

---

# **Derwent World Patents Index<sup>®</sup> on STN<sup>®</sup>**

**Chemical Fragmentation Codes**

---

© Fachinformationszentrum Karlsruhe, November 2009

Robert Austin  
FIZ Karlsruhe GmbH  
Hermann-von-Helmholtz Platz 1  
76344 Eggenstein-Leopoldshafen  
Germany

Email: [robert.austin@fiz-karlsruhe.de](mailto:robert.austin@fiz-karlsruhe.de)

Web: [www.fiz-karlsruhe.de](http://www.fiz-karlsruhe.de)

---

# Contents

	Page
<b><u>Workshop Slides</u></b>	<b>1</b>
Agenda	1
Derwent World Patents Index (DWPI <sup>SM</sup> ) on STN	2
Introduction to chemical fragmentation codes	3
Basics of fragmentation code searching	7
Editing fragmentation code strategies	14
Chemical Indexing User Guide	28
<b><u>Fragmentation Code Workbook</u></b>	<b>31</b>
<b><u>Appendix – Galenical codes (/M6)</u></b>	<b>67</b>



# STN<sup>®</sup>

Derwent World Patents Index<sup>®</sup>  
*Chemical Fragmentation Codes*

Robert Austin – FIZ Karlsruhe

## Agenda

2

- Introduction to fragmentation codes
- Basics of fragmentation code searching
- Editing fragmentation code strategies
- Fragmentation code workbook

**STN**

 FIZ Karlsruhe

## Derwent World Patents Index (DWPI<sup>SM</sup>)

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 file
- FILE WPIDS
  - Derwent Subscriber file
- FILE WPIX
  - Derwent Subscriber file with Extension Abstracts
- FILE LWPI
  - The DWPI learning file
- FILE WPIFV
  - 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 Chemical Fragmentation Codes

6

- 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

## Why use Fragmentation Codes?

7

- 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

## A brief history of Derwent World Patents Index Chemical Fragmentation Codes

8

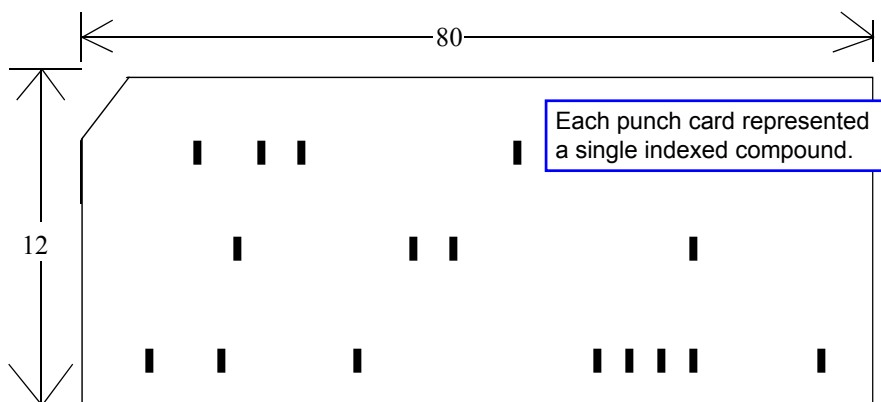
- 1963 Pharmaceuticals (B)
- 1965 Agricultural Chemicals (C)
- 1970 **Code revision** + General chemistry (E)
- 1972 **Code revision** + Ring Index Numbers
- 1981 **Code revision** + DRNs
- 1987 Markush graphical indexing + DCNs
- 1992 Autogeneration of codes
- 1999 Derwent Chemistry Resource

**STN**

 FIZ Karlsruhe

## Fragmentation codes were once positions on an IBM punch card

9



12 ROWS X 80 COLUMNS - TOTAL POSITIONS=960

STN

FIZ Karlsruhe

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

10

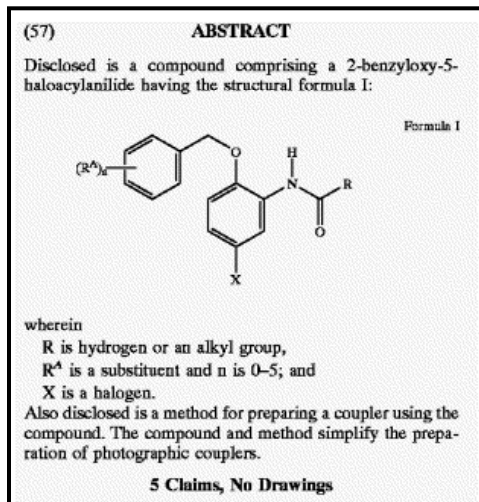
AN	2005-217884 [23]	WPIX	
TI	Recovery of solvent and styrene from polystyrene solution involves recovering solvent by evaporation and recovering styrene from polystyrene thermally decomposed by solvent.		
IT	UPIT 20050708		
	2113-DIS 2113-PRD; 368-CL 368-PRD		
CMC	UPB 20050708		
	DRN: 0708-P 0708-U 1119-P 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	Limonene
		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

11



STN

FIZ Karlsruhe

## Specific and Markush compounds are indexed with Fragmentation codes

12

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 G039 G040 G050 G051 G100 G563 H100 H141 H341 H401 H441 H541 H7 H715 H721 H722 H723 H724 H725 H726 H727 H728 H729 H730 H731 H732 H733 H734 H735 H736 H737 H738 H739 H740 H741 H742 H743 H744 H745 H746 H747 H748 H749 H750 H751 H752 H753 H754 H755 H756 H757 H758 H759 H760 H761 H762 H763 H764 H765 H766 H767 H768 H769 H770 H771 H772 H773 H774 H775 H776 H777 H778 H779 H780 H781 H782 H783 H784 H785 H786 H787 H788 H789 H790 H791 H792 H793 H794 H795 H796 H797 H798 H799 H800 H801 H802 H803 H804 H805 H806 H807 H808 H809 H810 H811 H812 H813 H814 H815 H816 H817 H818 H819 H820 H821 H822 H823 H824 H825 H826 H827 H828 H829 H830 H831 H832 H833 H834 H835 H836 H837 H838 H839 H840 H841 H842 H843 H844 H845 H846 H847 H848 H849 H850 H851 H852 H853 H854 H855 H856 H857 H858 H859 H860 H861 H862 H863 H864 H865 H866 H867 H868 H869 H870 H871 H872 H873 H874 H875 H876 H877 H878 H879 H880 H881 H882 H883 H884 H885 H886 H887 H888 H889 H890 H891 H892 H893 H894 H895 H896 H897 H898 H899 H900 H901 H902 H903 H904 H905 H906 H907 H908 H909 H910 H911 H912 H913 H914 H915 H916 H917 H918 H919 H920 H921 H922 H923 H924 H925 H926 H927 H928 H929 H930 H931 H932 H933 H934 H935 H936 H937 H938 H939 H940 H941 H942 H943 H944 H945 H946 H947 H948 H949 H950 H951 H952 H953 H954 H955 H956 H957 H958 H959 H960 H961 H962 H963 H964 H965 H966 H967 H968 H969 H970 H971 H972 H973 H974 H975 H976 H977 H978 H979 H980 H981 H982 H983 H984 H985 H986 H987 H988 H989 H990 H991 H992 H993 H994 H995 H996 H997 H998 H999

M3 \*05\* . . . . .  
 DCN: 0149-69501-K 0149-69501-U

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

## Agenda

13

- Introduction to fragmentation codes
- Basics of fragmentation code searching
- Editing fragmentation code strategies
- Fragmentation code workbook



## Basics of fragmentation code searching

14

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

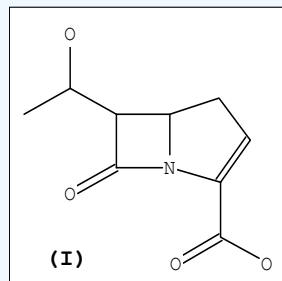


## Fragmentation code search example

15

### 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

16

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**

Structure Drawing  
File Edit Draw Template QueryDef Display Preferences! Window Help  
New  
Open... Ctrl+O  
Close...  
Browse...  
Save Ctrl+S  
Save As...  
Page Setup...  
Print Preferences...  
Print... Ctrl+P  
Exit

WPI.str  
str  
str  
str

Save the current structure with a new name

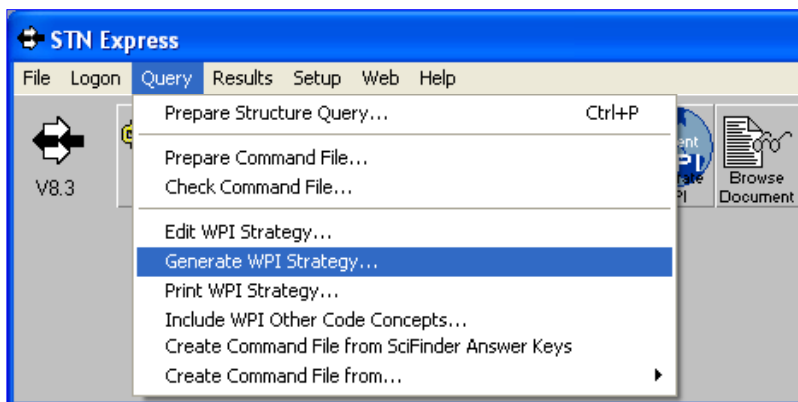
Mon 5/16/2005 1:29:38

STN

FIZ Karlsruhe

## Use STN Express to generate the fragmentation code script

17



STN

FIZ Karlsruhe

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

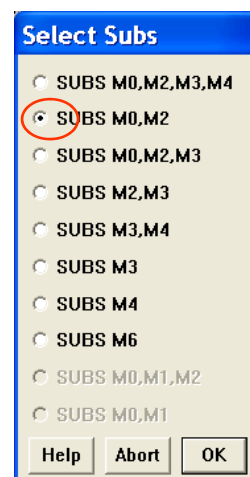
18

### 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

19

The screenshot shows the STN software interface with a script window titled 'C:\STNEXPWscripts\carbapenemwpi.sc'. The script contains DWPI fragmentation code strategy in Command File format, including lines like:
   
=>S (D690(P)H401(P)
   
=>S \_line1(P)(M511(P)
   
=>S \_line2(P)(M286(P)
   
=>S \_line3(P)41252
   
=>S \_line4(P)(D811
   
=>S (\_line1(P)M906
   
=>S \_line6(NOTP)(H
   
=>S \_line7(NOTP)(J
   
=>S \_line8(NOTP)(H
   
=>S \_line9(NOTP)(
   
\* J011 - TOTAL NU
   
\* from 812
   
\* M321 - ONE OF O
   
\* di- and
   
\* to 4.6 I
   
\* M342 - DIVALENT
   
\* heteroat
   
\* attached
   
\* (see 0.4
   
\* J011 - TOTAL NUMBER OF CARBOXYLIC ACID, ESTER AND AMIDE GROUPS = 1;
   
\* from 8127 only, including SUBS M3
   
\* M321 - ONE OF CODES M311 T
   
\* di- and polyvalent
   
\* to 4.6 IM3A)
   
\* M342 - DIVALENT CHAIN OF AD
   
\* heteroatom does not
   
\* attached to a ring C via an exo double bond - from 1972 only
   
\* (see 0.4.5 - 4.6 IM3A)

A callout box points to the 'Check Command File' function, stating: 'The Check Command File function includes mouse-over code definitions.'

STN

FIZ Karlsruhe

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

20

The screenshot shows the STN Online and Results software interface. The 'Query' menu is open, displaying several options:
   
Upload Structure Query (Ctrl+U)
   
Prepare Command File
   
Check Command File
   
Run Command File (Ctrl+R)
   
Use Predefined Search Strategy
   
Create Command File from SciFinder Answer Keys
   
Create and Run Command File from SciFinder Answer Keys
   
Create Command File from...

STN

FIZ Karlsruhe

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

21

```
=> 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

22

```
=> 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-penem-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

23

```

=> S L10(P)M720/M2
L11          26 L20(P)M720/M2
=> 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
  
```

M720 is the Role for synthesized/produced.

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

This hit is a Markush compound.

STN

FIZ Karlsruhe

## Fragmentation code roles

24

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

25

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

## DWPI compound selection and indexing guidelines

26

The following compounds are selected for indexing

- All novel compounds claimed or disclosed including both specific and Markush structures
- Known compounds from a new process
- Known compounds having a new use
- Components of compositions and formulations
- Compounds/materials detected, detecting agents and detection media
- Compounds recovered or purified in new ways
- Compounds removed and removing agents

**STN**

 FIZ Karlsruhe

## DWPI compound selection and indexing guidelines (cont.)

27

- Markush structures are indexed from:
  - the patent claims
  - the embodiment if a 'wider disclosure' is indicated
- Specific compounds are indexed from:
  - the claims
  - the main (best) example
  - further examples at the analyst's discretion
  - if the patent claims do not contain specific compounds, the analyst selects representative compounds from the examples and embodiment



## Agenda

28

- Introduction to fragmentation codes
- Basics of fragmentation code searching
- Editing fragmentation code strategies
- Fragmentation code workbook



## Editing fragmentation code strategies for substructure searches

29

- 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



## Editing fragmentation code strategies for substructure searches (cont.)

30

- 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

31

- Search fields (/M2, etc) – slide 18
- Fragmentation Code Sheet
- Colored **historical time periods**
- Standard search strategy format
- Time range codes
- Negation codes
- Ring Index Numbers (RINs)
- Discontinued fragmentation codes (#)

**STN**

 FIZ Karlsruhe

## Fragmentation code sheet

32




A,B,C	Inorganic parts of structure
D,E	Fused heterocyclic ring systems
F	Mononuclear heterocyclic rings
G	Carbocyclic rings
H,J,K,L	Chemical functional groups
M	Ring linkages, carbon chains and other non-chemical structure concepts
N	Processes and Apparatus
P,Q	Pharmaceutical (P) and other (Q) uses
R	Shape/form; diagnostic processes; apparatus
V	Natural products and Polymers
W	Dyes and pigments

**STN**

 FIZ Karlsruhe

## Fragmentation code sheet (cont.)

33

<b>F2:</b> S-SOLE HETERO.	<b>F200</b> Thiirene* Thiete	<b>F211</b> 	<b>F212</b> 	<b>F213</b> 
---------------------------------	------------------------------------	--	---	--

- 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

34

**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

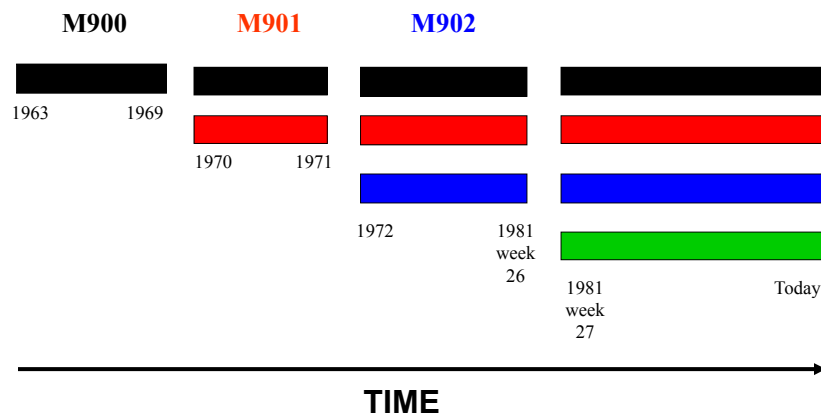
35

- 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

36



## Time range codes

37

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

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



## Negation codes

38

- 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)

39

- 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.

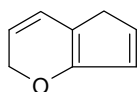
STN

FIZ Karlsruhe

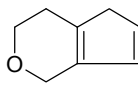
## Ring Index Numbers improve search precision for non-specific ring codes

40

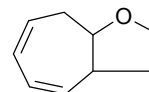
- All the ring systems below are searched with code D130\*



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

41

- 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

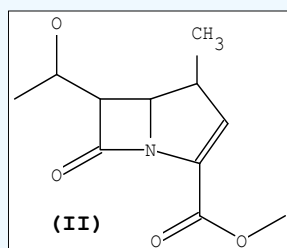
STN

FIZ Karlsruhe

## Fragmentation code editing example

42

**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

43

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. . .

44

A standard fragmentation code query was generated following the steps on slides 16-19.

```
=>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
```

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

45

```
=>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

46

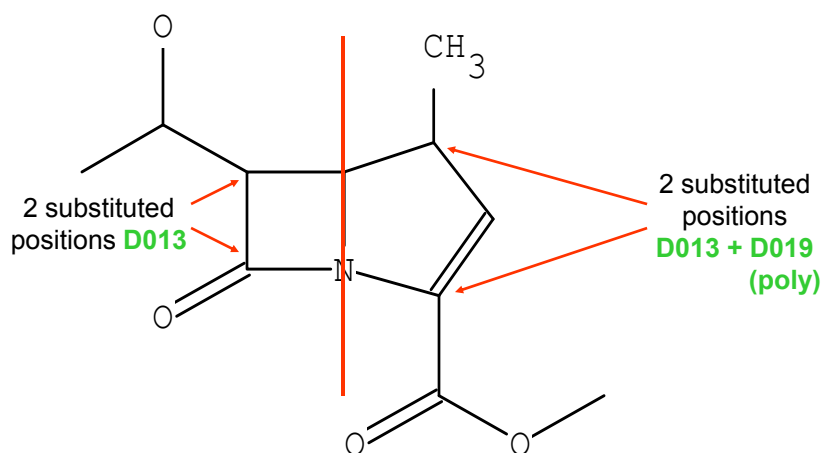
2. M412 OR-ed with M411 (e.g. to include salts)
3. 2-substitutions D013, OR-ed with 3, D014
4. 1-ester J211, OR-ed with >1 J212  
1-oxo J521, OR-ed with >1 J522, J523  
1-carboxy derivative J011 deleted
5. Hydroxyl H401/H481 deleted
6. Carbon chain (M2/M3) options OR-ed in
7. 1-ring M511, OR-ed with >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

47

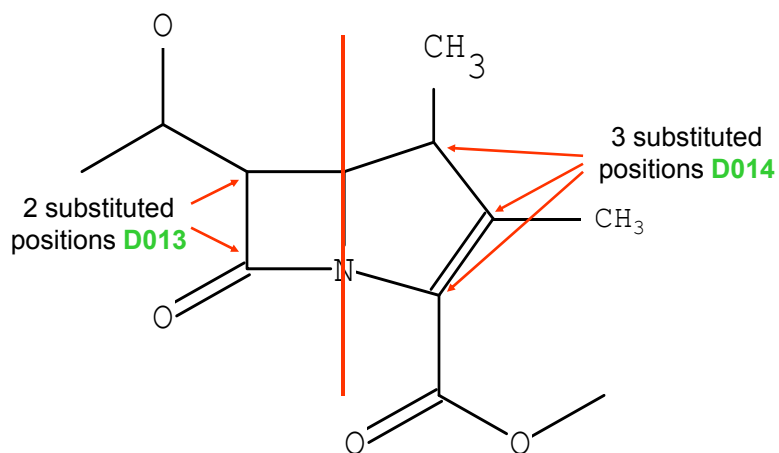


STN

FIZ Karlsruhe

## Fused-heterocyclic ring substitution codes

48



STN

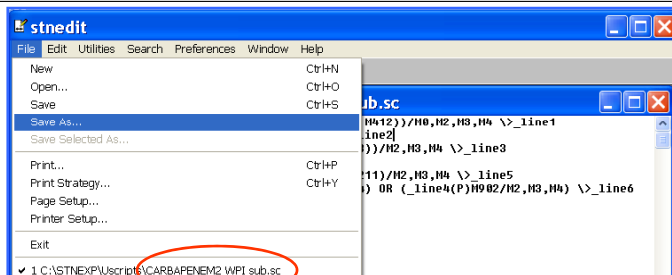
FIZ Karlsruhe

## Save the edited strategy. . .

49

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

50

**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

## Search the edited strategy. . .

51

```

=> FILE WPIDS
=> S (D690 (P) J21! (P) J52! (P) M240 (P) M331 (P) . . . ) /M0, M2, M3, M4
L1      2203 (D690 (P) J21! (P) J52! (P) M240 (P) (M411 OR M412) ) /M0, M2, M3, M4
=> S L1 (P) (M511 OR M512 OR M513) /M2, M3, M4
L2      2179 L1 (P) (M511 OR M512 OR M513) /M2, M3, M4
=> S L2 (P) (M210 (P) (M281 OR M282 OR M283) (P) . . . ) /M2, M3, M4
L3      2131 L2 (P) (M210 (P) (M281 OR M282 OR M283) (P) . . . ) /M2, M3, M4
=> S L3 (P) 41252 /RIN
L4      554 L3 (P) 41252 /RIN
=> S L4 (P) (D013 (P) (D019 OR D014) (P) "L941" (P) M211) /M2, M3, M4
L5      506 L4 (P) (D013 (P) (D019 OR D014) (P) "L941" (P) M211) /M2, M3, M4
=> S (L1 (P) M900 /M0) OR (L2 (P) M901 /M2, M3, M4) OR . . . .
L6      105 (L1 (P) M900 /M0) OR (L2 (P) M901 /M2, M3, M4) OR . . . .
=> S L6 OR L5
L7      582 L6 OR L5
    
```

Run the command file in WPIDS (or WPIX).

The Command File runs automatically line-by-line

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

This is a true *substructure* search (L7).

STN

FIZ Karlsruhe

## Review fragmentation code answers

52

```

=> D AN TI HITCMC FRAGHITSTR 60 254

L7 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-bicyc
   carboxylic acid allyl ester
SDCN RAR2T6
. . .
    
```

This hit is a specific compound.

DCN: RAR2T6-N

DCR: 1510822-N

STN

FIZ Karlsruhe

## Review fragmentation code answers (cont.)

53

```
L7 ANSWER 254 OF 582 WPIDS COPYRIGHT 2008 THOMSON REUTERS on STN
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 H1 The HITCMC format is often helpful for
H342 H343 H361 H362 H3 displaying hit fragmentation paragraphs.
H405 H461 H462 H463 H4
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 M904
RIN: 02714 41252
MCN: 0041-45701-M 0041-45701 This hit is a Markush compound.
```

STN

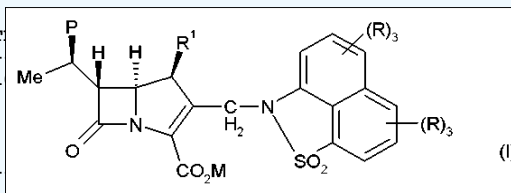
FIZ Karlsruhe

## If we explore this answer further...

54

=> D L7 IFULLG 254

```
L7 ANSWER 254 OF 582 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

## Agenda

55

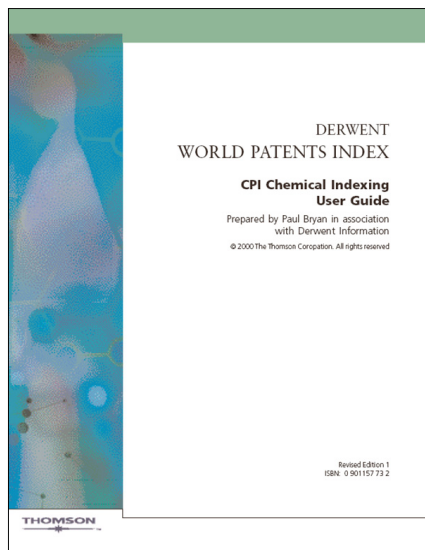
- Introduction to fragmentation codes
- Basics of fragmentation code searching
- Editing fragmentation code strategies
- Fragmentation code workbook

**STN**

 FIZ Karlsruhe

## Chemical Indexing User Guide

56



The complete guide to chemical fragmentation codes (PDF)

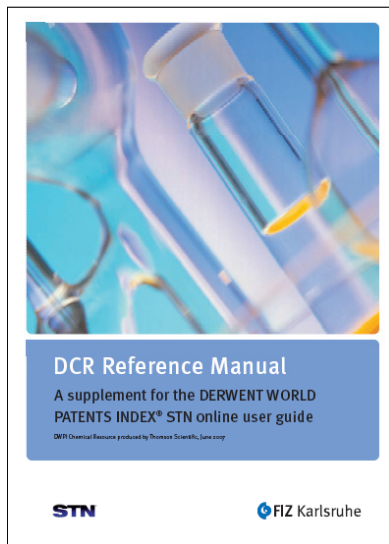
<http://scientific.thomson.com/support/userguides/chemistryguides/>

**STN**

 FIZ Karlsruhe

## DCR Reference Manual

57



Available at (PDF):

[http://www.stn-international.com/  
training\\_center/patents/dcr\\_rm.pdf](http://www.stn-international.com/training_center/patents/dcr_rm.pdf)

**STN**

FIZ Karlsruhe

FIZ Karlsruhe

## Acknowledgement

58

**Steve Hajkowski**

Thomson Scientific

[steve.hajkowski@thomsonreuters.com](mailto:steve.hajkowski@thomsonreuters.com)

**STN**

FIZ Karlsruhe

**STN**<sup>®</sup>

Derwent World Patents Index<sup>®</sup>  
*Chemical Fragmentation Codes*

[www.stn-international.com](http://www.stn-international.com)

## Derwent World Patents Index<sup>®</sup> Fragmentation code workbook

### Introduction

The general rules for indexing compounds with the fragmentation codes are shown below

**Inorganic compounds** : Any metals present are indexed in Part A and non-metals are indexed in Part C (halogens, oxygen, hydrogen, sulphur, nitrogen, carbon). Any other elements present are indexed in Part B. Finally the elements absent from the compound are indexed in Part C8: and the code M411 is added.

**Organic compounds** : Any ring systems present are indexed in Parts D & E (fused heterocyclics), Part F (monocyclic heterocyclics) and/or Part G (carbocyclics). The functional groups present (including double and triple bonds) are indexed in Parts H-L and the carbon chains, ring linkages, number of rings and basic group type are applied from Part M.

**Organometallic compounds** : The metal or metal(s) present require codes from Part A, if there is an inorganic counterion this will be indexed in Part B or C depending on the elements it contains and also if there is an inorganic anion the elements absent will be indexed in Section C8. Any ring systems present are indexed in Parts D & E (fused heterocyclics), Part F (monocyclic heterocyclics) and or Part G (carbocyclics). The functional groups present (including double and triple bonds) are indexed in Parts H-L and the carbon chains, ring linkages, number of rings present and basic group code M411 are applied from Part M.

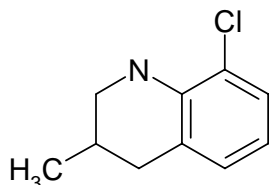
**Organophosphorus, organoboron, organosilicon and organoarsenic compounds** : The P, B, Si and / or As and all the bonds to it are indexed in Part B. Any ring systems present are indexed in Parts D & E (fused heterocyclics), Part F (monocyclic heterocyclics) and or Part G (carbocyclics). The functional groups present (including double and triple bonds) are indexed in Parts H-L and the carbon chains, ring linkages, number of rings and basic group type are applied from Part M.

**Organoselenium, organotellurium compounds** : Apply the appropriate code for Se or Te in Part B and then index the compound as if it were the corresponding organic compound containing sulphur but apply M411 instead of the M412 – M416 code.

## Part D and E codes - fused ring heterocycles

Part D and E contain codes D01: that describe the **number of positions substituted** on each individual heterocyclic ring in the fused system, and codes D02: that describe the **number of positions substituted** on each individual carbocyclic ring in the fused system (if present).

### Example 1

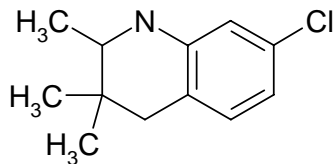


To index these codes, the ring is divided into its component heterocyclic and carbocyclic parts, and the two parts are indexed separately. In this example, the heterocyclic part of the ring has one position substituted, and this substituent is two positions away from the fusion (i.e. the  $\beta$ -position), so the code indexed is D012. The aromatic ring also has one position substituted, and the substituent is one position away from the fusion (i.e. the  $\alpha$ -position), so this is indexed as D021. These codes are indexed along with the code for the ring system itself, D622.

Note: any substituent at the ring fusion is not counted when assigning D01: and D02: codes. The code D030 is the only code used to indicate that substituents are present at the ring fusion.

### Example 2

Write down the codes indexed from parts D01: and D02: for this molecule

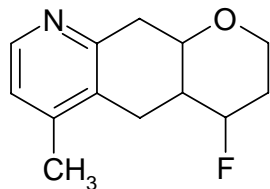


## Poly Codes

Many sections of the Fragmentation Code contain codes marked 'Poly'. These codes are indexed to indicate that a code in that set appears more than once in the molecule.

### Example 3

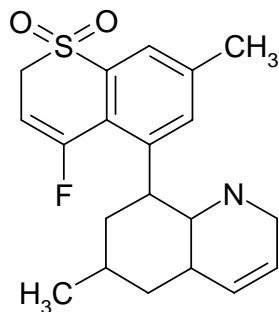
In this molecule, the fused system contains two heterocyclic rings, each with one substituent at the  $\alpha$ -position. Both rings require the code D011, so the indexing for the molecule contains both D011 and D019 codes (plus E520 for the ring system itself).



### Example 4

Write down all the part D and E codes for this structure:

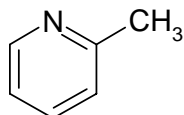
*(Note: parts D, E and F codes where 0 is the final digit of the code (e.g. D320) are indexed for all possible saturation states of the ring (system). Codes that do not have 0 as the final digit (e.g. E111) are indexed only for the ring shown. For part G, this rule also applies, but note that the system must contain at least one aromatic ring to be indexed in section G2: to G4: (a fully alicyclic system is indexed in sections G5: to G8:).)*



## Part F codes - mononuclear heterocycles

In contrast to parts D and E, the codes in part F01:, the codes indicate the **position of substitution** on the heterocycle.

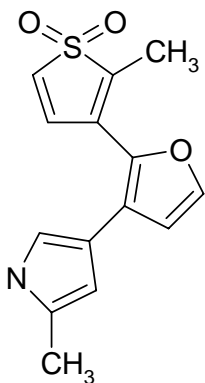
### Example 5



This is indexed F431 for the ring, along with code F012, indicating position 2 is substituted. The numbering system always numbers the heteroatom(s) first. If more than one heteroatom is present in the ring, any oxygen is numbered first (i.e. lowest), followed by any sulphur then nitrogen, phosphorus etc. The substituted positions are then numbered as low as possible.

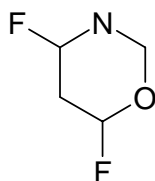
### Example 6

Write down all Part F codes for this structure



### Example 7

Write down all Part F codes for this structure



## Part G codes - carbocyclic rings

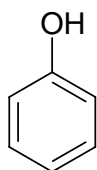
The codes indicating substitution are like those for parts D and E, i.e. the codes describe the **number of positions substituted**. The codes are in 3 sections:

G01: number of positions substituted on an isolated aromatic ring

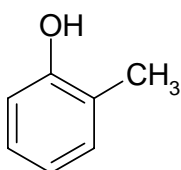
G02: number of positions substituted on a fused aromatic ring

G03: number of positions substituted on an alicyclic ring

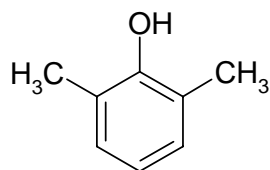
### Example 8



G010, G100



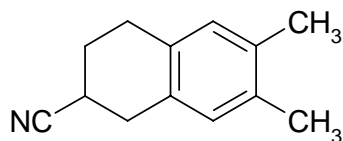
G011, G100



G014, G100

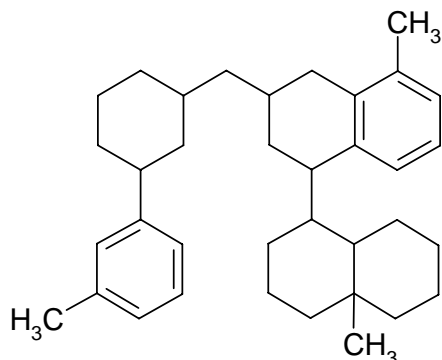
### Example 9

This structure is indexed with code G032 to indicate the cyano substituent on the alicyclic ring, and code G022 to indicate that 2 positions are substituted on the fused benzene. The ring system is indexed G223.



### Example 10

Write down the part G codes for this molecule:



## Part H and J codes – functional groups

### Main Headings

H1: Amine	J0: Number of carboxyl derivatives present
H2: Ring tertiary nitrogen	J1: Carboxylic acid, thiocarboxylic acid
H3: Nitro	J2: Ester, thioester
H4: Hydroxy, mercapto	J3: Amide, thioamide
H5: Ether, thioether	J4: Aldehyde, thioaldehyde
H6: Halogen (F, Cl, Br, I)	J5: Oxo, thioxo
H7: Olefinic, acetylenic	

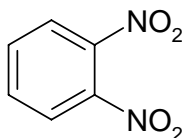
These codes describe functional groups in terms of:

- the type of functional group
- the type of carbon the functional group is attached to, i.e. heterocyclic carbon, aromatic carbon, alicyclic carbon or aliphatic carbon
- the number of these functional groups with these attachments present

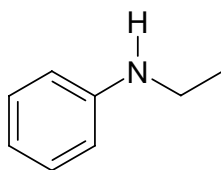
If the functional group is attached to more than one type of carbon, only the highest priority is indexed according to:

**heterocyclic > aromatic > alicyclic > aliphatic**

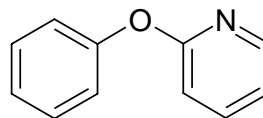
### Example 11



indexed H342



indexed H141 (not H181)

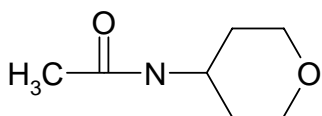


indexed H521 (not H541)

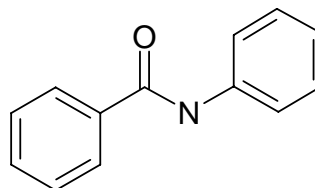
### Indexing of non-symmetrical, multivalent functional groups

These kind of groups (esters, amides etc.) are indexed according to the priority rules above, but the codes also describe the point of attachment of the highest priority group. If the non-symmetrical functional group is attached to carbons of equal priority, only the attachment to the C=O group is indexed.

### Example 12



indexed J321 (not J371)

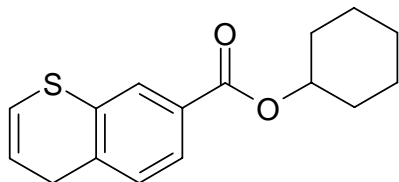


indexed J331 (not J341)

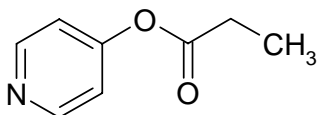
**Example 13**

Write down the part D, F, G and J codes for these structures:

**(1) Answer:**



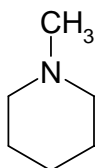
**(2) Answer:**



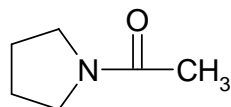
**Indexing ring tertiary nitrogen (H2: codes)**

A trivalent nitrogen atom in a ring, bonded to an atom other than H outside the ring system, is indexed with the ring tertiary nitrogen codes (part H2:). The nitrogen is indexed either as an amine (H20: codes), or as a non-amine (H21: codes).

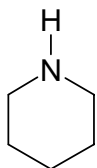
**Example 14**



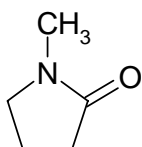
indexed H201, H181, F433, F011



indexed H211, J371, J011, F423, F011



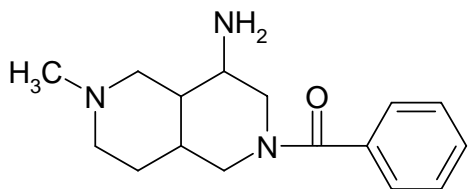
indexed F433 only (no part H: codes)



indexed H211, J521, L941, F423, F011, F012

**Example 15**

Write down the part D, G, H and J codes for this structure



## Part K and L codes – functional groups

Part K indexes bonds between heteroatoms in organic compounds

### Main Headings

K1: X-Y and X=Y bonds	K6: N-N bonds
K2: S-S and S=S bonds	K7: N=O bonds
K3: S-N and S=N bonds	K8: N-O bonds
K4: S-O and S=O bonds	K9: O-O bonds
K5: N=N and N≡N bonds	

X is F, Cl, Br or I; Y is O, S, Se, Te, N, F, Cl, Br, or I

Part L indexes complex functional groups

### Main Headings

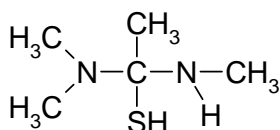
L1: -C≡N  
 L2: U=C=N- U-C(=N)-U  
 L3: Other C=N groups  
 L4: U-C(=T)-U  
 L5: Other C=T groups  
 L6: U-C-U  
 L7: Miscellaneous element descriptors – onium, free radicals, ions, valencies  
 L8: Sugars and derivatives  
 L9: Groups in rings

U is O, S, Se, Te or N; T is O, S, Se or Te

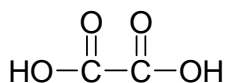
**General Indexing Rule for Part K and L – for large functional groups, the minimum number of functional group codes are assigned that between them specify all atoms and links between atoms. If a code definition includes variable element symbols, then additional functional group codes are added to identify these.**

### Example 16

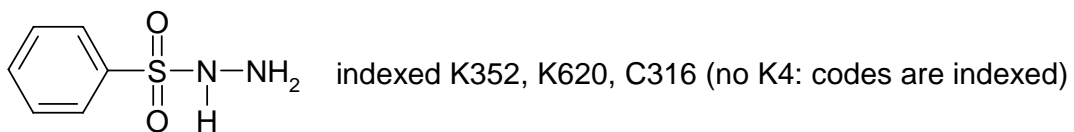
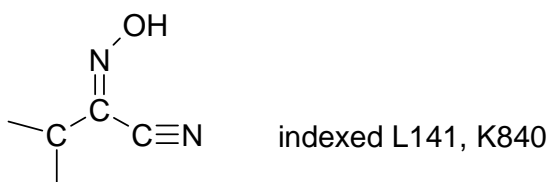
N-N-C≡N indexed as K620 N-N-C≡U and L110 N-C≡N



indexed as H102, H103, H182, H498, L610



indexed L560, J012, J172



## Part M codes - miscellaneous

The part M codes are miscellaneous descriptor codes, divided into the following sections:

- M1: linkages between rings
- M2: zero or monovalent carbon chains
- M3: di or polyvalent carbon chains
- M4: basic groups
- M5: ring systems present
- M6: miscellaneous codes
- M7: patent type
- M8: stereochemistry
- M9: time range codes

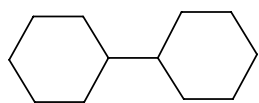
M1: codes indicate the linkages between rings, further divided as

- M11: - rings linked by bonds
- M12: - rings linked by atoms
- M13: - links containing carbon
- M14: - links containing no carbon

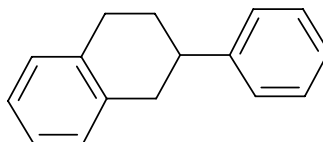
The following definitions are used for these codes:

- Benzene – unfused benzene
- Aryl – benzene fused to carbocyclic or heterocyclic
- Other – ring not covered by benzene or aryl

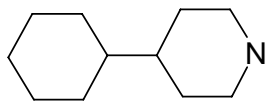
**Example 17**



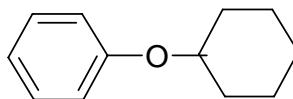
M116



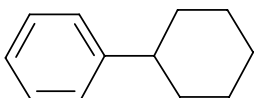
M113



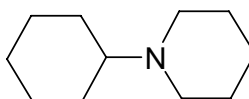
M116



M123, M141



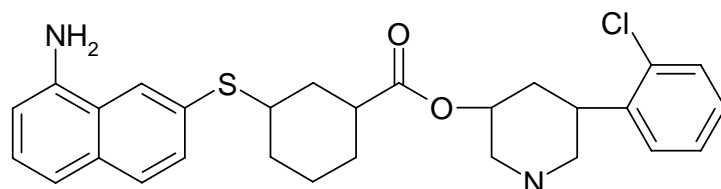
M113



No M1 code indexed

**Example 18**

Write down the codes from parts D to M1 for this structure, and group together the appropriate colour codes:



BLACK:

RED:

BLUE:

GREEN:

## Carbon chains - indexed with part M2: and M3: codes

Carbon chains are defined as an optionally branched, optionally unsaturated aliphatic hydrocarbon group.

The chain ends at the carbon bonded to the following terminating atoms:

- ring carbon
- carbon multiply bonded to a heteroatom
- any heteroatom

**The valency of the chain is the number of bonds linking the chain to terminating atoms**

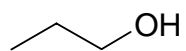
Part M2: codes index the features of **zero- or monovalent** carbon chains

Part M3: codes index the features of **di- or polyvalent** carbon chains

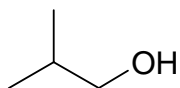
The codes indicate the following information:

- valency of the chain
- number of carbons in the chain
- type of group attached to the chain
- type of chain branching
- number of times the codes are used

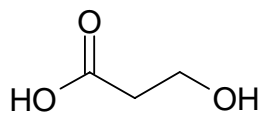
Example 19



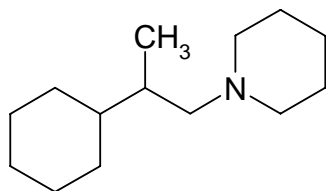
monovalent, M213, M231, M272, M281



monovalent, M214, M232, M272, M281



divalent, M312, M321, M332, M342, M381, M391



divalent, M313, M321, M331, M342, M373, M391

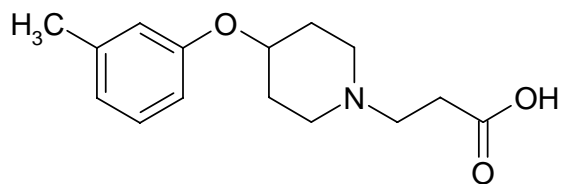
**Example 20**

Write down the part M1: to M3: codes for these structures:

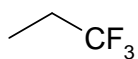
(a)



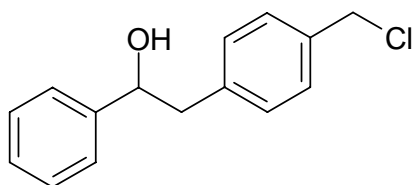
(b)



(c)



(d)



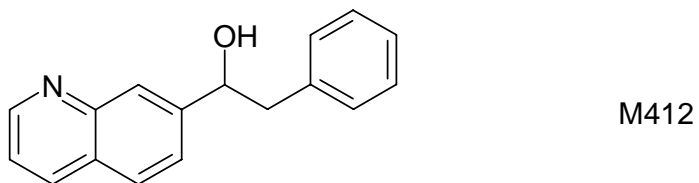
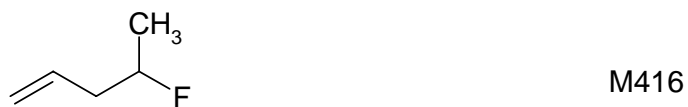
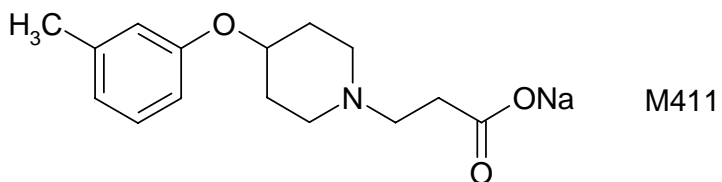
**Part M4: Basic group codes**

Only **one** code from part M4: is indexed for each structure, and the codes are listed below in order of priority (i.e. M411 > M412 > M413 > M414 > M415 > M416 > M417)

- M411 Miscellaneous; inorganic
- M412 Fused heterocyclic
- M413 Mononuclear heterocyclic
- M414 Carbocyclic aromatic
- M415 Alicyclic
- M416 Aliphatic
- M417 Incomplete structure

M411 includes organic compounds containing elements other than C, H, O, N, S and halogen.

**Example 21**



## Part M5: Ring systems present

This section describes the number of ring systems present in the molecule:

M51: number of fused heterocyclic ring systems present

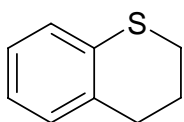
M52: number of monocyclic heterocyclic ring systems present

M53: number of carbocyclic aromatic ring systems

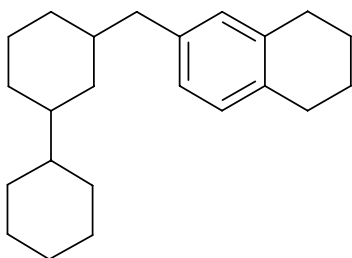
M54: number of carbocyclic alicyclic systems present

For this section, each whole ring system is counted, rather than individual rings within a fused system

Examples:



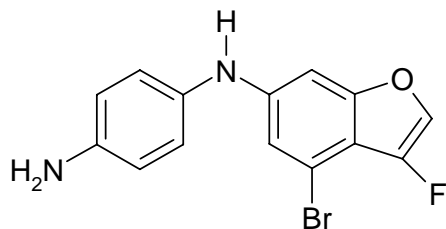
indexed M511 (not M531)



indexed M531, M542

**Example 22**

Write down the codes from parts D to M for this structure, and group together the appropriate colour codes:



BLACK:

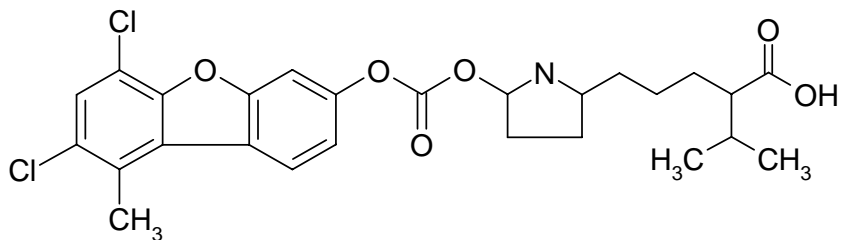
RED:

BLUE:

GREEN:

**Example 23**

Write down the codes from parts D to M for this structure, and group together the appropriate colour codes:



BLACK:

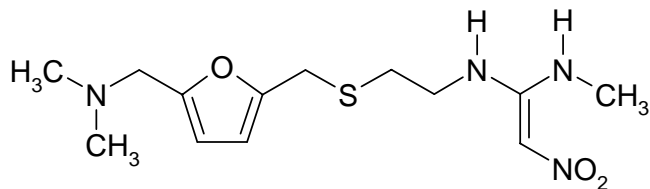
RED:

BLUE:

GREEN:

**Example 24**

Write down the codes from parts D to M for this structure, and group together the appropriate colour codes:



BLACK:

RED:

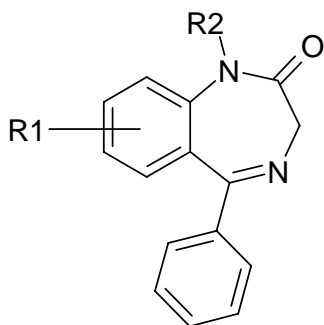
BLUE:

GREEN:

### Example 25

In this example we start to look at how to incorporate variable groups into the search strategy. In this simple case, the codes for the various halogens, and also the codes to indicate the different possible positions of substitution of the R1 group can be OR-ed in the search strategy

Write down the codes from parts D to M for this structure, and group together the appropriate colour codes:



R1 = Halogen  
 R2 = 1-6C alkyl

BLACK:

RED:

BLUE:

GREEN:

## 'OR'-ing codes in the standard search strategy

When the structure search query contains variable groups, special care must be taken when OR-ing the fragmentation codes for the alternatives. This is especially true when the alternative groups are searched with codes from different time periods (i.e. different colour codes).

When combining codes with OR, check the table below to see what line the codes should be entered.

CODE	Black	Red	Blue	Green	Green# (precursor Black)	Green# (precursor Blue)
Black	1	*	*	*	1	*
Red	*	2	*	*	*	*
Blue	*	*	3	*	*	3
Green	*	*	*	4	*	*
Green# (precursor Black)	1	*	*	*	1	*
Green# (precursor Blue)	*	*	3	*	*	3

1 = Line 1 (Black line) of the standard search strategy

2 = Line 2 (Red line) of the standard search strategy

3 = Line 3 (Blue line) of the standard search strategy

4 = Line 4 (Green line) of the standard search strategy

\* = see notes in (3) below

There are three general cases:

1. The alternatives are all searched with the same colour codes. In this case the codes can be OR-ed together on the same line.

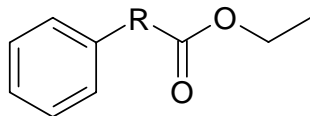
**Example 26:** CH<sub>3</sub>-CH<sub>2</sub>-X ; where X= Cl, Br, I

Cl and Br and I are all indexed with black codes, so these codes can be OR-ed together on the **black line** of the strategy

(H602 OR H603 OR H604)

2. The alternatives are searched with different colour codes, and some codes are marked with #. In this case, the colour of the codes should be noted and the table consulted to determine which line to add the codes.

**Example 27**



R= O or NH

For R=O, the black code for carbonate is searched (L472). For R=NH, the green code for urethane L462 is searched. L462 is marked #, so the older generic black code L460 can also be searched. The **black line** of the strategy contains:

(L472 OR L462 OR L460)

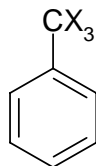
3. The alternatives are searched with different colour codes, and no codes are marked with #. This is often the most difficult case, and there are three possible alternative solutions:

(a) delete the codes for the variable group, and rely on the other codes for the structure to retrieve the hits required

(b) OR the codes on the **lowest line** of the strategy that any alternative appears. This will retrieve the hits required, but also some noise

(c) make completely separate search strategies for the alternative structures, and OR together the final answers online.

**Example 28**



X = F, Cl, Br, I

The following codes make up the strategy:

BLACK: G100, (H600, (H601 not indexed for CF3), H602, H603, H604), H685, M414, M531

RED: no codes

BLUE: M280, M311, M321, (M332 not indexed for 1C chain), M344, M391

GREEN: H686, M353#

The  $-CX_3$  substituent can be searched using the black code H685 ( $-CF_3$ ), and the green code H686 ( $-CX_3$ , X is Cl, Br, I).

Looking at the 3 alternatives:

(a) delete the codes – in this case this is not recommended, as the codes in question define the only functional group present

(b) OR the codes on the green line of the strategy. The search will run correctly, but some noise will result

(c) make two completely separate search strategies, one for  $X=F$ , the second for  $X=Cl, Br, I$ . This will produce the most accurate results. The two strategies are shown below, these are combined to form a complete answer set (note negation codes not shown):

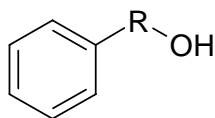
```
S (G100(P)H685(P)M414(P)M531)/M0,M2,M3,M4
S L1
S L2(P)(M280(P)M311(P)M321(P)M344(P)M391(P)(M353 OR M350))/M2,M3,M4
S L3(P)G010/M2,M3,M4
S (L1(P)M900/M0) OR (L2(P)M901/M2,M3,M4) OR (L3(P)M902/M2,M3,M4) OR L4

S (G100(P)(H600 OR H602 OR H603 OR H604)(P)M414(P)M531)/M0,M2,M3,M4
S L6
S L7(P)(M280(P)M311(P)M321(P)M344(P)M391(P)(M353 OR M350))/M2,M3,M4
S L8(P)(G010(P)H686)/M2,M3,M4
S (L6(P)M900/M0) OR (L7(P)M901/M2,M3,M4) OR (L8(P)M902/M2,M3,M4) OR L9

S L5 OR L10
```

### Example 29

Write the codes required to search for this molecule:



$R = -CH_2-CH_2-CH_2-$  or  $-CH_2-CH(CH_3)-$

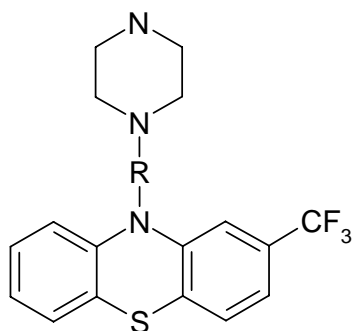
BLACK:

RED:

BLUE:

GREEN:

**Example 30**



R = direct bond, -O- or 2-6C alkylene

When R = direct bond, the -N-N- bond is searched with:

H212 (black code)  
 (K640 OR K600) (green code with black precursor)

When R = -O-, the -N-O-N- bond is searched with:

H212 (black code)  
 (K810 OR K800) (green code with black precursor)

When R = 2-6C alkylene, the -N-alkylene-N- is searched with:

H182, H202 (black codes)  
 (M312 OR M313 OR M314 OR M315), M342 (blue codes)

Looking at the table, the cases where R = direct bond and R = -O- can be OR-ed on line1 of the strategy, but the case R = 2-6C alkylene has blue codes which are best searched in a separate strategy.

STN strategy for R = direct bond and R = O (negation codes not shown):

```
=>S ("E800"(P)F553(P)H212(P)H685(P)M412(P)((K600 OR K640) OR (K800 OR
K810)))/M0,M2,M3,M4 \>_line1
=>S _line1(P)(M511(P)M521(P)M530(P)M540)/M2,M3,M4 \>_line2
=>S _line2(P)(M280(P)M311(P)M321(P)M344(P)M391(P)(M350 OR M353))/M2,M3,M4
\>_line3
=>S _line3(P)(D011(P)D022(P)F011)/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
```

STN strategy for R = 2-6C alkylene:

```
=>S ("E800"(P)F553(P)H182(P)H202(P)H685(P)M412)/M0,M2,M3,M4 \>_line1
=>S _line1(P)(M511(P)M521(P)M530(P)M540)/M2,M3,M4 \>_line2
=>S _line2(P)(M280(P)M311(P)(M312 OR M313 OR M314 OR M315)
(P)M342(P)M344(P) M391)/M2,M3,M4 \>_line3
=>S _line3(P)((M350 OR M353)(P)(M380 OR M383))/M2,M3,M4 \>_line4
=>S _line4(P)(D011(P)D022(P)F011)/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
```

These two separate strategies could then be run online, and the final answers from each combined with 'OR' to produce the final answer set.

## Incorporating free-sites into the search strategy

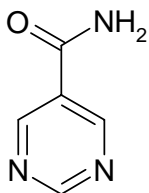
These examples show how to modify the specific compound strategy created by STN Express or Topfrag to incorporate free sites, or to make the query a sub-structure query. The basic technique is to generate a search strategy for the compound with no free sites, and then examine each code and modify as appropriate. For each code, we have 3 choices:

Option 1- leave the code unchanged, this can be done if the code is still suitable for the strategy with free sites.

Option 2 - delete the code – this is done if the code is not applicable to the structure when it contains free sites

Option 3 – add in extra codes to broaden the scope of the search. This option is used especially for codes that indicate the number of a particular feature present

### Example 31



Strategy 1, was generated by STN Express for the specific 5-amido pyrimidine structure.

#### Strategy 1

```
=>S (F541(P)J311(P)M413)/M0,M2,M3 \>_line1
=>S _line1(P)(M521(P)M530(P)M540)/M2,M3 \>_line2
=>S _line2(P)(M280(P)M320)/M2,M3 \>_line3
=>S _line3(P)(F015(P)J011)/M2,M3 \>_line4
=>S (_line1(P)M900/M0) OR (_line2(P)M901/M2,M3) OR (_line3(P)M902/M2,M3)
\>_line5
=>S _line5 OR _line4 \>_line6
=>S _line6(NOTP)(H1 OR H2 OR H3 OR H4 OR H5 OR H6 OR H7 OR H8 OR H9 OR
J1)/M2 \>_line7
=>S _line7(NOTP)(J2 OR J4 OR J5 OR J9 OR K0 OR M1)/M2 \>_line8
```

Strategy 2 is a modified version of strategy 1. This has been edited by hand - to make the original query into a sub-structure query.

## Strategy 2

```
=>S (F541(P)(J311 OR J312)(P)(M411 OR M412 OR M413))/M0,M2,M3 \>_line1
=>S _line1(P)(M521 OR M522 OR M523)/M2,M3 \>_line2
=>S _line2 \>_line3
=>S _line3(P)(F015(P)J01!)/M2,M3 \>_line4
=>S (_line1(P)M900/M0) OR (_line2(P)M901/M2,M3) OR (_line3(P)M902/M2,M3)
\>_line5
=>S _line5 OR _line4 \>_line6
```

### For strategy 2:

#### Line 1:

'J312' has been included, to allow for the possibility of there being more than one heterocyclic amide.

Also, 'M411' has been included because there may be a substituent that contains a P, Si or B which would make the whole cpd. 'Organometallic'. For a similar reason 'M412' has been included to allow for the possibility of a fused ring heterocyclic.

#### Line 2:

Codes 'M530' and 'M540' have been removed, to allow for the presence of those rings and, 'M522' and 'M523' included, to allow for the possible presence of 2 or more mono-heterocyclic rings.

#### Line 3:

'M280' and 'M320' have been removed to allow for the possible presence of carbon chains. The line number has been kept, otherwise all subsequent lines would need re-numbering and the time range line would need adjusting.

#### Line 4:

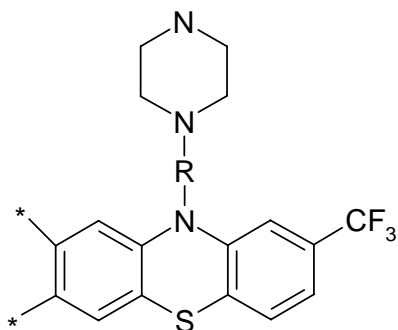
'J011' has been replaced with 'J01!' to incorporate the possibility of 2 or more carboxylic acid derivs. being present. Alternatively, this can be written: (J011 or J012 or J013 or J014).

Also, Line 7 and line 8 have been deleted, i.e. the negation codes, to make this into a sub-structure query - i.e. allow any kind of substituent(s) on the core structure.

The codes added or deleted in strategy makes this strategy as generic as possible without losing the specificity of the original query. This method is easier than the addition of free sites in the structure drawing for WPI.

**Example 32**

Use the strategies from example 30 to write full search strategies for this structure



R = direct bond, -O- or 2-6C alkylene

\* = free sites, may form a ring

## Parts A, B, and C fragmentation codes

Part A: indexes metals present in the compound, and how they are bonded to the rest of the structure

- A1: Alkali metals
- A2: Alkaline earth metals
- A3: Group IIIA to IVA
- A4: First transition series
- A5: Second transition series
- A6: Third transition series
- A7: Lanthanides
- A8: Actinides
- A9: General metal descriptors

Part B: indexes metalloids or noble gases present

- B0: Noble gases
- B1: Noble gas, B, Si, P, As, Se, Te as element, inorganic compound, ion or ligand
- B2: Noble gas, B, Si, P, As, Se, Te in inorganic anion of an organic compound
- B3: Noble gas, B, Si, P, As, Se, Te in inorganic cation of an organic compound
- B4: Noble gas, B, Si, P, As, Se, Te in aliphatic compound
- B5: Noble gas, B, Si, P, As, Se, Te in carbocyclic compound
- B6: Noble gas, B, Si, P, As, Se, Te in heterocyclic compound
- B7: Bonds to B, Si, P, As ("Z" atoms)
- B8: Valency and number of Z atoms

Part C: indexes non-metallic elements (i.e. Halogens, O, S, N, C and H) present in *inorganic* compounds. Additionally, valency codes are used to indicate high valence states of these elements in *organic* compounds.

- C0: Halogens (including organic halogens in Part K, and L511, L512, L760)
- C1: Halogen (valency I or II), H, C, N (valency II, III, IV), O, S (valency II or III), in inorganic compound, ion or ligand
- C2: Halogen (valency III or IV), S (valency IV or V)
- C3: Halogen (valency V or more), N (valency V), S (valency VI)
- C4: -O-O-, -S-S-, -S=S- in inorganic compound, ion or ligand
- C5: Inorganic compound, ion, ligand consisting solely of C, N, O and/or S
- C6: Halogen as a ring heteroatom
- C7: Inorganic compound or ion not containing noble gas, B, Si, P, As, Se, Te
- C8: Miscellaneous descriptors, including elements absent, isotopes

### Indexing rule:

**For ionic and co-ordinate compounds, each ion or ligand is indexed separately, in the same indexing paragraph. The ion or ligand is indexed as if only one of each is present.**

### Example 33

calcium acetate is indexed :

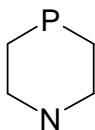
J011, J171 *one* –COOH group  
 M211, M262 *one* carbon chain  
 M630 metal salt  
 A960 metal to oxygen of organic ion  
 C710 inorganic cation  
 M411 basic group  
 A220 calcium

### Example 34

Pt(PPh<sub>3</sub>)<sub>4</sub> is indexed

A678 Platinum  
 A970 metal to other atom  
 B515 P in carbocyclic compound  
 B720 No Z-S bonds  
 B743 Z-C  
 B813 3 valent P  
 B831 1 Z atom  
 Plus part G: and M: codes

### Example 35



Indexed:

B615 P in heterocyclic compound  
 B720 No Z-S  
 B742 2 Z-C bonds  
 B760 Z-H bond  
 B813 3 valent P  
 B831 1 phosphorus  
 B840 Z as ring heteroatom  
 F000 unsubstituted heterocycle  
 F433 piperidene ring

Note: rings containing unusual heteroatoms are indexed with ring codes from part D, E, F, G, as if the heteroatom is carbon. The exceptions are Se and Te, which are assigned ring codes as if the heteroatom was sulphur.

M411 basic group code

**Example 36**

Butyl lithium

A103 Lithium  
 A910 One metal-carbon bond  
 Plus part M codes

**Example 37**

Sodium carbonate

A111 Sodium  
 A940 Metal-inorganic anion  
 C106 Carbon  
 C108 Oxygen  
 C530 carbonic acid derivative  
 C730 inorganic compound  
 C801 metalloid absent  
 C802 halogen absent  
 C803 hydrogen absent  
 C805 nitrogen absent  
 C807 sulphur absent

**Example 38**

Write down the codes for the following compounds:

(a)  $\text{H}_2\text{SO}_4$  (part A, B, C only)

C101, C108, C316, C540, C730, C800, C801, C802, C804, C805

(b) Lithium nitrate (A, B, C only)

A103, A940, C108, C307, C510, C730, C801, C802, C803, C804, C807

(c) Sodium ethoxide (all parts A-M)

A111, A960, C710, H401, H481, M210 OR M212, M272, M281, M320, M411, M620, M630

(d) Ferrocene (all parts A-M)

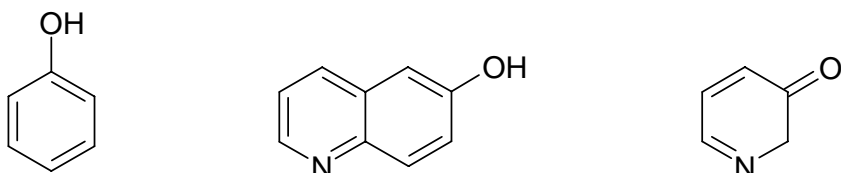
A426, A923, G030, G039, G551, G599, M126, M144, M280, M320, M411, M542

## Appendix 1 - Tautomerism

### 1. Keto-enol tautomerism

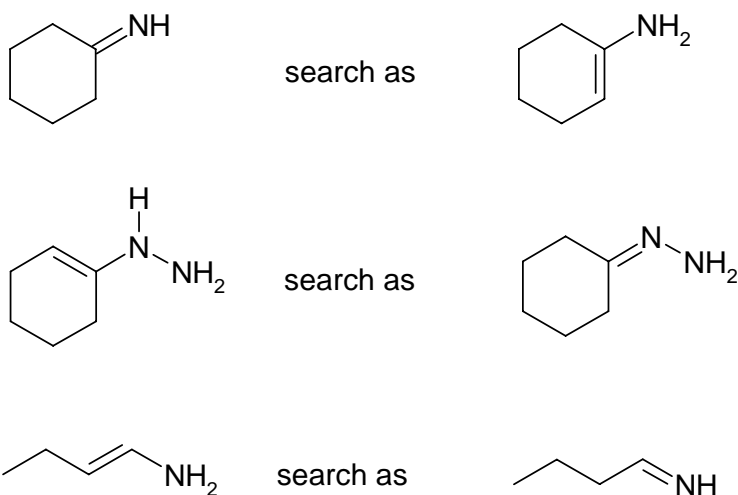
The structure is indexed in keto form, except when the –OH (or –SH) is bonded to a fully conjugated carbocyclic ring, e.g. benzene. In these cases, the enol form is indexed.

Below are some examples of the forms indexed:

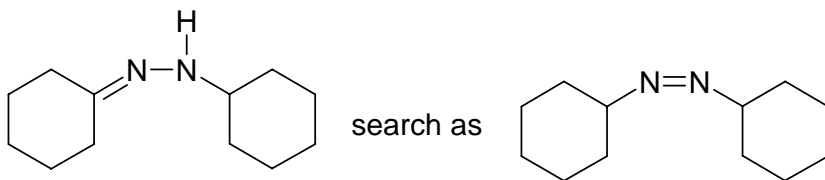


### 2. Imine-amine tautomerism

In this case, the imine form ( $>C=N-$ ) is indexed, except where the C atom is part of a ring, when the  $=C-NH-$  form is indexed, but only when this produces an amine or amide group. If this produces a group other than amine or amide, then the imine form is indexed.



Azo groups (i.e.  $R-N=N-R'$  where R and R' are not H) are an exception:



## Appendix 2 - Common functional groups

Listed below are the most common functional groups and the corresponding fragmentation and negation codes

Functional Group	Fragmentation codes	Negation codes
Acetal	L660	K0, L6
Acid anhydride, carboxylic	L543	K0, L5
Alcohol	H4xx	H4, H8
Aldehyde	J4xx	J4
Alkene	H7xx	H7
Alkyne	H73x	H7
Amide, carboxylic	J0xx, J3xx	J0, J3
Amide, sulfonic	(K353 OR K350), C316	K0, K3
Amidine	L340	K0, L3
Amine	H1xx	H1
Amine oxide	K742	K0, K7
Azide	K510	K0, K5
Azo	(K534 OR K530)	K0, K5
Biguanide	L240	K0, L2
Bromocarbons	H6xx	H6
Carbamic acid	(L46x OR L460)	K0, L4
Carbonate	L472	K0, L4
Carboxylic acid	J0xx, J1xx	J0, J1
Carboxylic acid amide	J0xx, J3xx	J0, J3
Carboxylic acid anhydride	L543	K0, L5
Carboxylic acid halide	L512, C0xx	K0, L5
Carboxylic acid ester	J0xx, J2xx	J0, J2
Chlorocarbons	H6xx	H6
Chlorofluorocarbons	H6xx	H6
Cyanate	L130	K0, L1
Cyanide	(L14x OR L140)	K0, L1
Cyclic anhydride	J522, L922	J5, L9
Ester, carboxylic	J0xx, J2xx	J0, J2
Ester sulfonic	K432	K0, K4
Ether	H5xx	H5, H8
Fluorocarbons	H6xx	H6
Guanidine	L250	K0, L2
Halocarbons	H6xx	H6
Halogen	H6xx	H6
Hydrazide	J0xx, J3xx, K620	J0, J3, K0, K6
Hydrazine	K640	K0, K6
Hydrazone	K630	K0, K6
Hydroperoxide	K920	K0, K9
Imide (cyclic)	J522, L930	J5, L9
Imide (non-cyclic)	L532	K0, L5
Imine	(L355 OR L350)	K0, L3
Isocyanate	L230	K0, L2
Isocyanide	(L146 OR L140)	K0, L1
Isothiocyanate	L220	K0, L2

Ketal	L660	K0, L6
Ketene	H712, H721, J581	H7, J5
Ketone	J5xx	J5
Nitrate	K710	K0, L8
Nitro	H3xx	H3
Lactam	J521, L941	J5, L9
Lactone	J521, L942	J5, L9
Olefin	H7xx	H7
Ortho-carboxylic acid ester	L630	K0, L8
Peracid	K910, J0xx, J1xx	J0, J1, K0, K9
Perester	K910, J0xx, J2xx	J0, J2, K0, K9
Peroxide	K930	K0, K9
Quinone	L951	K0, L9
Sulfate	K421	K0, K4
Sulfonamide	(K353 OR K350),C316	K0, K3
Sulfonic acid	K431	K0, K4
Thioaldehyde	J4xx, J490	J4, J9
Thiocarbamic acid, di	L440	K0, L4
Thiocarbamic acid, mono	L450	K0, L4
Thiocarbonate	L471	K0, L4
Thiocarboxylic acid	J19x	J1, J9
Thiocarboxylic amide	J3xx, J390	J3, J9
Thiocarboxylic ester	J2xx, J290	J2, J9
Thiocyanate	L120	K0, L1
Thioether	H59x	H5, H9
Thioketone	J59x	J5, J9
Thiol	H49x	H4, H9
Thiourea	L420	K0, L2
Urea	L432	K0, L4
Urethane	L46x	K0, L4

### Appendix 3 - generic fragmentation codes

There are some generic fragmentation codes available to include in a search. These generic codes are indexed to indicate a general reference in the patent. If the search is intended to be exhaustive, then these generic codes should be considered for inclusion in the search strategy.

Note that in most cases the Topfrag software does not generate these codes, so if they are required they should be added to the strategy by hand.

<b>Code</b>	<b>Definition</b>
A100	Alkali metal, general (Black)
A200	Alkaline earth metal, general (Green)
A300	Group IIIA-VA, general (Green)
A400	First transition metal series, general (Black)
A500	Second transition series, general (Black)
A600	Third transition series, general (Black)
A700	Lanthanides, general (Black)
A800	Actinides, general (Red)
B000	Noble gases, general (Black)
C000	Halogen, general (Green)
D010	Non-specific substitution on the heterocyclic ring in a fused system (Green)
D020	Non-specific substitution on the carbocyclic ring in a fused system (Green)
D040	Fused ring heterocycle, general (Black)
F010	Non-specific substitution on a mononuclear heterocycle (Green)
F020	Aromatic mononuclear heterocycle, general (Black)
F021	Non-aromatic mononuclear heterocycle, general (Black)
G001	Non-specific substitution on benzene (Green)
G002	Non-specific substitution on fused aromatic ring (Green)
G003	Non-specific substitution on fused or unfused alicyclic ring (Green)
G040	Aromatic carbocycle, general (Black)
G050	Unfused alicyclic ring, general (Black)
G051	Fused alicyclic ring system, general (Black)
H600	Halogen, general (Black)

# 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 Intra-aural	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 Intra-aural	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	R515 # Others (specific)
R52:-3: Other proc./apparatus (specific)	R520 Agglutination precipitation	R521 Biological; fermentation	R522 Coating	R523 Compressant; 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](mailto:helpdesk@fiz-karlsruhe.de)  
Internet: [www.stn-international.de](http://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](mailto:help@cas.org)  
Internet: [www.cas.org](http://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](mailto:helpdesk@jaici.or.jp) (Technical Service)  
[cas-stn@jaici.or.jp](mailto:cas-stn@jaici.or.jp) (Customer Service)  
Internet: [www.jaici.or.jp](http://www.jaici.or.jp)