
Structure Searching DWPI Using STN Express®

**PIUG North East Conference
Workshop, October 2008**

© Fachinformationszentrum Karlsruhe, October 2008

Robert Austin
FIZ Karlsruhe Inc
376 Carter Road
Princeton, NJ 08540-7422

Email: robert.austin@fiz-k.com

Web: www.fiz-k.com

Contents

	Page
<u>Workshop Slides</u>	1
Agenda	1
Derwent World Patents Index on STN	2
Overview of DWPI chemical indexing	3
Basics of structure searching	6
Multifile structure searching DCR and CAS databases	13
Basics of fragmentation code searching	19
Editing fragmentation code strategies	29
Multifile searching codes and CAS databases	40
Why use fragmentation codes?	55
Summary	55
User Guides for DWPI, DCR and MARPAT	56
<u>Appendix – Galenical codes (/M6)</u>	59



STN[®]

Structure searching Derwent World Patents
Index[®] (DWPISM) using STN Express[®]

Robert Austin – FIZ Karlsruhe

Agenda

2

- Overview of DWPI chemical indexing
- Basics of structure searching (DCR)
- Multifile searching DCR and CAS databases
- Basics of fragmentation code searching
- Editing fragmentation code strategies
- Multifile searching codes and CAS databases

STN

 FIZ Karlsruhe

Derwent World Patents Index (DWPI)

3

- The largest value-added patent database
 - 17+ million records and 10+ million selected images
 - 36+ million patents from 41 worldwide authorities
- Concise patent families which include
 - Basic & equivalent company names & patent classifications, non-convention equivalents
- Enhanced abstracts and titles
 - Improved relevance and easier scanning of answers
- Patent Assignee Codes
 - For efficient company searching and analysis



Derwent World Patents Index on STN

4

- FILE WPINDEX
 - Open access database
- FILE WPIDS
 - Subscriber database
- FILE WPIX
 - Subscriber database with Extension Abstracts
- FILE LWPI
 - The DWPI learning file
- FILE WPIFV
 - The DWPI First View preview database



Chemical indexing in DWPI

5

- DWPI Chemistry Resource (DCR)
 - Specific compound registry for DWPI
 - Standard STN structure searching
 - Available to all users of DWPI on STN
- Chemical Fragmentation Codes
 - Substances represented by codes
 - Code queries generated by STN Express
 - Available to DWPI Subscribers

STN

 FIZ Karlsruhe

DWPI Chemistry Resource (DCR)

6

- Integrated specific compound database for patent records in DWPI on STN
- Chemical structures and substance data
- Standard STN structure searching
- DCR numbers form the connection to and from DWPI patent records
- Indexing backfile to 1999

STN

 FIZ Karlsruhe

DWPI Chemical Fragmentation Codes

7

- Specific and Markush substance indexing searchable directly within DWPI records
- Substances are represented as separate paragraphs of linked codes in DWPI records
- Each code represents a part of a molecule and each code paragraph represents a substance
- STN Express can be used to generate and upload the appropriate combination of codes
- Indexing backfile to 1963

STN

FIZ Karlsruhe


DWPI on STN provides patent family and chemical substance records

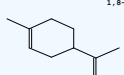
8

Patent families (DWPI)

L1 ANSWER 1 OF 1 WPIINDEX COPYRIGHT 2008 THOMSON REUTERS ON STN
AN 2005-217884 [23] WPIINDEX
TI Recovery of solvent and styrene from polystyrene solution involves recovering solvent by evaporation and recovering styrene from polystyrene thermally decomposed by solvent
DC A13; A35; B14; J01
IS HANG T; KYO T; OZUMA A
PA (YOSH-N) SOSHIBA PLANT KENSHITSU KK
PI JP 2005060471 A 20050310 (200523)* JA 10[2] CORJ0011-12
ADY JP 2005060471 A JP 2003-290004 20030808
PVAL JP 2003-290004 20030808
IPC8 B01D0001-22 [I,A]; B01D0001-22 [I,C]; B01D0003-00 [I,A]; B01D0003-00 [A,B]
AB JP 2005060471 A UPAB: 20050708
HOWEIT - Solvent from a polystyrene solution obtained by dissolving polystyrene in a solvent is evaporated and the solvent is recovered. The solvent thermally decomposes the separated polystyrene and styrene is recovered.
DETAILED DESCRIPTION - An INDEPENDENT CLAIM is included for equipment for recovering solvent and styrene from a polystyrene solution.
USE - Used for recovering solvent and styrene from a polystyrene solution.
ADVANTAGE - The solvent and styrene are recovered efficiently from the polystyrene solution. The styrene monomer of high purity is obtained with high yield.
DESCRIPTION OF DRAWINGS - The figure shows the thermal decomposition portion of the apparatus used for solvent and styrene recovery. (Drawing includes non-English language text).
Storage tank (1)
Transfer pump (2)
Solvent evaporator (3)
Piping (4)
Condenser (5)
TECH ORGANIC CHEMISTRY - Preferred Process: The cracked gas obtained by thermally decomposing polystyrene is condensed. The oil component is distilled and styrene of high purity is recovered.
FS CFI
MC CFI; A04-C02b; A10-E05c; A10-G01A; E10-J02A1; E10-J02B2; E11-Q01A; J01-A01
IV UPIT 20050708
2113-018 2113-PRD; 368-CL 368-PRD

Substances (DCR)

L2 ANSWER 1 OF 2 WPIINDEX COPYRIGHT 2008 THOMSON REUTERS ON STN
ACCESSION NUMBER: DCR-368
DERIVENT CHEM.RES.NO.: 368-0-0-0
PREF. CHEMICAL NAME: STYRENE
SYSTEMATIC NAME: Vinyl-benzene
SYNONYM: POLYSTYRENE (MONOMER); STYRENE

MOLECULAR FORMULA: C8 H8
MOLECULAR WEIGHT: 104.1512
DERIVENT COMPOUND NO.: 368008
DERIVENT REGISTRY NO.: 368

L2 ANSWER 2 OF 2 WPIINDEX COPYRIGHT 2008 THOMSON REUTERS ON STN
ACCESSION NUMBER: DCR-2113
DERIVENT CHEM.RES.NO.: 2113-0-0-0
PREF. CHEMICAL NAME: LIMONENE
SYSTEMATIC NAME: 4-Isopropenyl-1-methyl-cyclohexene
SYNONYM: (+)-LIMONENE; 1,8-P-MENTHADIENE; CALEPOTENE; CINENE; DIPENTENE; DL-LIMONENE; EULIMEN; KAUFSCHEIN; LIMONENE; MENTHADIENE, 1,8-P-; MENTHOLE

MOLECULAR FORMULA: C10 H16
MOLECULAR WEIGHT: 136.239
DERIVENT COMPOUND NO.: R01119
DERIVENT REGISTRY NO.: 1119

STN

FIZ Karlsruhe

DWPI on STN provides patent family and chemical substance records (cont.)

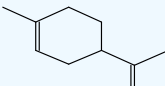
9

L1 ANSWER 1 OF 1 WPINDEX COPYRIGHT 2008 THOMSON REUTERS on STN
 AN 2005-217884 [23] WPINDEX
 TI Recovery of solvent and styrene from polystyrene solution involves recovering solvent by evaporation and recovering styrene from polystyrene thermally decomposed by solvent

 IT UPIT 20050708
 2113-DIS 2113-PRD; 368-CL 368-PRD


DCR numbers form the connection between patent and substance records.

AN.S DCR-2113
 DCSE 2113-0-0-0
 CN.P LIMONENE
 CN.S 4-Isopropenyl-1-methyl-cyclohexene
 SY (+)-LIMONENE; 1,8-P-MENTHADIENE; . . .



MF C10 H16

AN.S DCR-368
 DCSE 368-0-0-0
 CN.P STYRENE
 CN.S Vinyl-benzene
 SY POLYSTYRENE (MONOMER); STYRENE



MF C8 H8

STN FIZ Karlsruhe

Substances are also represented in DWPI by paragraphs of fragmentation codes

10

L1 ANSWER 1 OF 1 WPINDEX COPYRIGHT 2008 THOMSON REUTERS on STN
 AN 2005-217884 [23] WPINDEX
 TI Recovery of solvent and styrene from polystyrene solution involves recovering solvent by evaporation and recovering styrene from polystyrene thermally decomposed by solvent

 IT UPIT 20050708
 2113-DIS 2113-PRD; 368-CL 368-PRD

CMC UPB 20050708
 DRN: 0708-P 0708-U 1119-P 1119-U

M3 *01* G035 G562 H7 H721 M210 M211 M213 M232 M240 M282 M320 M415
 M424 M510 M520 M530 M541 M610 M720 M740 N163 N480 N513 Q431
 M905 M904 M910
 DCN: R01119-K R01119-P
 DCR: 130846-P 130846-U 2113-K 2113-P 2113-U

M3 *02* G010 G100 H7 H715 H721 M210 M212 M240 M281 M320 M414 M424
 M510 M520 M531 M540 M610 M720 M740 N163 N480 N513 Q110 Q431
 M905 M904 M910
 DCN: R00708-K R00708-P
 DCR: 368-K 368-P 368-U

More on this later...!

Limone.

Styrene.

STN FIZ Karlsruhe

Agenda

11

- Overview of DWPI chemical indexing
- **Basics of structure searching (DCR)**
- Multifile searching DCR and CAS databases
- Basics of fragmentation code searching
- Editing fragmentation code strategies
- Multifile searching codes and CAS databases

STN

 FIZ Karlsruhe

All the standard STN structure search options are available for searching DCR

12

- Exact (**EXA**)-Search
 - retrieves specific compounds and isotopes
- Family (**FAM**)-Search
 - retrieves specific compounds, isotopes, salts and mixtures
- Closed Substructure (**CSS**)-Search
 - allows for substitution at defined positions
- Substructure (**SSS**)-Search
 - allows for substitution at any position
- SAMPLE search: free of charge pre-search (default)
- SUBSET search: structure search based on subset
- BATCH search
 - for generic structure searches when system limits are reached

STN

 FIZ Karlsruhe

The basic steps for running a DCR structure search in DWPI on STN 13

1. Draw & save the structure query in standard STN format with STN Express
2. Upload the structure query to DWPI on STN
3. Run the structure search, e.g. **S L1 SSS FULL**
4. Retrieve DWPI patent records, e.g. **S L2 /DCR**
5. Display the patent records in-context with hit structures, e.g. **D L3 1- FULL HITSTR**

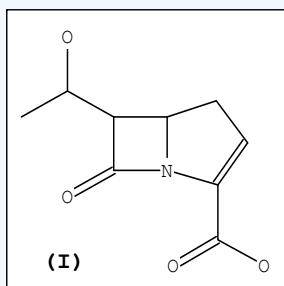
STN

FIZ Karlsruhe

DCR structure search example 14

Search Question:

Search for DWPI patent references to specific carbapenem derivatives of substructure (I)

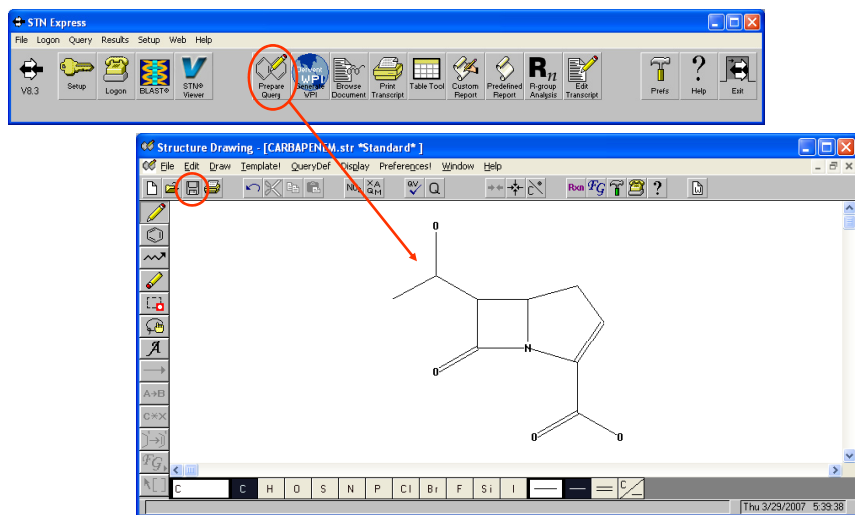


STN

FIZ Karlsruhe

1. Draw & save the structure query in standard STN format using STN Express

15



STN

FIZ Karlsruhe

2. Upload the structure query to DWPI on STN.

16

STN Online and Results - [STN Karlsruhe]

File Edit Online Query Results Preferences W

Upload the query with the 'Q' button.

Structure File:

Look in: Queries

<input checked="" type="checkbox"/> MMA	<input checked="" type="checkbox"/> ring
<input checked="" type="checkbox"/> HYDROXY	<input checked="" type="checkbox"/> CARBAPENEM
<input checked="" type="checkbox"/> BENZENE	<input checked="" type="checkbox"/> test
<input checked="" type="checkbox"/> diazepam	<input checked="" type="checkbox"/> bell2
<input checked="" type="checkbox"/> diazepam wpi	<input checked="" type="checkbox"/> bell1a
<input checked="" type="checkbox"/> test2	<input checked="" type="checkbox"/> bell1

File name: CARBAPENEM

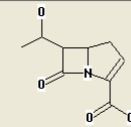
Files of type: Query Files (*.str)

Open Cancel Help

Please make sure you have switched to a file suitable for structure searching.

Modifiable Queries Use Filters

StandardJ



Transcript: WPIINDEX INS Hold On Print Off Online 00:14:03

STN Online and Results - [STN Karlsruhe] 3. Run the structure search. 17

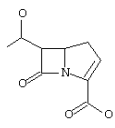
File Edit Online Query Results Preferences! Web Window Help

>>> FOR DETAILS ON THE NEW AND ENHANCED DERWENT WORLD PATENTS INDEX
PLEASE SEE
http://www.stn-international.de/stndatabases/details/dwpi_r.html <<<

=>
Uploading C:\Program Files\STNEXP\Queries\CARBAPENEM.str

L1 STRUCTURE UPLOADED The uploaded structure query (L1).

=> D
L1 HAS NO ANSWERS
L1 STR



Option: display the query (L1), to verify that the Upload was successful.

Structure attributes must be viewed using STN Express query preparation.

=> S L1 FULL Run a full file substructure search using the uploaded query (L1).

FULL SEARCH INITIATED 00:11:28 FILE 'WPINDEX'
FULL SCREEN SEARCH COMPLETED - 999 TO ITERATE

100.0% PROCESSED 999 ITERATIONS 790 ANSWERS
SEARCH TIME: 00.00.01

L2 790 SEA SSS FUL L1 790 compounds are retrieved (L2).

Transcript: WPINDEX INS Hold On Print Off Online 00:18:41

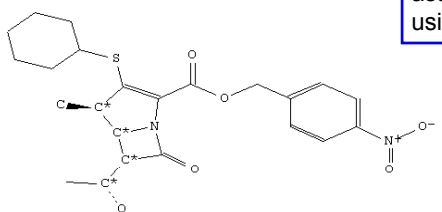
STN Online and Results - [STN Karlsruhe] Option: review answers with D SCAN. 18

File Edit Online Query Results Preferences! Web Window Help

=> D SCAN

L2 790 ANSWERS WPINDEX COPYRIGHT 2007 THE THOMSON CORP on STN

CN.S (4R,5S,6S)-3-Cyclohexylsulfanyl-6-((R)-1-hydroxy-ethyl)-4-methyl-7-oxo-1-aza-bicyclo[3.2.0]hept-2-ene-2-carboxylic acid 4-nitro-benzyl ester
MF C23 H28 N2 O6 S



The effectiveness of the search can be assessed by reviewing some answers, e.g. using the free-of-charge D SCAN format.

HOW MANY MORE ANSWERS DO YOU WISH TO SCAN? (1):2

L2 790 ANSWERS WPINDEX COPYRIGHT 2007 THE THOMSON CORP on STN

CN.S (5S,6R)-6-((R)-1-Hydroxy-ethyl)-7-oxo-3-(2-oxo-2,3-dihydro-1H-indol-5-yl)-1-aza-bicyclo[3.2.0]hept-2-ene-2-carboxylic acid

Transcript: WPINDEX INS Hold On Print Off Online 00:22:37

STN Online and Results - [STN Karlsruhe]

4. Retrieve DWPI patent records.

19

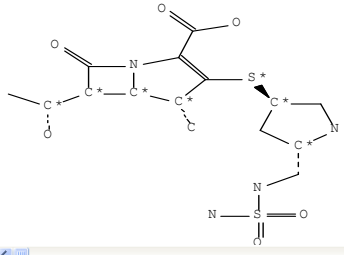
```

=> S L2/DCR
L3 305 L2/DCR
=> D BIB HITSTR 3
L3 ANSWER 3 OF 305 WPIINDEX COPYRIGHT 2007
AN 2007-184686 [18] WPIINDEX Full-text
DNC C2007-065698 [18]
TI New pyrrolidine derivative useful as intermediate in preparation of
   doripenem
DC B02; B03
IN LI B; LIU Z; TANG Y; XU D; YU Y; ZHOU W
PA (CHEN-N) CHENGDU DI'AO JIUHONG PHARM FACTORY
CYC 113
PIA WO 2007009354 A1 20070125 (200718)* ZH 21[0]
ADT WO 2007009354 A1 WO 2006-CN1664 20060713
PRAI CN 2005-10021270 20050715
AN.S DCR-111946
CN.P DORIPENEM
SDCN RASIXI

```

305 DWPI patent family records are retrieved (L3).

5. Display the patent records in context with hit structures (HITSTR).



The Hit Structure (HITSTR) display is useful for reviewing in-context results following a chemical structure search.

Transcript: WPIINDEX INS Hold On Print Off Online 00:29:01

Option: DCR searches can also be refined using a series of DCR Number Roles

20

- DCR Number Roles help describe the context of an indexed compound within the patent, e.g.
 - compound is prepared, purified or part of a mixture
 - compound is claimed or from the examples
- Use the (T)-Term proximity operator to refine DCR searches with DCR Number Roles
 - For example: => S L2 /DCR (T) PRD/DCR
 - Where PRD = the Produced (synthesized) Role

See **HELP ROLES** in DWPI on STN (file WPIINDEX/WPIDS/WPIX) for more information on effective searching using DCR Number Roles.

Example: refine the search with Roles

21

```
=> S L2/DCR(T) PRD/DCR  
L4      80 L2/DCR(T) PRD/DCR
```

PRD is the DCR Role for produced (synthesized).

```
=> D TI PA PN IT HITSTR
```

```
L4 ANSWER 1 OF 80 WPINDEX COPYRIGHT 2007 THE THOMSON CORP on STN  
TI Preparation of dendritic drug for treating disease involves chemically  
protecting reactive group in therapeutic drug; deprotecting and  
reacting the group formed with linker group
```

```
PA (UNMI-C) UNIV MICHIGAN CENT
```

```
PI WO 2007087256 A2 20070802 (200
```

```
US 20070190151 A1 20070816 (200
```

```
IT UPIT 20070907
```

```
. . . 86604-RCT; 108231-CL 108231-PRD 108231-RCT; 103208-CL 103208-PRD  
103208-RCT; 102303-CL 102303-PRD 102
```

```
AN.S DCR-103208
```

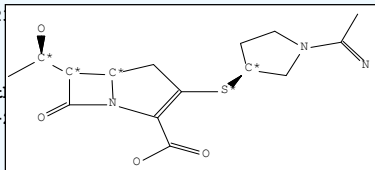
```
CN.P PANIPENEM
```

```
CN.S 6-(1-Hydroxy-ethyl)-3-[1-(1-imino-ethyl-  
oxo-1-aza-bi cyclo[3.2.0]hept-2-ene-
```

```
SDCN RA00WB
```

```
. . . .
```

The Hit Structure (HITSTR) display is useful for reviewing in-context results following a chemical structure search.



STN

FIZ Karlsruhe

DCR display formats

22

- D SCAN* Names, formula & structure (random)
- D TRIAL* Names, formula & structure
- D STD Number, Names, formula & structure
- D ALL STD + RIN, DCN and DRN numbers
- D MAX ALL + DDRN, CT and SS data
- D ISTD STD indented with text labels
- D IALL ALL indented with text labels
- D IMAX MAX indented with text labels

* Free-of-charge display formats in WPINDEX/WPIDS/WPIX.

STN

FIZ Karlsruhe

DWPI Chemistry Resource (DCR) coverage

23

- Specific chemical substances indexed by Thomson Reuters Analysts for patents in DWPI
- DWPI patents classified in Pharmaceutical (B), agrochemical (C) and/or general chemical (E)
- Comprehensive coverage began in 4/1999*
- Selective coverage for approximately
 - 20,000 substances from 1/1987 to date
 - 2,100 substances from 7/1981 to date

* Except Japanese patents which are covered from 9/2000 onwards.

STN

 FIZ Karlsruhe

Which compound types are included in DCR?

24

- Organic and inorganic specific compounds
- Specific peptides up to 30 peptide residues
- Oligomers up to 8 repeat units
- Natural products including proteins
- 'Standard' polymers (for chemical indexing)
- Chemically modified polysaccharides

STN

 FIZ Karlsruhe

Agenda

25

- Overview of DWPI chemical indexing
- Basics of structure searching (DCR)
- **Multifile searching DCR and CAS databases**
- Basics of fragmentation code searching
- Editing fragmentation code strategies
- Multifile searching codes and CAS databases

STN

 FIZ Karlsruhe

Structure searchable STN patent databases

26

File	Search Features	File Features
REGISTRY	Names MF Structures	Specific substances classified with a CAS RN
CAplus SM	CAS RN Substance class terms	Literature indexed to CAS RN's
MARPAT [®]	Structures	Substances described generically by a Markush structure
USPATFULL	CAS RN	US chemical patents indexed to CAS RN's
WPINDEX/ WPIDS/WPIX	Derwent fragment coding DCR - structures DCR - names DCR - MF DCR - substance class terms DRN	A merged structure and fragment code search database

STN

 FIZ Karlsruhe

Multifile structure searching using DWPI/DCR and CAS databases

27

1. Prepare a suitable standard structure query for CAS REGISTRYSM/MARPAT[®] and DCR
2. Run the search in REGISTRY/HCAPLUS and display records
3. Run the search in MARPAT, remove duplicates between HCAPLUS and MARPAT and display additional MARPAT records
4. Run the search in DCR/DWPI
5. Remove duplicates between HCAPLUS/MARPAT and DWPI, display additional DWPI records

STN

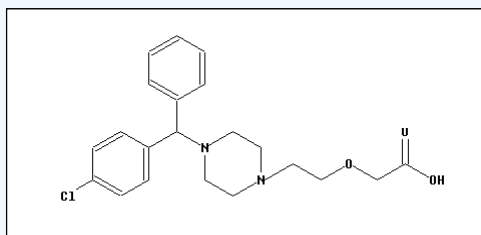
FIZ Karlsruhe

Multifile structure search example: DCR/DWPI and CAS databases

28

Search Question:

Search for all patent references of cetirizine, including salts and mixtures.



STN

FIZ Karlsruhe

Multifile search: HCAPLUS

31

```
=> FIL HCAPLUS
=> S L3 AND P/DT
      1097 L3
      5720347 P/DT
L4      379 L3 AND P/DT

=> D BIB ABS HITSTR 1-

L4 ANSWER 1 OF 379 HCAPLUS COPYRIGHT 2008 ACS on STN
AN 2005:963795 HCAPLUS
DN 143:222555
ED Entered STN: 02 Sep 2005
TI Treatment of behavioral disorders
IN Melamed, Isaac
PA USA
FAN.CNT 1
PATENT NO.          KIND  DATE          APPLICATION NO.      DATE
-----
PI US2005192290      A1  20050901  2005US-0036182      20050113
PRAI 2004US-536458P  P   20040113
AB The invention relates to a method for treating a behavior disorder
comprising the administration of a therapeutically effective amount
of antihistamine, such as ceterizine, fexofenadine; loratadine, and
desloratadine. The behavioral disorders may include ADHD, . . .
```

Crossover to HCAPLUS
retrieves 379 patent
(P/DT) records (L4).

Display HCAPLUS records in,
e.g. BIB ABS HITSTR format.

STN

FIZ Karlsruhe

Multifile search: MARPAT

32

```
=> FIL MARPAT
=> S L3 CSS SAM
FULL FILE PROJECTIONS: ONLINE **COMPLETE**
                        BATCH **COMPLETE**
PROJECTED ITERATIONS: 24139 TO 28221
PROJECTED ANSWERS:    0 TO 0

L5      0 SEA CSS SAM L1

=> S L3 CSS FULL
L6      16 SEA CSS FUL L1

=> S L6 NOT L4
      76 L4
L7      5 L6 NOT L4

=> D BIB ABS FQHIT 1-

L7 ANSWER 1 OF 5 MARPAT COPYRIGHT 2008
AN 133:281798 MARPAT
TI Preparation of diphenylmethylpiperazinyhydroxyureas and
related compounds for treatment of asthma, allergy and . . .
```

Use REGISTRY answer set
(L3) for MARPAT search.

In MARPAT there is no family search
option, we use CSS instead.

The MARPAT search retrieves 5
additional patent records (L7).

Display MARPAT answers with,
e.g. the BIB ABS FQHIT format.

STN

FIZ Karlsruhe

Multifile search: DCR/DWPI

33

=> **FIL WPINDEX**

=> **S L1 FAM SAM**

PROJECTED ITERATIONS: 2 TO 62
PROJECTED ANSWERS: 1 TO 40

L8 1 SEA FAM SAM L1

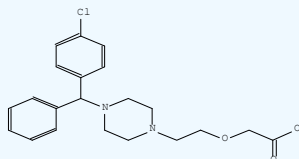
=> **D TRIAL**

L8 ANSWER 1 OF 1 WPINDEX COPYRIGHT 2008 THOMSON REUTERS on STN

AN.S DCR-1000746

CN.S (2-(4-[(4-Chloro-phenyl)-phenyl-methyl]-piperazin-1-yl)-ethoxy)-acetic acid; monohydrochloride

MF C21 H25 Cl N2 O3 . H Cl



=> **S L1 FAM FULL**

L9 10 SEA FAM FUL L1

Repeat the family structure search in DWPI.

The full-file structure search retrieves **10** DCR-records (L9).

STN

FIZ Karlsruhe

Multifile search: DCR/DWPI

34

=> **S L9/DCR**

L10 273 L11/DCR

10 DCR-records (L9) retrieve 273 patent records (L10).

=> **D TRIAL HITSTR**

L10 ANSWER 1 OF 273 WPINDEX COPYRIGHT 2008

THOMSON REUTERS on STN

AN 2007-291662 [28] WPINDEX

CR 2001-138233; 2001-367536; 2001-602548; 2006-27

DNC C2007-106907 [28]

DNN N2007-214309 [28]

TT TT: TREAT RHINITIS COMPRISE INFUSION AGENT GAS VARIOUS FACE ORIFICE

DC B05; B07; P35

IPCI A62B0007-00 [I,A]; A62B0007-00 [I,C]

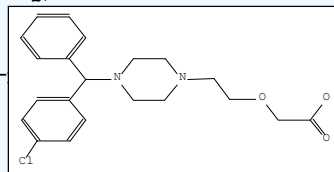
MC CPI: B01-B03; B05-C03; B06-H; B07-D05; B07-D09; B07-D11; B12-M12D; B12-M12G; B12-M12H; B12-M12N; B12-M12Q; B14-N04

AN.S DCR-90453

CN.P CETIRIZINE

CN.S (2-(4-[(4-Chloro-phenyl)-phenyl-methyl]-ethoxy)-acetic acid

SDCN R14937; R16291



Use the free format **TRIAL HITSTR** for a fast relevance check.

STN

FIZ Karlsruhe

Multifile-search: duplicate removal

35

```
=> DUP IDE L4 L7
L11          384 DUP IDE L4 L7
```

The command **DUP IDE** is useful to merge the HCAPLUS (L4) and MARPAT (L7) answer sets (L11).

```
=> FILE WPINDEX
```

Enter FILE WPINDEX again.

```
=> TRANSFER L11 PN 1-
L12          TRANSFER L11 1- PN
L13          452 L12
```

Use the **TRANSFER** command to transfer all patent numbers from MARPAT/HCAPLUS to DWPI.

```
=> S L10 NOT L13
L14          44 L10 NOT L13
```

The DCR search has retrieved 44 additional inventions (L14).

STN

FIZ Karlsruhe

Multifile-search: DCR/DWPI

36

```
=> D FULL HITSTR 1-
```

```
L14 ANSWER 1 OF 44 WPINDEX COPYRIGHT 20
AN 2005-356035 [36] WPINDEX
CR 2005-356033 [36]; 2005-356034 [36]; 2
DNC C2005-110122
```

Display the additional records from the DWPI search in full using the **FULL HITSTR** format.

```
TI Substance for improving oral absorption e.g. gastrointestinal tract
absorption, transdermal or subcutaneous absorption of drug, comprises
complex containing drug moiety and transport moiety.
```

```
DC B05 B07
```

```
IN GUITTARD, G V; WONG, P S L; YAN, D
```

```
PA (GUIT-I) GUITTARD G V; (WONG-I) WONG P S L; (YAND-I) YAN D; . . . .
```

```
PI WO2005041925 A2 20050512 (200536)* EN 92 A61K-009-00
```

```
US2005165102 A1 20050728 (200550) A61K-031-185
```

```
. . . . .
```

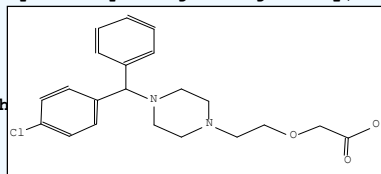
```
AB NOVELTY - A substance comprises a complex comprising a drug moiety,
and a transport moiety. . . . .
```

```
AN.S DCR-140237
```

```
CN.P CETIRIZINE HYDROCHLORIDE
```

```
CN.S (2-{4-[(4-Chloro-phenyl)-phenyl-meth
acid; dihydrochloride
```

```
SDCN RA1JYL; RA46F
```



STN

FIZ Karlsruhe

Agenda

37

- Overview of DWPI chemical indexing
- Basics of structure searching (DCR)
- Multifile searching DCR and CAS databases
- **Basics of fragmentation code searching**
- Editing fragmentation code strategies
- Multifile searching codes and CAS databases

STN

 FIZ Karlsruhe

Chemical indexing in DWPI

38

- DWPI Chemistry Resource (DCR)
 - Specific compound registry for DWPI
 - Standard STN structure searching
 - Available to all users of DWPI on STN
- **Chemical Fragmentation Codes**
 - Substances represented by codes
 - Code queries generated by STN Express
 - Available to DWPI Subscribers

STN

 FIZ Karlsruhe

DWPI Chemical Fragmentation Codes

39

- Specific and Markush substance indexing searchable directly within DWPI records
- Substances are represented as separate paragraphs of linked codes in DWPI records
- Each code represents a part of a molecule and each code paragraph represents a substance
- STN Express can be used to generate and upload the appropriate combination of codes
- Indexing backfile to 1963



A brief history of Derwent World Patents Index Chemical Fragmentation Codes

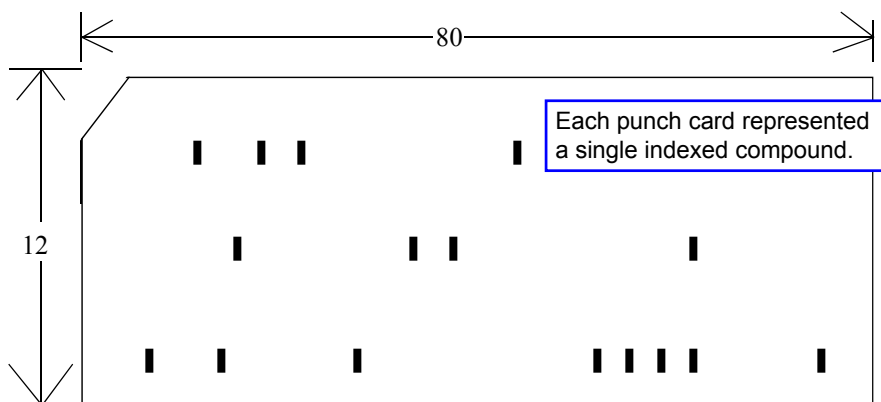
40

- 1963 Pharmaceuticals (B)
- 1965 Agricultural Chemicals (C)
- 1970 **Code revision** + General chemistry (E)
- 1972 **Code revision** + Ring Index Numbers
- 1981 **Code revision** + DRNs
- 1987 Markush graphical indexing + DCNs
- 1992 Autogeneration of codes
- 1999 Derwent Chemistry Resource



Fragmentation codes were once positions on an IBM punch card

41



12 ROWS X 80 COLUMNS - TOTAL POSITIONS=960

STN

FIZ Karlsruhe

Fragmentation codes are grouped into paragraphs with (P)-proximity

42

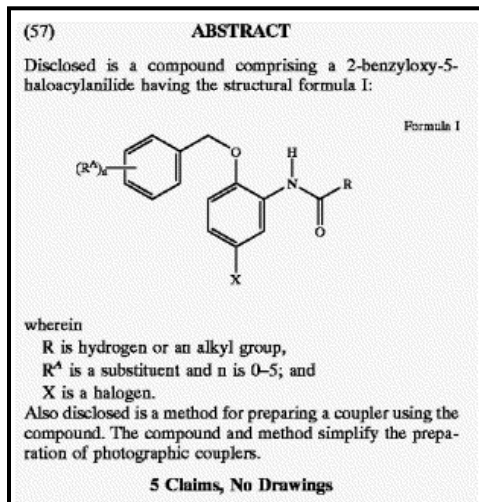
AN	2005-217884 [23]	WPIX	
TI	Recovery of solvent and styrene from polystyrene solution involves recovering solvent by evaporation and recovering styrene from polystyrene thermally decomposed by solvent.		
IT	UPIT 20050708		
	2113-DIS 2113-PRD; 368-CL 368-PRD		
CMC	UPB 20050708		
	DRN: 0708-P 0708-U 1119-P 1119-U		
	DCR: 2113-P 2113-U 368-P 368-U		
M3	*01*	G035 G562 H7 H721 M210 M211 M213 M232 M240 M282 M320 M415 M424 M510 M520 M530 M541 M610 M720 M740 N163 N480 N513 Q431 M905 M904 M910	
(P)		DCN: R01119-K R01119-P	Limone
		DCR: 2113-K 2113-P 2113-U	
M3	*02*	G010 G100 H7 H715 H721 M210 M212 M240 M281 M320 M414 M424 M510 M520 M531 M540 M610 M720 M740 N163 N480 N513 Q110 Q431 M905 M904 M910	
(P)		DCN: R00708-K R00708-P	Styrene
		DCR: 368-K 368-P 368-U	

STN

FIZ Karlsruhe

Generic representations in patents are called Markush structures

43



STN

FIZ Karlsruhe

Specific and Markush compounds are indexed with Fragmentation codes

44

AN 2005-141457 [15] WPIX

TI Inhibiting polymerization of styrene in the dehydrogenation section of a styrene production system involves contacting a quaternary ammonium salt polymerization inhibitor with the dehydrogenation section of an ethylbenzene dehydrogenation section

CMC UPB 20060121
 DRN: 0707-S 0707-U 0708-P 0708-U
 DCR: 368-P 368-U 576-S 576-U

M3 *01* G010 G100 H7 H715 H721 M210 M212 M240 M281 M320 M414 M510 M520
 M531 M540 M610 M720 N209 N224 N312 N450 Q110 M905 M904 M910
 DCN: R00708-K R00708-P
 DCR: 368-K 368-P 368-U

M3 *02*
 M3 *03*

M3 *04* G001 G002 G010 G011 G012 G013 G015 G017 G020 G021 G022 G023 G024 G025 G026 G027 G028 G029 G030 G031 G032 G033 G034 G035 G036 G037 G038 G039 G040 G041 G042 G043 G044 G045 G046 G047 G048 G049 G050 G051 G052 G053 G054 G055 G056 G057 G058 G059 G060 G061 G062 G063 G064 G065 G066 G067 G068 G069 G070 G071 G072 G073 G074 G075 G076 G077 G078 G079 G080 G081 G082 G083 G084 G085 G086 G087 G088 G089 G090 G091 G092 G093 G094 G095 G096 G097 G098 G099 G100 G101 G102 G103 G104 G105 G106 G107 G108 G109 G110 G111 G112 G113 G114 G115 G116 G117 G118 G119 G120 G121 G122 G123 G124 G125 G126 G127 G128 G129 G130 G131 G132 G133 G134 G135 G136 G137 G138 G139 G140 G141 G142 G143 G144 G145 G146 G147 G148 G149 G150 G151 G152 G153 G154 G155 G156 G157 G158 G159 G160 G161 G162 G163 G164 G165 G166 G167 G168 G169 G170 G171 G172 G173 G174 G175 G176 G177 G178 G179 G180 G181 G182 G183 G184 G185 G186 G187 G188 G189 G190 G191 G192 G193 G194 G195 G196 G197 G198 G199 G200 G201 G202 G203 G204 G205 G206 G207 G208 G209 G210 G211 G212 G213 G214 G215 G216 G217 G218 G219 G220 G221 G222 G223 G224 G225 G226 G227 G228 G229 G230 G231 G232 G233 G234 G235 G236 G237 G238 G239 G240 G241 G242 G243 G244 G245 G246 G247 G248 G249 G250 G251 G252 G253 G254 G255 G256 G257 G258 G259 G260 G261 G262 G263 G264 G265 G266 G267 G268 G269 G270 G271 G272 G273 G274 G275 G276 G277 G278 G279 G280 G281 G282 G283 G284 G285 G286 G287 G288 G289 G290 G291 G292 G293 G294 G295 G296 G297 G298 G299 G300 G301 G302 G303 G304 G305 G306 G307 G308 G309 G310 G311 G312 G313 G314 G315 G316 G317 G318 G319 G320 G321 G322 G323 G324 G325 G326 G327 G328 G329 G330 G331 G332 G333 G334 G335 G336 G337 G338 G339 G340 G341 G342 G343 G344 G345 G346 G347 G348 G349 G350 G351 G352 G353 G354 G355 G356 G357 G358 G359 G360 G361 G362 G363 G364 G365 G366 G367 G368 G369 G370 G371 G372 G373 G374 G375 G376 G377 G378 G379 G380 G381 G382 G383 G384 G385 G386 G387 G388 G389 G390 G391 G392 G393 G394 G395 G396 G397 G398 G399 G400 G401 G402 G403 G404 G405 G406 G407 G408 G409 G410 G411 G412 G413 G414 G415 G416 G417 G418 G419 G420 G421 G422 G423 G424 G425 G426 G427 G428 G429 G430 G431 G432 G433 G434 G435 G436 G437 G438 G439 G440 G441 G442 G443 G444 G445 G446 G447 G448 G449 G450 G451 G452 G453 G454 G455 G456 G457 G458 G459 G460 G461 G462 G463 G464 G465 G466 G467 G468 G469 G470 G471 G472 G473 G474 G475 G476 G477 G478 G479 G480 G481 G482 G483 G484 G485 G486 G487 G488 G489 G490 G491 G492 G493 G494 G495 G496 G497 G498 G499 G500 G501 G502 G503 G504 G505 G506 G507 G508 G509 G510 G511 G512 G513 G514 G515 G516 G517 G518 G519 G520 G521 G522 G523 G524 G525 G526 G527 G528 G529 G530 G531 G532 G533 G534 G535 G536 G537 G538 G539 G540 G541 G542 G543 G544 G545 G546 G547 G548 G549 G550 G551 G552 G553 G554 G555 G556 G557 G558 G559 G560 G561 G562 G563 G564 G565 G566 G567 G568 G569 G570 G571 G572 G573 G574 G575 G576 G577 G578 G579 G580 G581 G582 G583 G584 G585 G586 G587 G588 G589 G590 G591 G592 G593 G594 G595 G596 G597 G598 G599 G600 G601 G602 G603 G604 G605 G606 G607 G608 G609 G610 G611 G612 G613 G614 G615 G616 G617 G618 G619 G620 G621 G622 G623 G624 G625 G626 G627 G628 G629 G630 G631 G632 G633 G634 G635 G636 G637 G638 G639 G640 G641 G642 G643 G644 G645 G646 G647 G648 G649 G650 G651 G652 G653 G654 G655 G656 G657 G658 G659 G660 G661 G662 G663 G664 G665 G666 G667 G668 G669 G670 G671 G672 G673 G674 G675 G676 G677 G678 G679 G680 G681 G682 G683 G684 G685 G686 G687 G688 G689 G690 G691 G692 G693 G694 G695 G696 G697 G698 G699 G700 G701 G702 G703 G704 G705 G706 G707 G708 G709 G710 G711 G712 G713 G714 G715 G716 G717 G718 G719 G720 G721 G722 G723 G724 G725 G726 G727 G728 G729 G730 G731 G732 G733 G734 G735 G736 G737 G738 G739 G740 G741 G742 G743 G744 G745 G746 G747 G748 G749 G750 G751 G752 G753 G754 G755 G756 G757 G758 G759 G760 G761 G762 G763 G764 G765 G766 G767 G768 G769 G770 G771 G772 G773 G774 G775 G776 G777 G778 G779 G780 G781 G782 G783 G784 G785 G786 G787 G788 G789 G790 G791 G792 G793 G794 G795 G796 G797 G798 G799 G800 G801 G802 G803 G804 G805 G806 G807 G808 G809 G810 G811 G812 G813 G814 G815 G816 G817 G818 G819 G820 G821 G822 G823 G824 G825 G826 G827 G828 G829 G830 G831 G832 G833 G834 G835 G836 G837 G838 G839 G840 G841 G842 G843 G844 G845 G846 G847 G848 G849 G850 G851 G852 G853 G854 G855 G856 G857 G858 G859 G860 G861 G862 G863 G864 G865 G866 G867 G868 G869 G870 G871 G872 G873 G874 G875 G876 G877 G878 G879 G880 G881 G882 G883 G884 G885 G886 G887 G888 G889 G890 G891 G892 G893 G894 G895 G896 G897 G898 G899 G900 G901 G902 G903 G904 G905 G906 G907 G908 G909 G910 G911 G912 G913 G914 G915 G916 G917 G918 G919 G920 G921 G922 G923 G924 G925 G926 G927 G928 G929 G930 G931 G932 G933 G934 G935 G936 G937 G938 G939 G940 G941 G942 G943 G944 G945 G946 G947 G948 G949 G950 G951 G952 G953 G954 G955 G956 G957 G958 G959 G960 G961 G962 G963 G964 G965 G966 G967 G968 G969 G970 G971 G972 G973 G974 G975 G976 G977 G978 G979 G980 G981 G982 G983 G984 G985 G986 G987 G988 G989 G990 G991 G992 G993 G994 G995 G996 G997 G998 G999

M3 *05*

AN.S DCR-368
 DCSE 368-0-0-0
 CN.P STYRENE
 CN.S Vinyl-benzene
 SY POLYSTYRENE (MONOMER); STYRENE

MF C8 H8
 SDRN 0708
 SDCN R00708

DCR record.

Specific compound.

Markush compound

All Markush options, are "over-coded" into a single paragraph.

STN

FIZ Karlsruhe

Basics of fragmentation code searching

45

1. Draw & save the structure query in *WPI format*
2. Select Query and *Generate WPI strategy* from the main STN Express window
3. Select the fields to be searched, e.g. /M0,M2, and the strategy is generated and saved
4. Logon to STN, access WPIDS or WPIX, then select Query and Run Command File
5. Select the command file – this runs line-by-line automatically – and review the results

STN

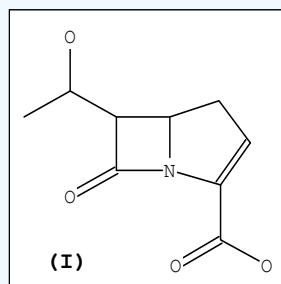
FIZ Karlsruhe

Fragmentation code search example

46

Search Question:

Search for DWPI patent references to carbapenem structure (I) using chemical fragmentation codes.

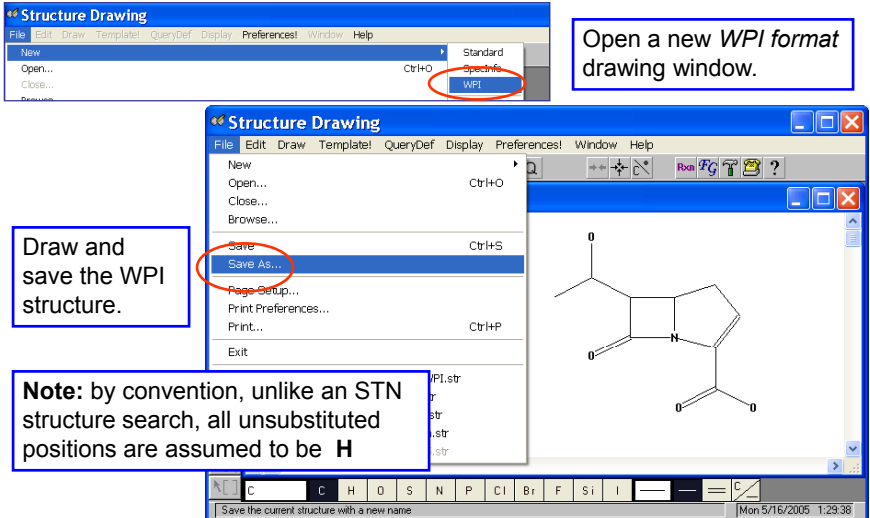


STN

FIZ Karlsruhe

Open, Draw and save the query structure in "WPI" format

47



Open a new *WPI format* drawing window.

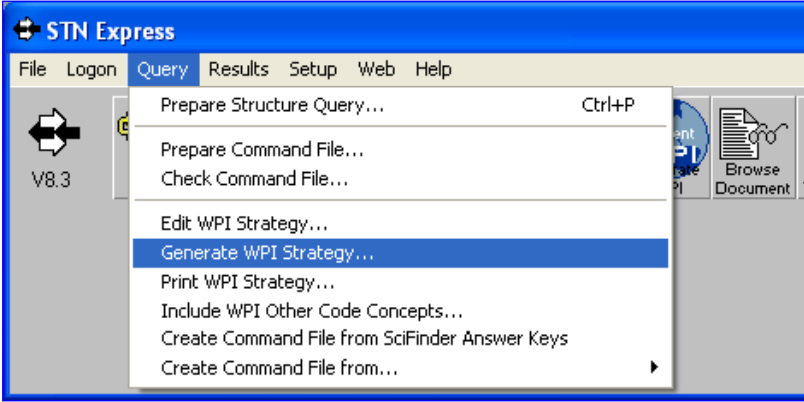
Draw and save the WPI structure.

Note: by convention, unlike an STN structure search, all unsubstituted positions are assumed to be **H**

STN **FIZ Karlsruhe**

Use STN Express to generate the fragmentation code script

48



STN Express

File Logon Query Results Setup Web Help

- Prepare Structure Query... Ctrl+P
- Prepare Command File...
- Check Command File...
- Edit WPI Strategy...
- Generate WPI Strategy...**
- Print WPI Strategy...
- Include WPI Other Code Concepts...
- Create Command File from SciFinder Answer Keys
- Create Command File from...

STN **FIZ Karlsruhe**

Select the search fields (subheadings) and generate the fragmentation codes

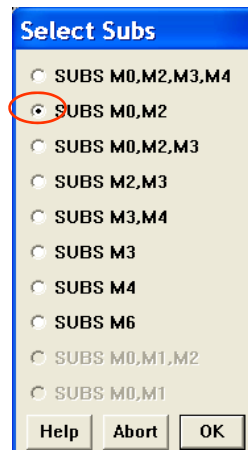
49

STN Express search field options:

/M0 Pre-1970 Pharma/agrochem
 /M2 1970-date Pharma/agrochem
 /M3 1970-date Other chemicals (excl. M4)
 /M4 1970-date Dyes & pigments

Other fragmentation code fields

/M1 1970-date Natural products
 /M5 1963-1999 Steroids
 /M6 1976-date Galenicals/formulations

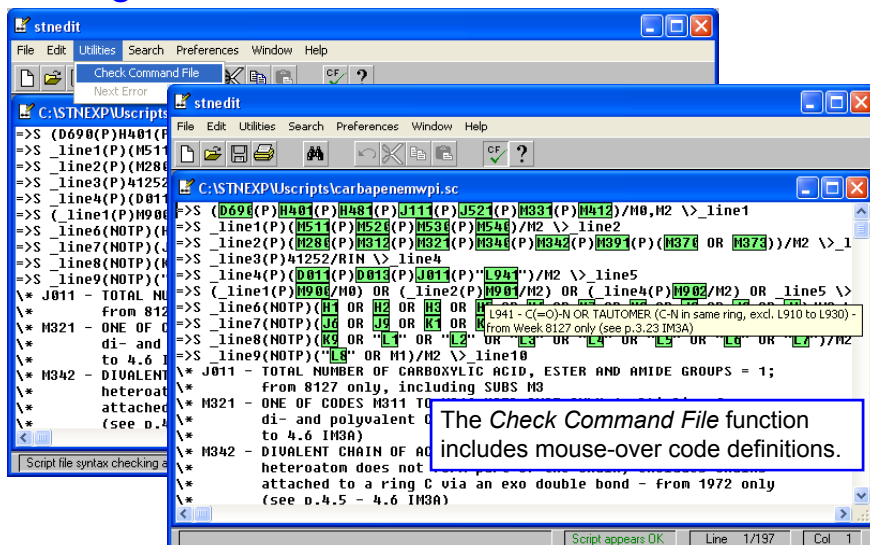


STN

FIZ Karlsruhe

DWPI fragmentation code strategy is generated in Command File format

50

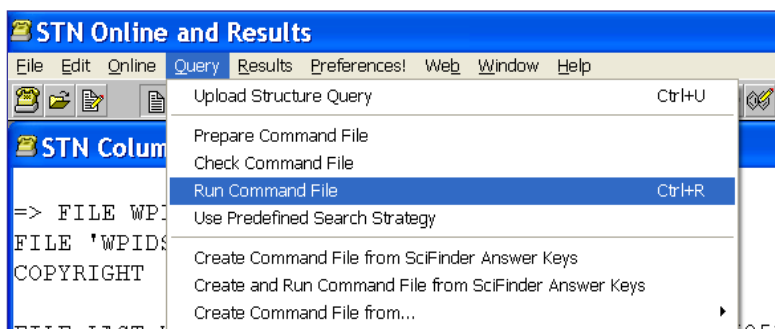


STN

FIZ Karlsruhe

Use *Run Command File* to execute the fragmentation code search online

51



STN

FIZ Karlsruhe

The *Command File* code query runs automatically line-by-line

52

```
=> FILE WPIDS
=> SET POSTINGS OFF
. . .
=> S (D690 (P)H401 (P)H481 (P)J111 (P)J521 (P)M331 (P)M412) /M0 ,M2
L1      1611 (D690 (P)H401 (P)H481 (P)J111 (P)J521 (P)M331 (P)M412) /M0 ,M2

=> S L1 (P) (M511 (P)M520 (P)M530 (P)M540) /M2
L2      955 L1 (P) (M511 (P)M520 (P)M530 (P)M540) /M2

=> S L2 (P) (M280 (P)M312 (P)M321 (P)M340 (P)M342 (P)M391 . . .) /M2
L3      628 L2 (P) (M280 (P)M312 (P)M321 (P)M340 (P)M342 (P)M391 . . .) /M2

=> S L3 (P) 41252/RIN
L4      224 L3 (P) 41252/RIN
. . .
=> S L9 (NOTP) ("L8" OR M1) /M2
L10     91 L9 (NOTP) ("L8" OR M1) /M2
```

STN Express links and groups codes together into correctly formatted queries.

Note: a relatively small answer set (L10) was retrieved because this is **not** a substructure search.

STN

FIZ Karlsruhe

Review answers retrieved

53

=> D AN TI 1-

```
L10 ANSWER 1 OF 91 WPIDS COPYRIGHT 2008 THOMSON REUTERS on STN
AN 2007-135316 [14] WPIDS
TI Manufacture of nitrogen-containing bicyclo compound for manufacturing
beta-lactam type compound, involves reacting specific silyl ether
compound or its salt with complex of titanium tetrachloride and nitro
compound, in solvent
. . .
L10 ANSWER 3 OF 91 WPIDS COPYRIGHT 2008 THOMSON REUTERS on STN
AN 2006-078961 [08] WPIDS
TI New beta-methyl carbapenem compounds useful for preventing or treating
an infection by gram negative bacteria or by a drug resistant
bacterial strain
. . .
L10 ANSWER 84 OF 91 WPIDS COPYRIGHT 2008 THOMSON REUTERS on STN
AN 1978-36645A [21] WPIDS
TI 1-Carba-2-penam-3-carboxylic acid derivs. - useful as broad spectrum
antibacterials for human or veterinary use.
. . .
```

STN

FIZ Karlsruhe

Fragmentation Code Searches can be refined with Role codes

54

=> S L10(P)M720/M2

L11 26 L20(P)M720/M2

M720 is the Role for synthesized/produced.

=> D AN TI HITCMC

```
L11 ANSWER 1 OF 26 WPIDS COPYRIG
AN 2007-135316 [14] WPIDS
TI Manufacture of nitrogen-containing bicyclo compound for manufacturing
beta-lactam type compound, involves reacting specific silyl ether
compound or its salt with complex of titanium tetrachloride and nitro
compound, in solvent
CMC UPB 20070227
M2 *01* C316 D011 D012 D013 D014 D016 D019 D690 D790 D800 E670 E680
E690 F011 F012 F013 F014 F015 F113 F410 F423 F710 H121 H201
H211 H4 H401 H481 H592 H716 H721 H8 J0 J011 J012 J111 J211
J311 J5 J521 J522 J592 K352 K399 L463 L9 L922 L941 L943 L999
M116 M126 M142 M210 M211 M212 M213 M214 M215 M216 M220 M221
M222 M223 M224 M225 M226 M231 M232 M233 M240 M271 M272 M273
M280 M281 M282 M312 M321 M331 M340 M342 M373 M391 M412 M511
M512 M520 M521 M522 M530 M540 M720 N282 N362 N511 Q431 M905
M904
RIN: 00862 00932 41252 45743
MCN: 0343-43401-K 0343-43401-P
```

The HITCMC format is often helpful for displaying hit fragmentation paragraphs.

This hit is a Markush compound.

STN

FIZ Karlsruhe

Fragmentation code roles

55

<u>Role*</u>	<u>Definition</u>
M710	New compound
M720	Produced/synthesized
M730	Used in synthesis
M740	Apparatus
M750	Detected/removed
M760	Medium (e.g. diagnostics)
M781	Use of one compound
M782	Use of >1 compound

* Role codes are **red** - i.e. introduced in 1970

STN

 FIZ Karlsruhe

Display formats for reviewing chemical fragmentation codes in DWPI

56

D CODE	All codes, indexing and classes
D CMC	All chemical codes, M0-M6
D M2, etc	Pharma/Agrochem only, etc
D RIN	Ring Index Numbers only
D HITCMC	The hit fragmentation code paragraph(s) found in a fragmentation code search
D FRAGHITSTR	The DCR hit structure (HITSTR) for specific compounds retrieved in a fragmentation code search

STN

 FIZ Karlsruhe

Agenda

57

- Overview of DWPI chemical indexing
- Basics of structure searching (DCR)
- Multifile searching DCR and CAS databases
- Basics of fragmentation code searching
- **Editing fragmentation code strategies**
- Multifile searching codes and CAS databases

STN

 FIZ Karlsruhe

Editing fragmentation code strategies for substructure searches

58

- STN Express incorporates the TOPFRAG program to generate fragmentation code queries from a WPI-format structure
- TOPFRAG always attempts to generate the most specific (narrow) query possible
- As such, to perform an effective code-based substructure search it is wise to learn how to edit TOPFRAG strategies

STN

 FIZ Karlsruhe

Editing fragmentation code strategies for substructure searches (cont.)

59

- The TOPFRAG software also typically cannot handle queries with complex variable groups or several free sites available for substitution
- In addition the following codes can not be generated by STN Express and, if required, need to be added by manually to the query
 - Sugar stereochemistry codes L8:
 - Steroid codes Parts S, T, U
 - Parts N, P, Q, R, V, W
 - Part M codes after M620



Important concepts for effective editing of code strategies




60

- Search fields (/M2, etc) – slide 49
- Fragmentation Code Sheet
- Colored historical time periods
- Standard search strategy format
- Time range codes
- Negation codes
- Ring Index Numbers (RINs)
- Discontinued fragmentation codes (#)



Fragmentation code sheet example

61

F2: S-SOLE HETERO.	F200 Thiirene* Thiete	F211 	F212 	F213 
---------------------------------	------------------------------------	--	---	--

- Part F - mononuclear heterocyclics
- Shaded boxes describe the code section
 - Set F2 - sulphur sole heteroatom
- Unshaded boxes show the searchable codes
 - F200 thiirene
 - F211 thiophene
 - F212 dihydrothiophene
 - F213 tetrahydrothiophene

STN

 FIZ Karlsruhe

Fragmentation code sheet includes colored time periods

62

BLACK CODES FROM 1963

RED CODES FROM 1970

BLUE CODES FROM 1972

GREEN CODES FROM DW 198127

STN

 FIZ Karlsruhe

STN Express code strategies use standard search strategy format

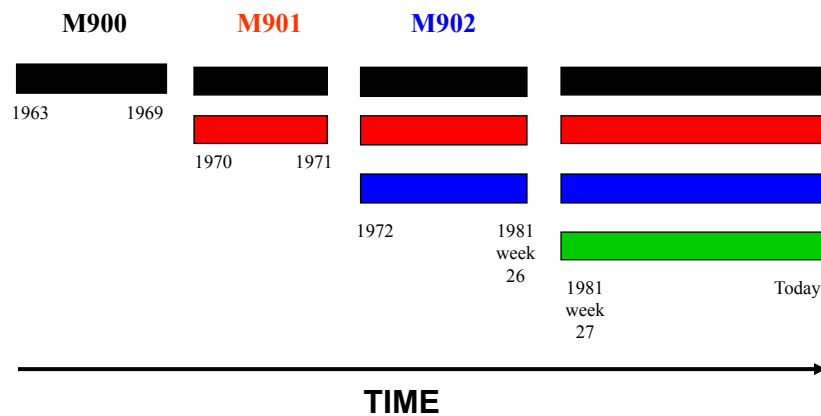
63

- L1: **BLACK CODES (P) TOGETHER**
- L2: L1 (P) **RED CODES**
- L3: L2 (P) **BLUE CODES**
- L4: L3 (P) **RING INDEX NUMBER (RIN)**
- L5: L4 (P) **WITH GREEN CODES**
- L6: (L1 (P) **M900**) OR (L2 (P) **M901**)
OR (L4 (P) **M902**) OR **L5**



STN Express code strategies incorporate time range codes

64



Time range codes

65

M900	Pre-1970
M901	1970 – 1971
M902	1972 – Update 198126
M903*	Update 198127 to 1999/2000
M905*	1999/2000 to date (Steroid STU and natural product V codes discontinued)
M904*	MMS indexing available (1987-date)

(* note that these codes are not used in a standard search strategy)



Negation codes

66

- Negation codes are listed as the last few lines of an STN Express code strategy
- They have a 2-digit format, e.g. H1 Amine
- And are used to “not out” unwanted functionalities with the (NOTP) operator
- Negation codes should typically be deleted from an STN Express code strategy for an effective substructure search



Ring Index Numbers (RINs)

67

- Derived from the Patterson Ring Index with later number ranges added by Thomson Reuters
- 5 digit code numerical codes - /RIN
- STN Express code strategies include RINs for non-specific ring codes, e.g. D130*
- RINs (and ring codes) only retrieve the defined ring, i.e. they do NOT retrieve fused system hits within which the defined ring is a smaller part

Note: if the possibility of further ring fusion on to the specified ring is desired, all potential alternate ring codes and/or RINs of interest need to be identified and included in the search query.

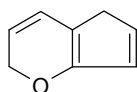
STN

FIZ Karlsruhe

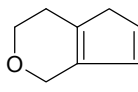
Ring Index Numbers improve search precision for non-specific ring codes

68

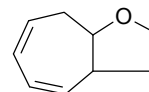
- All the ring systems below are searched with code D130*



RIN = 01382



RIN = 01384



RIN = 01431

- Ring Index Numbers define the rings more precisely

Note: Ring Index Numbers are also used to index spiro systems and their individual non-spiro parts.

STN

FIZ Karlsruhe

Discontinued fragmentation codes are not present on the code sheet

69

- Some generic codes have been discontinued and replaced by more specific codes
 - e.g. L140 (black code), was replaced by codes L141 - L146 (green codes) in 1981
- New codes that replaced an older generic codes are indicated on the sheet with # e.g. L142
- Although discontinued codes are not present on the fragmentation code sheet STN Express includes them on the correct line of the strategy

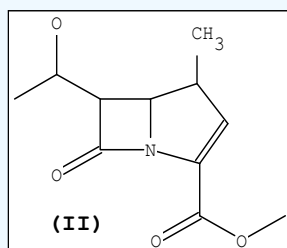
STN

FIZ Karlsruhe

Fragmentation code editing example

70

Challenge: prepare a comprehensive DWPI fragmentation code strategy for carbapenem derivatives of substructure (II)



STN

FIZ Karlsruhe

General tips for editing code strategies for substructure searches

71

1. Remove (or edit) negation codes, e.g. H1
2. Include higher basic group codes, e.g. M411
3. Include higher ring substitution counts for ring systems in D01/D02 and/or G01/G02/G03
4. Include higher count codes for common functional groups present in the query, e.g. J111, J112
5. Can hydroxyl become ether/ester? Can amine become amide? If so, OR in or delete these codes
6. Edit/remove carbon chain codes*, e.g. M280, M320
7. Include higher ring systems count code options, e.g. M511, M512, etc
8. Include additional ring linkage M1 code options, e.g. M131 (>C=O), M132 (other carbon), etc

(* Carbon chain codes are probably the most confusing aspect of effective editing. One option, especially when first learning how to edit fragmentation code queries, is to completely delete M2/M3 codes from the query.)



Generate the fragmentation code command file and edit. . .

72

A standard fragmentation code query was generated following the steps on slides 45-54.

```
=>S (D690 (P)H401 (P)H481 (P)J211 (P)J521 (P)M240 (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)M210 (P)M281 (P)M312 (P)M321 (P)M340 (P)M342 (P)M391)/M2,M3,M4 \>_line3
=>S _line3 ((M270 OR M272) (P) (M370 OR M373))/M2,M3,M4 \>_line4
=>S _line4 (P)41252/RIN \>_line5
=>S _line5 (P)D013 (P)D019 (P)J011 (P)"L941" (P)M211)/M2,M3,M4 \>_line6
=>S (_line1 (P)M900/M0) OR (_line2 (P)M901/M2,M3,M4) OR (_line5 (P)M902/M2,M3,M4) \>_line7
=>S _line7 OR _line6 \>_line8
=>S _line8 (NOTP) (H1 OR H2 OR H3 OR H5 OR H6 OR H7 OR H9 OR J1 OR J3 OR J4)/M2,M3,M4 \>_line9
=>S _line9 (NOTP) (J6 OR J9 OR K1 OR K2 OR K3 OR K4 OR K5 OR K6 OR K7 OR K8)/M2,M3,M4 \>_line10
=>S _line10 (NOTP) (K9 OR "L1" OR "L2" OR "L3" OR "L4" OR "L5" OR "L6")/M2,M3,M4 \>_line11
=>S _line11 (NOTP) ("L7" OR "L8" OR M1)/M2,M3,M4 \>_line12
```

For this example we will be searching all areas of chemical indexing, M0 – M4.

Note: Line1 = Black; Line2 = Red; Line3-5 = Blue; Line6 = Green



Step 1: remove negation codes

73

```
=>S (D690 (P)H401 (P)H481 (P)J211 (P)J521 (P)M240 (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) (M210 (P)M281 (P)M312 (P)M321 (P)M340 (P)M342 (P)M391)/M2,M3,M4 \>_line3
=>S _line3 (P) ((M270 OR M272) (P) (M370 OR M373))/M2,M3,M4 \>_line4
=>S _line4 (P) 41252/RIN \>_line5
=>S _line5 (P) (D013 (P)D019 (P)J011 (P)"L941" (P)M211)/M2,M3,M4 \>_line6
=>S (_line1 (P)M900/M0) OR (_line2 (P)M901/M2,M3,M4) OR (_line5 (P)M902/M2,M3,M4) \>_line7
=>S _line7 OR _line6 \>_line8

=>S _line8 (NOTP) (H1 OR H2 OR H3 OR H5 OR H6 OR H7 OR H9 OR J1 OR J3 OR J4)/M2,M3,M4 \>_line9
=>S _line9 (NOTP) (J6 OR J9 OR K1 OR K2 OR K3 OR K4 OR K5 OR K6 OR K7 OR K8)/M2,M3,M4 \>_line10
=>S _line10 (NOTP) (K9 OR "L1" OR "L2" OR "L3" OR "L4" OR "L5" OR "L6")/M2,M3,M4 \>_line11
=>S _line11 (NOTP) ("L7" OR "L8" OR M1)/M2,M3,M4 \>_line12
```

Negation codes are listed as the last few lines of an STN Express code strategy. They have a 2-digit format, e.g. H1 Amine.

STN

FIZ Karlsruhe

Steps 2-8: editing the codes

74

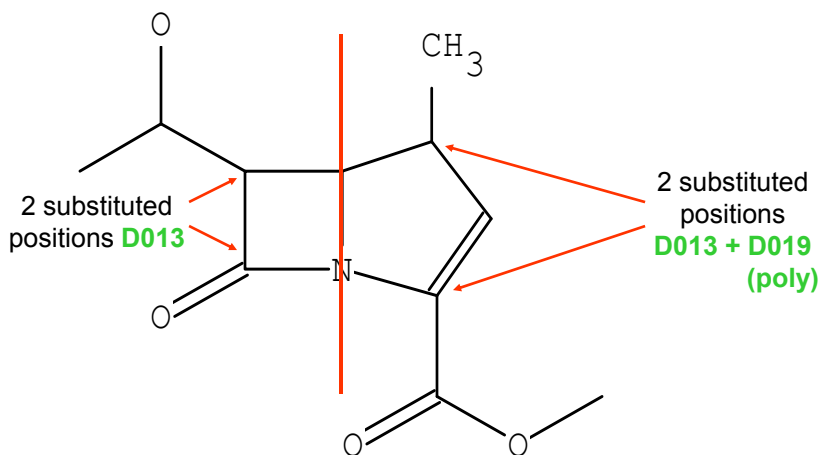
2. M412 OR-ed with M411 (e.g. to include salts)
3. 2-substitutions D013, OR-ed with 3, D014
4. 1-ester J211, OR-ed with >1 J212
1-oxo J521, OR-ed with >1 J522, J523
1-carboxy derivative J011 deleted
5. Hydroxyl H401/H481 deleted
6. Carbon chain (M2/M3) options OR-ed in
7. 1-ring M511, OR-ed with >1 M512, M513
No-ring M520, M530, M540 deleted
8. There are no M1 codes in this query

STN

FIZ Karlsruhe

Fused-heterocyclic ring substitution codes

75

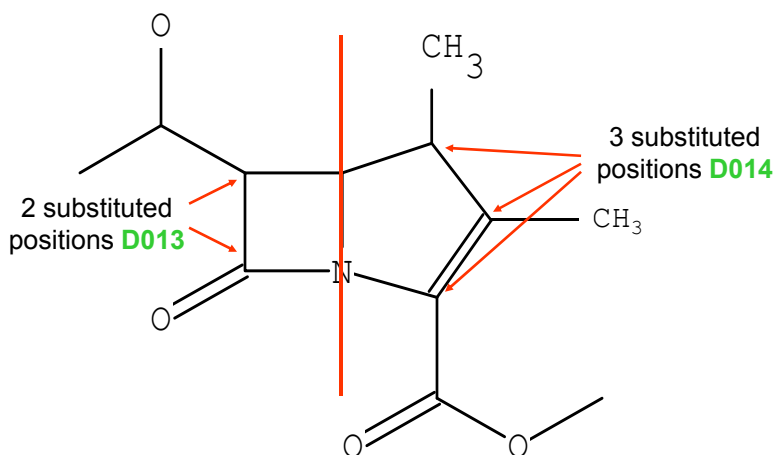


STN

FIZ Karlsruhe

Fused-heterocyclic ring substitution codes

76



STN

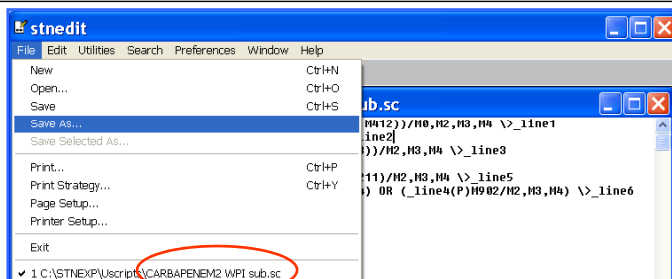
FIZ Karlsruhe

Save the edited strategy. . .

77

This is the edited STN command file.

```
=>S (D690 (P) J21! (P) J52! (P) M240 (P) (M411 OR M412)) /M0, M2, M3, M4 \> _line1
=>S _line1 (P) (M511 OR M512 OR M513) /M2, M3, M4 \> _line2
=>S _line2 (P) (M210 (P) (M281 OR M282 OR M283) (P) (M321 OR M322 OR M323)) /M2, M3, M4 \> _line3
=>S _line3 (P) 41252 /RIN \> _line4
=>S _line4 (P) (D013 (P) (D019 OR D014) (P) "L941" (P) M211) /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
```



STN

FIZ Karlsruhe

Notes on M2/M3 carbon chain code choices

78

Originally selected by STN Express TOPFRAG (see slide 72):

BLACK: M240 M331
BLUE: M210 M281 M312 M321 M340 M342 M391 (M270 OR M272#) (M370 OR M373#)
GREEN: M211

Choices for the final strategy (as given in the final query on slide 77):

BLACK: M240
BLUE: M210 (M281 OR M282 OR M283) (M321 OR M322 OR M323)
GREEN: M211

Notes on choices for this query:

1. Polyvalent 2-carbon carbon chain top left of the structure. Assumption: hydroxyl does not become a ketone. Deleted M331 M391 M342 M340 M370 M373, since we do not know what the free positions for substitution will become. However, there will always be at least one polyvalent carbon chain of some kind, so M321 is retained and ORed in with M322 and M323.
2. Ring bound methyl group top right of structure. Assumption: it is closed for substitution, so M210, M211 and M240 are retained. As such, there will always be at least one monovalent ring bound carbon chain, so M281 is retained and ORed in with M282 and M283.
3. Ester-oxygen bound methyl group bottom right of structure. Assumption: it is open for substitution. M270 M273 are deleted since we do not know what the free positions for substitution will become.
4. It might be a little better to OR M211 with M210 on the BLUE line rather than narrow M210 (BLUE) with M211 (GREEN/BLUE). However the query has been left as it is in this instance.

STN

FIZ Karlsruhe

Agenda

79

- Overview of DWPI chemical indexing
- Basics of structure searching (DCR)
- Multifile searching DCR and CAS databases
- Basics of fragmentation code searching
- Editing fragmentation code strategies
- Multifile searching codes and CAS databases

STN

 FIZ Karlsruhe

Steps for comprehensive structure searching. . .

80

1. Build a query for REGISTRY/MARPAT/DWPI
2. Upload structure in one database, e.g. REGISTRY
3. Run the query in each database as appropriate
4. Use DUP REMOVE to remove duplication or DUP IDENTIFY to merge bibliographic results
5. Use FSORT & D PFAM to display by invention
6. Generate & edit the code query and run in DWPI
7. TRANSFER patent numbers from step 4 to DWPI, then identify and review unique answers

STN

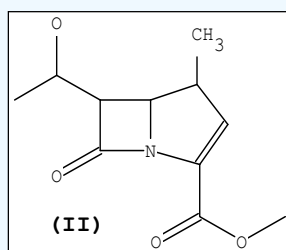
 FIZ Karlsruhe

Comprehensive multifile chemical patent search example

81

Search Question:

Find patent references for carbapenem derivatives of substructure (II) using all relevant CAS and DWPI tools



STN

FIZ Karlsruhe

Prepare and save a standard format structure query

82

Structure Drawing

File Edit Draw Template QueryDef Display Preferences Window Help

CARBAPENEM2.str *Standard*

Match Level:

- Class
- Atom
- Any
- Mixed

Element Count Level:

- Limited

OK Cancel

For this query, these carbon atoms have been changed to ring/chain nodes.

In MARPAT using the default *class match* will retrieve Alkyl (Ak) as well as a methyl group.

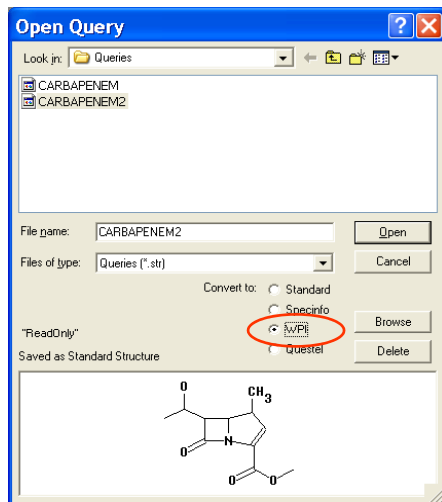
Wed 5/18/2005 11:04:25

STN

FIZ Karlsruhe

A standard format structure can be converted to a *WPI format* structure

83



A standard STN structure query can easily be converted to *WPI format*, for generating a fragmentation code query.

However, the code query has already been prepared for this example – see slide 77.

STN

FIZ Karlsruhe

Upload the structure query to STN

84

=> FILE REGISTRY

=>

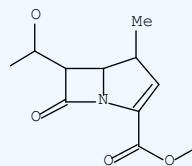
Uploading C:\Documents and Settings\Robert Austin\
My Documents\STN Express 8.3\Queries\carbapenem2.str

L1 STRUCTURE UPLOADED

=> D

L1 HAS NO ANSWERS

L1 STR



Option: display the query (L1), to verify that the Upload was successful.

STN

FIZ Karlsruhe

Run a sample search in REGISTRY

85

=> FILE REGISTRY

Run a free-of-charge **SAMPLE** search to confirm the query will run OK.

=> S L1 SSS SAM

SAMPLE SEARCH INITIATED 18:59:05 FILE 'REGISTRY'

SAMPLE SCREEN SEARCH COMPLETED - 375 TO ITERATE

100.0% PROCESSED 375 ITERATIONS 50 ANSWERS

SEARCH TIME: 00.00.01

FULL FILE PROJECTIONS: ONLINE **COMPLETE**

BATCH **COMPLETE**

PROJECTED ITERATIONS: 6339 TO 8661

PROJECTED ANSWERS: 4688 TO 6712

L2 50 SEA SSS SAM L1

STN

FIZ Karlsruhe

Review REGISTRY answers

86

=> D SCAN

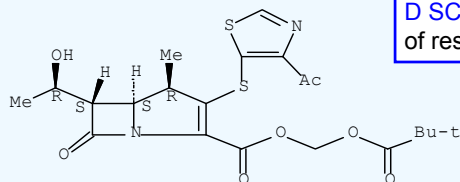
L2 50 ANSWERS REGISTRY COPYRIGHT 2008 ACS on STN

IN 1-Azabicyclo[3.2.0]hept-2-ene-2-carboxylic acid, 3-[(4-acetyl-5-thiazolyl)thio]-6-[(1R)-1-hydroxyethyl]-4-methyl-7-oxo-, (2,2-dimethyl-1-oxopropoxy)methyl ester, (4R,5S,6S)-(9CI)

MF C21 H26 N2 O7 S2

Absolute stereochemistry.

Review **SAMPLE** search answers with **D SCAN** to confirm the anticipated type of results are retrieved by the query.



HOW MANY MORE ANSWERS DO YOU WISH TO SCAN? (1): 10 . . .

STN

FIZ Karlsruhe

Run the full-file REGISTRY search and retrieve the HCAPLUS patent references

87

```
=> S L1 SSS FULL
FULL SEARCH INITIATED 19:10:28 FILE 'REGISTRY'
FULL SCREEN SEARCH COMPLETED -      7262 TO ITERATE

100.0% PROCESSED      7262 ITERATIONS      5352 ANSWERS
SEARCH TIME: 00.00.01

L3          5352 SEA SSS FUL L1

=> FILE HCAPLUS

=> S L3 AND P/DT
L4          448 L3 AND P/DT
```

5352 REGISTRY substance records (L3) are retrieved.

Crossover from L3 retrieves 448 HCAPLUS patent references (L4).

Tip: Limit to patents with P/DT in HCAPLUS.

STN

FIZ Karlsruhe

Run a sample search in DCR

88

```
=> FILE WPIDS

=> S L1 SSS SAM
SAMPLE SEARCH INITIATED 19:11:07 FILE 'WPIDS'
SAMPLE SCREEN SEARCH COMPLETED -      19 TO ITERATE

100.0% PROCESSED      19 ITERATIONS      15 ANSWERS
SEARCH TIME: 00.00.01

FULL FILE PROJECTIONS:  ONLINE  **COMPLETE**
                        BATCH   **COMPLETE**
PROJECTED ITERATIONS:   60 TO   320
PROJECTED ANSWERS:     34 TO   266

L5          15 SEA SSS SAM L1
```

Option: run a free-of-charge **SAMPLE** search to confirm the query will run OK.

STN

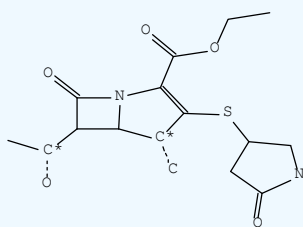
FIZ Karlsruhe

Review DCR answers

89

=> D SCAN

L5 15 ANSWERS WPIDS COPYRIGHT 2008 THOMSON REUTERS on STN
CN.S 6-(1-Hydroxy-ethyl)-4-methyl-7-oxo-3-(5-oxo-pyrrolidin-3-ylsulfanyl)-1-aza-bicyclo[3.2.0]hept-2-ene-2-carboxylic acid ethyl ester
MF C16 H22 N2 O5 S



Review **SAMPLE** search answers with **D SCAN** to confirm the anticipated type of results are retrieved by the query.

HOW MANY MORE ANSWERS DO YOU WISH TO SCAN? (1): 10 . . .

STN

FIZ Karlsruhe

Run the full-file DCR search and retrieve the DWPI patent references

90

=> S L1 SSS FULL

FULL SEARCH INITIATED 19:13:18 FILE 'WPIDS'
FULL SCREEN SEARCH COMPLETED - 186 TO ITERATE

100.0% PROCESSED 186 ITERATIONS 155 ANSWERS
SEARCH TIME: 00.00.01

L6 155 SEA SSS FUL L1

155 DCR substance records (L6) are retrieved.

=> S L6/DCR

L7 48 L6/DCR

Crossover of DCR numbers from L6 retrieves 48 DWPI patent references (L7).

STN

FIZ Karlsruhe

Run a sample search in MARPAT

91

=> S L1 SSS SAM

Run a free-of-charge **SAMPLE** search to confirm the query will run OK.

SAMPLE SEARCH INITIATED 19:15:55 FILE 'MARPAT'

SAMPLE SCREEN SEARCH COMPLETED - 337 TO ITERATE

100.0% PROCESSED 337 ITERATIONS 15 ANSWERS
SEARCH TIME: 00.00.01

FULL FILE PROJECTIONS: ONLINE ****COMPLETE****
BATCH ****COMPLETE****
PROJECTED ITERATIONS: 5647 TO 7833
PROJECTED ANSWERS: 67 TO 533

L8 15 SEA SSS SAM L1

STN

FIZ Karlsruhe

Review MARPAT answers

92

=> D SCAN

L8 15 ANSWERS MARPAT COPYRIGHT 2008 ACS on STN

IC ICM C07D0477-00

ICS C07D0207-16; C07F0009-568; C07D0403-12; C07C0229-64;
C07C0229-62; C07C0229-56; C07C0233-54; C07C0309-61;
A61K0031-40

CC 26-5 (Biomolecules and The
Section cross-reference(s)

TI Preparation of [(carboxy
carbapenems as antibiotics

ST carbapenem carboxyphenylcarbamoylepyrrolidinylthio prepn
antibiotic

IT Antibiotics

(carbapenem, [(carboxyphenyl) carbamoylepyrrolidinylthio]-)

.

Review **SAMPLE** search answers with **D SCAN** to confirm the anticipated type of results are retrieved by the query.

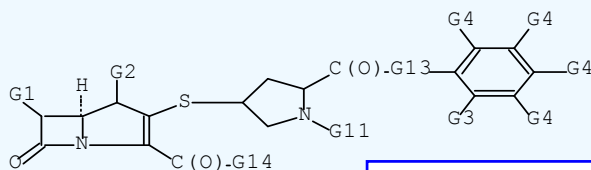
STN

FIZ Karlsruhe

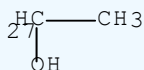
Review MARPAT answers (cont.)

93

MSTR 2A



G1 = 27



Review **SAMPLE** search answers with **D SCAN** to confirm the anticipated type of results are retrieved by the query.

STN

FIZ Karlsruhe

Run the full-file search in MARPAT

94

=> S L3 SSS FULL

Use the REGISTRY answer set (L3) to get a reduced price for the MARPAT search.

THE ESTIMATED SEARCH COST FOR FILE 'MARPAT' IS 64.50 U.S. DOLLARS
DO YOU WANT TO CONTINUE WITH THIS REQUEST? (Y)/N or END: Y

FULL SEARCH INITIATED 20:04:53 FILE 'MARPAT'
FULL SCREEN SEARCH COMPLETED - 6964 TO ITERATE

100.0% PROCESSED 6964 ITERATIONS 322 ANSWERS
SEARCH TIME: 00.00.04

L9 322 SEA SSS FUL L1

322 MARPAT patent records are retrieved (L9).

Tip: use the **SET NOTICE** feature to be warned about search or display costs before commands are sent to STN.

STN

FIZ Karlsruhe

Specify the preferred database, remove duplicates and sort by patent family

95

```

=> SET DUPORDER FILE
SET COMMAND COMPLETED
=> DUP REM L4 L7 L9
. . . . .
L10          529 DUP REM L4 L7 L9 (289 DUPLICATES REMOVED)
              ANSWERS '1-448' FROM FILE HCAPLUS
              ANSWERS '449-455' FROM FILE WPIDS
              ANSWERS '456-529' FROM FILE MARPAT

=> FSORT L10
. . . . .
L11          529 FSO L10
              33 Multi-record Families   Answers 1-69
                Family 1                 Answers 1-2
                Family 2                 Answers 3-4
                . . . . .
                Family 32                Answers 66-67
                Family 33                Answers 68-69
              460 Individual Records     Answers 70-529
              0 Non-patent Records
    
```

L4 = HCAPLUS
L7 = WPIDS
L9 = MARPAT

289 duplicate references are removed (L10).

The remaining records belong to 493 (33 + 460) patent families (L11).



Create a custom default format for each database *before* the multifile search

96

Database	Suggested custom format
HCAPLUS	IBIB ABS HITSTR
DWPI	IBIB ABS IT HITSTR
MARPAT	IBIB ABS FQHIT

Note: custom format names must start with a dot, for example: `.BIBABSFQHIT`



Create the custom format and set it to be the default in that database, e.g. MARPAT

97

```

=> FILE MARPAT
=> SET FORMAT
ENTER FORMAT NAME OR (?): .BIBABSFQHIT
ENTER FORMAT DEFINITION OR (?): IBIB ABS FQHIT
SET COMMAND COMPLETED

=> SET DFORMAT .BIBABSFQHIT
SET COMMAND COMPLETED

=> FILE MARPAT
FILE 'MARPAT' ENTERED AT 16:33:00
COPYRIGHT (C) 2008 American Chemical Society
    
```

1. Use SET FORMAT to define a preferred custom format for each database.

2. Use SET DFORMAT to make the new custom display format the file default.

Note: STN reminds you what the new custom file default format is each time you enter the database.

' .BIBABSFQHIT ' IS DEFAULT FORMAT FOR 'MARPAT' FILE

STN

FIZ Karlsruhe

An HCAPLUS answer with HITSTR

98

```

=> D PFAM=1-
L11 ANSWER 1 OF 529 HCAPLUS COPY
ACCESSION NUMBER: 2003:68591
DOCUMENT NUMBER: 138:137088
TITLE: Preparation of sulfamides and pyrrolidines as
intermediates for pyrrolidylthiocarbapenem
antibiotic
INVENTOR(S): Nishino, Yutaka; Yuasa, Tetsuya; Komurasaki,
Tadashi; Kakinuma, Makoto; Masui, Toshiaki;
Kobayashi, Makoto
PATENT ASSIGNEE(S): Shionogi and Co., Ltd., Japan
SOURCE: Japan Kokai Tokkyo Koho, 36 pp.
CODEN: JKXXAF
LANGUAGE: Japanese
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:
    
```

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
JP 2003026680	A	20030129	JP 2002-129301	20020430
JP 2008120830	A	20080529	JP 2008-29734	20080208
JP 2008184466	A	20080814	JP 2008-29733	20080208
PRIORITY APPLN. INFO.:			JP 2001-140782	A 20010510
			JP 2002-129301	A3 20020430

This displays the first member of each multi-record family, and all the *individual records* in the *file default* display format.

STN

FIZ Karlsruhe

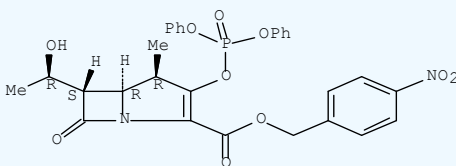
An HCAPLUS answer with HITSTR (cont.)

99

IT 90776-59-3
RL: RCT (Reactant); RACT (Reactant or reagent)
(preparation of sulfamides and pyrrolidines as
intermediates for pyrrolidylthiocarbapenem antibiotic)

RN 90776-59-3 HCAPLUS
CN 1-Azabicyclo[3.2.0]hept-2-ene-2-carboxylic acid,
3-[(diphenoxyphosphinyl)oxy]-6-[(1R)-1-hydroxyethyl]-4-
methyl-7-oxo-,
(4-nitrophenyl)methyl ester, (4R,5R,6S)- (CA INDEX NAME)

Absolute stereochemistry. Rotation (+).

**STN**

FIZ Karlsruhe

A unique DWPI answer with HITSTR

100

L11 ANSWER 467 OF 529 WPIDS COPYRIGHT 2008 THOMSON REUTERS on STN
ACCESSION NUMBER: 2006-468056 [48] WPIDS Full-text
DOC. NO. CPI: C2006-147053 [48]
TITLE: Manufacture of (beta)-lactam compound useful as
antimicrobial agent, involves cyclizing hydroxyl
group protected compound in presence of strong base
and reacting with chlorinated diphenyl phosphoric
acid
DERWENT CLASS: B02
INVENTOR: KOGA T; NISHINO K
PATENT ASSIGNEE: (KANF-C) KANEKA CORP
COUNTRY COUNT: 1

PATENT INFO ABBR.:

PATENT NO	KIND	DATE	WEEK	LA	PG	MAIN	IPC
JP 2006176418	A	20060706	(200648)*	JA	40	[0]	

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
JP 2006176418	A	JP 2004-368886	20041221

PRIORITY APPLN. INFO: JP 2004-368886 20041221

STN

FIZ Karlsruhe

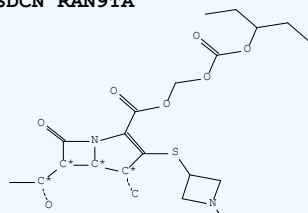
A unique DWPI answer with HITSTR (cont.) 101

.
 IT UPIT 20060727
 0330-10701-CL 0330-10701-NEW 0330-10701-PRD; 0330-10702-CL
 0330-10702-NEW 0330-10702-PRD; 0330-10703-CL 0330-10703-NEW
 0330-10703-PRD; 0330-10704-CL 0330-10704-PRD; **1332401-EX**
1332401-PRD; 11791-CL 11791-RCT; 77-CL 77-RGT 77-USE; 78-CL
 78-RGT 78-USE

AN.S DCR-1332401

CN.S (4R,5S,6R)-3-[1-(4,5-Dihydro-thiazol-2-yl)-azetidin-3-ylsulfanyl]-6-((R)-1-hydroxy-ethyl)-4-methyl-7-oxo-1-azabicyclo[3.2.0]hept-2-ene-2-carboxylic acid 1-ethyl-propoxycarbonyloxymethyl ester

SDCN RAN9TA



STN

FIZ Karlsruhe

A MARPAT answer with FQHIT 102

L11 ANSWER 510 OF 529 MARPAT COPYRIGHT 2008 ACS on STN
 ACCESSION NUMBER: 117:212221 MARPAT Full-text
 TITLE: Preparation of 2-(9-fluorenonyl)carbapenems as antibacterial agents
 INVENTOR(S): Cama, Lovji D.; Greenlee, Mark L.; Dininno, Frank P.; Heck, James V.
 PATENT ASSIGNEE(S): Merck and Co., Inc., USA
 SOURCE: Eur. Pat. Appl., 91 pp.
 CODEN: EPXXDW
 DOCUMENT TYPE: Patent
 LANGUAGE: English
 FAMILY ACC. NUM. COUNT: 2
 PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
EP 469844	A1	19920205	EP 1991-306954	19910730
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE				
US 5025007	A	19910618	US 1990-561541	19900801
US 5144028	A	19920901	US 1990-594809	19901009
PRIORITY APPLN. INFO.:			US 1990-561541	19900801
			US 1990-594809	19901009

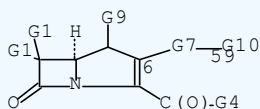
STN

FIZ Karlsruhe

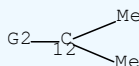
A MARPAT answer with FQHIT (cont.)

103

MSTR 1A



G1 = 12



G2 = OH

G4 = 28

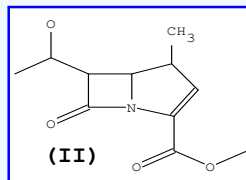


G5 = CH₂Ph (opt. substd.)

G9 = Me

Patent location: claim 1

Recall the query (slide 81).



STN

FIZ Karlsruhe

Recall the *edited* fragmentation code query from slide 77. . .

104

=> FILE WPIDS

Run the command file in WPIDS (or WPIX).

=> S (D690 (P) J21! (P) J52! (P) M240 (P) M331 (P) . . .) /M0, M2, M3, M4
L12 2203 (D690 (P) J21! (P) J52! (P) M240 (P) (M411 OR M412)) /M0, M2, M3, M4

=> S L12 (P) (M511 OR M512 OR M513) /M2, M3, M4
L13 2179 L12 (P) (M511 OR M512 OR M513) /M2, M3,

The Command File runs automatically line-by-line

=> S L13 (P) (M210 (P) (M281 OR M282 OR M283) (P) . . .) /M2, M3, M4
L14 2131 L13 (P) (M210 (P) (M281 OR M282 OR M283) (P) . . .) /M2, M3, M4

=> S L14 (P) 41252 /RIN
L15 554 L14 (P) 41252 /RIN

STN Express links and groups codes together into correctly formatted queries.

=> S L15 (P) (D013 (P) (D019 OR D014) (P) "L941" (P) M211) /M2, M3, M4
L16 506 L15 (P) (D013 (P) (D019 OR D014) (P) "L941" (P) M211) /M2, M3, M4

=> S (L12 (P) M900 /M0) OR (L13 (P) M901 /M2, M3, M4) OR
L17 105 (L12 (P) M900 /M0) OR (L13 (P) M901 /M2, M3, M4) OR

=> S L17 OR L16
L18 582 L17 OR L16

This is a true *substructure* search (L18).

STN

FIZ Karlsruhe

Review fragmentation code answers

105

=> D AN TI HITCMC FRAGHITSTR 60

L18 ANSWER 60 OF 582 WPIDS COPYRIGHT 2008 THOMSON REUTERS on STN

AN 2007-572301 [55] WPIDS

TI Novel carbapenem compound useful as intermediate for producing
2-substituted mercapto-1-beta-methyl carbapenem antibiotic

CMC UPB 20070827

M2 *01* B615 B701 B713 B720 B815 B831 D013 D014 D030 D690 H4 H401
H481 H6 H602 H609 H686 H689 H7 H716 H721 H8 J0 J011 J2 J211
J5 J521 L9 L941 M210 M211 M213 M231 M240 M272 M281 M312
M323 M331 M332 M340 M342 M344 M362 M373 M391 M392 M411
M511 M520 M530 M540 M710 M905 M904

RIN: 41252

DCN: RAR2T6-N

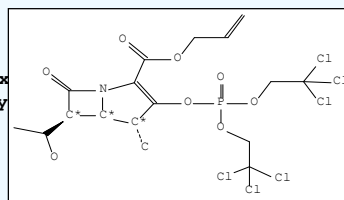
DCR: 1510822-N

AN.S DCR-1510822

CN.S (4R,5R,6S)-3-[Bis-(2,2,2-trichloro-ethoxy-
hydroxy-ethyl)-4-methyl-7-oxo-1-aza-bicyclo-
carboxylic acid allyl ester

SDCN RAR2T6

...



STN

FIZ Karlsruhe

Identify the unique answers retrieved by the chemical fragmentation code in DWPI

106

=> SET AUDIT ON

SET COMMAND COMPLETED

L11 = combined HCAPLUS,
WPIDS and MARPAT answer set.

=> TRANSFER L11 PN.B 1- /PN

L19 TRANSFER L11 1- PN.B : 529 TERMS

L20 503 L19/PN

L21 QUE TERMS FROM L19/PN WITH NO HITS: 7 TERMS

285 DWPI records were uniquely
retrieved by the fragmentation
code search (L22).

=> S L18 NOT L20

L22 285 L18 NOT L20

=> S L18 (P)M710/M2,M3,M4 NOT L20

L23 219 L18 (P)M710/M2,M3,M4 NOT L20

Role codes can help improve
the relevance of the results,
e.g. M710 = New Compound
(see slide 55 for the full list).

STN

FIZ Karlsruhe

Review the answers retrieved...

107

```

=> D AN TI HITCMC 70
L23 ANSWER 70 OF 219 WPIDS COPYRI
AN 2001-373534 [39] WPIDS
TI Antibiotic composition useful in treating and preventing enterococcal
infections includes 2-(naphthosultamyl)methyl-carbapenem compound and
another antibiotic
CMC UPB 20050525
M2 *09* C316 D013 D014 D016 D019 D021 D022 D023 D024 D029 D690 E810
F010 F019 F020 F021 F029 G003 G030 G039 G050 G553 G563 H100
H101 H102 H103 H161 H162 H163 H181 H182 H183 H2 H211 H341
H342 H343 H361 H362 H363 H381 H382 H383 H401 H402 H403 H404
H405 H461 H462 H463 H464 H481 H482 H483 H484 H496 H498 H600
H601 H608 H609 H641 H642 H643 H682 J0 J011 J012 J013 J014 J111
J211 J212 J261 J262 J271 J272 J273 J361 J362 J371 J372 J373 J5
J521 J523 J581 J582 J583 K0 K353 K399 K4 K441 L143 L144 L145
L199 L250 L299 L340 L399 L432 L499 L9 L941 L970 M115 M116 M119
M125 M129 M131 M132 M135 M136 M139 M210 M211 M212 M213 M214
M215 M216 M220 M221 M222 M223 M224 M225 M226 M231 M232 M233
M240 M272 M280 M281 M282 M283 M311 M312 M313 M314 M315 M321
M322 M323 M331 M332 M333 M340 M342 M353 M372 M373 M391 M392
M393 M412 M431 M512 M520 M521 M522 M523 M530 M540 M541 M542
M543 M710 M782 P220 M905 M9
RIN: 02714 41252
MCN: 0041-45701-M 0041-45701-N 0041-45701-T
    
```

The HITCMC format is often helpful for displaying hit fragmentation paragraphs.

This hit is a Markush compound.

STN

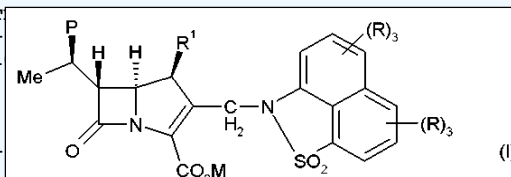
FIZ Karlsruhe

If we explore this answer further...

108

```

=> D L23 IFULLG 70
L23 ANSWER 70 OF 219 WPIDS COPYRIGHT 2008 THOMSON REUTERS on STN
ACCESSION NUMBER: 2001-373534 [39] WPIDS
DOC. NO. CPI: C2001-114038 [39]
TITLE: Antibiotic composition useful in treating and
preventing enterococcal infections includes 2-
(naphthosultamyl)methyl-carbapenem compound and
another antibiotic
DERWENT CLASS: B02
INVENTOR: DORSO K L; GILL C J; JACKSON J J; KOHLER J; SILVER L L
PATENT ASSIGNEE: (MERI-C) MERCK & CO INC
PATENT INFORMATION:
PATENT NO KIND DAT
-----
US 6221859 B1 2001
APPLICATION DETAILS:
PATENT NO KIND
-----
US 6221859 B1
PRIORITY APPLN. INFO: US 1999-387709 19990827
. . .
    
```



STN

FIZ Karlsruhe

Why use Fragmentation Codes?

109

- Oldest available Markush chemical structure indexing from patents online: back to 1963
- Only comprehensive source of Markush structure indexing on STN from 1963–1977
- Specific substances from patents (pre-DCR)
- Often provide a simple solution for very broad search queries, e.g. a pyridine ring (F431)
- Can easily be combined with other text search terms in DWPI, e.g. IPC, NCL, ECLA and/or MC
- There is no structure search fee

STN

 FIZ Karlsruhe

Summary

110

- DWPI Chemistry Resource (DCR)
 - Specific compound registry for DWPI
 - Standard STN structure searching
 - Available to all users of DWPI on STN
- Chemical Fragmentation Codes
 - Substances represented by codes
 - Code queries generated by STN Express
 - Available to DWPI Subscribers

STN

 FIZ Karlsruhe

Summary (cont.)

111

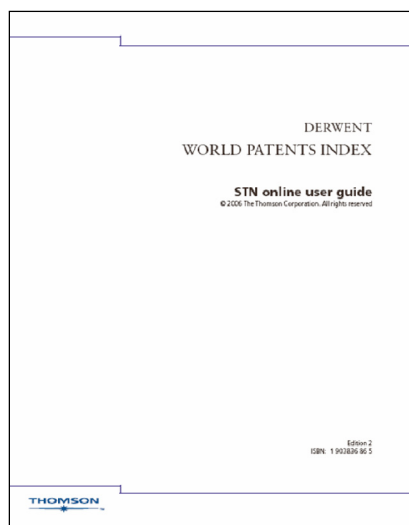
- A comprehensive structure search requires a multifile approach
 - Specific structures for REGISTRY and DWPI databases
 - Generic graphical structures for MARPAT (1978-)
 - Generic fragment codes for WPIDS/WPIX (1963-)
 - Other prior-art databases may also be included (e.g. Beilstein)
- Patent records from all databases can be organized using TRANSFER, DUP REM (or IDE) and FSORT
 - Patent family (PFAM) display allows for displaying first record from each family for reviewing the initial results
 - Additional records can be displayed as needed
 - File default formats allow for file specific displays

STN

 FIZ Karlsruhe

DWPI on STN User Guide

112



Available at (PDF):

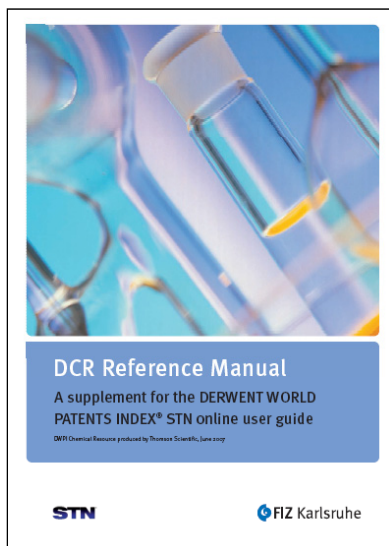
http://www.stn-international.com/training_center/patents/stn_guide.pdf

STN

 FIZ Karlsruhe

DCR Reference Manual

113



Available at (PDF):

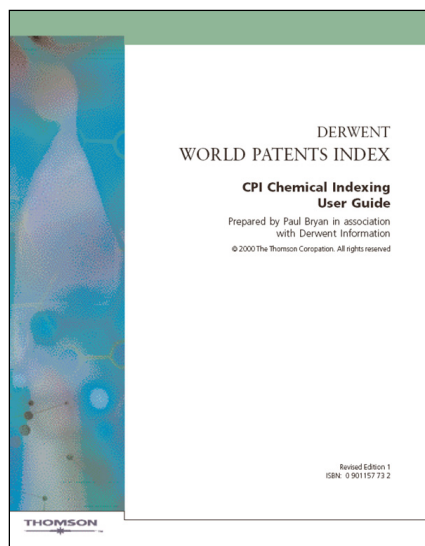
http://www.stn-international.com/training_center/patents/dcr_rm.pdf

STN

FIZ Karlsruhe

Chemical Indexing User Guide

114



The complete guide to chemical fragmentation codes (PDF):

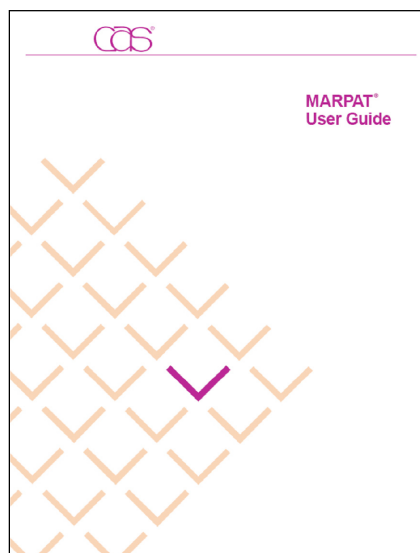
<http://scientific.thomson.com/support/userguides/chemistryguides/>

STN

FIZ Karlsruhe

MARPAT User Guide

115



The complete guide to the CAS MARPAT database (PDF):

http://www.stn-international.com/training_center/workshop_material/user_guides/marpatug.pdf

STN

 FIZ Karlsruhe

STN[®]

Structure searching Derwent World Patents Index[®] (DWPISM) using STN Express[®]

www.stn-international.com

Section R Fragmentation Codes

R	R0: FORMULATION TYPE	R01: Gaseous form	R011 # Aerosol	R012 # Inhalant	R013 # Other (smoke, gas, etc.)	R02: Liquid or semi-solid	R021 Cream; paste	R022 Emulsion	R023 Solution; syrup	R024 Suspension; dispersion of solids	R03: Solid form; moulded (excl. R04)	R031 Capsule	R032 Grain; granule; crystal	R033 Micro-capsule	R034 Pellet	R035 Powder (dusting)	R036 Powder (excl. R035)	R037 Suppository	R038 Tablet	
R04: Foam; other solid form	R041 Bandage; Dressing	R042 Filament; suture; wire	R043 Film; sheet	d R044 Foam; expanded/porous foam	d R045 Laminate	R046 Prosthesis	R047 Surgical sponge; tampon	R05: Timed release	b R051 Delayed release	b R052 Sustained release										
R1: GENERAL TERMS	R111 Dosage form novel	R112 Prodn. of dosage form novel	R120 Coated form	R130 Disposable	R140 Ink; label	R150 Multi-component	R160 Reference standard	R170 Special shape	R252 Intraarterial	R253 Intraocular	R261 Intra-muscular	R262 Sub-cutaneous	j R263 Trans-dermal	R271 Intrauterine	R272 Intravaginal	R280 Oral	R290 Rectal			
R2: MODE OF ADMINIST.	R200 Buccal; sublingual	R210 External; topical	R220 Implant	R231 Injection	R232 Infusion	R241 Intraarterial	R242 Intravenous	R251 Intraaural	R252 Intraarterial	R253 Intraocular	R261 Intra-muscular	R262 Sub-cutaneous	R263 Trans-dermal	R271 Intrauterine	R272 Intravaginal	R280 Oral	R290 Rectal			
R3: FORMULAT. ADDITIVE	R301 Absorbent; adsorbent	R302 Accelerator	R303 Adhesive; binder	R304 Anticaking; disintegrating	R305 Buffer	R306 Chelating agent	R307 Coating	R308 Diluent; filler	R309 Dye; Indicator	R310 Flavour	R311 Gelling agent	R312 Lubricant	R313 Plasticiser	R314 Propellant	R315 Stabiliser general	R316 Stabiliser to heat	R317 Stabiliser to oxidation	R318 Stabiliser to radiation	R319 Surfactant	
R4: ADMINIST. DEVICE	R410 Catheter	R420 Dosage into orifice	R430 Implanting	R440 Inhalation	R450 Injection gun	R46: Syringe	R460 Syringe; general	R461 Syringe; hypodermic	R462 Syringe; remote control	R463 Syringe needle	R5: PROCESS; APPARATUS	R50: General & others	R501 Physical process/apparatus	R502 Chemical process/apparatus	R51: Analytical; Diagnostic; testing	R511 # Automatic	R512 # Chromatographic	R513 # Isotope detection	R514 # Spectroscopic	R320 Suspending agent
R52:-3: Other proc./apparatus (specific)	R520 Agglutination precipitation	R521 Biological; fermentation	R522 Coating	R523 Compressnt; vacuum	R524 Cooling; freezing	R525 Cutting; grinding	R526 Distilling; heating	R527 Drying; lyophilisation	R528 Electrical; magnetic	R529 Filling	R530 Flow control; transporting (excl. R700)	R531 Forming; moulding	R532 Impregnating	R533 Microscopy; optical	R534 Mixing	R535 Purification; separation (excl. R527)	R536 Radiation	R537 Sampling	R538 Spraying	
R6: DIAGNOSTIC METHOD	R61: Material tested	R611 Blood or blood plasma	R612 Urine	R613 Other body fluid	R614 Other material	R62: Reagent used	R621 Antibody	R622 Antigen	R623 Colour indicator	R624 Enzyme	R625 Fluorescent labelled	R626 Isotope/radioactive label	R627 Other							
R63: Substance detected	R630 Antibody	R631 Antigen	R632 Enzyme	R633 Drug (not specified in R63)	R634 Fat	R635 Micro-organism	R636 Pesticides/agrochem.	R637 Protein	R638 Steroid	R639 Other	R7: PACKAGING (pre 1976 see M01)	R700 Aerosol can and valve	R710 Ampoule	R720 Bottle	R730 Box; carton	R740 Bag; blister/bubble pack	R750 Cap; closure	R760 Dispenser	R770 Safety container	

From 1963 for R0: codes in fields M0-M5 and from 1976 for R0: to R7: codes in field M6.

From Derwent Week 198127 for R0: codes in fields M0-M5 and for R0: to R7: codes in field M6.

SPECIAL NOTES:

- # in the top right hand of the code box indicates that for pre-1981 references, discontinued generic codes should be searched
- (b) Pre-1981 Farmdoc and Agdoc (M0, M1 & M2) also
- (d) Pre-1981 Chemdoc (M3 & M4) also
- (j) From Derwent Week 198337 onwards

STN Service Centers

FIZ Karlsruhe

STN Europe
P.O. Box 2465
76012 Karlsruhe
Germany

Phone: +49 7247 808 555
Fax: +49 7247 808 259
E-mail: helpdesk@fiz-karlsruhe.de
Internet: www.stn-international.de

CAS

STN North America
P.O. Box 3012
Columbus, Ohio 43210-0012

CAS Customer Care:
Phone: 800-753-4227 (North America)
614-447-3700 (worldwide)
Fax: 614-447-3751
E-mail: help@cas.org
Internet: www.cas.org

Japan Association for International Chemical Information (JAICI)

STN Japan
Nakai Building
6-25-4 Honkomagome, Bunkyo-ku
Tokyo 113-0021, Japan

Phone: +81-3-5978-3601 (Technical Service)
+81-3-5978-3621 (Customer Service)
Fax: +81-3-5978-3600
E-mail: helpdesk@jaici.or.jp (Technical Service)
cas-stn@jaici.or.jp (Customer Service)
Internet: www.jaici.or.jp