AN - 61100839
EN - 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:

AN 1986-095329 [15] WPIDS
TI New use of bi cyclic diazepine derivs. - as inhibitors of platelet activating factor.
PA (BOEH) BOEHRINGER INGELHEIM
PI EP 176928 A 19860409 (198615)* DE 27p

AN 1980-80636C [45] WPIDS
TI Sub-lingual and buccal admin. of benzodiazepine(s) - for rapid absorption to affect the CNS.
IN PORTIER, W R
PA (AMHP) AMERICAN HOME PROD CORP
PI US 4229447 A 19801021 (198045)*

INPI unique versus Derwent:

AN 1998-002241 [01] WPIDS
TI Use of antipyretic and benzodiazepine derivative stabilised by benzoic acid, salt or derivative - to treat e.g. infant convulsions, arthrosis and other painful disorders..
PA (CRIN-N) LAB CRINEX SA
PI FR 2747923 A1 19971031 (199801)* 14p A61K047-12

AN 1985-013506 [03] WPIDS
TI 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..
PA (ELIL) LILLY & CO ELI; (NITL) NITTO ELECTRIC IND CO
PI DE 3424057 A 19850110 (198503)* 31p

Derwent Fragmentation Codes (1963-1986)

Derwent unique versus CAS:

AN 1986-335074 [51] WPIDS
TI Drug-contg. porous medical gels - produced without heating or using crosslinking agent to avoid damaging drug.
PA (NITL) NITTO ELECTRIC IND CO
PI JP 61249920 A 19861107 (198651)* 7p

AN 1966-41591F [00] WPIDS
TI Process for 2 3-dihydro-1 4-benzodiazepine-2-one derivs anti-convulsant - ts muscle relaxants sedatives.
PA (HOFF) HOFFMANN LA ROCHE & CO AG F
PI NL 6908658 A (196800)*

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:

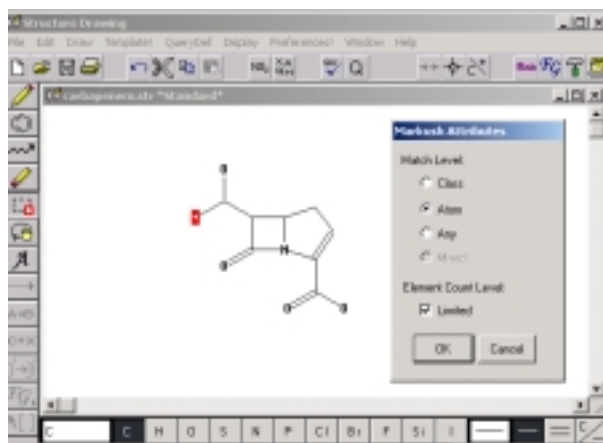
- **Diazepam:** a search for specific references to diazepam.
- **Carbapenem:** a search for carbapenem derivatives based on a common substructure.
- **Tyrosine Kinase Inhibitor:** a search for a class of *tyrosine kinase inhibitors* with a common substructure.
- **Sartan:** a more complex search for *Sartan* derivatives, based on a common substructure, but including a generic heterocycle substituent.
- **Nonanoyloxybenzenesulphonate (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.

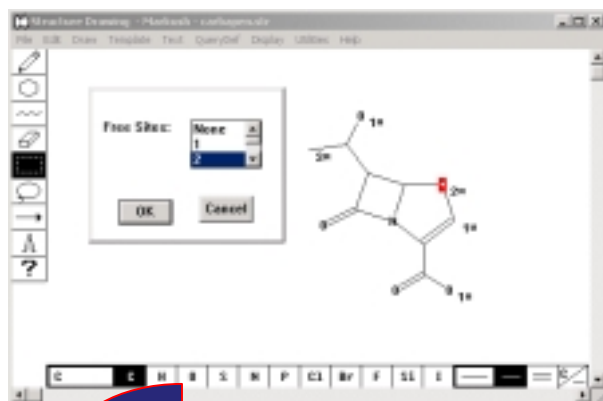
STN Express query for CASLINK



Notes:

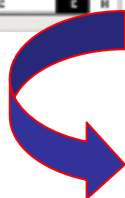
1. For CASLINK all undefined positions are assumed to be free for substitution.
2. The MARPAT default Match Level for chain atoms is Class (i.e. to retrieve Alkyl as well as C). For this query this is set to Atom to find an exact match only (i.e. one carbon)

Markush TOPFRAG query for MMS



Notes:

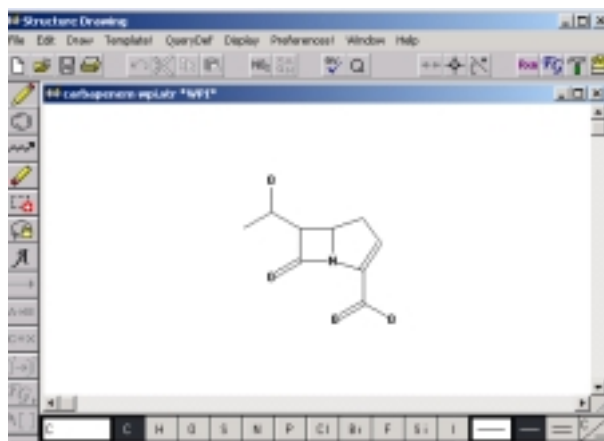
1. For MMS queries the number of substitutable positions (Free Sites) must be specified.
2. The default Translation is Exact for all real atoms in MMS. This is equivalent to MARPAT Atom Match Level (as set above).



```
117011 - MD44C - c:\ref\carbapenem\PM4
File Edit Format Search Help
#1
5-1-5
1-11-12
2-6-9
3-7
5-14
8-10
11-13
#2
0 5,12,13,14
#3
4# 1-5,6-14,11-12
#4
1 5,5,13
2 4
3 10
```

See page 21 for how to upload this query via STN Express.

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):

```
=>S (D690(P)H401(P)H481(P)J111(P)J521(P)M331(P)M412)/M0,M2,M3,M4 \>_line1
=>S _line1(P)(M511(P)M520(P)M530(P)M540)/M2,M3,M4 \>_line2
=>S _line2(P)(M280(P)M312(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:

```
=>S (D690(P)(J11! OR J21!)(P)J52!(P)(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)D013(P)(D011 OR D013(P)D019 OR D014)(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
```

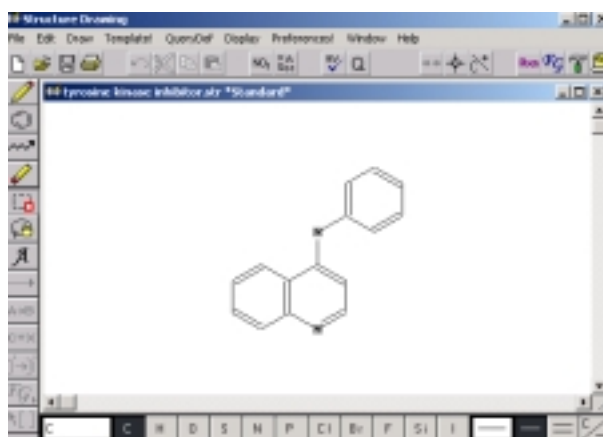
note

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.

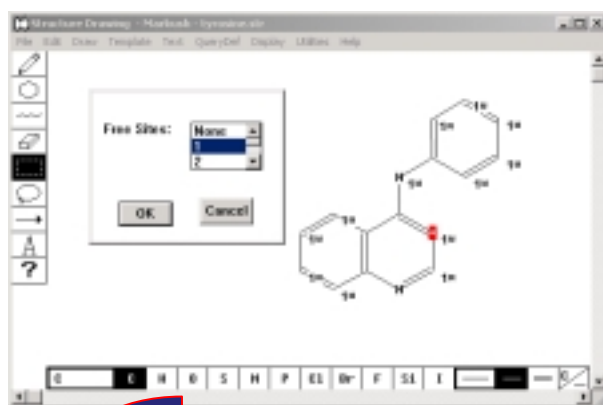
STN Express query for CASLINK



Notes:

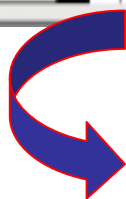
1. For CASLINK all undefined positions are assumed to be free for substitution.

Markush TOPFRAG query for MMS



Notes:

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



```
MFPEDIT - HDARC - c:\mf\tyrosine.MAR
File Edit Revert Search Help
gr
6-1:6
4-2:10
5-10
7-11:17
12-17

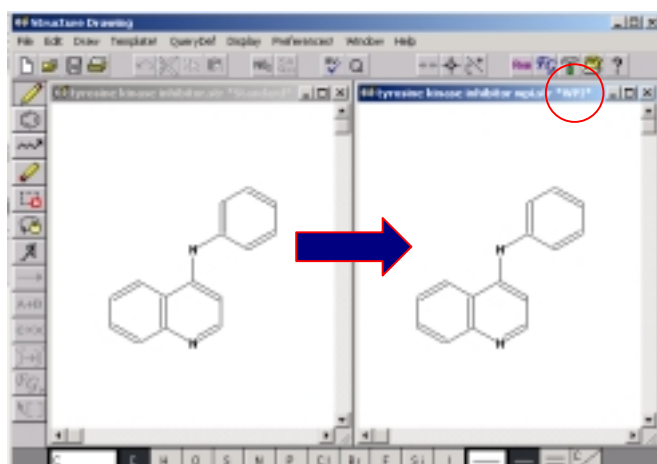
at
N 10,11

bo
no 1-2,1-5,2-3,3-4,4-5,4-7,5-5,5-10,7-8,0-9
no 9-10,12-13,12-17,13-14,14-15,15-16,16-17

fo
1 1,2,3,6,8,9,11,13,14,15,16,17
```

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):

```
=>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(NOIP)(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(NOIP)(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:

```
=>S (D621(P)(H12! OR J31! OR J32!)(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 J32!))/M2 \>_line4
=>S (_line1(P)M900/M0) OR (_line2(P)M901/M2) OR (_line3(P)M902/M2) \>_line5
=>S _line5 OR _line4 \>_line6
```

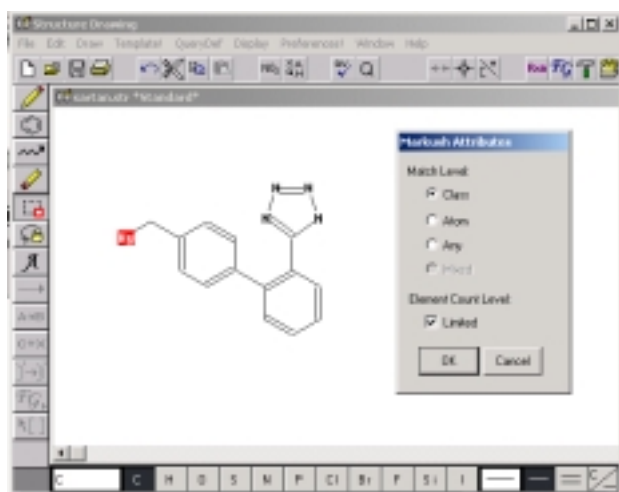
note

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.

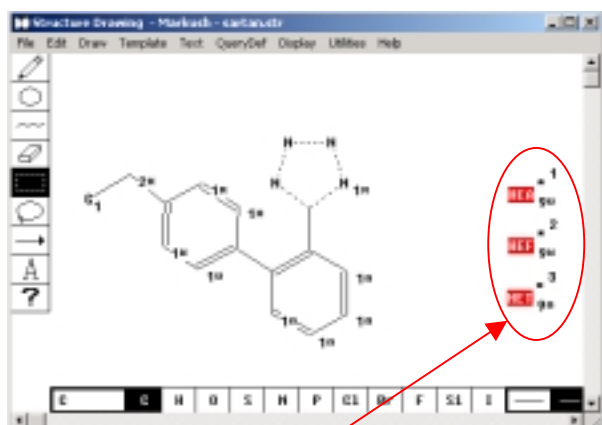
STN Express query for CASLINK



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



Please specify translation attributes:

NT (narrower)

BT (broader)

EQ (equal only)

ANY

Clear

OK Cancel

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.

```

MTEEDIT MDARC c:\mf\arlan.MAR
File Edit Revert Search Help
gr
8-1:6
3-12:17
4-7:11
7-11
12-17
15-18-19

at
N 8,9,10,11
G1 19

ba
no 1-2,1-6,2-3,3-4,4-5,5-6,12-13,12-17,13-14
x cy 7-8,7-11,8-9,9-10,10-11
no 14-15,15-16,16-17

fs
1 1,2,5,6,11,13,14,16,17
2 18

gm
1
gr
1,2,3

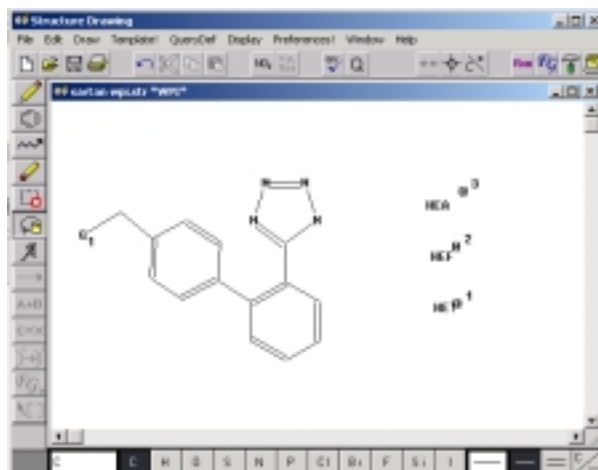
at
HEA 1
HEF 2
HET 3

fs
9 1,2,3

atrr
wa
at 1,2,3
    
```

See page 21 for how to upload this query via STN Express.

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):

```

=>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
    
```

(Continued...)

The Complete Markush Structure Search: Mission Impossible?

```
=>S _line7(P)M413/M0,M2,M3,M4 \>_line8
=>S _line8(P) (M123(P)M522)/M2,M3,M4 \>_line9
=>S _line9(P) (F011 OR F012 OR F013 OR F014 OR F016 OR F019)/M2,M3,M4 \>_line10
=>S _line7(P)M412/M0,M2,M3,M4 \>_line11
=>S _line11(P) (M511(P)M521(P) (M122 OR M123))/M2,M3,M4 \>_line12
=>S _line12(P) (D011 OR D012 OR D021 OR D022 OR D030)/M2,M3,M4 \>_line13
=>S (M900(P)(_line8 OR _line11)) OR (M901(P)(_line9 OR _line12)) \>_line14
=>S _line14 OR (M902(P)(_line9 OR _line12)) OR _line10 OR _line13 \>_line15
=>S _line15(NOTP) (H1 OR H2 OR H3 OR H4 OR H5 OR H6 OR H7 OR H8 OR H9)/M2,M3,M4 \>_line16
=>S _line16(NOTP) (J0 OR J1 OR J2 OR J3 OR J4 OR J5 OR J6 OR J9 OR K0)/M2,M3,M4 \>_line17
```

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:

```
=>S (F570(P) (M532 OR M533) (P) (M411 OR M412 OR M413) )/M0,M2,M3,M4 \>_line1
=>S _line1(P) (M111(P)M113(P) (M122 OR M123) (P) (M131 OR M132) (P) (M521 OR M522 OR M523) )/M2,M3,M4\>_line2
=>S _line2(P) (M311(P)M342)/M2,M3,M4 \>_line3
=>S _line3(P)00061/RIN \>_line4
=>S _line4(P) (G1!!(P)F015)/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
```

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

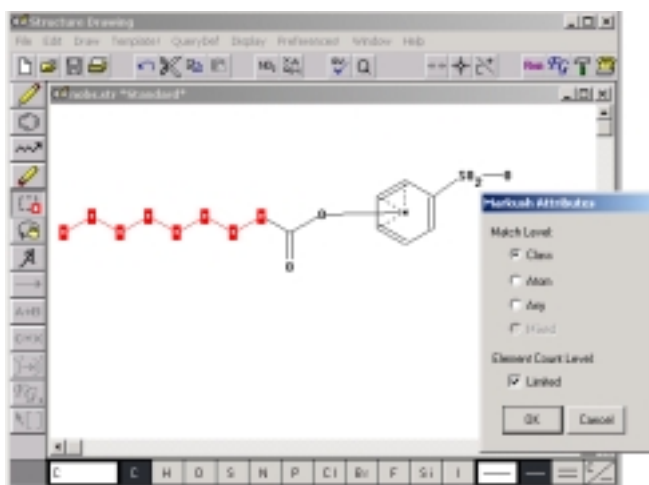
note

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.

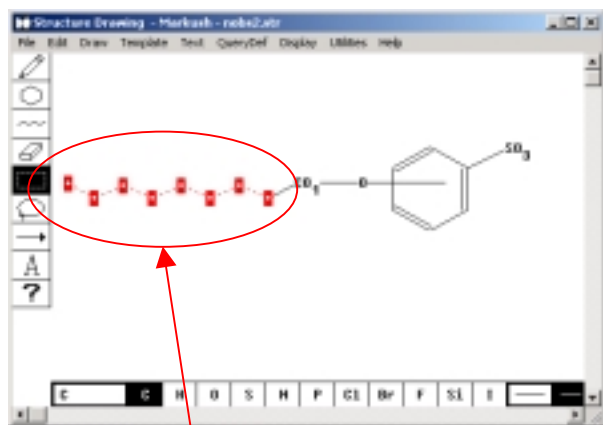
STN Express query for CASLINK



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



Please specify translation attributes:

NT (narrowed)

BT (broader)

EQ (equal only)

ANY

Clear

OK Cancel

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.

```

MTFEDIT - MDARC - c:\mft\nobs2.MAR
File Edit Revert Search Help
gr
6-1:6
5-7
16-17-8:10
18-11:16

at
S03 7
C01 8
O 9
# 10

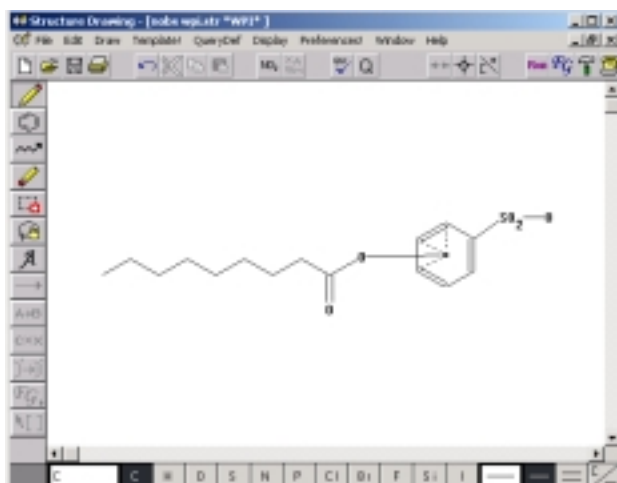
vp
2,3,4

bo
no 1-2,1-6,2-3,3-4,4-5,5-6

attr
tra
bt 11,12,13,14,15,16,17,18
    
```

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. 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)M531)/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(NOIP)(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(NOIP)(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(NOIP)(K7 OR K8 OR K9 OR "L1" OR "L2" OR "L3" OR "L4" OR "L5" OR "L6")/M2,M3,M4 \>_line8
=>S _line8(NOIP)("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:

```
=>S (G100(P)J241(P)K431(P)M414(P)M531)/M0,M2,M3,M4 \>_line1
=>S _line1(P)((M222 OR M220)(B)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
```

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

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)

Result description ¹	Lx	Diazepam	Carbapenem	Tyrosine	Sartan	NOBS ²
MMS						
Total CNs		34	1704	495	1250	258
Derwent WPI CNs		16	919	326	712	256
INPI Pharmsearch CNs		18	785	169	538	2
INPI PHARMSEARCH						
INPI Pharmsearch PNs		151	497	154	329	2
DERWENT WPI (STN)						
Records from Derwent CNs	L1	241	386	204	396	261
Records from INPI PNs	L2	149	481	146	310	2
INPI PNs not in DWPI		1	0	0	0	0
Frag. Code (1963-date)	L3	741	826	2108	2249	1091
Frag. Code (1963-1986)	L4	487	429	1198	1540	693
Total Derwent + INPI	L5	791	981	1460	2023	894
CASLINK						
CASLINK (dups. removed)	L13	11069 (-4)	4838 (-338)	1328 (-51)	2450 (-297)	265 (-43)
CASLINK (Patents)	L19	562	891	464	546	257
COMBINED RESULTS						
Total (duplicates removed)	L20	1132 (-221)	1257 (-615)	1794 (-130)	2157 (-412)	1011 (-140)
FSORT patent families	L21	908	1024	1499	1955	958
MMS & CAS (dup. rem.) ³		-	-	667 (-104)	623 (-410)	-
FSORT (MMS/CAS)		-	-	585	540	-
DERWENT WPI (STN)						
CAS DWPI records (via PNs)	L29	522	843	426	525	245
CAS Patents not in DWPI	L34	34	11	13	4	1
Total Derwent/INPI/CAS	L35	997	1163	1692	2122	995
Derwent/INPI not CAS	L36	475	320	1266	1597	750
CAS not Derwent/INPI	L37	205	182	232	99	101
Total MMS in DWPI	L38	344	710	307	487	263
MMS Derwent not MMS INPI	L39	195	229	161	177	261
MMS INPI not MMS Derwent	L40	103	324	103	91	2
CAS not MMS	L41	335	299	291	100	109
MMS not CAS	L42	157	166	172	62	127
MMS Derwent not CAS	L43	113	58	120	50	125
MMS INPI not CAS	L44	54	132	70	20	2
Pre-1980 CAS not MMS	L45	153	42	13	0	0
1980-date CAS not MMS	L46	182	257	159	100	109
Frag Code (63-86) not CAS	L47	334	176	1102	1535	625
CAS not Frag Code (63-86)	L49	87	65	21	0	9

¹ All searches were all conducted on 02/10/01, except *Diazepam* conducted on 28/09/01.

² The general CAS Registry Number for Nonanoyloxybenzenesulfonate (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.

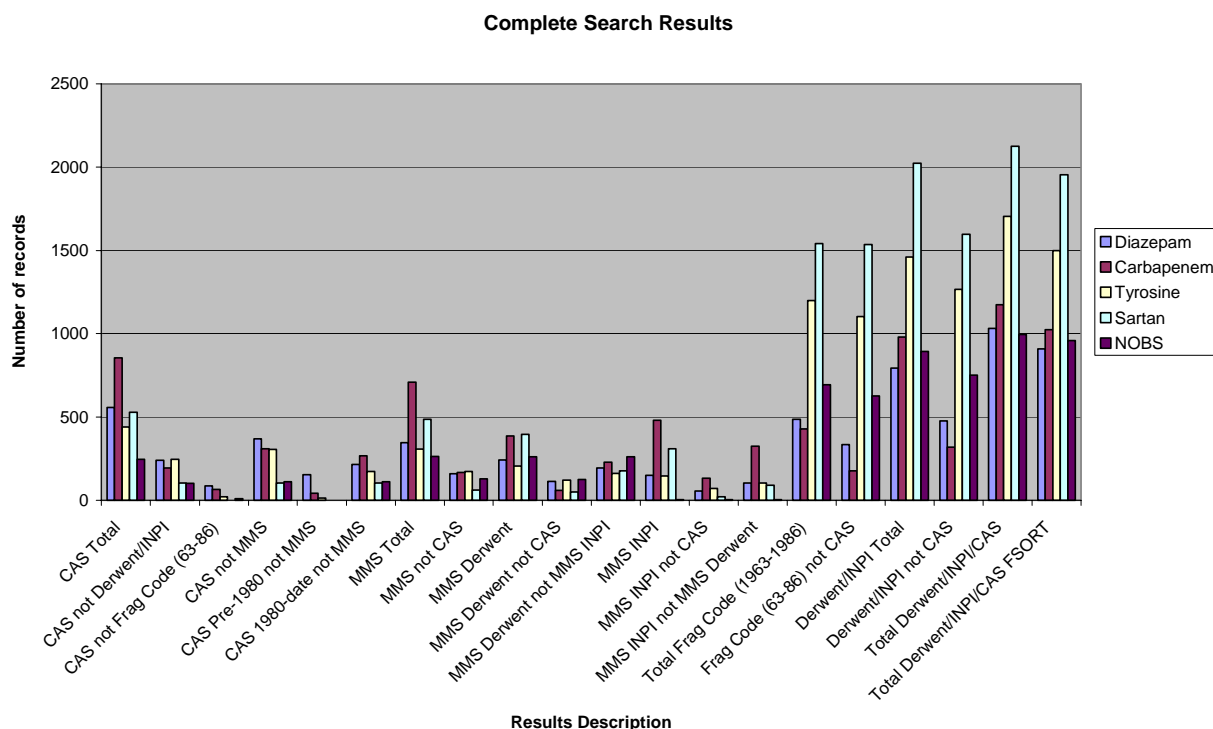
³ In these two cases the Fragmentation Code substructure search clearly retrieves a large number of incorrect answers (as well as the correct ones). Thus a total featuring just MMS and CAS is a useful reference point.

Search Results

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

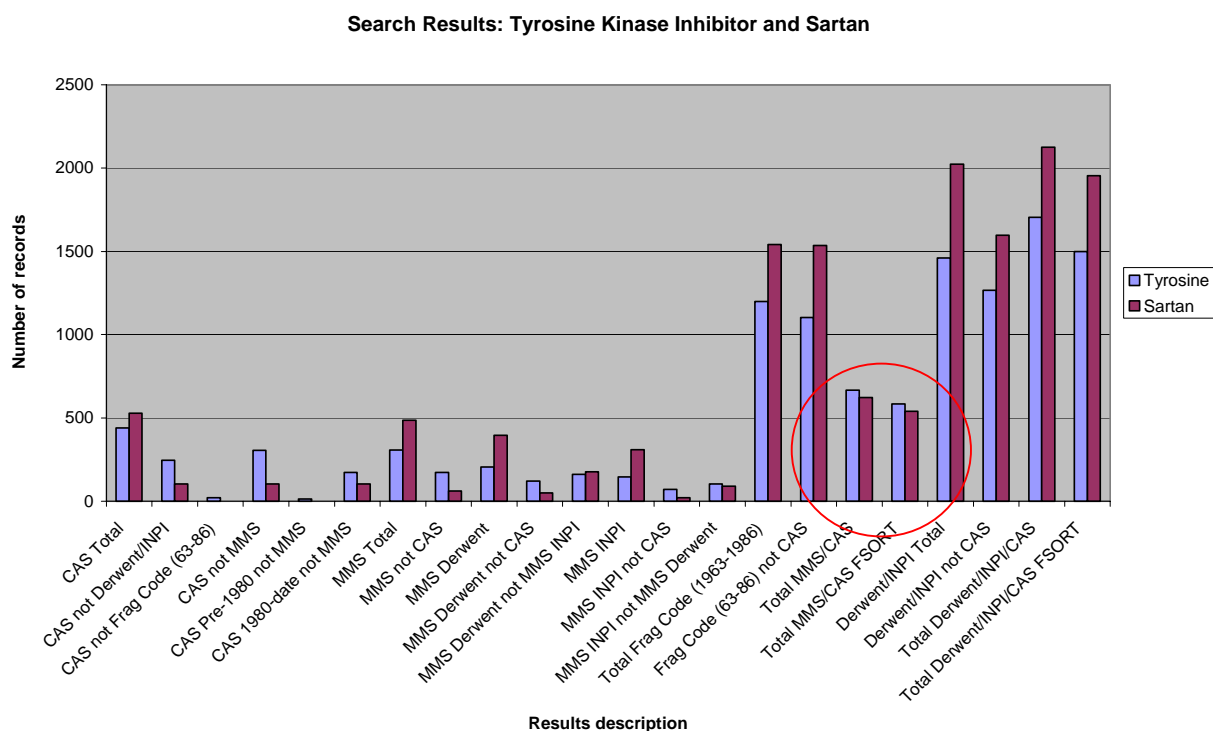
Result description	Diazepam	Carbapenem	Tyrosine	Sartan	NOBS
CAS Total	556	854	439	529	246
CAS not Derwent/INPI	239	193	245	103	102
CAS not Frag Code (63-86)	87	65	21	0	9
CAS not MMS	369	310	304	104	110
CAS Pre-1980 not MMS	154	42	13	0	0
CAS 1980-date not MMS	216	268	172	104	110
MMS Total	345	710	307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

This table of results is converted into the chart below.

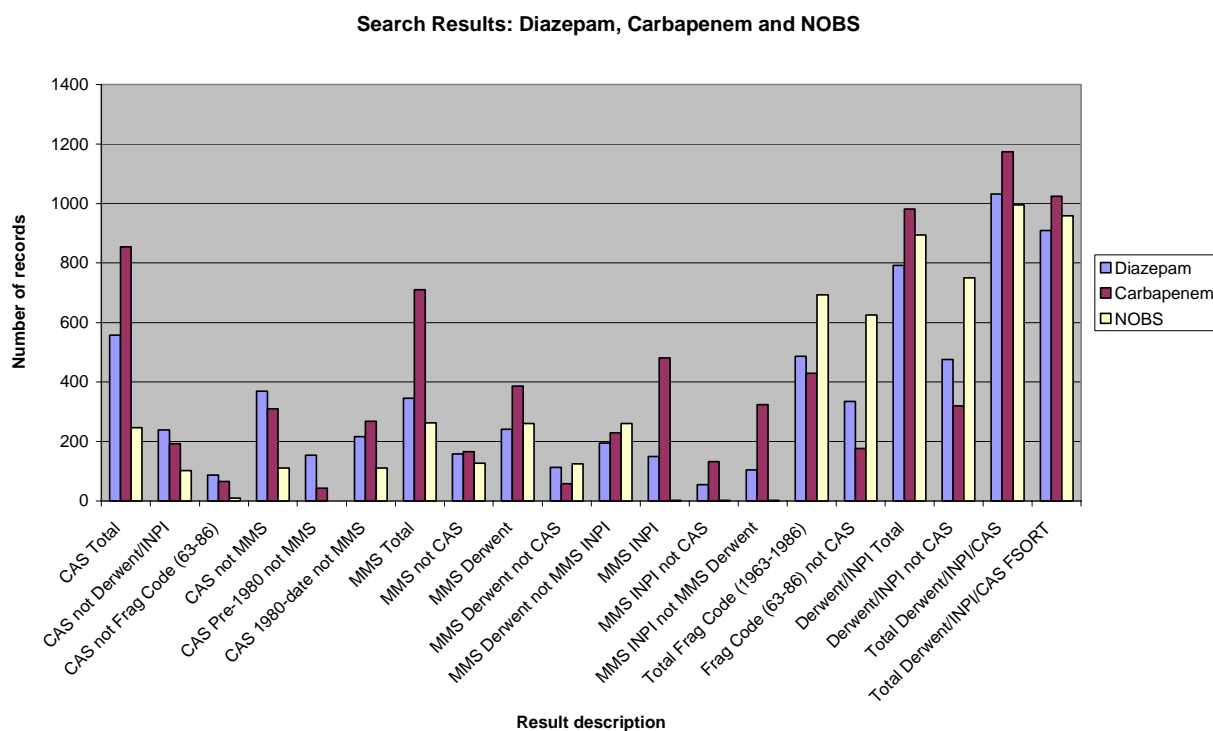


The Complete Markush Structure Search: Mission Impossible?

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.

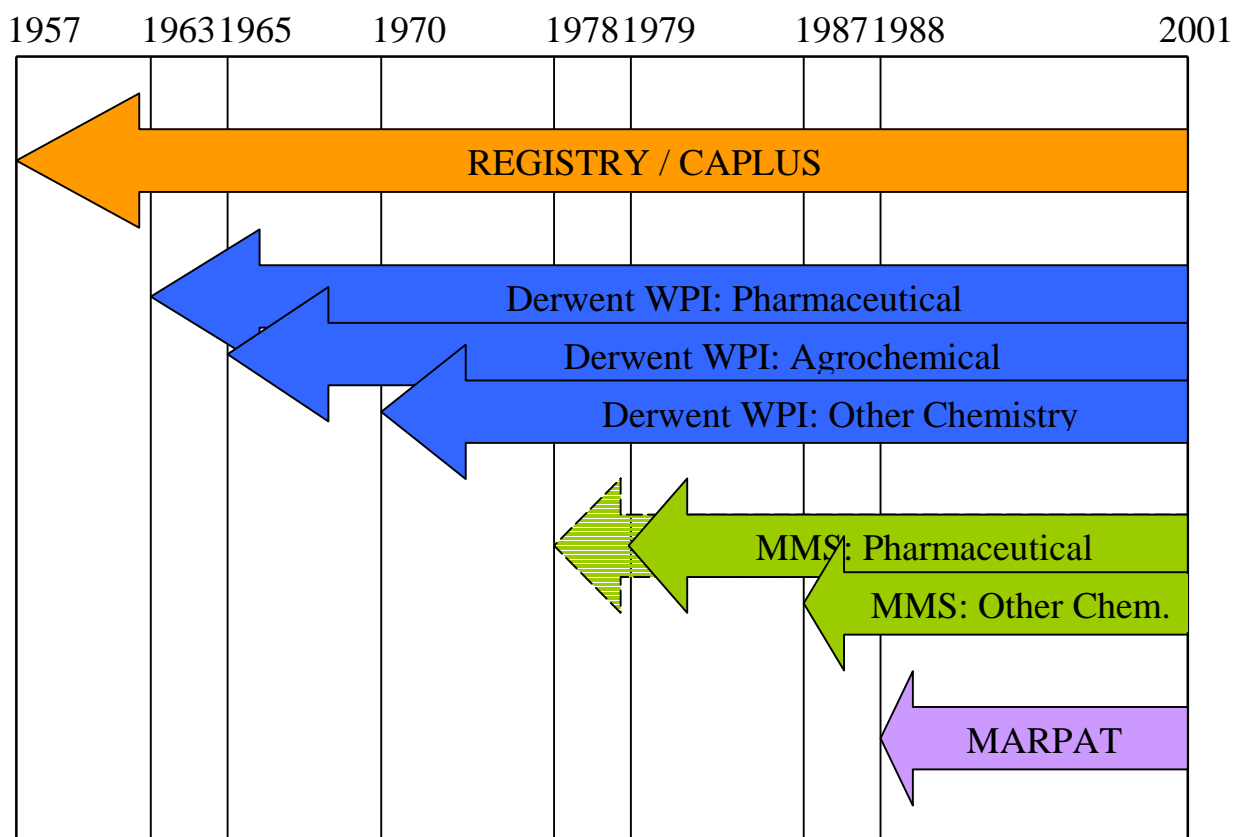
```
AN 1983-820415 [46] WPIDS
DNC C1983-112603
TI Bleaching compsns. effective over wide temp. ranges - contg. per oxygen
bleaching cpd. and specific bleach activator.
DC D25 E12 F09
IN SPADINI, G L; YAT-MING CHUNG, S; CHUNG, S Y
PA (PROC) PROCTER & GAMBLE CO
CYC 19
PI.B US 4412934 A 19831101 (198346)* 1lp
AB US 4412934 A UPAB: 19980617
Bleaching compsn. 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 inclusive; 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.
Compsns. 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 compsns. are also detergent compsns.), 0.1-60% pref. 0.5-
40% (II) (esp. 0.5-20% when detergent compsns.). Opt. compsns. 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% detergency 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 compsns. are extremely effective in surface bleaching of textiles, and are esp. useful
in removing 'dingy soils' from textiles. The compsns. are effective over a wide temp. range.
Dwg.0/0
```

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.



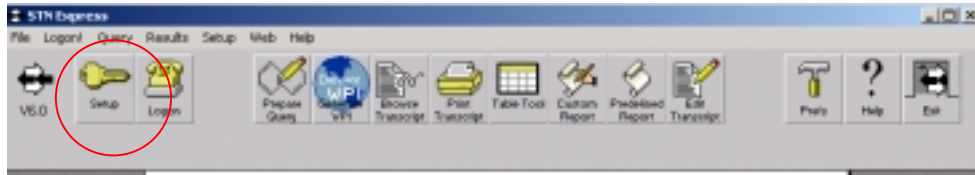
Conclusions

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.

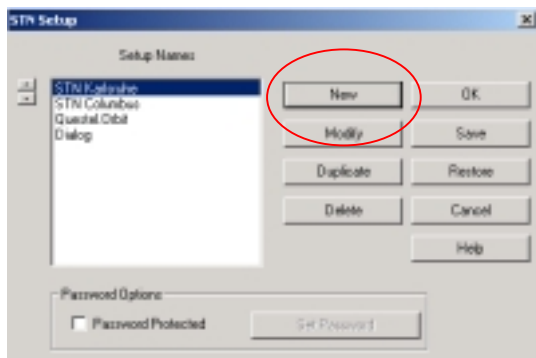
Connecting to MMS via STN Express

Setup STN Express to connect to MMS

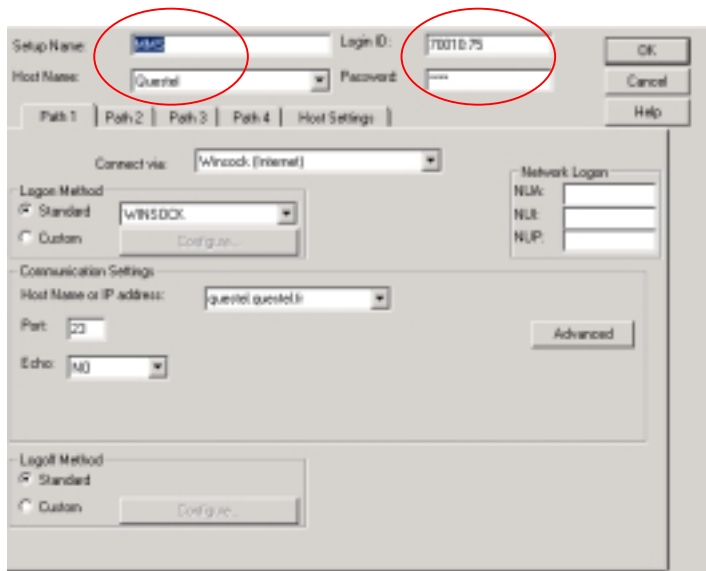
Set up a new connection option in the “Setup” part of STN Express.



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

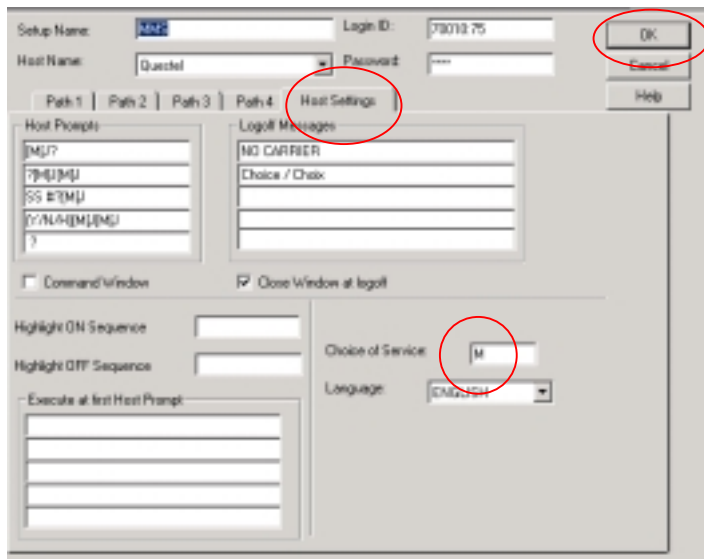


Give your new Set up a name, e.g. “MMS”; choose “Questel” as the Host Name¹; enter in your Login ID (optional) and Password (optional)

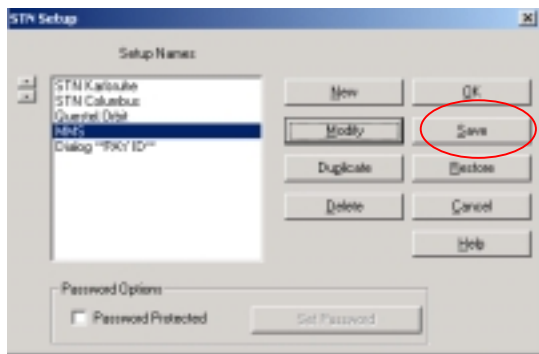


¹ The standard logon method is via the Internet (Winsock). For connection using other routes please contact your STN or Questel.Orbit representative for expert advice.

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-NSE03

SELECTED SEGMENT(S):

```
BACKF
MPHARM
WPIM
FRONTF
```

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

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

-ST- (BA,CN,QT,QG,RF,RE,AA,SB,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:

```
FRANCAIS(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 (TEKTRO MODE) -----> 4
5. VT 640 ( OR PCPLOT-IV EMULATOR ) -----> 5
6. HEWLETT-PACKARD 2647A,2648A -----> 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. SIN EXPRESS -----> S

PLEASE ENTER YOUR TERMINAL CODE -----> S
```

3. Select the transmission rate and format for MMS bibliographic images

TRANSMISSION RATE (BAUD) ? **300**
NO IMAGE (0) / GIF FORMAT (G) / TIFF FORMAT (T) ? **T**

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.

© Fachinformationszentrum Karlsruhe, October 2001

Fachinformationszentrum (FIZ) Karlsruhe
Hermann von Helmholtz Platz 1
76344 Eggenstein-Leopoldshafen
Germany

Tel: +49 7247 808 555

Fax: +49 7247 808 131

Email: helpdesk@fiz-karlsruhe.de

Web: www.fiz-karlsruhe.de