CAS STNext® Webinar

BOOST YOUR STRUCTURE SEARCH WITH THE DERWENT CHEMISTRY RESOURCE (DCR) DATABASE

Jim Brown, FIZ Karlsruhe

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Agenda

- History of Derwent substance indexing
- Tour of DCR database
  - Searchable fields
  - Structure search
- Relationship with Derwent World Patents Index
  - Use of roles
- Search examples

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

History of Derwent substance indexing

- Started in 1963 with pharmaceutical patents
  - Added Agricultural chemical patents in 1965 and rest of chemistry in 1970
- Substances were indexed using a rudimentary fragmentation system
  - Fragmentation system became more robust over the years, still used today
- 1978 - MMS – covered Markush structures and specific substances with structure searching capabilities
  - Became DWPIM on CAS STNext platform, with some cleanup of original data

History of Derwent substance indexing

- Derwent Chemistry Resource(s) (aka DRN/DCN/DCR)
- Started in 1981 with 2,100 substances (DRN)
- Added additional substance records in 1987 to create a total of ~20,000 (DCN)
- Comprehensive coverage April 1999 (DCR)
- All DRN and DCN records were updated with DCR numbers
  - DRNs and-or DCNs remain in records
- Specific substances indexed by Derwent from basic patents in DWPI
  - DWPI indexing includes roles associated with DCR substances indexed

Derwent Chemistry Resource (DCR) coverage

- Novel specific compounds from the claims
- Known specific compounds from the claims
- Main examples and other representative examples from the description
- Up to 999 substances can be indexed from the DWPI basic patent

Agenda

- History of Derwent substance indexing
- Tour of DCR
  - Searchable fields
  - Structure search
- Relationship with Derwent World Patents Index
  - Use of roles
- Search examples

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

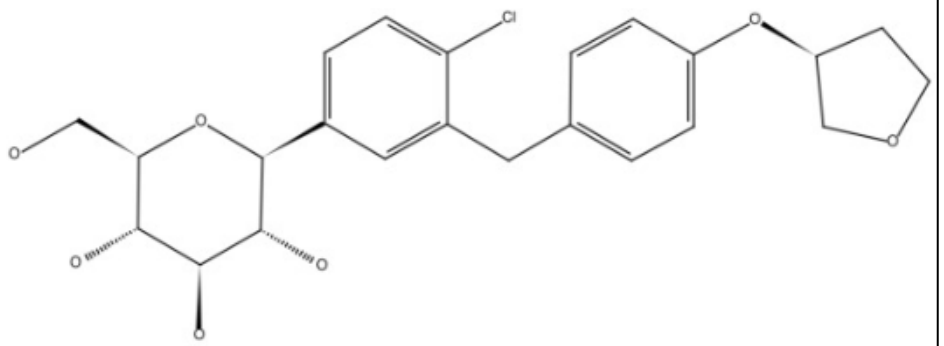
CAS
A division of the American Chemical Society

Derwent Chemistry Resource (DCR) database

- Substance based database
- Most, but not all, records have a searchable structure, molecular formula, molecular weight, etc.
- Some records do not have a structure
  - Antibodies, etc.
  - For these records, consider searching /CMT field for additional text on substance
- Records also contain Derwent classification codes/subject descriptors

Sample DCR record

L3 ANSWER 1 OF 1 DCR COPYRIGHT 2024 CLARIVATE on STN.
AN DCR-1395204 DCR
DCSE 1395204-1-0-0
CN.P EMPAGLIFLOZIN
CN.S (2S,3R,4R,5S,6R)-2-(4-Chloro-3-{4-[(S)-(tetrahydro-furan-3-yl)oxy]-benzyl}-phenyl)-6-hydroxymethyl-tetrahydro-pyran-3,4,5-triol
SY EMPAGLIFLOZIN; JARDIANCE



MF C23 H27 Cl O7
SMF C23 H27 Cl O7 *1; TYPE *1; TOTAL *1
MW 450.9204
SDCN RAOMA7
SD CARBOHYDRATES
ED Entered STN: 8 Dec 2006
Last updated on STN: 11 Oct 2022
Update DWPI Cross Ref.: 14 Aug 2024

CN.P – Preferred chemical name
CN.S – Systematic chemical name
SY- Synonyms

Use /CN to search all three fields

SD – Subject descriptor

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

DCR subject descriptors

```
=> E A/SD 25

**** START OF FIELD ****
E3            0 --> A/SD
E4        22088     ALKALOIDS/SD
E5          584     ALLOYS/SD
E6         1431     ANTHRACYCLINES/SD
E7          947     ANTIBODIES/SD
E8         2400     BARBITURATES/SD
E9        14894     BENZODIAZEPINES/SD
E10        4172     BETA LACTAMS/SD
E11         179     BORANES/SD
E12           1     CARBOHYDRATE/SD
```

```
E13      122275     CARBOHYDRATES/SD
E14        1046     CARBORANES/SD
E15         784     CROWN ETHERS/SD
E16       16405     CYCLIC PEPTIDES/SD
E17         395     CYCLODEXTRINS/SD
E18         239     DENDRIMERS/SD
E19        3231     ENZYMES/SD
E20       57302     FATTY ACIDS/SD
E21       14134     FLAVONOIDS/SD
E22          12     FULLERENE/SD
E23         742     FULLERENES/SD
E24         391     GLYCOPROTEINS/SD
E25        7010     HALOCARBONS/SD
```


FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

DCR subject descriptors

```
=> e 25

E26          381     HETEROFULLERENES/SD
E27          405     HETEROPOLY ACIDS/SD
E28           30     LIPOPROTEINS/SD
E29        18922     METALLOCENES/SD
E30        23863     NUCLEOSIDES/SD
E31        22778     NUCLEOTIDES/SD
E32          265     OLIGONUCLEOTIDES/SD
E33            1     OTHER NATURAL PRODUCT/SD
E34        15588     OTHER NATURAL PRODUCTS/SD
E35            1     PEPTIDE/SD
E36       106840     PEPTIDES/SD
E37         1971     PHOSPHOLIPIDS/SD
```

```
E38         2112     POLYMERS/SD
E39         1732     POLYSACCHARIDES/SD
E40         3483     PROSTAGLANDINS/SD
E41         5727     PROTEINS/SD
E42         1754     RETINOIDS/SD
E43         5229     SAPONINS/SD
E44          267     SILICONES/SD
E45        46701     STEROIDS/SD
E46         2108     TAXANES/SD
E47        25546     TERPENES/SD
E48         3609     TETRACYCLINES/SD
E49        48238     UNSATURATED FATTY ACIDS/SD
E50          102     ZEOLITES/SD
```


FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

ANTIBODIES subject descriptor search & display

```
=> S ANTIBODIES/SD

L1          947 ANTIBODIES/SD

=> D 1-3 ALL


L1      ANSWER 1 OF 947 DCR  COPYRIGHT 2024 CLARIVATE on STN.
AN      DCR-7272902   DCR
DCSE    7272902-0-0-0
CN.P    Q-1802
SY      Q-1802

          Substance image not available

CMT     Bispecific antibody targeting Claudin-18.2 and programmed death-ligand 1
MF      Unknown
SD      ANTIBODIES; PROTEINS
ED      Entered STN: 29 Jul 2024
        Last updated on STN: 29 Jul 2024
        Update DWPI Cross Ref.: 29 Jul 2024
```

The CMT field contains useful information about substances, especially those without a structure.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

ANTIBODIES subject descriptor search & display

```
L1      ANSWER 2 OF 947 DCR  COPYRIGHT 2024 CLARIVATE on STN.
AN      DCR-7128783   DCR
DCSE    7128783-0-0-0
CN.P    TEBOTELIMAB
SY      MGD-013; TEBOTELIMAB

        Substance image not available

CMT     Tetravalent humanized immunoglobulin G4 kappa monoclonal antibody
        (Fc-bearingdual-affinity re-targeting (DART) molecule) targeting D-1 and
        LAG-3
MF      Unknown
SDCN    RDM0RT
SD      PROTEINS; ANTIBODIES
ED      Entered STN: 19 Jul 2024
        Last updated on STN: 19 Jul 2024
        Update DWPI Cross Ref.: 23 Aug 2024
```

```
L1      ANSWER 3 OF 947 DCR  COPYRIGHT 2024 CLARIVATE on STN.
AN      DCR-6911121   DCR
DCSE    6911121-0-0-0
CN.P    MIPTENALIMAB
SY      BI-754111; MIPTENALIMAB

        Substance image not available

CMT     Humanized monoclonal immunoglubulin G4 kappa antibody targeting
        lymphocyteactivation gene-3 (LAG-3)
MF      Unknown
SDCN    RDJTK4
SD      ANTIBODIES; PROTEINS
ED      Entered STN: 19 Apr 2024
        Last updated on STN: 19 Apr 2024
        Update DWPI Cross Ref.: 14 Aug 2024
```

The CMT field contains useful information about substances, especially those without a structure.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Chemical Name vs. Chemical Name Segment

- Chemical name ~ complete name - /CN
  - Phrase parsed field
- EXPAND around chemical name to see alphabetical name variations
- Consider alternate namings/synonyms for /CN
- Chemical name segment – /CNS
  - Word parsed field

Chemical Name vs. Chemical Name Segment

```
=> E PANTOPRAZOLE/CN 25

E1          1      PANTOPENIL/CN
E2          2      PANTOPON/CN
E3          1  --> PANTOPRAZOLE/CN
E4          1      PANTOPRAZOLE (18F)/CN
E5          1      PANTOPRAZOLE (2H)/CN
E6          1      PANTOPRAZOLE (5-DIFLUOROMETHOXY-2H)/CN
E7          1      PANTOPRAZOLE (BIS-2H)/CN
E8          1      PANTOPRAZOLE (DECA-2H)/CN
E9          1      PANTOPRAZOLE (DODECA-2H)/CN
E10         1      PANTOPRAZOLE (HEXA-2H)/CN
E11         1      PANTOPRAZOLE (NONA-2H)/CN
E12         2      PANTOPRAZOLE (OCTA-2H)/CN
```

```
E13         3      PANTOPRAZOLE (PENTA-2H)/CN
E14         1      PANTOPRAZOLE (PENTADECA-2H)/CN
E15         1      PANTOPRAZOLE (PER-2H)/CN
E16         1      PANTOPRAZOLE (TETRA-2H)/CN
E17         3      PANTOPRAZOLE (TRIS-2H)/CN
E18         1      PANTOPRAZOLE -SODIUM- TRIHYDRATE/CN
E19         1      PANTOPRAZOLE HYDROXY SESQUIHYDRATE/CN
E20         2      PANTOPRAZOLE HYDROXY TETRAHYDRATE/CN
E21         1      PANTOPRAZOLE LITHIUM/CN
E22         1      PANTOPRAZOLE MAGNESIUM/CN
E23         1      PANTOPRAZOLE MAGNESIUM DIHYDRATE/CN
E24         1      PANTOPRAZOLE MAGNESIUM DIMETHANOLATE/CN
E25         1      PANTOPRAZOLE MAGNESIUM HEMIPENTAHYDRATE/CN
```


FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Chemical Name vs. Chemical Name Segment

```
=> E 25

E26          1      PANTOPRAZOLE MAGNESIUM HYDROXIDE MONOHYDRATE/CN
E27          1      PANTOPRAZOLE MAGNESIUM HYDROXIDE SESQUIHYDRATE/CN
E28          1      PANTOPRAZOLE MAGNESIUM TETRAHYDRATE/CN
E29          1      PANTOPRAZOLE POTASSIUM/CN
E30          2      PANTOPRAZOLE SODIUM/CN
E31          1      PANTOPRAZOLE SODIUM CHLORIDE/CN
E32          1      PANTOPRAZOLE SODIUM DIHYDRATE/CN
E33          1      PANTOPRAZOLE SODIUM DISEMIHYDRATE/CN
E34          1      PANTOPRAZOLE SODIUM HYDRATE/CN
E35          1      PANTOPRAZOLE SODIUM HYDRATE (2:2:5)/CN
E36          1      PANTOPRAZOLE SODIUM MONOHYDRATE/CN
E37          1      PANTOPRAZOLE SODIUM SESQUIHYDRATE/CN
```

```
E38          1      PANTOPRAZOLE SODIUM TRIHYDRATE/CN
E39          1      PANTOPRAZOLE SODIUM-MONOHYDRATE/CN
E40          1      PANTOPRAZOLE SULFONE N-OXIDE SODIUM/CN
E41          1      PANTOPRAZOLE ZINC/CN
E42          1      PANTOPRAZOLE-SODIUM/CN
E43          1      PANTOPRAZOLE-SODIUM-DIHYDRATE/CN
E44          1      PANTOPRAZOLE-SODIUM-MONOHYDRATE/CN
E45          1      PANTOS/CN
E46          1      PANTOSAL/CN
E47          1      PANTOSEDIV/CN
E48          1      PANTOSEPT/CN
E49          1      PANTOSIN/CN
E50          1      PANTOTENATE-DE-CLORAMFENICOL-COMPASE/CN
```


FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Chemical Name vs. Chemical Name Segment

```
=> S PANTOPRAZOLE/CNS

L1          59 PANTOPRAZOLE/CNS

=> S L1 NOT E3-E44

             1 PANTOPRAZOLE/CN
             1 "PANTOPRAZOLE (18F)"/CN
             1 "PANTOPRAZOLE (2H)"/CN
             1 "PANTOPRAZOLE (5-DIFLUOROMETHOXY-2H)"/CN
             1 "PANTOPRAZOLE (BIS-2H)"/CN
             1 "PANTOPRAZOLE (DECA-2H)"/CN
             1 "PANTOPRAZOLE (DODECA-2H)"/CN
             1 "PANTOPRAZOLE (HEXA-2H)"/CN
```

. . .

```
             1 PANTOPRAZOLE-SODIUM/CN
             1 PANTOPRAZOLE-SODIUM-DIHYDRATE/CN
             1 PANTOPRAZOLE-SODIUM-MONOHYDRATE/CN
L2          19 L1 NOT (PANTOPRAZOLE/CN OR "PANTOPRAZOLE (18F)"/CN OR "PANTOPRAZ
               OLE (2H)"/CN OR "PANTOPRAZOLE (5-DIFLUOROMETHOXY-2H)"/CN OR "PAN
               TOPRAZOLE (BIS-2H)"/CN OR "PANTOPRAZOLE (DECA-2H)"/CN OR "PANTOP
               RAZOLE (DODECA-2H)"/CN OR "PANTOPRAZOLE (HEXA-2H)"/CN OR "PANTOP
               RAZOLE (NONA-2H)"/CN OR "PANTOPRAZOLE (OCTA-2H)"/CN OR "PANTOPRA
               ZOLE (PENTA-2H)"/CN OR "PANTOPRAZOLE (PENTADECA-2H)"/CN OR "PANT
               OPRAZOLE (PER-2H)"/CN OR "PANTOPRAZOLE (TETRA-2H)"/CN OR "PANTOP
               RAZOLE (TRIS-2H)"/CN OR "PANTOPRAZOLE -SODIUM- TRIHYDRATE"/CN
               OR "PANTOPRAZOLE HYDROXY SESQUIHYDRATE"/CN OR "PANTOPRAZOLE HYDR
               OXY TETRAHYDRATE"/CN OR "PANTOPRAZOLE LITHIUM"/CN OR "PANTOPRAZO
               LE MAGNESIUM"/CN OR "PANTOPRAZOLE MAGNESIUM DIHYDRATE"/CN OR
               "PANTOPRAZOLE MAGNESIUM DIMETHANOLATE"/CN OR "PANTOPRAZOLE MAGNE
               SIUM HEMIPENTAHYDRATE"/CN OR "PANTOPRAZOLE MAGNESIUM HYDROXIDE
               MONOHYDRATE"/CN OR "PANTOPRAZOLE MAGNESIUM HYDROXIDE SESQUIHYDRA
               TE"/CN OR "PANTOPRAZOLE MAGNESIUM TETRAHYDRATE"/CN OR "PANTOPRAZ
               OLE POTASSIUM"/CN OR "PANTOPRAZOLE
```


FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Chemical Name vs. Chemical Name Segment

```
=> D SY 1-19


L2      ANSWER 1 OF 19 DCR  COPYRIGHT 2024 CLARIVATE on STN.
SY      (S)-PANTOPRAZOLE DISODIUM PENTAHYDRATE; (S)-PANTOPRAZOLE SODIUM
        HYDRATE (2:2:5)

L2      ANSWER 2 OF 19 DCR  COPYRIGHT 2024 CLARIVATE on STN.
SY      (S)-PANTOPRAZOLE SODIUM TETRAHYDRATE

L2      ANSWER 3 OF 19 DCR  COPYRIGHT 2024 CLARIVATE on STN.
SY      S-PANTOPRAZOLE SODIUM SESQUIHYDRATE; S-PANTOPRAZOLE SODIUM-SESQUIHYDRATE

L2      ANSWER 4 OF 19 DCR  COPYRIGHT 2024 CLARIVATE on STN.
SY      4-CHLORO-PANTOPRAZOLE
```

```
L2      ANSWER 5 OF 19 DCR  COPYRIGHT 2024 CLARIVATE on STN.
SY      6-CHLORO-PANTOPRAZOLE

L2      ANSWER 6 OF 19 DCR  COPYRIGHT 2024 CLARIVATE on STN.
SY      7-CHLORO-PANTOPRAZOLE

L2      ANSWER 7 OF 19 DCR  COPYRIGHT 2024 CLARIVATE on STN.
SY      (L)-PANTOPRAZOLE SODIUM

L2      ANSWER 8 OF 19 DCR  COPYRIGHT 2024 CLARIVATE on STN.
SY      S-PANTOPRAZOLE SODIUM TRIHYDRATE

L2      ANSWER 9 OF 19 DCR  COPYRIGHT 2024 CLARIVATE on STN.
SY      (S)-PANTOPRAZOLE SODIUM

L2      ANSWER 10 OF 19 DCR  COPYRIGHT 2024 CLARIVATE on STN.
SY      (S)-PANTOPRAZOLE
```


FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Chemical Name vs. Chemical Name Segment

```
L2     ANSWER 11 OF 19 DCR  COPYRIGHT 2024 CLARIVATE on STN.
SY     LITHIUM (S)-PANTOPRAZOLE

L2     ANSWER 12 OF 19 DCR  COPYRIGHT 2024 CLARIVATE on STN.
SY     POTASSIUM (S)-PANTOPRAZOLE

L2     ANSWER 13 OF 19 DCR  COPYRIGHT 2024 CLARIVATE on STN.
SY     ALUMINUM (S)-PANTOPRAZOLE

L2     ANSWER 14 OF 19 DCR  COPYRIGHT 2024 CLARIVATE on STN.
SY     (S)-PANTOPRAZOLE CALCIUM; CALCIUM (S)-PANTOPRAZOLE

L2     ANSWER 15 OF 19 DCR  COPYRIGHT 2024 CLARIVATE on STN.
SY     ZINC (S)-PANTOPRAZOLE
```

```
L2     ANSWER 16 OF 19 DCR  COPYRIGHT 2024 CLARIVATE on STN.
SY     ALUMINUM PANTOPRAZOLE

L2     ANSWER 17 OF 19 DCR  COPYRIGHT 2024 CLARIVATE on STN.
SY     CALCIUM PANTOPRAZOLE

L2     ANSWER 18 OF 19 DCR  COPYRIGHT 2024 CLARIVATE on STN.
SY     S-(-)-PANTOPRAZOLE

L2     ANSWER 19 OF 19 DCR  COPYRIGHT 2024 CLARIVATE on STN.
SY     R-(+)-PANTOPRAZOLE
```


FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Chemical name synonyms search – helpful hint!

- Use substance databases (CAS REGISTRY, DCR, etc.) to find synonyms for substance
- Include synonyms of interest in free-text search
- Beware of false drops!
  - Example on next slide – SORTIS synonym could be problematic

Chemical name synonyms search

```
=> S LIPITOR/CN

L1            2 LIPITOR/CN

=> D CN 1-2


L1      ANSWER 1 OF 2 DCR  COPYRIGHT 2024 CLARIVATE on STN.
CN.P    ATORVASTATIN-LACTONE
CN.S    5-(4-Fluoro-phenyl)-1-[2-(4-hydroxy-6-oxo-tetrahydro-pyran-2-yl)-ethyl]-
        2-isopropyl-4-phenyl-1H-pyrrole-3-carboxylic acid phenylamide
SY      ATORIS; ATORVASTATIN (LACTONE); ATORVASTATIN LACTONE;
        ATORVASTATIN-LACTONE; CARDYL; CI-981; LIPITOR; LIPRIMAR; SORTIS;
        STORVAS; TAHOR; TORVACARD; ZARATOR

L1      ANSWER 2 OF 2 DCR  COPYRIGHT 2024 CLARIVATE on STN.
CN.P    ATORVASTATIN CALCIUM
CN.S    Calcium; 7-[2-(4-fluoro-phenyl)-5-isopropyl-3-phenyl-4-phenylcarbamoyl-
        pyrrol-1-yl]-3,5-dihydroxy-heptanoate
SY      ATORVASTATIN CALCIUM; ATORVASTATIN HEMICALCIUM; CALCIUM ATORVASTATIN;
        LIPITOR; SORTIS; ZARATOR
```

In this example, no additional records were found by searching LIPITOR/CNS.

```
=> FILE USPATALL
```

```
=> S LIPITOR

L2          4725 FILE USPATFULL
L3             0 FILE USPATOLD
L4          2256 FILE USPAT2

TOTAL FOR ALL FILES
L5          6981 LIPITOR

=> S ATORVASTATIN

L6         20002 FILE USPATFULL
L7             0 FILE USPATOLD
L8          9900 FILE USPAT2

TOTAL FOR ALL FILES
L9         29902 ATORVASTATIN
```

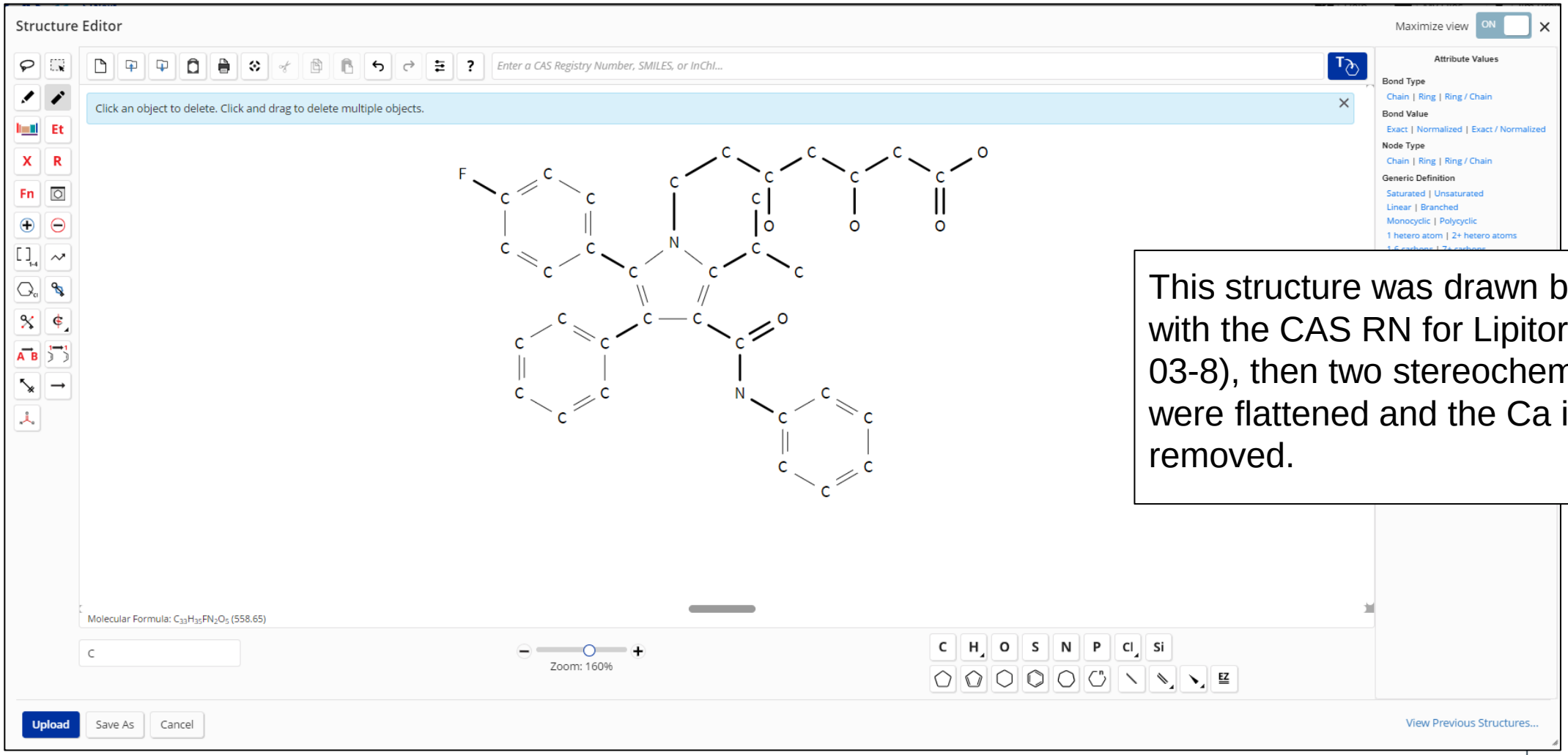

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Search considerations

- For comprehensive searching, consider *all* search options
  - Chemical name/chemical name segment search
  - Chemical structure search
    - For more information about structure searching on STNext, please visit Structure-Based Searching Overview – CAS STNext (zendesk.com)

Family structure search for Lipitor type substances



This structure was drawn by starting with the CAS RN for Lipitor (134523-03-8), then two stereochemistry bonds were flattened and the Ca ion removed.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

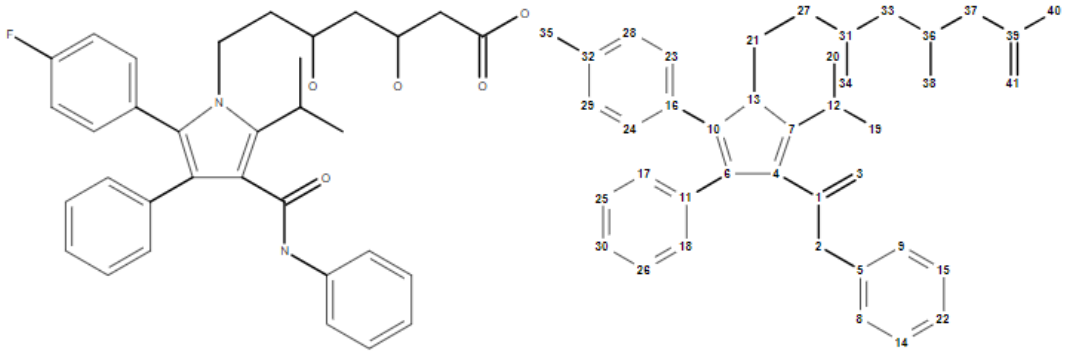
CAS
A division of the American Chemical Society

Structure search options

- Exact (EXA)-Search
  retrieves specific compounds and isotopes
- Family (FAM)-Search
  retrieves specific compounds, isotopes, salts and mixtures
- Closed Substructure (CSS)-Search
  allows for substitution at defined positions
- Substructure (SSS)-Search
  allows for substitution at any position

Family structure search for Lipitor type substances

Uploading structure file: Lipitor



Node Attributes
Ring Nodes : 4 5 6 7 8 9 10 11 13 14 15 16 17 18 22 23 24 25 26 28 29 30 32
Chain Nodes : 1 2 3 12 19 20 21 27 31 33 34 35 36 37 38 39 40 41
Bond Attributes
Ring Bonds : 4-6 4-7 5-8 5-9 6-10 7-13 8-14 9-15 10-13 11-17 11-18 14-22 15-22 16-23 16-24 17-25 18-26 23-28 24-29 25-30 26-30 28-32 29-32
Chain Bonds : 1-2 1-3 1-4 2-5 6-11 7-12 10-16 12-19 12-20 13-21 21-27 27-31 31-33 31-34 32-35 33-36 36-37 36-38 37-39 39-40 39-41
Exact Bonds : 1-4 6-11 7-12 10-16 12-19 12-20 21-27 27-31 31-33 32-35 33-36 36-37 37-39
Normalized Bonds : 5-8 5-9 8-14 9-15 11-17 11-18 14-22 15-22 16-23 16-24 17-25 18-26 23-28 24-29 25-30 26-30 28-32 29-32
Exact/Normalized Bonds : 1-2 1-3 2-5 4-6 4-7 6-10 7-13 10-13 13-21 31-34 36-38 39-40 39-41
Markush Attributes
Match Level (ATOM) : 4 5 6 7 8 9 10 11 13 14 15 16 17 18 22 23 24 25 26 28 29 30 32
Match Level (CLASS) : 1 2 3 12 19 20 21 27 31 33 34 35 36 37 38 39 40 41
Element Count Level (LIMITED) : 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41

L11 STRUCTURE UPLOADED

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Family structure search for Lipitor-type substances

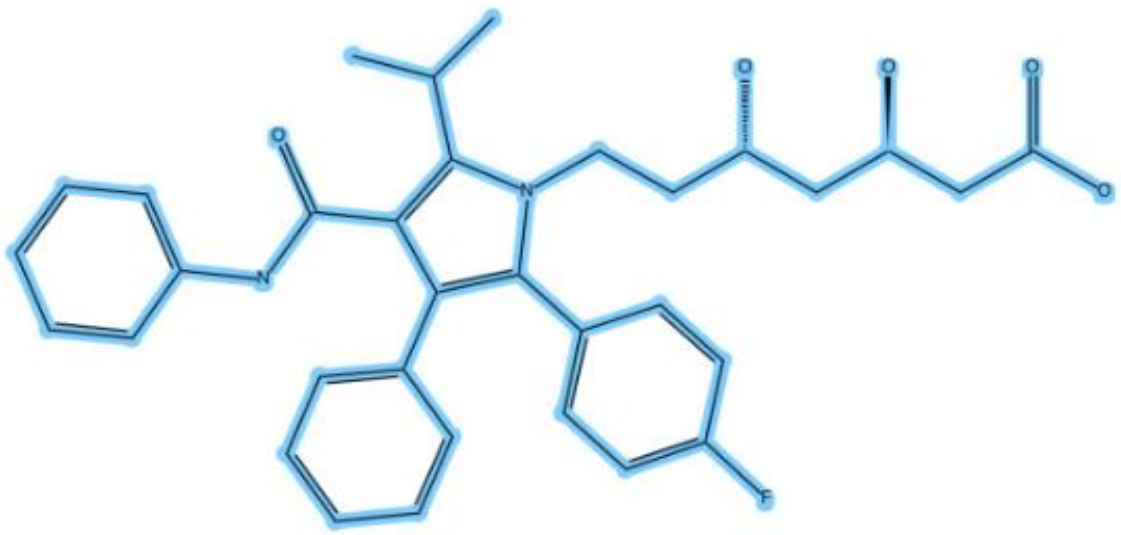
```
=> S L11 FAM FULL

FULL SEARCH INITIATED 09:41:42
FULL SCREEN SEARCH COMPLETED -        0 TO ITERATE

  0.0% PROCESSED    5684501 ITERATIONS                          148 ANSWERS
SEARCH TIME: 00.00.01

L12          148 SEA FAM FUL L11

=> S L12 NOT L1

L13          147 L12 NOT L1

=> D 1-3
```


FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Family structure search for Lipitor-type substances

L13 ANSWER 3 OF 147 DCR COPYRIGHT 2024 CLARIVATE on STN.
AN DCR-4998910 DCR
DCSE 4998910-1-0-0
CN.S (3R,5S)-7-[2-(4-Fluoro-phenyl)-5-isopropyl-3-phenyl-4-phenylcarbamoyl-pyrrol-1-yl]-3,5-dihydroxy-heptanoic acid



MF C33 H35 F N2 O5
ED Entered STN: 19 Jun 2020
Last updated on STN: 14 Aug 2020
Update DWPI Cross Ref.: 19 Jun 2020

This structure looks just like LIPITOR but maybe has different stereochemistry.

Chemical name search vs. structure search in DWPI

```
=> FILE WPIN
```

```
=> S L1

L14          1322 L1

=> S DCR-4998910/DCR

L15             1 DCR-4998910/DCR

=> S L14 AND L15

L16             0 L14 AND L15
```

L1 = Lipitor/CN

DCR-4998910 is the unique structure found on the previous slide.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Agenda

- History of Derwent substance indexing
- Tour of DCR database
  - Searchable fields
  - Structure search
- Relationship with Derwent World Patents Index
  - Use of roles
- Search examples

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Relationship between DCR and DWPI

- Derwent World Patents Index (DWPI) is the value-added patent family database produced by Clarivate
- To search for patent family records where a DCR substance is indexed
  - **S DCR-xxxxxxx/DCR**
- To search for patent family records where any DCR substances are indexed from a DCR set
  - **S Lx** (where Lx is the DCR set created)
    - This is the same search strategy that is used in CAS REGISTRY/CAplus and REAXYSFILESUB/REAXYSFILEBIB searches

Relationship between DCR and DWPI

- DCR substances indexed in DWPI also have roles associated with them
- Derwent has used multiple roles systems over the years, consider using all options
- Use /RL to search roles
- Link roles to the DCR indexing with the (T) operator
  - **S DCR-xxxxx/DCR (T) xxx/RL**
  - **S Lx(T) xxx/RL**
- For more information about the Derwent roles, use **HELP ROLES** in WPIN/WPIDS/WPIX

DCR roles in DWPI

- DWPI uses three partly overlapping sets of roles
  - DRN number roles – one letter roles from the chemical coding section assigned to DRN numbers (1981 – date)
  - DCN number roles – single letter roles from the chemical coding section assigned to DCR and DCN indexing (1987 – date)
  - DCR number roles – 2 or 3 letter roles from the IT field assigned to DCR numbers (1999 – date)
  - Always check the start and stop dates for roles!!!

DCR Number roles

DCR-Number Roles The following DCR-number 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).
DIS	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.

DCR Number roles (cont.)

USE	USE	Use of substance or apparatus is claimed or described.
DET	DETECTED	Applied to the keyword for a con- dition 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 (1987-2000)
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., hydrohalides, acetates.

DCN Number roles

The following DCN roles are available from 1987 onwards, except as indicated.

Role	Definition/Notes
A	Analysed or detected
C	Catalyst
D	Detecting agent
R	Removing or purifying agent
S	Intermediate or starting material
X	Substance removed
N	New Compound

DCN Number roles (cont.)

The following DCN roles are available from 1987 onwards, except as indicated.

Role	Definition/Notes
P	Known compound produced
Q	Product defined by its starting material(s)
M	Component of a Mixture
U	Use of a single compound
E	Excipient (from 1998)
T	Therapeutically active agent or prodrug (from 1998)
V	Reagent (from 1998)
K	Known compound (from 1998)

DRN Number roles

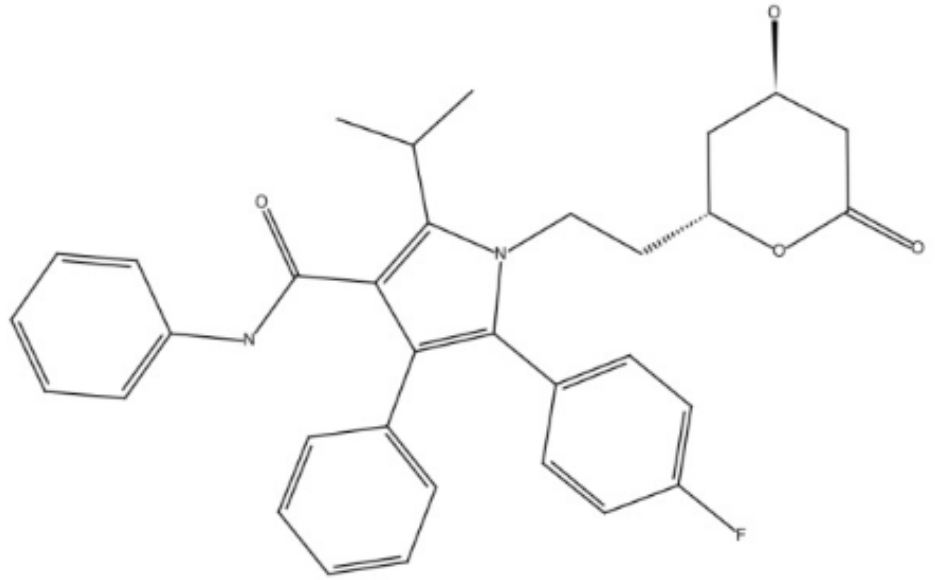
The following DCN roles are available from 1981 onwards.

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

Use of roles in DWPI

```
=> FILE DCR
```

```
=> S LIPITOR/CN

L1          2 LIPITOR/CN

=> D


L1     ANSWER 1 OF 2 DCR  COPYRIGHT 2024 CLARIVATE on STN.
AN     DCR-167146   DCR
DCSE   167146-1-0-0
CN.P   ATORVASTATIN-LACTONE
CN.S   5-(4-Fluoro-phenyl)-1-[2-(4-hydroxy-6-oxo-tetrahydro-pyran-2-yl)-ethyl]-
       2-isopropyl-4-phenyl-1H-pyrrole-3-carboxylic acid phenylamide
SY     ATORIS; ATORVASTATIN (LACTONE); ATORVASTATIN LACTONE;
       ATORVASTATIN-LACTONE; CARDYL; CI-981; LIPITOR; LIPRIMAR; SORTIS;
       STORVAS; TAHOR; TORVACARD; ZARATOR
```



```
CMT    The name "atorvastatin" also applies to the free acid.
MF     C33 H33 F N2 O4
ED     Entered STN: 24 Jun 1999
       Last updated on STN: 25 Jul 2016
       Update DWPI Cross Ref.: 8 Jan 2024
```


FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Use of roles in DWPI

```
=> FILE WPIN
```

```
=> S L1 (T) (PRD OR P)/RL

          1322 L1
       1091143 PRD/RL
       1293984 P/RL
L2          246 L1 (T) (PRD OR P)/RL

=> S DCR-167146/DCR (T) (PRD OR P)/RL

           739 DCR-167146/DCR
       1091143 PRD/RL
       1293984 P/RL
L3           41 DCR-167146/DCR (T) (PRD OR P)/RL
```


FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Agenda

- History of Derwent substance indexing
- Tour of DCR database
  - Searchable fields
  - Structure search
- Relationship with Derwent World Patents Index
  - Use of roles
- Search examples

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Structure searching on CAS STNext

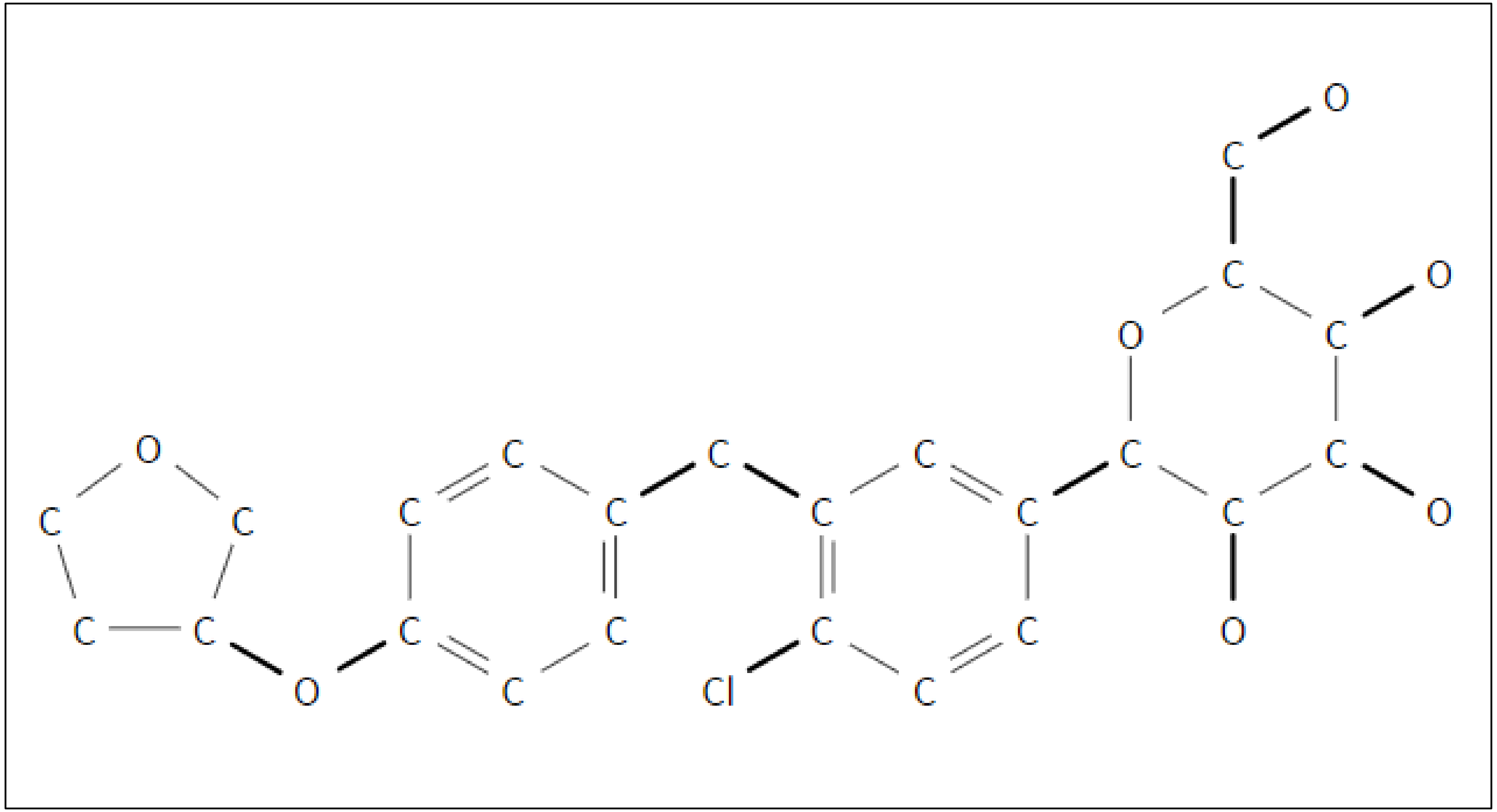
- CAS STNext has multiple structure searchable databases
- For comprehensiveness, search multiple databases
  - Consider databases from different producers
    - Different indexing criteria, different coverage, etc.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Multi-file structure search on CAS STNext

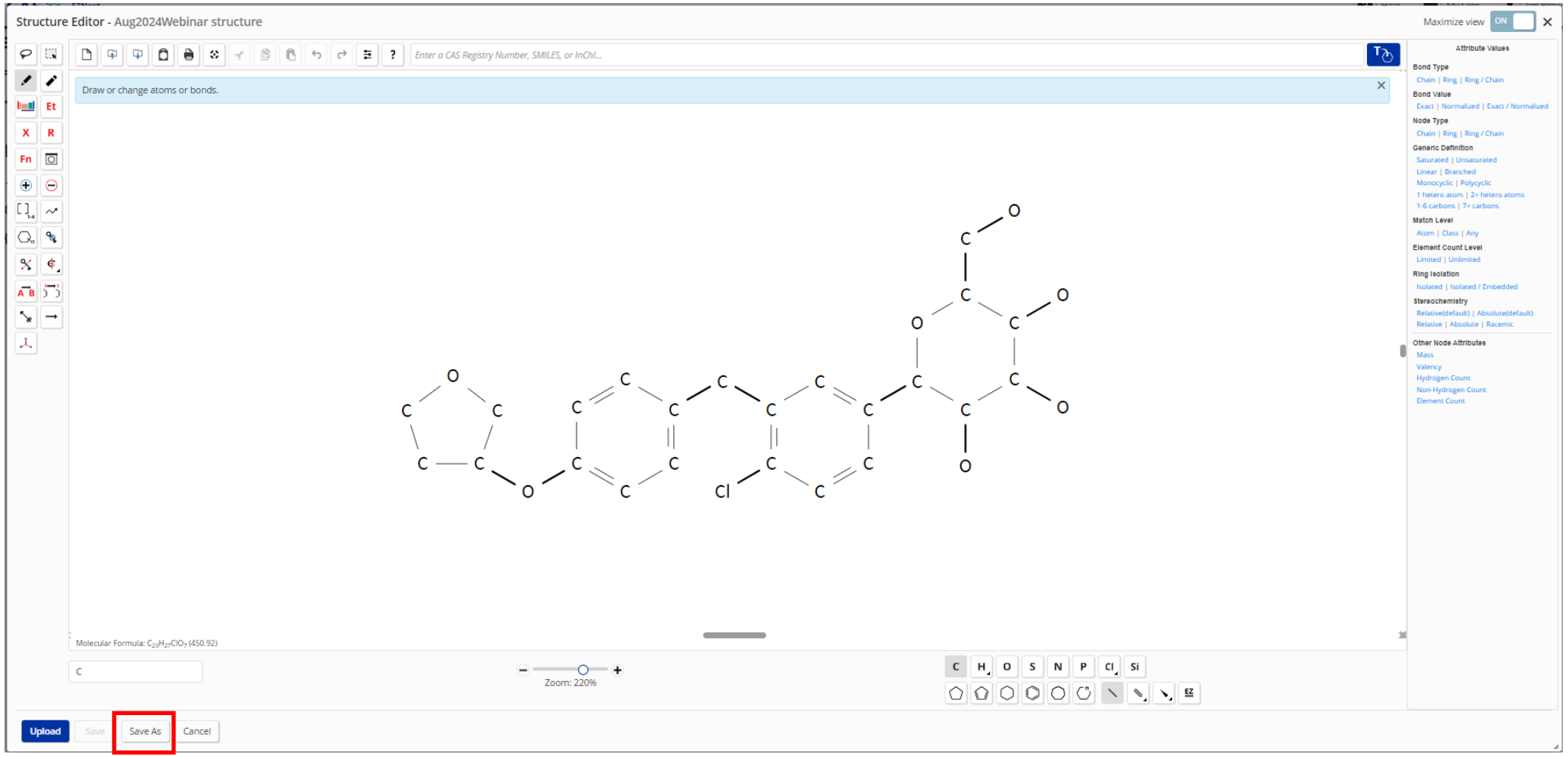
Find patent family records for this specific compound



Multi-file structure search

- Draw structure using STNext structure Editor
- Choose Save As (to name structure before uploading)
- Upload structure into CAS REGISTRY
- Run EXACT structure search in CAS REGISTRY
- Find corresponding patent families in (H)CAplus
- Run EXACT search in DCR
- Find corresponding patent families in DWPI
- TRANSFER patent numbers from (H)CAplus search into DWPI
- NOT them from the DWPI set to find unique patent families
- If there are any unique records, limit using Derwent product roles

STNext Structure Editor



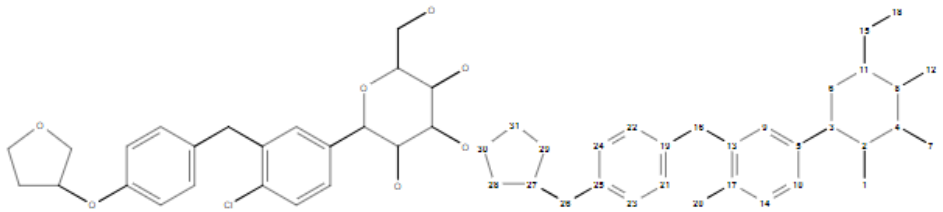
FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Multi-file structure searching on CAS STNext

```
=> FILE REGISTRY
```

```
Uploading structure file: Aug2024Webinar structure
```



```
Node Attributes
Ring Nodes : 2  3  4  5  6  8  9  10  11  13  14  17  19  21  22  23  24  25  27  28  29  30  31
Chain Nodes : 1  7  12  15  16  18  20  26
Bond Attributes
Ring Bonds : 2-3  2-4  3-6  4-8  5-9  5-10  6-11  8-11  9-13  10-14  13-17  14-17  19-21  19-22  21-23  22-24  23-25  24-25  27-28  27-29  28-30  29-31  30-31
Chain Bonds : 2-1  3-5  4-7  8-12  11-15  13-16  15-18  16-19  17-20  25-26  27-26
Exact Bonds : 3-5  11-15  13-16  16-19  17-20
Normalized Bonds : 5-9  5-10  9-13  10-14  13-17  14-17  19-21  19-22  21-23  22-24  23-25  24-25
Exact/Normalized Bonds : 2-1  2-3  2-4  3-6  4-7  4-8  6-11  8-11  8-12  15-18  25-26  27-26  27-28  27-29  28-30  29-31  30-31
Markush Attributes
Match Level (ATOM) : 2  3  4  5  6  8  9  10  11  13  14  17  19  21  22  23  24  25  27  28  29  30  31
Match Level (CLASS) : 1  7  12  15  16  18  20  26
Element Count Level (LIMITED) : 1  2  3  4  5  6  7  8  9  10  11  12  13  14  15  16  17  18  19  20  21  22  23  24  25  26  27  28  29  30  31


L1      STRUCTURE UPLOADED
```

```
=> S L1 EXA FULL

FULL SEARCH INITIATED 14:52:02
FULL SCREEN SEARCH COMPLETED -       80 TO ITERATE

100.0% PROCESSED         80 ITERATIONS                                22 ANSWERS
SEARCH TIME: 00.00.01

L2          22 SEA EXA FUL L1
```

An EXACT structure search captures exact structures plus those substances with isotopes.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Multi-file structure searching on CAS STNext

```
=> FILE HCAPLUS
```

```
=> S L2 AND P/DT

         2845 L2
     20567521 P/DT
L3        637 L2 AND P/DT

=> D BIB HITRN


L3   ANSWER 1 OF 637  HCAPLUS  COPYRIGHT 2024 ACS on STN
PatentPak PDF
AN   2024:1750368  HCAPLUS Full-text
TI   Combination of SGLT2 inhibitor and menthol and use thereof in treatment of
     heart disease
IN   Li, Jian; Xu, Yixiang; Huang, Yunyuan; Zhang, Chao; Yao, Yue
PA   East China University of Science and Technology, Peop. Rep. China
SO   PCT Int. Appl., 28pp.
     CODEN: PIXXD2
DT   Patent
LA   Chinese
FAN.CNT 1
```

P/DT limits the set to patent family records.

```
PPPI
     PATENT NO.          KIND  DATE      LANGUAGE   PatentPak
     ---------------     ----  --------  ---------- ------------------------
     WO 2024164183        A1   20240815  Chinese    PDF
PI
     PATENT NO.          KIND  DATE        APPLICATION NO.          DATE
     ---------------     ----  --------    ---------------------    --------
     WO 2024164183        A1   20240815    WO 2023-CN75028          20230208
PRAI WO 2023-CN75028           20230208
PSPI
     PATENT NO.          KIND  STATUS          STATUS DATE
     ---------------     ----  -------------   -----------
     WO 2024164183        A1   Alive           20240822
IT   864070-44-0, Empagliflozin
     RL: PAC (Pharmacological activity); THU (Therapeutic use); BIOL
     (Biological study); USES (Uses)
        (combination of SGLT2 inhibitor and menthol and use thereof in
        treatment of heart disease)
```


FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Multi-file structure searching on CAS STNext

```
=> FILE DCR
```

```
=> S L1 EXA FULL

FULL SEARCH INITIATED 15:08:51
FULL SCREEN SEARCH COMPLETED -        0 TO ITERATE

  0.0% PROCESSED    5684501 ITERATIONS                          12 ANSWERS
SEARCH TIME: 00.00.01

L4           12 SEA EXA FUL L1

=> FILE WPIN
```

No need to re-upload the structure into the same STNext session; L1 can be used here too.

```
=> S L4

L5          518 L4

=> TRA L3 PN 1-

L6           TRANSFER L3 1- PN :    2902 TERMS
L7          637 L6
L8              QUE  TERMS FROM L6 WITH NO HITS:   265 TERMS

=> S L4 NOT L7

            518 L4
L9           50 L4 NOT L7
```

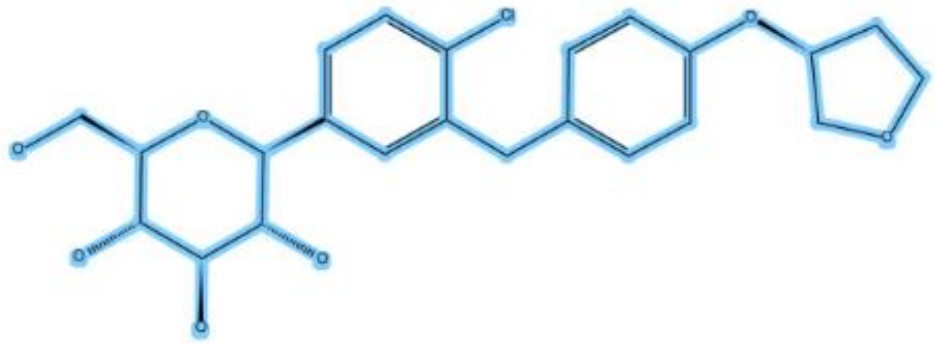

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Multi-file structure searching on CAS STNext

```
=> D BIB HITSTR


L9   ANSWER 1 OF 50  WPINDEX COPYRIGHT 2024   CLARIVATE on STN
AN   2024-82087A [2024068]   WPINDEX Full-text
TI   Empagliflozin-metformin sustained-release tablet, comprises a
     sustained-release tablet core comprising metformin hydrochloride, organic
     acid, surfactant, sustained-release material, excipient and lubricant, and
     immediate-release coat layer comprising empagliflozin and film-forming
     material
DC   A96; B04; B05; B07
IN   GENG F; GONG R; SHI G; WANG X; WANG Z
PA   (SHAN-N) SHANDONG KANMEILE PHARM TECHNOLOGY CO
CYC  1
PI   CN 118384120     A  20240726 (2024068)* ZH
ADT  CN 118384120  A  CN 2024-10567561  20240509
PRAI CN 2024-10567561      20240509
```

```
AN.S DCR-1395204
CN.P EMPAGLIFLOZIN
CN.S (2S,3R,4R,5S,6R)-2-(4-Chloro-3-{4-[(S)-(tetrahydro-furan-3-yl)oxy]-benzyl}-
     phenyl)-6-hydroxymethyl-tetrahydro-pyran-3,4,5-triol
MF   C23 H27 Cl O7
STR
```



Unique DWPI hit.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Considerations

- CAS and Derwent have different indexing criteria
- CAS will create REGISTRY records for mixtures; Derwent usually does not, they will index each component separately
- That unique Derwent record looks like a mixture; maybe changing the structure search from EXA to FAM will capture additional records of interest in REG/(H)CAplus portion of search

Multi-file structure searching on CAS STNext

```
=> FILE REG
```

```
=> S L1 FAM FULL

FULL SEARCH INITIATED 15:14:59
FULL SCREEN SEARCH COMPLETED -      172 TO ITERATE

100.0% PROCESSED        172 ITERATIONS                          84 ANSWERS
SEARCH TIME: 00.00.01

L10           84 SEA FAM FUL L1

=> FILE HCAPLUS
```

```
=> S L10 AND P/DT

          2848 L10
      20567521 P/DT
L11        637 L10 AND P/DT

=> FILE DCR
```

Re-running the structure searches using FAM instead of EXA.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Multi-file structure searching on CAS STNext

```
=> S L1 FAM FULL

FULL SEARCH INITIATED 15:15:35
FULL SCREEN SEARCH COMPLETED -        0 TO ITERATE

  0.0% PROCESSED    5684501 ITERATIONS                          48 ANSWERS
SEARCH TIME: 00.00.01

L12          48 SEA FAM FUL L1

=> FILE WPIN
```

```
=> S L12

L13        519 L12

=> TRA L11 PN 1-

L14         TRANSFER L11 1- PN :    2902 TERMS
L15        637 L14
L16            QUE  TERMS FROM L14 WITH NO HITS:   265 TERMS

=> S L13 NOT L15

L17         50 L13 NOT L15

=> S L9 AND L17

L18         50 L9 AND L17
```

Re-running the structure searches using FAM instead of EXA.

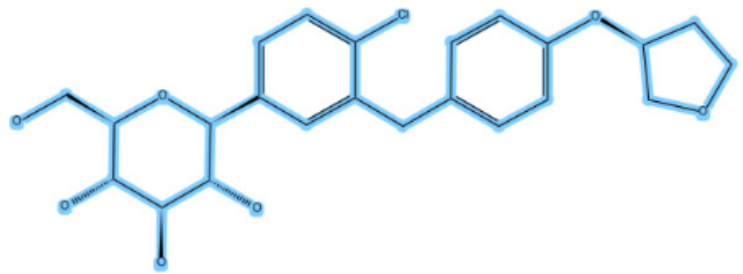
FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Multi-file structure searching on CAS STNext

```
=> S L12(T)(P OR PRD OR N OR NEW)/RL

          519 L12
      1293984 P/RL
      1091143 PRD/RL
       546905 N/RL
       430330 NEW/RL
L19       106 L12(T)(P OR PRD OR N OR NEW)/RL

=> S L18 AND L19

L20         5 L18 AND L19
```

```
=> D BIB HITSTR

L20  ANSWER 1 OF 5  WPINDEX COPYRIGHT 2024   CLARIVATE on STN
AN   2024-718533 [2024062]   WPINDEX Full-text
TI   Preparing Empagliflozin for treating type-2 diabetes involves reacting
     D-Gluconolactone compound with trimethylsilane, condensing
     octadecyltrimethoxysilane intermediate with benzylbenzene intermediate,
     reacting chloro derivative with methanol, reducing O-methylated
     Empagliflozin compound
DC   B03
IN   ATKARI A K; BADAM S; KATARI V K; KESINENI S R; SYAMALA N; UPPARAPALLI S;
     VANGARU S; VELISOJU M
PA   (GRAN-N) GRANULES INDIA LTD
CYC  1
PI   IN 202241077059 A  20240705 (2024062)* EN
ADT  IN 202241077059 A IN 2022-41077059 20221230
PRAI IN 2022-41077059      20221230
AN.S DCR-1395204
CN.P EMPAGLIFLOZIN
CN.S (2S,3R,4R,5S,6R)-2-(4-Chloro-3-{4-[(S)-(tetrahydro-furan-3-yl)oxy]-benzyl}-
     phenyl)-6-hydroxymethyl-tetrahydro-pyran-3,4,5-triol
MF   C23 H27 Cl O7
STR
```



Use the Derwent roles to limit to records where the substance had been indexed as NEW or PRODUCED (using all role systems.)

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Considerations

- Should structure search be expanded?
  - CSS or SSS
- Should other specific structure databases be searched?
  - REAXYSFILESUB/REAXYSFILEBIB
- Should Markush databases be searched?
  - MARPAT, DWPIM

Summary

- The Derwent Chemistry Resource database (DCR) is a valuable tool when substance searching in the Derwent databases
- Roles can help focus search results
- Multi-file searching should always be considered for comprehensive searching

Between problems and progress are connections that matter

CONTACT

CAS
help@cas.org
cas.org

EMEA Help
EMEAhelp@cas.org
stn-international.de

CAS
A division of the American Chemical Society