DCR Reference Manual

A supplement for the DERWENT WORLD PATENTS INDEX® STN online user guide

DWPI Chemical Resource produced by Thomson Scientific, August 2007

STN

FIZ Karlsruhe

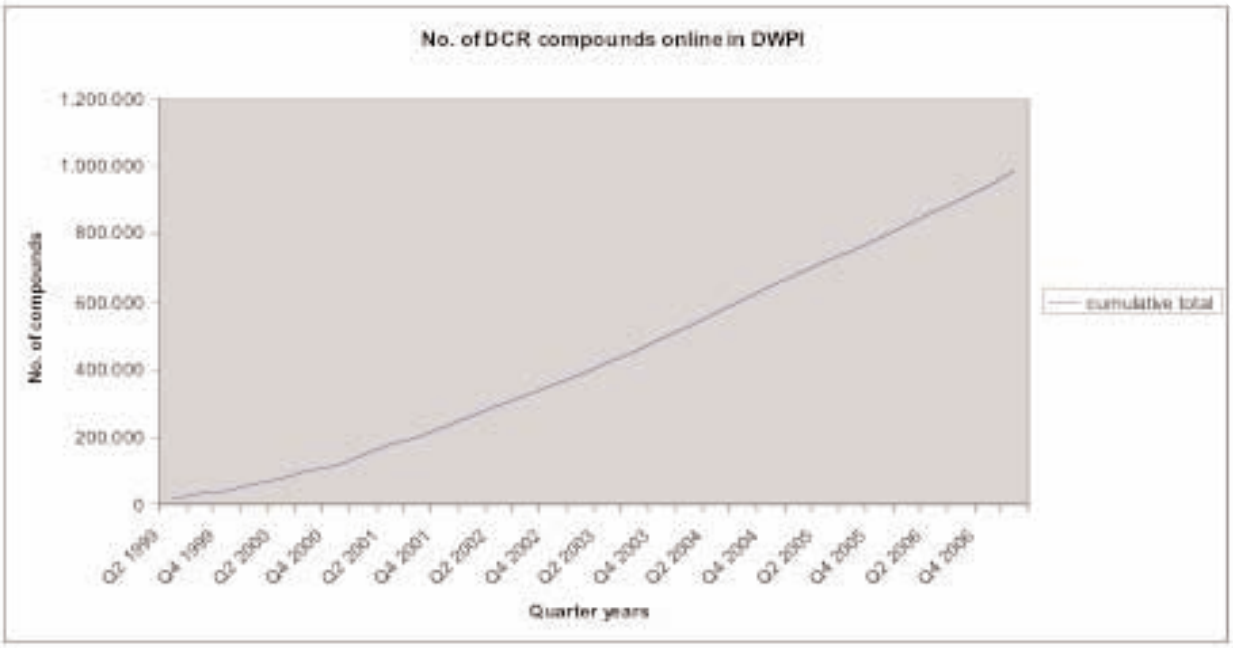
Content

Introduction

The DWPI Chemistry Resource is a chemical structure database for searching specific chemical structures indexed in Derwent World Patent Index® (DWPI) bibliographic records. DWPI Chemistry Resource was released on STN in August 1999 as an integral part of the DWPI suite of files. As of spring 2007 there are about one million compounds in DCR. The database is searchable both by chemical structure and by various text fields, allowing simple access to the DWPI database requiring only a minimum of specialist knowledge. DWPI Chemistry Resource indexing commenced in DPWI update 199916, and runs in parallel to, and to a certain extent replicates, current subscriber Chemical Indexing (Fragmentation Codes) for patents classified in Chemical Patents Index (CPI) Sections B (Pharmaceuticals), C (Agrochemicals) and/or E (General Chemicals).

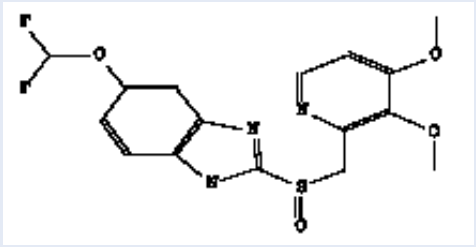
DCR Numbers, which are unique identifiers for specific chemical compounds, form the link between the DWPI Chemistry Resource chemical structure database and corresponding bibliographic indexing in DWPI.

The DCR on STN is available to all searchers, using standard structure searching techniques via either STN Express, command line or STN on the Web. The database is provided as a seamless part of DWPI files WPINDEX, WPIDS and WPIX, rather than a separate file.



A Sample Record

```
L7    ANSWER 17 OF 17  WPIX COPYRIGHT 2007        THE THOMSON CORP on STN
AN.S  DCR-111250
DCSE  111250-0-0-0
CN.P  PANTOPRAZOLE
CN.S  5-Difluoromethoxy-2-(3,4-dimethoxy-pyridin-2-ylmethanesulfinyl)-1H-
      benzoimidazol e
SY    CONTROLOC; INIPOMP; PANTOLOC; PANTOPRAZOLE; PANTOZOL; PEPTAZOL;
      PROTIUM; PROTONIX; PROTONIX-IV; RIFUN; SKF-96022; SOMAC
```

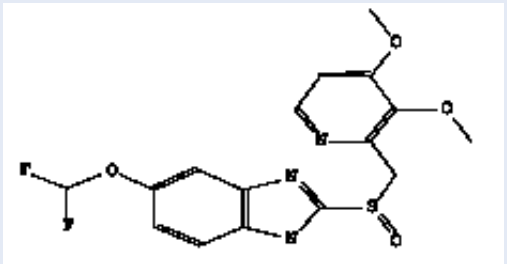


```
MF    C16 H15 F2 N3 O4 S
SMF   C16 H15 F2 N3 O4 S *1; TOTAL *1; TYPE *1
MW    383.3765
SDCN  R22667
```

AN.S contains the DCR Number primary key which is the unique and unambiguous structure identifier. This is also used for crossing over to the bibliographic (DWPI) file segment.

DCSE contains the Enhanced DCR Number which contains information about stereochemistry, isotopes or charges and may serve to aggregate related compounds by masking parts of it.

```
L7    ANSWER 16 OF 17  WPIX COPYRIGHT 2007        THE THOMSON CORP on STN
AN.S  DCR-159347
DCSE  111250-0-1-0
CN.P  PANTOPRAZOLE SODIUM
SY    PANTOPRAZOLE SODIUM; PROTIUM; SOMAC

      CM    1

      Na

      CM    2
```



```
MF    C16 H15 F2 N3 O4 S . Na
SMF   C16 H15 F2 N3 O4 S *1; TYPE *2; TOTAL *2; Na *1
MW    406.3635
SDCN  RA10NM
```

Structure Searching

Structure searching is available on STN Express, command line or STN on the Web. The following are features relating to the structure searching of DCR (additional manuals covering general structure searching techniques are available from STN).

Structure search modes available in DCR are:

- Substructure (SSS)
- Closed Substructure (CSS)
- Family (FAM)
- Exact Match (EXA)

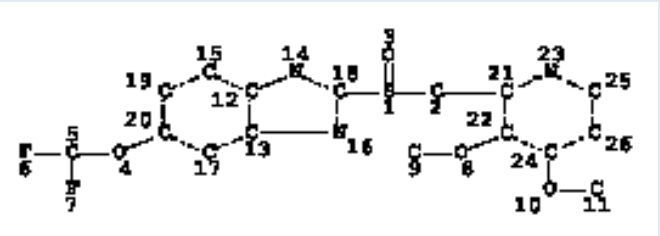
Structure search scopes available are

- Sample (SAM)
- Full Substructure Search (FUL)
- Subset Search (based on answer sets resulting from structure and text searches)
- Range can be set.

Structure Modelling and a Simple Start

In DCR either the templates built-in at STN or the structures in the database can be used as templates on the command line. Here an already known chemical compound is called by its DCR number and used as a template:

```
=> stru 111250
ENTER (DIS), GRA, NOD, BON OR ?:.
```



```
ENTER (DIS), GRA, NOD, BON OR ?:end
L10  STRUCTURE CREATED

=> s l10 full
FULL SEARCH INITIATED 11:15:14
FULL SCREEN SEARCH COMPLETED -        20 TO ITERATE

100.0% PROCESSED        20 ITERATIONS                                  19 ANSWERS
SEARCH TIME: 00.00.05

L11           19 SEA SSS FUL L10
```

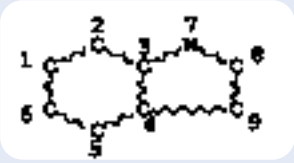
Subset Searching

Subset searching is suitable for refining structure searches or when combing text searches with structure searches. If the subset search is based on a previously conducted substructure search the charge incurred is considerably reduced compared to a full substructure search.

Subset searching a text search answer set

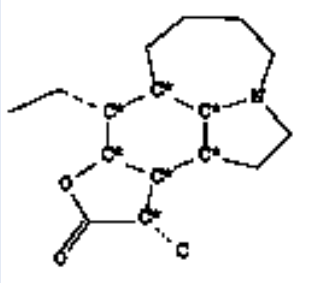
Please note that the text search needs to be pointing to the DCR file segment text data otherwise the subset search will not yield any results.

```
=> stru none
ENTER (DIS), GRA, NOD, BON OR ?:gra r65
ENTER (DIS), GRA, NOD, BON OR ?:nod 7 n
ENTER (DIS), GRA, NOD, BON OR ?:.
```




```
ENTER (DIS), GRA, NOD, BON OR ?:end
L1    STRUCTURE CREATED


=> e alkaloid/cc
E#      FILE             FREQUENCY   TERM
--      ----             ---------   ----
**** START OF FIELD ****
E3      WPIX                     0 --> ALKALOID/CC
E4      WPIX                  3129     ALKALOIDS/CC
E5      WPIX                   123     ALLOYS/CC
E6      WPIX                   478     ANTHRACYCLINES/CC
E7      WPIX                   125     ANTIBODIES/CC
E8      WPIX                  1047     BARBITURATES/CC
E9      WPIX                  5556     BENZODIAZEPINES/CC
E10     WPIX                  1369     BETA LACTAMS/CC
E11     WPIX                    24     BORANES/CC
E12     WPIX                     6     CARBOHYDRATE/CC

=> s e4
L2          3129 ALKALOIDS/CC


=> s l1 sss ful sub=l2
FULL SUBSET SEARCH INITIATED 10:36:07
FULL SUBSET SCREEN SEARCH COMPLETED -      2171 TO ITERATE

100.0% PROCESSED      2171 ITERATIONS                                  584 ANSWERS
SEARCH TIME: 00.00.05

L3            584 SEA SUB=L2 SSS FUL L1
```

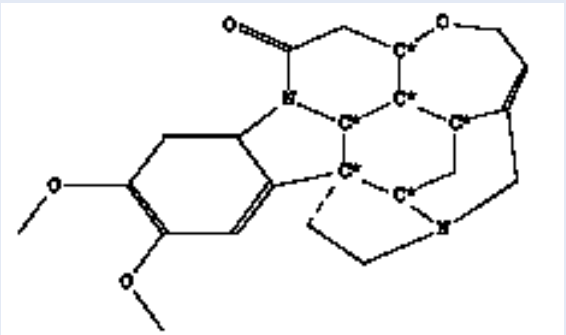
```
=> d 43 all

L3   ANSWER 43 OF 584  WPIX COPYRIGHT 2007       THE THOMSON CORP on STN
AN.S DCR-1195457
DCSE 107809-5-0-0
CN.P NEOSTENINE
SY   NEOSTENINE
```



```
MF   C17 H27 N O2
SMF  C17 H27 N O2 *1; TOTAL *1; TYPE *1
MW   277.4102
SRIN 13070
SDCN RAKC9Q
CC   ALKALOIDS
```

Subset searching a structure search answer set

```
=> stru none
ENTER (DIS), GRA, NOD, BON OR ?:gra r7
ENTER (DIS), GRA, NOD, BON OR ?:nod 1 o
ENTER (DIS), GRA, NOD, BON OR ?:end
L4   STRUCTURE CREATED

=> s l4  sss full sub=l3
FULL SUBSET SEARCH INITIATED 10:46:08
FULL SUBSET SCREEN SEARCH COMPLETED -      168 TO ITERATE

100.0% PROCESSED      168 ITERATIONS                          11 ANSWERS
SEARCH TIME: 00.00.01

L5            11 SEA SUB=L3 SSS FUL L4


=> d 11

L5   ANSWER 11 OF 11  WPIX COPYRIGHT 2007       THE THOMSON CORP on STN
AN.S DCR-3457
DCSE 3457-1-0-0
CN.P BRUCINE
SY   2,3-DIMETHOXYSTRICHNIDIN-10-ONE; BRUCINE; BRUZIN; CANIRAMIN;
     DIMETHOXYSTRYCHNINE; VOMICINUM
```



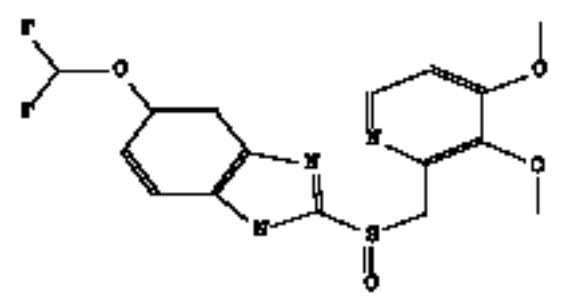
```
MF   C23 H26 N2 O4
```

Crossing over into the bibliographic segment

Starting from the results of the Pantoprazol search the corresponding bibliographic documents can be retrieved by requalifying with /DCR. The DCR references in the chemical and enhanced polymer coding fields are then searched for the structure identifiers laid down in the answer set of the structure search.

```
=> s l11/dcr
L12         332 L11/DCR

=> d max hitstr


L12  ANSWER 12 OF 332  WPIX COPYRIGHT 2007        THE THOMSON CORP on STN
AN   2006-767310 [ 78]    WPIX  Full-text
ED   20061204
DNC  C2006-237782 [ 78]
TI   Use of a proton pump inhibitor e.g. omeprazole, lanosprazole in the
     treatment of sleeping disturbance due to silent gastroesophageal reflux
DC   B02
IN   FERNSTROEM P; HASSELGREN G
PA   (ASTR-C) ASTRAZENECA AB
CYC  111
PI   WO 2006118534   A1 20061109 (200678)* EN  22[ 0]
ADT  WO 2006118534 A1 WO 2006-SE535 20060503
PRAI US 2005-680932P       20050512
     SE 2005-1041          20050504
IPCI A61K0031-4164 [ I,C] ; A61K0031-4184 [ I,A] ; A61K0031-4427 [ I,C] ;
     A61K0031-4439 [ I,A] ; A61P0001-00 [ I,C] ; A61P0001-04 [ I,A] ; C07D0235-00
     [ I,C] ; C07D0235-28 [ I,A] ; C07D0401-00 [ I,C] ; C07D0401-12 [ I,A] ;
     C07D0401-14 [ I,A] ; C07D0471-00 [ I,C] ; C07D0471-04 [ I,A]
AB    WO 2006118534 A1   UPAB: 20061204
      NOVELTY - In the treatment of sleeping disturbance due to silent gastroesophageal
reflux, a proton pump inhibitor (PPI) is administered. ACTIVITY - Hypnotic;
Antiinflammatory; Gastrointestinal-Gen.. Patients suffering from sleeping disturbance due
to silent gastroesophageal reflux were evaluated. A total of 53 reflux events, which were
associated with 41 awakenings and 128 arousals were observed. All reflux events were
associated with either an arousal or awakening or both. Subjects with reflux were analy-
zed pre- and post-treatment with omeprazole. After treatment with omeprazole the number
of awakenings preceded by reflux events decreased from 3.7 plus minus 0.9 - 1.3 plus
minus 0.5. The number of arousals proceeded by reflux events decreased from 11.6 plus
minus 3.8 - 1.5 plus minus 0.8 and the total time (pH less than 4) decreased from 38.7
plus minus 13.7 - 5.3 plus minus 1.6 minutes.
      MECHANISM OF ACTION - Proton pump inhibitor; H+ ATPase inhibitor; K+ ATPase inhibi-
tor.
      USE - For treating sleeping disturbance due to silent gastroesophageal reflux
(claimed).
      ADVANTAGE - The use of proton pump inhibitor improves sleep; reduces risk of deve-
loping esophagitis; prevents development of Barett's esophagus/adenocarcinoma and reduces
the use of hypnotics in this group of patients. It also limits the amount of fluid excre-
ted by the stomach, reduces intervariability between patients and shows more effective
acid secretion inhibition than therapeutic amounts of other drugs with this effect.
TECH ORGANIC CHEMISTRY - Preferred Compound: The proton pump inhibitor is a
     substituted benzoimidazole compound of formula (Ia).
ABEX  SPECIFIC COMPOUNDS - Use of omeprazole, lansoprazole, pantoprazole,
     rabeprazole, esomeprazole, tenatoprazole, ilaprazole, leminoprazole their
     salts and/or enantiomer as the PPI, are specifically claimed.
      EXAMPLE - No suitable example is given.
IT   UPIT 20061204
     76120-CL 76120-USE; 99135-CL 99135-USE; 111250-CL 111250-USE; 269446-CL
     269446-USE; 109574-CL 109574-USE; 730862-CL 730862-USE; 99239-CL
     99239-USE; 93863-CL 93863-USE; 1393483-CL 1393483-USE; 1393484-CL
     1393484-USE; 1393485-CL 1393485-USE; 1393486-CL 1393486-USE
FS   CPI
MC   CPI: B05-A01B; B06-D05; B14-D03; B14-E10A; B14-J01B1; B14-L12
```

```
CMC   UPB    20061204
      M2 *01*    C216 D012 D022 D711 F012 F013 F014 F015 F431 H5 H521 H541 H8 K0
                  K4 K442 L922 M210 M211 M240 M272 M282 M311 M321 M342 M373 M391
                  M412 M431 M511 M521 M530 M540 M781 M782 P420 P445 P616 P617 P714
                  M905 M904
                  DCN: R04401-K R04401-M R04401-U
                  DCR: 76120-K 76120-M 76120-U
      M2 *02*    C216 D013 D711 F012 F013 F014 F431 H5 H521 H6 H685 H8 K0 K4 K442
                  L922 M210 M211 M240 M281 M311 M312 M321 M332 M342 M344 M362 M373
                  M391 M412 M431 M511 M521 M530 M540 M781 M782 P420 P445 P616 P617
                  P714  M905 M904
                  DCN: R22683-K R22683-M R22683-U
                  DCR: 99135-K 99135-M 99135-U
      M2 *03*    C216 D012 D022 D711 F012 F013 F014 F431 H5 H522 H541 H6 H601
                  H608 H684 H8 K0 K4 K442 L922 M210 M211 M272 M282 M311 M322 M342
                  M343 M362 M373 M391 M412 M431 M511 M521 M530 M540 M781 M782 P420
                  P445 P616 P617 P714  M905 M904
                  DCN: R22667-K R22667-M R22667-U
                  DCR: 111250-K 111250-M 111250-U
      M2 *04*    C216 D012 D022 D711 F012 F013 F014 F015 F431 H5 H521 H541 H8 K0
                  K4 K442 L922 M210 M211 M240 M272 M282 M311 M321 M342 M373 M391
                  M412 M431 M511 M521 M530 M540 M781 M782 P420 P445 P616 P617 P714
                  M905 M904
                  DCN: RA1IY2-K RA1IY2-M RA1IY2-U
                  DCR: 269446-K 269446-M 269446-U
      ...
AN.S  DCR-111250
CN.P  PANTOPRAZOLE
CN.S  5-Difluoromethoxy-2-(3,4-dimethoxy-pyridin-2-ylmethanesulfinyl)-1H-
      benzoimidazol e
SDCN  R22667
```



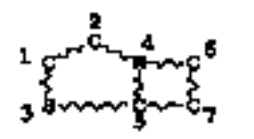
In order to allow for effective crossing over from the chemical repository to the bibliography segment in DWPI, Thomson Scientific has equipped DCN and DRN numbers with the corresponding DCR numbers for chemical and enhanced polymer coding. This was part of the 2006 DWPI reload, where the backfile for these numbers was populated and run against a DCN and DRN correspondence list. Please note that some DRN entries in particular in the polymer area do not have a DCR equivalent. Therefore in special cases leveraging the DRN numbers is still being called for. For instance 5214 (Fibre, Glass), 5111 (Fuller's Earth) or 5234 (Jute) don't have a DCR number since they are not well-defined specific chemical compounds.

Current Awareness Searching

Structure searching for current awareness purposes can be conducted in the DCR segment of DWPI either by setting up an SDI or running your own scripts. Structure SDIs can be set up to deliver the results in hardcopy or softcopy form or as an online answer set delivered to your online account. The latter is recommended if subsequent crossover into the bibliography segment is required. Below the procedure to accomplish this is illustrated. Please note that after crossing over into the bibliography segment you'll probably need to confine the results to those DWPI documents having been updated and containing a reference to the chemical compound from your structure search result set.

Setting up the SDI:

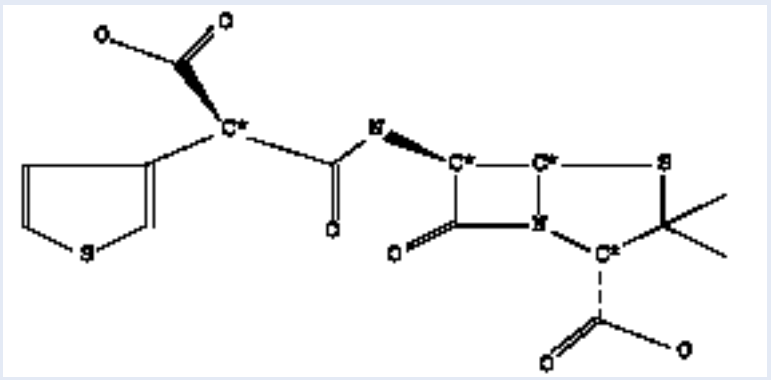
```
=> fil wpix
...
=> stru penicl
ENTER (DIS), GRA, NOD, BON OR ?:.
```




```
ENTER       (DIS), GRA, NOD, BON OR ?:end
L1          STRUCTURE CREATED

=> sdi
ENTER       QUERY L# FOR SDI REQUEST OR (END):l1
ENTER       SDI REQUEST NAME, (AA001/S), OR END:PENICL/S
ENTER       COST CENTER (NONE) OR NONE:.
ENTER       TYPE OF SEARCH (SSS), CSS, FAMILY, OR EXACT:.
ENTER       TITLE (NONE):Carbapenam SSS
ENTER       METHOD OF DELIVERY (OFFLINE), ONLINE, OR EMAIL:online
ELIMINATE   PREVIOUSLY SEEN ANSWERS WITH EACH SDI RUN?   Y/(N):.
HIGHLIGHT   HIT TERMS? (Y)/N:y
ENTER       SDI RUN FREQUENCY - WEEKLY, (EVERYUPDATE), MONTHLY, OR ?:.
ENTER       SDI EXPIRATION DATE 'YYYYMMDD' OR (NONE):.
QUERY       L1 HAS BEEN SAVED AS SDI REQUEST 'PENICL/S'
```

Collecting the results:

```
=> d sav/a
NAME              CREATED       NOTES/TITLE
--------------    -----------   -----------------------------------
PENICL20/A        24 MAR 2007   8 ANSWERS IN FILE WPIX


=> act penicl20/a
TITLE: CARBAPENAM SSS
L1                STR
L2              8 SEA FILE=WPIX SSS SDI L1

=> d

L2          ANSWER 1 OF 8  WPIX COPYRIGHT 2007        THE THOMSON CORP on STN
AN.S        DCR-108920
DCSE        108920-1-0-0
CN.P        TICARCILLIN
CN.S        6-(2-Carboxy-2-thiophen-3-yl-acetylamino)-3,3-dimethyl-7-oxo-4-thia-
            1-aza-bicyclo[ 3.2.0] heptane-2-carboxylic acid
SY          AERUGIPEN; TICARCILLIN; TICARPEN; TRIACILLIN
```



```
MF          C15 H16 N2 O6 S2
```

Crossing over into the bibliography segment:

```
=> s l2/dcr
L3          1607 L2/DCR
```

Confine to those bibliographic documents actually having been updated at the desired time, here run #20 :

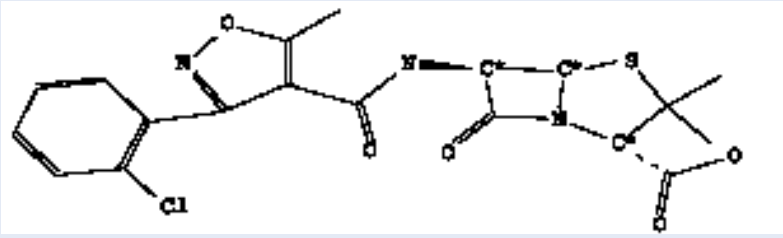
```
=> s l3 and  200720/dw.b
            8045 200720/DW.B
L9             3 L3 AND  200720/DW.B
```

or

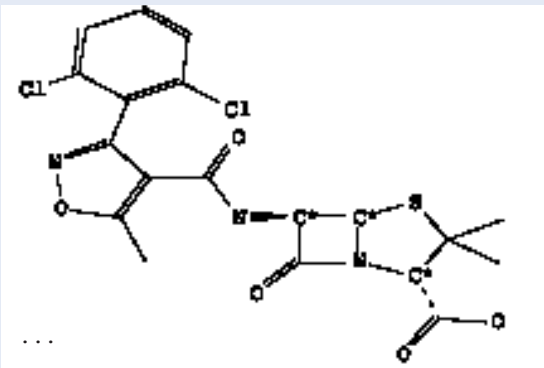
```
=> s l3 and  20070322/upit
             915 20070322/UPIT
                   (20070322/UPIT)
L10            3 L3 AND  20070322/UPIT
```

Display the full record including the hit structures:

```
=> d full hitstr

L10   ANSWER 1 OF 3  WPIX COPYRIGHT 2007        THE THOMSON CORP on STN
AN    2007-200390 [ 20]    WPIX  Full-text
DNC   C2007-073120 [ 20]
TI    Biocompatible release system useful for e.g. drug delivery system
      comprises inorganic component dispersed inside polymer matrix and has
      lamellar structure with neutralized charge to intercalate within the
      structure of active principle
DC    A96; B05; B07; C03; C07; D22
IN    BOLOGNESE A; CALIFANO L; CALIGNANO A; COSTANTINO U; MARENZI G;
      SAMMARTINO G; VITTORIA V
PA    (BOLO-I) BOLOGNESE A; (CALI-I) CALIFANO L; (CALI-I) CALIGNANO A;
      (COST-I) COSTANTINO U; (MARE-I) MARENZI G; (SAMM-I) SAMMARTINO G;
      (VITT-I) VITTORIA V
CYC   113
PI    WO--2007010584  A2 20070125 (200720)* EN  31[ 1]
ADT   WO--2007010584 A2 2006WO-IT0000556  20060721
PRAI  2005IT-RM0000393  20050722
IPCI  A61K [ ,S]
AB    WO 2007010584 A2   UPAB: 20070322
      NOVELTY - Biocompatible release system comprises a polymer matrix; an inorganic
      component dispersed inside the matrix and having a lamellar structure with a
      neutralized net positive or negative charge able to intercalate within the
      lamellar structure of a pharmacologically active principle, establishing an
      ionic type bond with the inorganic component. The combination of the inorganic
      component and of the active principle constitutes the intercalation compound.
      DETAILED DESCRIPTION - An INDEPENDENT CLAIM is included for preparation of  the
      release system involving treating the inorganic component in such a way as to
      confer a net positive or negative charge, then combining it with the active
      principle also in an ionic form or transformed in such a way as to confer an
      ionic nature, thereby obtaining an intercalation compound which is then mixed
      with the polymer matrix.
      ...
AN.S  DCR-91382
CN.P  CLOXACILLIN
CN.S  6-{[ 3-(2-Chloro-phenyl)-5-methyl-isoxazole-4-carbonyl] -amino} -3,3-
      dimethyl-7-oxo -4-thia-1-aza-bicyclo[ 3.2.0] heptane-2-carboxylic acid
SDCN  R00225
SDRN  0225
```



```
AN.S  DCR-92977
CN.P  DICLOXACILLIN
CN.S  6-{[ 3-(2,6-Dichloro-phenyl)-5-methyl-isoxazole-4-carbonyl] -amino} -3,3-
      dimethyl-7-oxo-4-thia-1-aza-bicyclo[ 3.2.0] heptane-2-carboxylic acid
SDCN  R07291
```



```
...
[ five more compounds]
```

Reference Part

The Individual Fields

The set of fields for search, select, sort and display available for DCR is listed below. There aren't any 'mixed' fields where search terms for DCR and the bibliography segment of DWPI are available side by side. Hence there aren't any items from DCR indexed in the basic index (/BI) of DWPI.

Field Code

Code_	Name	
AN.S	SEA/DIS	DWPI Chemistry Resource Number, DCR Segment
CC	SEA/DIS	Classification Code
CMF	SEA	Component Molecular Formula
CMF.CNT	SEA	Component Molecular Formula Count
CMT	SEA/DIS	Comment
CN	SEA/DIS	Chemical Name
CN.P	SEA/DIS	Chemical Name Preferred
CN.S	SEA/DIS	Systematic Chemical Name
CNS	SEA	Chemical Name Segment
CT*	SEA/DIS	Controlled Term
CT.DA*	SEA/DIS	Controlled Term Drug Activity
CT.MA*	SEA/DIS	Controlled Term Mechanism
DDRN*	SEA/DIS	Derwent Drug Registry Name
DCSE	SEA/DIS	DWPI Chem. Res. Number Enhanced, DCR Segment
EDCR	SEA/DIS	Entry Date DWPI Chemical Repository
ELS	SEA	Element Symbol
ELS.CNT	SEA	Element Symbol Count
FRAGMF	SEA	Fragment Molecular Formula
FRAGMF.CNT	SEA	Fragment Molecular Formula Count
MF	SEA/DIS	Molecular Formula
MW	SEA/DIS	Molecular Weight
NC	SEA	Number of Components
NFRAG	SEA	Number of Fragments
SCR	SEA/DIS	Structure Cross Reference
SDCN	SEA/DIS	Structure Segment Derwent Compound Number
SDRN	SEA/DIS	Structure Segment Derwent Registry Number

SMF	SEA/DIS	Standardized Molecular Formula
SRIN	SAE/DIS	Structure Segment Ring Index Number
SS*	SEA/DIS	Substructure Term
STR	DIS	Chemical Structure Display
SY	SEA/DIS	Synonym Name
UPCR	SEA/DIS	Update Date DWPI Chemical Repository
UPWX	SEA/DIS	Update Date DWPI Cross Reference

* Sparsely occupied fields

The Predefined Display Formats

FORMAT

Default format: STD

TRIAL – CN, CN.S, MF, STR

SCAN – CN, CN.S, MF, STR

STD – AN.S, DCSE, CN, CN.S, STR, SCR, CMT, MF
Syn
IDE

ISTD – AN.S, DCSE, CN, CN.S, STR, SCR, CMT, MF

ALL – AN.S, DCSE, CN, CN.S, STR, SCR, CMT, SMF, MW
syn SDCN, SDRN
FULL

IALL – AN.S, DCSE, CN, CN.S, STR, SCR, CMT, SMF, MW, SRIN,
syn SDCN, SDRN
IFULL

MAX – AN.S, DCSE, CN, CN.S, STR, SCR, CMT, MF, SMF, MW, SRIN,
SDCN, SDRN, DDRN*, CC, CT, SS

IMAX – AN.S, DCSE, CN, CN.S, STR, SCR, CMT, MF, SMF, MW, SRIN,
SDCN, SDRN, DDRN*, CC, CT*, SS

Related DWPI format :

HITSTR – The DCR hit record which led to the retrieval of the bibliographic record.

Identifiers

DWPI Chemical Resource Number (AN.S)

DCR Numbers, the unique and unambiguous compound identifiers, have been assigned on a regular basis from 1999, for some backlog compounds reaching back to 1987, and are found in both segments of WPIDS, WPIX and WPINDEX: in the AN.S field (DCR segment) and in the Indexing Terms (IT), Chemical Coding (Mo-6) and Polymer Indexing (PLE) fields (bibliographic segment).

The DCR number can be up to ten-digits long providing scope for many new additions. In the primary key index field for the DCR segment (/AN.S) it is indexed with a 'DCR-' prefix :

```
=> e 0/an.s
E#      FILE   FREQUENCY   TERM
--      ----   ---------   ----
****  START OF FIELD ****
E3      WPIX         0 --> 0/AN.S
E4      WPIX         1     DCR-1/AN.S
E5      WPIX         1     DCR-10/AN.S
E6      WPIX         1     DCR-100/AN.S
E7      WPIX         1     DCR-1000/AN.S
E8      WPIX         1     DCR-10000/AN.S
E9      WPIX         1     DCR-100000/AN.S
E10     WPIX         1     DCR-1000001/AN.S
E11     WPIX         1     DCR-1000002/AN.S
E12     WPIX         1     DCR-1000003/AN.S
```

In the bibliographic segment, the DCR Numbers are also indexed along with their appropriate Role Qualifiers if available. Role Qualifiers can be used to refine the search further, if this is required. There are two distinct sets of roles available depending on the field. The (T) proximity operator should be used to link the DCR Numbers to the chosen Role Qualifier in the Indexing Terms (IT) or Chemical Coding (Mo-6) fields. Entries in the Polymer Indexing don't carry a role.

All references to DCR reference entries in the bibliography segment of DWPI are indexed in a universal search field (/DCR) which can be used to cross over from the structure to the bibliography segment of the DWPI file regardless whether the references stem from Indexing Terms, Chemical Coding or Polymer Indexing . The entries in /DCR have their roles indexed if available, with different role types depending on the origin of the DCR number. Those numbers genuinely indexed in the Indexing Terms field have their DCR type roles attached, and those generated from DCN and DRN numbers have the roles attached taken over from DCN or DRN.

DWPI Chemistry Resource Number (DCR) Roles

The DCR numbers occur in the index terms (IT) section, in the chemical coding, and in the polymer coding section. The DCR numbers are indexed in /DCR, /IT, /Mo-M6 and /PLE index fields. DCR Roles can be searched on their own, or linked with DCR Numbers in the /IT or the /DCR and /Mo-M6 fields. The (T) proximity operator is used to link Roles to DCR Numbers, e.g. => S (87874(T)PRD)/IT; S L2/DCR(T)NEW/IT.

See HELP DCR for further background information.

There are two different versions of roles potentially attached to the DCR numbers: The single letter roles in the chemical coding, and the multiple letter roles in the index term section. Both are indexed in /DCR accordingly.

[1] This had been different in the previous version of the DWPI file, where different structure identifiers had to be employed for comprehensive retrieval. This has been improved upon by Thomson Scientific through back file indexing of DCR numbers for DCN and DRN entries by correspondence lists.

The following DCR roles are available from 1999 onwards in the /IT section.

Role	Definition	Scope Notes
CL	CLAIM	Applied to compounds present in the patent claims (1999-date).
EX	EXAMPLE	Applied to compounds present in the examples, but not in the claims (from update 200253).
DISC	DISCLOSURE	Applied to compounds present in the disclosure, but not in the claims nor in the examples (from update 200253)
NEW	NEW	Substance, process, or apparatus claimed or described as new. (Before 1999 rarely applied.)
PRD	PRODUCED	Production or manufacture of substance or apparatus is claimed or described.
USE	USE	Use of substance or apparatus is claimed or described.
DET	DETECTED	Applied to the keyword for a condition or substance which has been detected as a result of testing.
RCT	REACTANT	Applied to starting materials or products defined in terms of starting materials (1987-date)
RGT	REAGENT	Applied to reaction components apart from starting materials e.g. catalysts, purifying agents (1987-date)
CMP	COMPONENT	Applied to components of a mixture (1987-date)
PUR	PURIFIED	
REM	REMOVED	
TES	TESTED	
ST	SALT	Applied to alkali or alkaline earth metal salts of organic acids; also to certain salts of organic bases e.g. hydro halides, acetates.

In the wake of the DCR back-propagation effort , documents prior to 1999 have been algorithmically equipped with DCR numbers in the indexing terms field as well. For this purpose the following relationship between single-letter and IT roles has been assumed.

```
A DET
C RGT
D RGT
M CMP
N NEW
P PRD
Q RCT
R RGT
S RCT
U USE
X REM
Z DIS

If the source DCN has no role or an
invalid role, the DCR number receives
the role DIS (as the safest default).
```

The set of single-letter roles which is also available for DCR numbers, is identical to those available for DWPI Compound Numbers (DCN). (see below).

Here is an example for linking a role with the structure identifier (CL= claimed/):

```
=> s l1/dcr(t)cl/it
           31 L1/DCR
        66622 CL/it
L3           31 L1/DCR(T)CL/it    <-- Corresponding bibliographic records
```

Structured DCR Number (DCSE)

The structured DCR Number (/DCSE) has a logical format, so that isomers and salts share a common 1-8 digit numerical stem - which can be searched without needing to use truncation.

For example: 3-Methyl-cyclotetradec-5-enone isomers (structured DCR number stem 270633)

```
=> s 270633/dcse
L1             6 270633/DCSE

=>  e 270633/dcse
E1             1      270630/DCSE
E2             1      270630-0-0-0/DCSE
E3             6 --> 270633/DCSE
E4             1      270633-1-0-0/DCSE
E5             1      270633-2-0-0/DCSE
E6             1      270633-3-0-0/DCSE

E7             1      270633-4-0-0/DCSE
E8             1      270633-5-0-0/DCSE
E9             1      270633-6-0-0/DCSE
E10            2      270638/DCSE
E11            1      270638-1-0-0/DCSE
E12            1      270638-2-0-0/DCSE

=> s e3
L2             6 270633/DCSE
```

Structured DCR Number Format

The stem of a structured DCR Number is the same for many related compounds (see above), but with suffixes to indicate, e.g. stereochemistry, salts, isotopes and physical forms. The format is as follows:

ooooooooo-oo-oo-oo

ooooooooo	1 to 8-figure sequential number (allows up to 100 million compounds)
oo	First suffix for stereo isomers (number from 1-99)
oo	Second suffix for salts (number from 1-99)
oo	Third suffix to deal with other cases such as physical forms, isotopes, tautomers, etc. (number from 1-99)

Numbers are not filled out with leading "0"s so relevant numbers appear online with a minimum of 4 digits, i.e. 1-0-0-0, etc.

The characters after the first hyphen (the first suffix) are for stereochemistry: 0 the default, used for compounds with no stereo centres or where the stereochemistry is not defined. For any compounds with stereo centres, the next available number is used, i.e. 1 for the first stereoisomer encountered, 2 for the next one, etc up to 99.

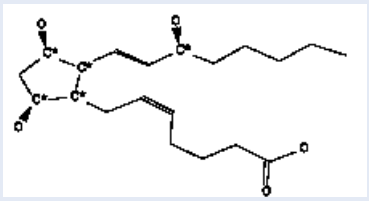
The characters after the second hyphen (the second suffix) are for salts: 0 The default, used for the free acid or free base. The next available number is then used for the next salt encountered. The use is restricted to Group I and II or "simple" metal or amine salts of acids, simple (inorganic) salts of bases e.g. halogens etc. N.B. Inorganics and organometallic complexes will have unique identifiers, as will most organic salts consisting of an organic acid and organic base.

The last characters (the third suffix) are for other cases where related compounds might be associated: These include isotopes, tautomers, different physical forms and other cases not covered above. This section is also used if there is a need for a special version of a structure for a particular file or service. Sequential numbers will be assigned whenever different forms appear. 0 is the default, used for the parent compound.

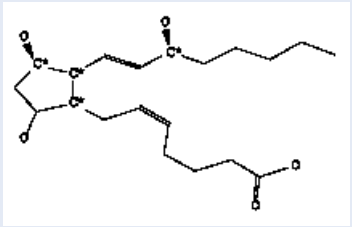
Structure Cross Reference (SCR syn XCR)

When chemical structures have related structures in DCR the related compounds can be cross referenced. The format is DCR number followed by colon and a descriptor of the relationship, e.g. SEE ALSO or ISOMER. There can be multiple cross references in one DCR record.

```
=> e
E13     WPIX                    1      101946 : SEE ALSO/SCR
E14     WPIX                    1      10240 : SEE ALSO/SCR
E15     WPIX                    1      102739 : SALT PARENT/SCR
E16     WPIX                    1      102861 : SEE ALSO/SCR
E17     WPIX                    1      103181 : DERIVATIVE OR PARTIAL/SCR
E18     WPIX                    1      103245 : SEE ALSO/SCR
E19     WPIX                    1      103524 : SEE ALSO/SCR
E20     WPIX                    1      103537 : SEE ALSO/SCR
E21     WPIX                    1      103743 : ISOMER/SCR
E22     WPIX                    1      103781 : SEE ALSO/SCR
E23     WPIX                    1      103843 : DERIVATIVE OR PARTIAL/SCR
E24     WPIX                    1      103918 : PRECURSOR/SCR

=> s e21
L3             1 "103743 : ISOMER"/SCR

=> d all scr
L3    ANSWER 1 OF 1  WPIX COPYRIGHT 2007        THE THOMSON CORP on STN
AN.S  DCR-111370
DCSE  93389-5-0-0
CN.P  PGF2-ALPHA-EPI-8
CN.S  7-[ 3,5-Dihydroxy-2-(3-hydroxy-oct-1-enyl)-cyclopentyl]-hept-5-enoic acid
SY    PGF2-ALPHA-EPI-8
```



```
MF    C20 H34 O5
SMF   C20 H34 O5 *1; TOTAL *1; TYPE *1
MW    354.4836
SDCN  RA03RZ
CC    PROSTAGLANDINS
SCR   103743 : ISOMER

=> s DCR-103743/an.s
L4             1 DCR-103743/AN.S

=> d all

L4    ANSWER 1 OF 1  WPIX COPYRIGHT 2007        THE THOMSON CORP on STN
AN.S  DCR-103743
DCSE  93389-2-0-0
CN.P  PGF2
CN.S  7-[ 3,5-Dihydroxy-2-(3-hydroxy-oct-1-enyl)-cyclopentyl]-hept-5-enoic acid
SY    GLANDIN-N; HORSAFERTIL; PANACELAN-F; PGF2; PROSTAGLAN;
      PROSTAGLANDIN-F2-ALPHA; PROSTAMODIN-F; U-14583
```



```
MF    C20 H34 O5
SMF   C20 H34 O5 *1; TOTAL *1; TYPE *1
MW    354.4836
SDCN  RA0CZ6
CC    PROSTAGLANDINS
```

Other structure identifiers (SDCN, SDRN, SRIN)

DCR records can also include, where applicable, other (older) compound numbering systems which exist in WPIDS,WPIX and WPINDEX. Three DCR fields are used for these systems: SDCN (Compound Number, DCR segment), SDRN (Registry Number, DCR segment), and SRIN (Ring Index Number, DCR segment). The corresponding fields in the bibliographic segment are DCN (DWPI Compound Number), DRN (DWPI Registry Number) and RIN (Ring Index Number).

In order to extract DCNs, DRNs and RINs from the DCR segment and then search them in bibliographic segment of WPIDS, WPIX or WPINDEX either the SELECT or TRANSFER commands can be used.

Structure Segment DWPI Compound Number (SDCN)

DWPI Compound Numbers (DCN) are Merged Markush Service (MMS) Compound Numbers, for specific compound entries in the MMS database on Questel.Orbit. MMS compound number indexing is available in DWPI on all hosts from 1987 onwards for patents classified in Sections B (Pharmaceuticals), C (Agrochemicals) and/or E (General Chemicals).

DWPI Compound Number (DCN) Roles

DCN roles are searchable appended to individual DWPI Compound Numbers in the DCN field of the bibliographic segment of the DWPI file, e.g. S R00708-P/DCN. They can be searched on their own, or linked to an L-numbered answer set with a proximity operator. The following DCN roles are available from 1987 onwards, except as indicated.

Role	Definition/Notes
A	Substance Analysed/Detected
C	Catalyst
D	Detecting Agent
E	Excipient
K	Known Compound
M	Component of a Mixture
N	New Compound
P	Known Compound Produced
Q	Product Defined in Terms of Starting Materials
R	Removing/Purifying Agent
S	Starting Material
T	Therapeutically Active
U	Use of a Single Compound
V	Reagent
X	Substance Removed
Z	Miscellaneous

Structure Segment DWPI Registry Number (SDRN)

About 2100 commonly occurring chemicals encountered in the claims and examples of patent specifications in DWPI sections B, C, and E have been indexed with unique Registry Numbers since 1981 (DWPI update 198127).

From 1984 (DWPI update 198401) the use of Registry Numbers was extended to cover DWPI sections A, D and H; and from DWPI update 198407, to the remaining chemical sections F, G, and J-M.

Section A (Plasdoc) has a separate list of Registry Numbers for about 750 compounds (or groups of compounds). Of these, approximately 350 are identical to those used in the other CPI sections and have the same numbers. The 400 additional section A compounds have been allocated numbers in the 5,000 series. These numbers in the 5,000 series were discontinued from DWPI update 199501 on the introduction of the Enhanced Polymer Indexing system.

Registry Numbers are searchable with or without the role letter.
Since Registry Numbers are only applied to specific compounds in claims and examples, a search by Registry Number alone does not retrieve unspecified compounds contained within a Markush structure. Registry Numbers do, however, give retrieval of high relevance.

DCR numbers which have been auto generated from the corresponding Registry Numbers are available in the Chemical Coding field.

DWPI Registry Number (DRN) Roles

The Registry Numbers are indexed in the bibliographic segment of DWPI with and without the following roles:

Role	Definition/Scope Notes
S	Intermediate or starting material
P	Compound produced
U	Use of a compound (single use or as a mixture)

The roles are indexed on their own as well.

Searching the DWPI Registry Numbers field (/DRN) in the bibliographic part of DWPI requires the appropriate level of subscription.

Structure Segment Ring Index Number (SRIN)

Ring Index Numbers (RIN) are codes assigned to chemical ring systems that are not precisely defined by appropriate DWPI Chemical Fragmentation Codes (Mo-M6). They are searchable in DWPI from 1972 onwards, for patents classified in Sections B (Pharmaceuticals), C (Agrochemicals) and/or E (General Chemicals).

Patents sometimes mention general types of rings rather than specifying the exact ring system involved in an invention e.g. "aryl" or "aromatic heterocyclic ring system". To enable more specific searches on ring systems, Derwent began assigning ring numbers from The Ring Index (Patterson, Capell and Walker, 2nd edition, American Chemical Society, and its supplements) to patent indexing records in 1972.
These Ring Index Numbers are five digit numbers that appear in the (S)RIN fields of the Derwent World Patents Index database. Although the "Patterson Ring Index" is used as a guide, not all of the Ring Index Numbers are used, since Thomson Scientific does not distinguish between levels of unsaturation or different tautomers.
Ring systems encountered in patent documents but not found in the "Patterson Ring Index" are assigned to RINs by Thomson Scientific numbering from 40,000 onwards.
In the same field "Rarer Fragment Numbers" are included. They were used during the period 1972-1975 to describe less common chemical fragments and were given numbers from 70,000 onwards.
Thomson Scientific has now stopped assigning new RINs (update 199901) but continues to apply existing RINs.
Ring index numbers are searchable in the bibliographic part of DWPI by eligible subscribers only.

Formula Fields

Molecular Formula (MF)

This formula has been calculated from the topological structure data. Molecular formula fragments are separated by dots in this type of Molecular Formula. Individual atoms plus stoichiometric factors are separated by spaces. This molecular formula can also contain words like 'complex'.

```
AN.S DCR-151227
DCSE 49376-1-1-0
CN.P PHENOXYMETHYLPENICILLIN POTASSIUM
...
MF   C16 H18 N2 O5 S . K

SMF  C16 H18 N2 O5 S *1; K *1;
     TOTAL *2; TYPE *2
```

Standardized Molecular Formula (SMF)

This type of molecular formula was introduced in order to improve the indexing for compounds not adequately searchable by structure searching. Hence it was mainly designed for retrieval of co-ordination compounds and salts, but a structured molecular formula is available for all chemical compounds for consistency.

It is a searchable text field that contains terms corresponding to chemical fragments. Each formula fragment represents the molecular formula of the ion or ligand, arranged according to the Hill standard. Individual fragments are separated by semicolons. Stoichiometry factors are linked to each formula fragment by an asterisk. If the stoichiometry is unknown, the factor is left out. The total number of fragments in the compound is shown as 'TOTAL* #', the total number of different types of fragments is indicated by 'TYPE* #'. Elements within a fragment are separated by spaces. There are no rules for the ordering of formula fragments within the SMF.

```
AN.S  DCR-186734
DCSE  186734-0-0-0
CN.P  COBALT TRIS-ETHYLENEDIAMINETRI-
      CHLORIDE
...
MF    C2 H8 N2 . 3 Cl . Co
SMF   C2 H8 N2 *1; Cl *3; Co *1;
      TOTAL *5; TYPE *3
MW    154.4852
SDCN  R07658
```

Component Molecular Formula

Multi-component compounds have the molecular formulae of their contributing fragments or components from the structured molecular formula (SMF) indexed in a separate field.

```
=> s e6
L6     5871 "NA *1"/CMF

=> d max

L6     ANSWER 1 OF 5871  WPIX COPYRIGHT
       2007   THE THOMSON CORP on STN
AN.S   DCR-1438098
DCSE   8975-0-1-0

    CM   1

    Na



    CM   2
```

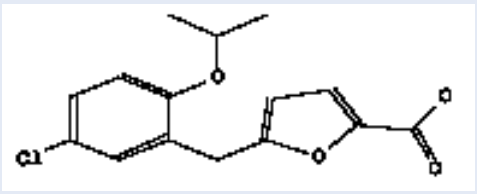

```
MF     C6 H11 Br O2 . Na
SMF    C6 H11 Br O2 *1; Na *1; TOTAL *2;
       TYPE *2
MW     218.0472
SDCN   RAPIDK
```

/FRAGMF is a synonym for /CMF and can be used in lieu.

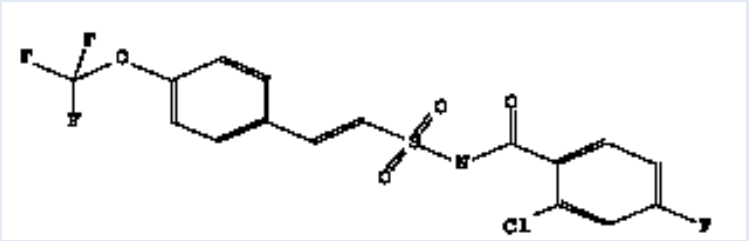
Element Symbol (ELS) and Element Symbol Count (ELS.CNT)

Compounds can be retrieved in DCR by searching the element components of the molecular formula.
For example (ELS= Element Symbol; Na= Sodium; Cl= Chlorine):

```
=> s (na and cl)/els
          9868 NA/ELS
        184521 CL/ELS
L20       1284 (NA AND CL)/ELS
=> d scan
L20  1284  ANSWERS  WPIX         COPYRIGHT 2007 THE THOMSON CORP on STN
MF   C15 H15 Cl O4 . Na
     CM   1
     Na

     CM   2
```



```
HOW MANY MORE ANSWERS DO YOU WISH TO SCAN? (1):2
L20  1284  ANSWERS  WPIX         COPYRIGHT 2007 THE THOMSON CORP on STN
MF   C16 H10 Cl F4 N O4 S . Na
     CM   1
     Na

     CM   2
```



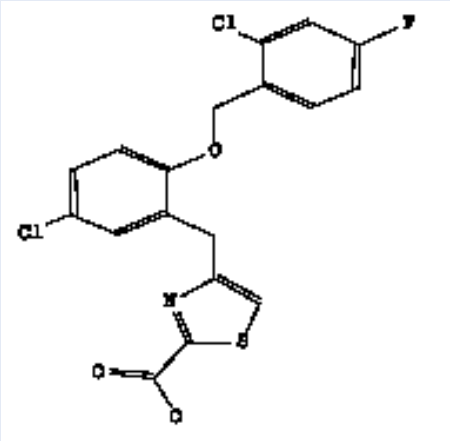
L20 1284 ANSWERS WPIX COPYRIGHT 2007 THE THOMSON CORP on STN

MF C18 H12 Cl2 F N O3 S . Na

CM 1

Na

CM 2

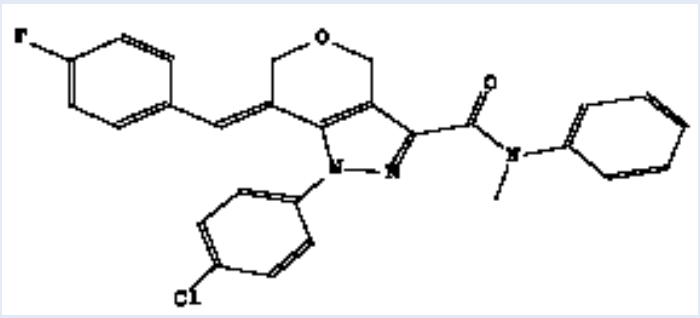


HOW MANY MORE ANSWERS DO YOU WISH TO SCAN? (1):end

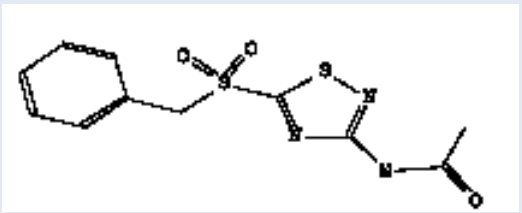
Elements can also be searched by numerical count within the formula using the Element Symbol Count Field (ELS.CNT). This makes use of the (T) proximity operator (default), and is also range searchable.

For example (ELS.CNT= Element Symbol Count; O= Oxygen)

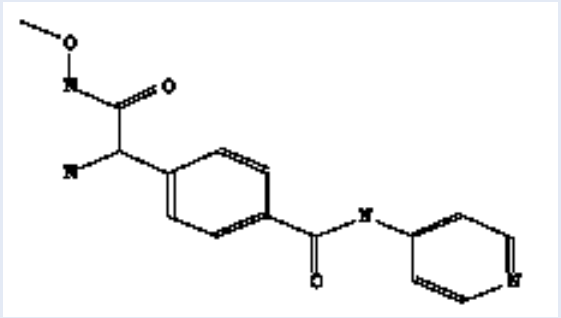
```
=> s o 2-3/els.cnt
         870298 O/ELS
         637224 2-3/ELS
L21      345323 O 2-3/ELS.CNT
                  (O/ELS (T) 2-3/ELS)

=> d scan

L21   345323  ANSWERS  WPIX         COPYRIGHT 2007 THE THOMSON CORP on STN

CN.S 1-(4-CHLORO-PHENYL)-7-[ 1-(4-FLUORO-PHENYL)-METHYLIDENE] -1,4,6,7-TETRAHYDRO-
      PYRANO[ 4,3-C] PYRAZOLE-3-CARBOXYLIC ACID METHYL-PHENYL-AMIDE
MF    C27 H21 Cl F N3 O2
```



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

L21   345323  ANSWERS  WPIX         COPYRIGHT 2007 THE THOMSON CORP on STN

CN.S N-(5-Phenylmethanesulfonyl-1,2,4-thiadiazol-3-yl)-acetamideN-(5-
      Phenylmethanesulfonyl-[ 1,2,4] thiadiazol-3-yl)-acetamide
MF    C11 H11 N3 O3 S2
```



```
L21   345323  ANSWERS  WPIX         COPYRIGHT 2007 THE THOMSON CORP on STN

CN.S 4-(Amino-methoxycarbamoyl-methyl)-N-pyridin-4-yl-benzamide
MF    C15 H16 N4 O3
```

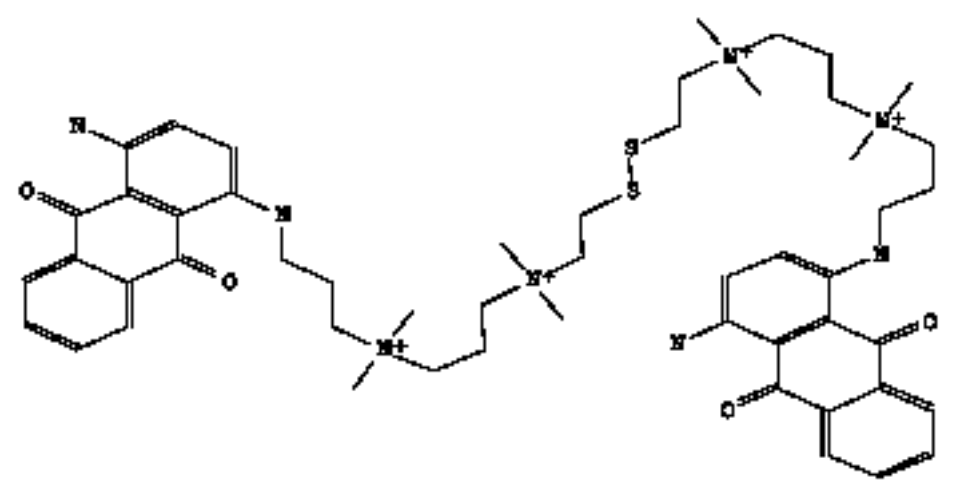


```
HOW MANY MORE ANSWERS DO YOU WISH TO SCAN?  (1):end
```

Number of Components (NC)

The number of components in a structured molecular formula can be numerically searched.
The value is visible in SMF with the heading 'TOTAL'. It will be highlighted there if the value had been searched for.

```
=> e 0/nc
E#      FILE              FREQUENCY   TERM
--      ----              ---------   ----
**** START OF FIELD ****
E3      WPIX                     0 --> 0/NC
E4      WPIX                873481     1/NC
E5      WPIX                 52833     2/NC
E6      WPIX                 17121     3/NC
E7      WPIX                  9057     4/NC
E8      WPIX                  6615     5/NC
E9      WPIX                  1616     6/NC
E10     WPIX                  1472     7/NC
E11     WPIX                   696     8/NC
E12     WPIX                   515     9/NC

=> s e8
L22         6615 5/NC

=> d max

L22  ANSWER 1 OF 6615  WPIX COPYRIGHT 2007        THE THOMSON CORP on STN
AN.S DCR-1435821
DCSE 1435821-0-1-0

      CM   1

      Br


      CM   2
```



```
CMT  1:4 ratio
MF   4 Br . C52 H74 N8 O4 S2
SMF  Br *4; C52 H74 N8 O4 S2 *1; TOTAL *5; TYPE *2
MW   1019.2528
SRIN 03618
SDCN RAPHC6
```

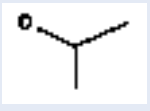
Number of Fragments (NFRAG)

The number of unique fragments in a structured molecular formula can be numerically searched.
The value is visible in SMF with the heading 'TYPE'. It will be highlighted there if the value had been searched for.

```
=> e 0/nfrag
E#      FILE              FREQUENCY   TERM
--      ----              ---------   ----
**** START OF FIELD ****
E3      WPIX                    0 --> 0/NFRAG
E4      WPIX               873517     1/NFRAG
E5      WPIX                72390     2/NFRAG
E6      WPIX                15036     3/NFRAG
E7      WPIX                 3264     4/NFRAG
E8      WPIX                  701     5/NFRAG
E9      WPIX                  136     6/NFRAG
E10     WPIX                   39     7/NFRAG
E11     WPIX                   12     8/NFRAG
E12     WPIX                    4     9/NFRAG

=> e 0/nfrag
E#      FILE              FREQUENCY   TERM
--      ----              ---------   ----
**** START OF FIELD ****
E3      WPIX                    0 --> 0/NFRAG
E4      WPIX               873517     1/NFRAG
E5      WPIX                72390     2/NFRAG
E6      WPIX                15036     3/NFRAG
E7      WPIX                 3264     4/NFRAG
E8      WPIX                  701     5/NFRAG
E9      WPIX                  136     6/NFRAG
E10     WPIX                   39     7/NFRAG
E11     WPIX                   12     8/NFRAG
E12     WPIX                    4     9/NFRAG

=> s e8
L23          701 5/NFRAG

=> d max

L23  ANSWER 1 OF 701  WPIX COPYRIGHT 2007        THE THOMSON CORP on STN
AN.S DCR-1436372
DCSE 1436372-0-0-0

      CM   1

      Al


      CM   2

      Ti


      CM   3
```

```
      CM   4
```

```
      CM   5
```



```
CMT   2:1:2:1:1 ratio
MF    2 C4 H11 N O2 . Al . C3 H8 O . 2 C4 H10 O . Ti
SMF   Al *1; C3 H8 O *1; C4 H10 O *2; C4 H11 N O2 *2; TOTAL *7; TYPE *5; Ti
      *1
MW    314.2396
SDCN  RAPHRB
```

Chemical Name Fields

Chemical Name (CN)

The CN field provides one step searching for names appearing in both the CN.P and SY fields (see below). Any multiple segment names appearing in this field are searchable and expandable as a single bound phrase. If you wish to search or expand the individual fragments of chemical names the Chemical Name Segment (CNS) field should be used instead.

Chemical Name Preferred (CN.P)

This is often, but not always, the first name encountered for the compound by Thomson Scientific Editorial staff. It has real no search significance over and above those names which appear in the SY field, so for complete retrieval CN.P should be searched in combination with the SY field, using the CN search field (see above). Names can originate from any Thomson Scientific product dealing with chemical substances, e.g. the Thomson Scientific Drug File (DDF) database (file DRUGU/DDFU).

Consequently names which appear here do not just originate from patent references in DWPI. Any multiple segment names appearing in this field are searchable and expandable as a single bound phrase. If you wish to search or expand the individual fragments of chemical names the Chemical Name Segment (CNS) field should be used instead.

Systematic Chemical Name (CN.S)

This field is populated by many multiple segment systematic names, separated by hyphens and spaces. The names are generated automatically from the structure drawing using Beilstein AUTONOM® software. Each systematic name is searchable and expandable as a single bound phrase. If you wish to search or expand the individual fragments of chemical names the Chemical Name Segment (CNS) field should be used instead.

Chemical Name Segment (CNS)

The CNS field provides one step searching for name segments appearing in the CN.P, SY or CN.S fields. Multiple segment names are searchable and expandable in this field, as the separate name segments. Chemical names are fragmented for this purpose, at all non-alphanumeric characters, e.g. a space or a hyphen. Simultaneous left and right truncation (SLART) can be used to search for name fragments. This can be particularly useful in conjunction with the term operator (T). If you prefer to search or expand chemical names as a bound phrase the Chemical Name (CN) and/or Systematic Chemical Name (CN.S) fields should be used instead.

Synonym Name (SY)

Synonym names to the preferred name (CN.P), as encountered by Thomson Scientific editorial staff, are recorded here. Names can originate from any Thomson Scientific product dealing with chemical substances, e.g. DDF (file DRUGU/DDFU). Consequently names which appear here do not just originate from patent references in DWPI. Any multiple segment names appearing in this field are searchable and expandable as a single bound phrase. If you wish to search or expand the individual fragments of chemical names the Chemical Name Segment (CNS) field should be used instead.

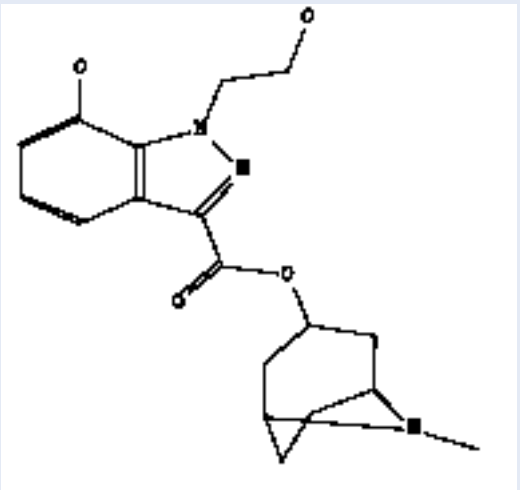
Text Data Fields

Classification Codes (CC)

Substance Descriptors are keywords that relate to classes of compounds, designed for retrieving groups of substances which are difficult, or impossible, to retrieve using a structure query, e.g. general Alkaloids. They are searched in the Classification Code (/CC) field of WPIDS, WPIX or WPINDEX. An alphabetical list of available Substance Descriptors is given below. These are also available by online by entering HELP SDC at the STN command prompt (=›).

For example: Searching 'Alkaloids' in the CC field will retrieve all DCR references to alkaloid compounds. Crossover from DCR to DWPI will therefore retrieve all patents which have references to alkaloid compounds.

```
=> s alkaloids/cc
L1          493 ALKALOIDS/CC

=> d scan

L1    493  WPIDS COPYRIGHT 2007        THE THOMSON CORP on STN

AN.S  DCR-334564
CN.S  7-Hydroxy-1-(2-hydroxy-ethyl)-1H-indazole-3-carboxylic acid
      8-methyl-8-aza-bicyclo[ 3.2.1] oct-3-yl ester
MF    C18 H23 N3 O4
```



List of available Substance Descriptors

This is the complete list of all controlled term keywords which are searchable in the /CC field.

ALKALOIDS

ALLOYS

ANTHRACYCLINES

ANTIBODIES

BARBITURATES

BENZODIAZEPINES

BETA LACTAMS

BORANES

CARBOHYDRATES

- glycoproteins
- polysaccharides
  - cyclodextrins

CARBORANES

CROWN ETHERS

CYCLIC PEPTIDES see PEPTIDES

CYCLODEXTRINS see CARBOHYDRATES

DENDRIMERS

ENZYME see PROTEINS

FATTY ACID see also UNSATURATED FATTY ACIDS

FLAVONOIDS

FULLERENES

GLYCOPROTEINS see CARBOHYDRATES and PROTEINS

HALOCARBONS

HETEROFULLERENES

HETEROPOLY ACIDS

LIPOPROTEINS

METALLOCENES

NOBLE GASES

NUCLEOSIDES

NUCLEOTIDES

- oligonucleotides

OLIGONUCLEOTIDE see NUCLEOTIDES

OTHER NATURAL PRODUCTS

PEPTIDES

- cyclic peptides

PHOSPHOLIPIDS

POLYMERS

POLYSACCHARIDES see CARBOHYDRATES

PROSTAGLANDINS

PROTEINS

- enzymes
- glycoproteins

RETINOIDS

SAPONINS

SILICONES

STEROIDS see SAPONINS

TAXANES

TERPENES

TETRACYCLINES

UNSATURATED FATTY ACIDS see also FATTY ACIDS

ZEOLITES

Comment (CMT)

This is a free text field containing structure descriptions. This is usually provided if there is no structure available to display.

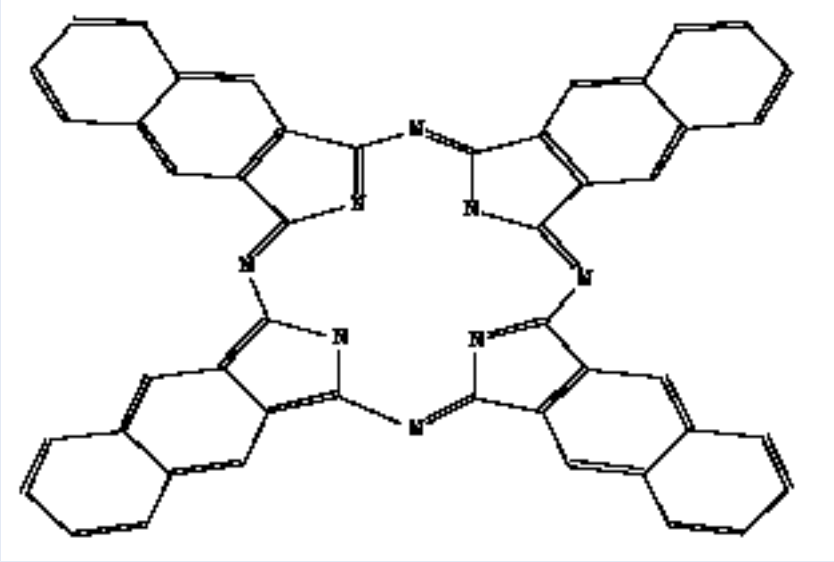
```
=> e peptidase/cmt
E#      FILE              FREQUENCY   TERM
--      ----              ---------   ----
E1      WPIX                     8      PEPPER/CMT
E2      WPIX                     1      PEPSIN/CMT
E3      WPIX                     8 -->  PEPTIDASE/CMT
E4      WPIX                   232      PEPTIDE/CMT
E5      WPIX                    16      PEPTIDES/CMT
E6      WPIX                     1      PEPTIDO/CMT
E7      WPIX                     2      PEPTIDOGLYCAN/CMT
E8      WPIX                     1      PEPTIDOGLYCANS/CMT
E9      WPIX                     1      PEPTIDOMIMETIC/CMT
E10     WPIX                     6      PEPTIDYL/CMT
E11     WPIX                     1      PEPTIDYLGLYCINE/CMT
E12     WPIX                     2      PEPTIDYLPROLINE/CMT

=> s e3
L1             8 PEPTIDASE/CMT

=> d

L1    ANSWER 1 OF 8  WPIX COPYRIGHT 2007        THE THOMSON CORP on STN
AN.S DCR-1231534
DCSE 1231534-1-0-0
CN.P FURIN
SY    DIBASIC PROCESSING ENZYME; DIBASIC-PROCESSING-ENZYME; FURIN; PAIRED
      BASIC AMINO ACID RESIDUE CLEAVING ENZYME; PAIRED-BASIC-AMINO-ACID-
      RESIDUE-CLEAVING-ENZYME; PROHORMONE CONVERTASE

      NO STRUCTURE DIAGRAM AVAILABLE FOR THIS ACCESSION NUMBER

CMT   Belongs to peptidase family, cleaves paired basic amino acid
      residues.
MF    Unknown
```

Structure Display

Structure (STR)

Topological structures in DCR are displayed using the structure display software employed across STN (standard structure conventions at STN apply). However, it is important to note that the structures are drawn at Thomson Scientific employing a different set of software and that the coordinates from the connection tables (also provided by Thomson Scientific) are used as the basis for the displays rather than algorithmically calculated coordinates, as with other files on STN. There may be differences therefore in the displays for chemical compounds between DCR and other STN structure databases like Beilstein or CAS Registry.

```
AN.S DCR-134037
DCSE 134037-0-0-0
CN.P 2,3-NAPHTHOPHTHALOCYANINE
SY   2,3-NAPHTHOPHTHALOCYANINE; NAPHTHALOCYANINE; NAPHTHOPHTHALOCYA-
NINE,
     2,3-; NAPHTHOPHTHALOCYANINE,2,3-; PHTHALOCYANINE,2,3-NAPHTHO-
```



There are limits of sizes fitting on the screen, and if the chemical structure cannot be represented as a topological structure, an error message will be displayed.

```
AN.S DCR-7659
DCSE 7659-0-0-0
CN.P BUCKMINSTERFULLERENE
SY   BUCKMINSTERFULLERENE; BUCKMINSTERFULLERENE C60; FULLERENE C60;
     FULLERENE-C60

     NO STRUCTURE DIAGRAM AVAILABLE FOR THIS ACCESSION NUMBER

CMT  A carbon sixty fullerene
MF   C60
```

Update Dates

Entry Date DWPI Chemical Repository (EDCR)

When a new structure record enters the database it receives a ‘time stamp’, in this case the entry date.

Update Date DWPI Chemical Repository (UPCR)

Whenever a structure record enters the database or a structure record is amended, an update date is created.

Update Date DWPI Cross Reference (UPWX)

Whenever a DCR structure record is referenced in the bibliographic part of DWPI the structure record receives a ‘time stamp’, the DWPI cross reference update date.

All three update dates can be different for any given DCR structure record, for example:

```
 > d an.s upcr edcr upwx

L1   ANSWER 1 OF 71  WPIX COPYRIGHT 2007
THE THOMSON CORP on STN
AN.S DCR-1401361
UPCR 20070115
EDCR 20070102
UPWX 20070119
```

For structure SDIs UPWX is employed.

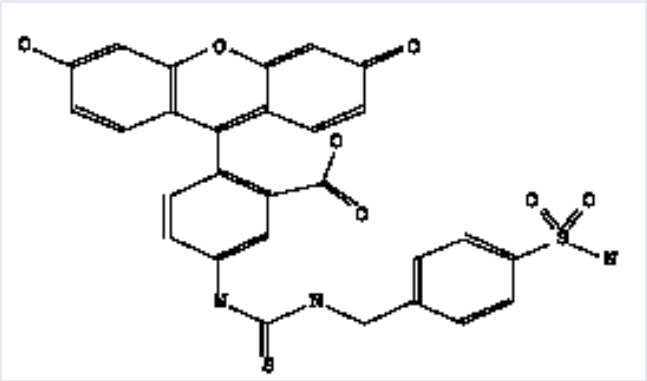
Supplementary Fields

There are some supplementary fields available which don't fit one of the former categories. Some of them are sparsely populated and hence of limited value, but may be useful on occasion.

Controlled term (CT)

The controlled terms field is the sum of its constituents CR.DA and CT.MA (see below) created for Thomson Scientific's Drug File (DDF) database, which is also available on STN.

```
=> e carbonic/ct
E#      FILE              FREQUENCY   TERM
--      ----              ---------   ----
E1      WPIX                    4      CARBOHYDRATE-METABOLISM-STIMULANT/CT
E2      WPIX                    1      CARBOHYDRATE-METABOLISM-STIMULANT./CT
E3      WPIX                    0 --> CARBONIC/CT
E4      WPIX                    1      CARBONIC-ANHYDRASE-I-INHIBITOR/CT
E5      WPIX                    2      CARBONIC-ANHYDRASE-I-INHIBITORS/CT
E6      WPIX                    2      CARBONIC-ANHYDRASE-II-INHIBITOR/CT
E7      WPIX                    2      CARBONIC-ANHYDRASE-II-INHIBITORS/CT
E8      WPIX                    1      CARBONIC-ANHYDRASE-III-INHIBITOR/CT
E9      WPIX                   40      CARBONIC-ANHYDRASE-INHIBITOR/CT
E10     WPIX                    1      CARBONIC-ANHYDRASE-INHIBITOR./CT
E11     WPIX                   39      CARBONIC-ANHYDRASE-INHIBITORS/CT
E12     WPIX                    1      CARBONIC-ANHYDRASE-IX-INHIBITOR/CT

=> s e9
L4            40 CARBONIC-ANHYDRASE-INHIBITOR/CT

=> d max ct

L4   ANSWER 1 OF 40  WPIX COPYRIGHT 2007        THE THOMSON CORP on STN
AN.S DCR-1122772
DCSE 1122772-0-0-0
CN.S 2-(6-Hydroxy-3-oxo-3H-xanthen-9-yl)-5-[ 3-(4-sulfamoyl-benzyl)-
     thioureido] -benzoic acid
```

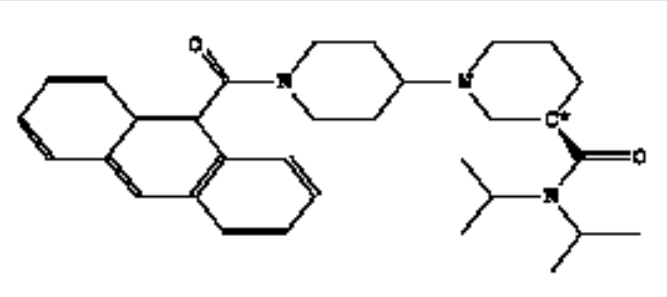


```
MF   C28 H21 N3 O7 S2
SMF  C28 H21 N3 O7 S2 *1; TOTAL *1; TYPE *1
MW   575.6235
SDCN RAM2LB
CT   CARBONIC-ANHYDRASE-II-INHIBITORS; CARBONIC-ANHYDRASE-INHIBITORS;
     CARBONIC-ANHYDRASE-IX-INHIBITORS
     CARBONIC-ANHYDRASE-II-INHIBITOR; CARBONIC-ANHYDRASE-IX-INHIBITOR;
     CARBONIC-ANHYDRASE-INHIBITOR
```

Controlled Term, Drug Activity (CT.DA)

This field contains controlled drug activity terms lifted from the DDF. Since this requires compounds appearing both in DWPI and DDF, the number of compounds in DCR having this field occupied is limited.

```
=> e a/ct.da
E#      FILE             FREQUENCY   TERM
--      ----             ---------   ----
E1      WPIX                     2      5-HT-7-ANTAGONISTS/CT.DA
E2      WPIX                     8      5-HT-UPTAKE-INHIBITORS/CT.DA
E3      WPIX                     0 --> A/CT.DA
E4      WPIX                    12      ABL-TYROSINE-KINASE-INHIBITORS/CT.DA
E5      WPIX                     8      ABORTIFACIENTS/CT.DA
E6      WPIX                     2      ABRASIVES/CT.DA
E7      WPIX                    73      ACARICIDES/CT.DA
E8      WPIX                    10      ACAT-INHIBITORS/CT.DA
E9      WPIX                   108      ACE-INHIBITORS/CT.DA
E10     WPIX                    10      ACETYL-COA-CARBOXYLASE-INHIBITORS/CT.DA
E11     WPIX                     5      ACIDIFIERS/CT.DA
E12     WPIX                    16      ACTH-AGONISTS/CT.DA

=> s e10
L2           10 ACETYL-COA-CARBOXYLASE-INHIBITORS/CT.DA

=> d max ct

L2    ANSWER 1 OF 10  WPIX COPYRIGHT 2007        THE THOMSON CORP on STN
AN.S DCR-785697
DCSE 785697-1-0-0
CN.P CP-640188
CN.S 1'-(Anthracene-9-carbonyl)-[ 1,4'] bipiperidinyl-3-carboxylic acid
      diisopropylamide
SY    CP-640188
```



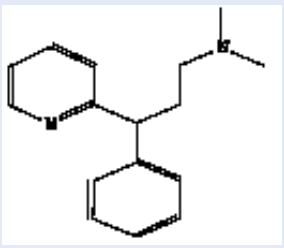
```
MF    C32 H41 N3 O2
SMF   C32 H41 N3 O2 *1; TOTAL *1; TYPE *1
MW    499.7025
SDCN RABV0N
CT    ACETYL-COA-CARBOXYLASE-INHIBITORS; ANTIARTERIOSCLEROTICS;
      TRIAL-PREP.
      ACETYL-COA-CARBOXYLASE-INHIBITOR
```

Controlled Term, Mechanism of Action (CT.MA)

This field contains controlled mechanism of action terms lifted from DDF. Since this requires compounds appearing both in DWPI and DDF, the number of compounds in DCR having this field occupied is limited.

```
=> e serotonin/ct.ma
E#      FILE              FREQUENCY   TERM
--      ----              ---------   ----
E1      WPIX                    1      SEROTININERGIC-1/CT.MA
E2      WPIX                    1      SEROTININERGIC-1D/CT.MA
E3      WPIX                    0 -->  SEROTONIN/CT.MA
E4      WPIX                    1      SEROTONIN-1A SEROTONINERGIC ACTIVITY./CT.MA
E5      WPIX                    1      SEROTONIN-1A-RECEPTOR-LIGAND/CT.MA
E6      WPIX                    1      SEROTONIN-2-LIGAND/CT.MA
E7      WPIX                    1      SEROTONIN-2B-LIGAND/CT.MA
E8      WPIX                   10      SEROTONIN-ANTAGONIST/CT.MA
E9      WPIX                    1      SEROTONIN-ANTAGONIST./CT.MA
E10     WPIX                    2      SEROTONIN-DEPLETOR/CT.MA
E11     WPIX                    2      SEROTONIN-RECEPTOR PARTIAL-AGONIST./CT.MA
E12     WPIX                    6      SEROTONIN-RECEPTOR-LIGAND/CT.MA

=> s e8
L3           10 SEROTONIN-ANTAGONIST/CT.MA


=>  d max ct

L3    ANSWER 1 OF 10  WPIX COPYRIGHT 2007        THE THOMSON CORP on STN
AN.S  DCR-151191
DCSE  103826-0-1-0
CN.P  PHENIRAMINE MALEATE
CN.S  Dimethyl-(3-phenyl-3-pyridin-2-yl-propyl)-amine; compound with
      but-2-enedioic acid
SY    ALLER-G; ALTERGIAN; ANTIHISTONE; AVIL; AVIL-RETARD; DANERAL; DANERAL-SA;
      FENAMINE; FENAMINE-SLOW; HEMARIL; INHISTON; LARIL; MALEATE-PHENIRAMINE;
      MEDOPHEN; METRON; PHENIL; PHENIRAMIN; PHENIRAMINE MALEATE;
      PHENIRAMINE-MALEATE; PHYLLAXENE; PIRIEX; PM-241; QUIL; S-108; TRIMETON;
      TRIPOTON

      CM    1
```

```
      CM    2
```



```
MF    C4 H4 O4 . C16 H20 N2
SMF   C16 H20 N2 *1; TOTAL *2; TYPE *2; C4 H4 O4 *1
MW    356.4253
SDCN  R17806
CT    ANTIHISTAMINES-H1
      Antihistamine-H1; serotonin-antagonist; enhances effects of
      adrenaline.
```

Drug Registry Name (DDRN)

If there is a cross-reference between the DDF and the DWPI available, it will be located in the DDRN (Drug Registry Name) field. This can be used to cross-over between both files.

```
=> e dr0121037/ddrn
E#      FILE              FREQUENCY   TERM
--      ----              ---------   ----
E1      WPIX                    1      DR0120945/DDRN
E2      WPIX                    1      DR0121029/DDRN
E3      WPIX                    1  -->  DR0121037/DDRN
E4      WPIX                    1      DR0121039/DDRN
E5      WPIX                    1      DR0121051/DDRN
E6      WPIX                    1      DR0121052/DDRN
E7      WPIX                    1      DR0121054/DDRN
E8      WPIX                    1      DR0121056/DDRN
E9      WPIX                    1      DR0121059/DDRN
E10     WPIX                    1      DR0121062/DDRN
E11     WPIX                    1      DR0121063/DDRN
E12     WPIX                    1      DR0121064/DDRN

=> d ddrn

DDRN DR0121037
```

A corresponding DRUGU record:

```
AN     2004-10239  DRUGU   B P
TI     Isozyme-nonselective N-substituted bipiperidylcarboxamide acetyl-CoA
       carboxylase inhibitors reduce tissue malonyl-CoA concentrations, inhibit
       fatty acid synthesis, and increase fatty acid oxidation in cultured cells
       and in experimental animals.
AU     Harwood H J Jr; Petras S F; Shelly L D; Zaccaro L M; Perry D A; Makowski
       M R; Hargrove D M; Martin K A; Tracey W R; Chapman J G
CS     Pfizer
LO     Groton, Conn., USA
SO     J.Biol.Chem. (278, No. 39, 37099-111, 2003) 6 Fig. 3 Tab. 56 Ref.
       CODEN: JBCHA3        ISSN:  0021-9258
AV     Dept. of Cardiovascular + Metabolic Diseases, Pfizer Global Research +
       Development, Groton Labs., Pfizer, Inc., Groton, CT 06340, U.S.A. (16
       authors; e-mail: h_james_harwood@groton.pfizer.com).
LA     English
DT     Journal
AB     The effects of acetyl-CoA carboxylase (ACC)1 and ACC2 inhibition by
       CP-640186, CP-640188 and CP-610431 were studied.  The pharmacokinetics of
       CP-640-188 was determined in rats.  The ACC inhibitors inhibited fatty
       acid synthesis and increased fatty acid oxidation in liver, adipose,
       heart and muscle tissue in-vitro and in rats and mice in-vivo after i.p.
       and p.o. administration.  The results suggest that isozyme-non-selective
       inhibition may reduce risk factors associated with metabolic syndrome.
SH     B Biochemistry
       P Pharmacology
CC     8  Pharmacokinetics
       22 Endogenous Compounds
       58 Vasoactive
       72 New Drugs
       73 Trial Preparations
CT        OBESITY *OC; BODY-WEIGHT *OC; CL-316243 *RC; RAT *FT; MOUSE *FT;
          IN-VIVO *FT; HEP-G2-CELL *FT; LIVER *FT; HEART *FT; MUSCLE *FT;
          ADIPOSE-TISSUE *FT; IN-VITRO *FT; LIPID-METAB. *FT; DRUG-COMPARISON
          *FT; I.P. *FT; P.O. *FT; ANTIARTERIOSCLEROTIC *FT;
          ACETYL-COA-CARBOXYLASE-INHIBITOR *FT; TRIAL-PREP. *FT;
          ACETYL-COA-CARBOXYLASE-INHIBITORS *FT; ANTIARTERIOSCLEROTICS *FT; NEW
          *FT; LAB.ANIMAL *FT; HEPATOBLASTOMA *FT; TUMOR-CELL *FT;
          TISSUE-CULTURE *FT; INJECTION *FT
    [ 01] CP-640186 *PH; CP-640186 *DM; DR0121039 *RN; I.V. *FT;
          PHARMACOKINETICS *FT; INJECTION *FT; PH *FT; DM *FT
    [ 02] CP-610431 *PH; DR0121035 *RN; PH *FT
    [ 03] CP-640188 *PH; DR0121037 *RN; PH *FT
FA     AB; LA; CT
FS     Literature
```

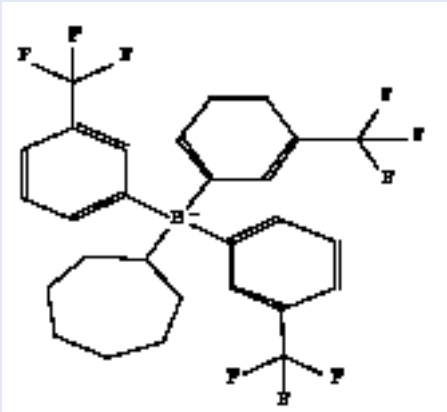
Molecular Weight (MW)

A fully range searchable molecular weight field is available, which may prove useful, e.g. in refining large answer sets retrieved using the Element Symbol (ELS) or Element Symbol Count (ELS.CNT) fields.

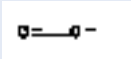
For example (MW<100= Molecular Weight of less than 100; ELS= Element Symbol; K= Potassium):

```
=> s mw<100 and k/els
            1604 MW<100
             308 K/ELS
L3            36 MW<100 AND K/ELS

=> d tri mw 1-3
L3    ANSWER 5 OF 36  WPIDS COPYRIGHT 2007        THE THOMSON CORP on STN
AN.S  DCR-208582
MF    C28 H25 B F9 . C4 H12 N
                      CM   1
```

```
        CM   2

MW    56.1049

L3    ANSWER 6 OF 36  WPIDS COPYRIGHT 2007        THE THOMSON CORP on STN
AN.S  DCR-133148
CN.P  POTASSIUM PEROXIDE
MF    K . O2
        CM   1
```



```
                CM   2

MW    71.096

L3    ANSWER 7 OF 36  WPIDS COPYRIGHT 2007        THE THOMSON CORP on STN
AN.S  DCR-131846
CN.P  POTASSIUM SULFIDE update if required.
MF    H K . H2 S
        CM   1
```

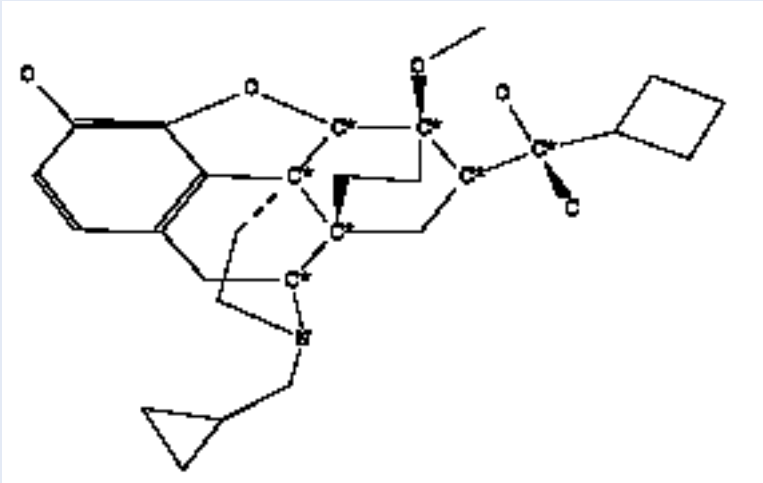


```
                CM   2
MW    71.164
```

Substructure Search Terms (SS)

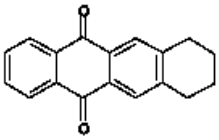
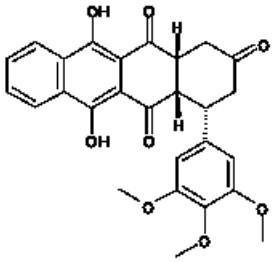
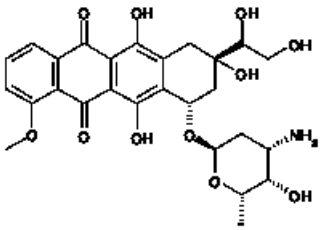
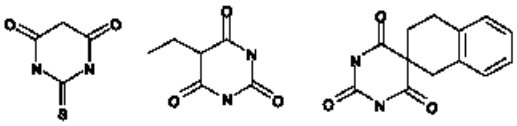
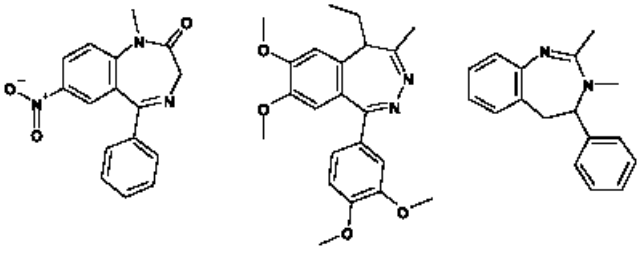
This field contains substructure search terms lifted from DDF. Since this requires compounds appearing both in DWPI and DDF, the number of compounds in DCR having this field occupied is limited.

```
=> e alkaloid/ss
E#      FILE              FREQUENCY   TERM
--      ----              ---------   ----
E1      WPIX                 239      ALDEHYDE/SS
E2      WPIX                 137      ALDIMINE/SS
E3      WPIX                 952  --> ALKALOID/SS
E4      WPIX                  28      ALKANE/SS
E5      WPIX                 121      ALKYLBROMIDE/SS
E6      WPIX                 452      ALKYLCHLORIDE/SS
E7      WPIX                1066      ALKYLFLUORIDE/SS
E8      WPIX                  36      ALKYLIODIDE/SS
E9      WPIX                   2      ALUMINUM/SS
E10     WPIX                  12      ALUMINUM-COMPLEX/SS
E11     WPIX                  22      ALUMINUM-SALT/SS
E12     WPIX                   1      AMERICIUM/SS

=> s e3
L2          952 ALKALOID/SS

=> d max ss

L2    ANSWER 1 OF 952  WPIX COPYRIGHT 2007        THE THOMSON CORP on STN
AN.S DCR-1271357
DCSE 1271357-1-0-0
```

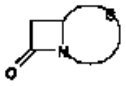


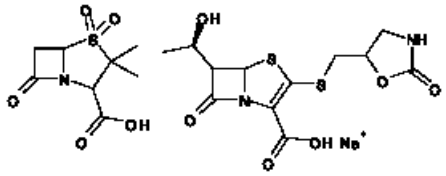
```
MF    C29 H39 N O4
SMF   C29 H39 N O4 *1; TOTAL *1; TYPE *1
MW    465.6385
SRIN 06766
SDCN RALXWI
CC    ALKALOIDS
SS    ALKALOID; BRIDGE-STRUCT.; COND.RING; CYCLOHEXANE; MORPHINAN;
      PHENOL; ETHER; ISOBENZOFURAN; CYCLOPROPANE; CYCLOBUTANE; BENZOFURAN;
      AMINOALCOHOL
```

Appendix Definitions of Substance Descriptors/Classification Codes

SUBSTANCE DESCRIPTOR	DESCRIPTION
ALKALOIDS	Organic nitrogen-containing bases, mainly of plant origin. This descriptor is only used when identified as such in the source document. Examples are morphine, caffeine, atropine, and strychnine.
ALLOYS	A metal that consists of an intimate mixture of two or more metallic elements.
ANTHRACYCLINES	A class of compounds containing the following ring system, the degree of saturation and substitution can vary.  e.g.  
ANTIBODIES	A blood serum protein of the globulin fraction which is formed in response to the introduction of an antigen. Only used when identified as such in the source document.
BARBITURATES	Used for ALL derivatives of barbituric acid, including thio analogues. e.g. 
BENZODIAZEPINES	Used when a benzene ring is condensed to a 7 membered ring containing 2 nitrogen atoms (in any position), other atoms in the ring being carbon. Regardless of the degree of saturation or substitution. e.g. 

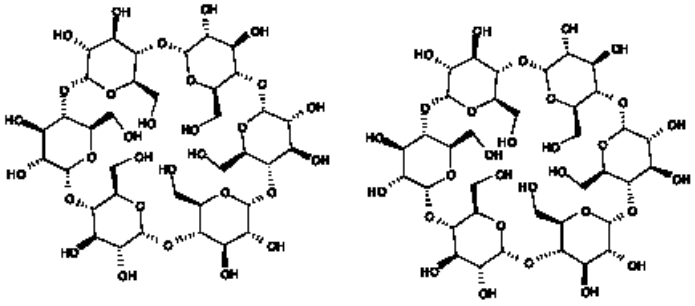
SUBSTANCE DESCRIPTOR	DESCRIPTION
ALKALOIDS	Organic nitrogen-containing bases, mainly of plant origin. This descriptor is only used when identified as such in the source document. Examples are morphine, caffeine, atropine, and strychnine.
ALLOYS	A metal that consists of an intimate mixture of two or more metallic elements.
ANTHRACYCLINES	A class of compounds containing the following ring system, the degree of saturation and substitution can vary.



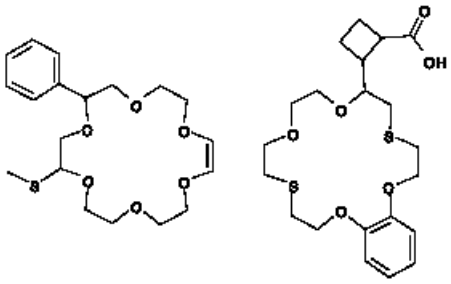
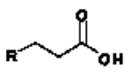
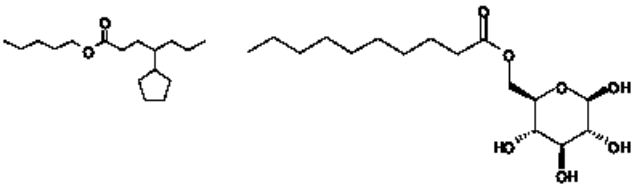
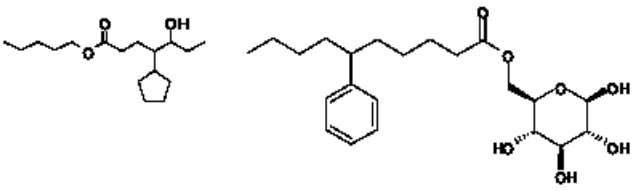


e.g.

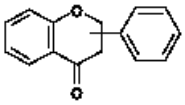
ANTIBODIES	A blood serum protein of the globulin fraction which is formed in response to the introduction of an antigen. Only used when identified as such in the source document.
BARBITURATES	Used for ALL derivatives of barbituric acid, including thio analogues. e.g.



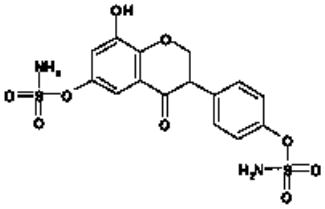
BENZODIAZEPINES	Used when a benzene ring is condensed to a 7 membered ring containing 2 nitrogen atoms (in any position), other atoms in the ring being carbon. Regardless of the degree of saturation or substitution. e.g.

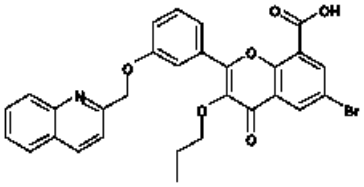
SUBSTANCE DESCRIPTOR	DESCRIPTION
BETA LACTAMS	Keyword applied to compounds containing the beta lactam group condensed to thiazine or thiazole ring i.e. cephalosporins, penicillins, regardless of the degree of saturation or substitution. Basic ring structure shown.
e.g.	
BORANES	Group of compounds that contain boron and hydrogen only. The simplest example is diborane B2H6. The larger borane molecules have open or closed polyhedra of boron atoms.
CARBOHYDRATES	Polyhydroxyaldehydes (or polyhydroxyketones) or substances that yield these on hydrolysis. The general molecular formula of carbohydrates is Cx(H2O)y.
at	Any compound containing a sugar moiety is assigned the keyword carbohydrate, the definition for a sugar sets the lower limit of size, such that compounds must contain least 2 stereocentres. Therefore glycoaldehyde ($HOCH_2CHO$) and glyceraldehyde ($HOCH_2CHOHCHO$) are both excluded because neither contains 2 stereocentres.
polysaccharides	A polysaccharide is a compound which contains at least 5 adjacent sugar residues (or their derivatives) linked via ether or thioether linkages. The term carbohydrate is also used.
	 **cyclodextrins** Cyclic oligomers of glucose in which the individual glucose units are connected by 1,4-bonds. The terms carbohydrate and polysaccharide are also used.
e.g.	
	

glycoproteins Any protein with carbohydrate group attached. The terms protein and carbohydrate are also used.



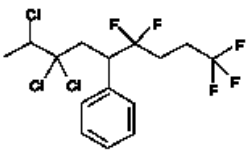
CARBORANES Boron cluster compounds with one or more of the polyhedral vertices replaced by carbon, e.g. $C_2H_{12}B_{10}$





SUBSTANCE DESCRIPTOR DESCRIPTION

CROWN ETHERS Macrocyclic compounds with O or S hetero atoms as the donor atoms in their ring structure and having the property of incorporating cations into their cavities. The best known crown ethers are the macrocyclic polyethers containing the repeating unit $(-OCR_2CR_2)n$, where R is most commonly H and are named in the form: x-crown-y, where x is the total number of atoms in the ring and y is the number of oxygens. e.g.



DENDRIMERS Globular structures in which well-defined branches radiate from a central core, becoming more branched and crowded as they extend out to the periphery. Some dendrimers have a diameter of more than 10 nm and a molecular weight exceeding 1 million Daltons.

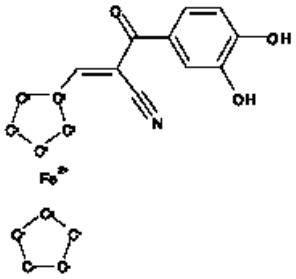
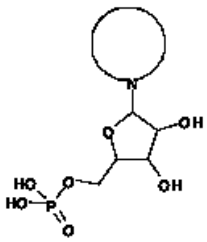
The second type of dendritic structure is the hyperbranched polymer. This type of polymer also has a fractal pattern of chemical bonds, but its branches don't emanate from a central core. Hyperbranched polymers can have either random or fairly regular architectures.

The term also applies to organometallics with dendrimer ligands.

FATTY ACIDS Any straight- or branched-chain, unsubstituted, saturated monocarboxylic acid with a total of 3 or more C atoms, includes derivatives such as esters and amides, and includes analogues with the cycloalkyl substituents in the chain. For unsaturated use UNSATURATED FATTY ACID

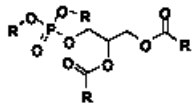
e.g.

The keyword is NOT applied to the following due to the substitution on the alkyl chain (only cycloalkyl substituents are allowed)

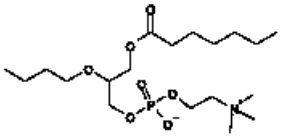
SUBSTANCE DESCRIPTOR	DESCRIPTION
FLAVONOIDS	Compounds containing a benzopyran ring substituted at C-2 or C-3 by an aryl group, the degree of saturation and substitution can vary, basic structure shown.  e.g.
FULLERENES	Giant closed-cage molecules that are formed entirely of carbon in the sp2 hybridised state and are arranged to form adjoining pentagonal and hexagonal rings. Number of C atoms = 2(10 + m) with 12 pentagonal rings and m hexagonal rings Number of rings = 12 + (n - 20)/2 where n = number of C atoms The nanotubes are very large tubular fullerenes and are not considered as a separate class of molecule. The tubular shape is the result of the large number carbon atoms which form hexagonal rings. The tube is sealed at each end due to the presence of pentagonal rings. See also HETEROFULLERENE
HALOCARBONS	A compound containing a carbon skeleton which is poly-substituted with, one or more halogens, no other heteroatoms are present. e.g. 
HETEROFULLERENES	Fullerenes where one or more carbon atoms have been replaced by another atom. See also FULLERENE
HETEROPOLY ACIDS	Definition: These are compounds that satisfy the following formula: $H_xA_yM_zOW$ where A = phosphorous, silicon, boron or arsenic M = transition metal (normally molybdenum, vanadium or tungsten) x = › o y = › o

z = › o

This substance descriptor also includes salts of the acids in which some or all of the hydrogen atoms are replaced by cations, most commonly ammonium or alkali metal cations. It is also possible to have structures in which some of atoms M are replaced by a second transition metal (niobium being the most common one) so that the hete ropolyacid contains two metals plus the metalloid A.



Examples of heteropolyacids are: $H_3PW_{12}O_{40}$, $H_4PVW_{11}O_{40}$, $Na_9PW_{15}Nb_3O_{62}$



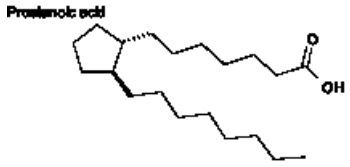
SUBSTANCE DESCRIPTOR **DESCRIPTION**

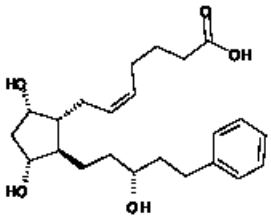
LIPOPROTEINS Any compound containing a protein and a lipid moiety. This descriptor is only used when identified as such in the source document.

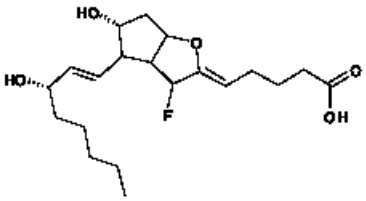
METALLOCENES

An organometallic compound that contains at least one cyclopentadienyl group, or its derivative, bonded to the central metal atom. Derivatives of the cyclopentadiene ligand which are also included within this definition are those with rings fused onto the cyclopentadienyl ring e.g. indene and fluorene.

e.g.







NOBLE GASES

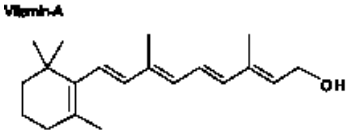
Helium, Neon, Argon, Krypton, Xenon, Radon.

NUCLEOSIDES

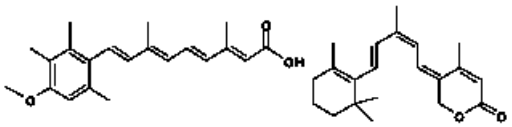
A nucleoside is a compound which contains a sugar residue attached via N to a cyclic base group. The base group is usually derived from purine or pyrimidine groups, or their ring modified derivatives including the thia derivatives. The more usual base groups are adenine, cytosine, thymine, uracil, and guanine residues.

Below is the basic structure for a nucleoside system, the sugar moiety can be substituted and the keyword is still applied for deoxy/dideoxy analogues.

NUCLEOTIDES A nucleoside with a phosphate group attached to the sugar moiety.

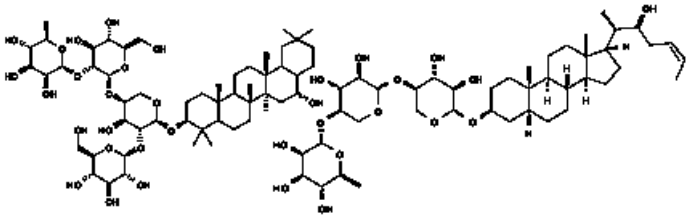


oligonucleotides
Compounds containing 3 or more nucleotide residues which are linked via the

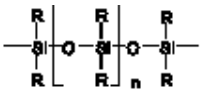


phosphate groups. Usually denoted with single letter codes representing the nucleoside bases e.g. TTUUGGCATU

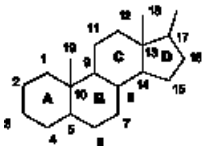
PEPTIDES
A compound formed by the linking of two or more amino acids by CO-NH groups.



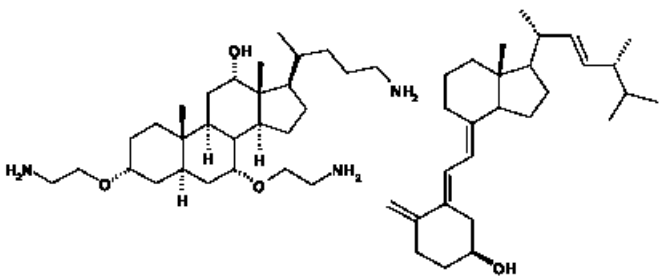
For peptides containing fifty or more residues use the term PROTEIN instead.

SUBSTANCE DESCRIPTOR	DESCRIPTION
	**cyclic peptides** As for peptide, but part or all of the peptide chain forms a ring.
	PHOSPHOLIPIDS Esters of fatty acids formed with alcohol components containing a phosphate group.

e.g.



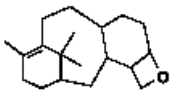
POLYMERS
A macromolecule with five or more structural repeat units.

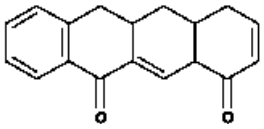


PROSTAGLANDINS
Compounds that are derived from 20-carbon unsaturated carboxylic

acids with a cyclopentane ring i.e. analogues of prostanoic acid. The degree of saturation and substitution can vary.

e.g.





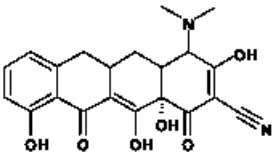
PROTEINS

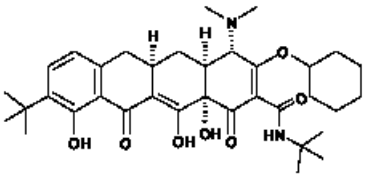
Peptides with a specific sequence of 50 or more residues. The term PEPTIDE is not also applied.

enzymes

Any of a large class of protein substances produced by living cells, which act as biocatalysts in biochemical reactions. Enzymes are typically composed of a protein part (the apoenzyme) and a non-protein part (the coenzyme) necessary for activity. The term protein is also used.

This descriptor is only used when identified as such in the source document or if





the enzyme name is obviously identifiable i.e. name ending in -ase.

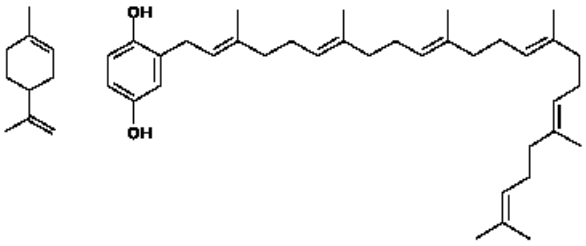
glycoproteins

Any protein with carbohydrate group attached. The terms protein and carbohydrate are also used.

SUBSTANCE DESCRIPTOR DESCRIPTION

RETINOIDS

Synthetic analogues of vitamin A. The keyword is applied regardless of the degree



of saturation and substitution.

e.g.

FIZ Karlsruhe
STN Europe
Hermann-von-Helmholtz-Platz 1
76344 Eggenstein-Leopoldshafen
Germany

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

FIZ Karlsruhe