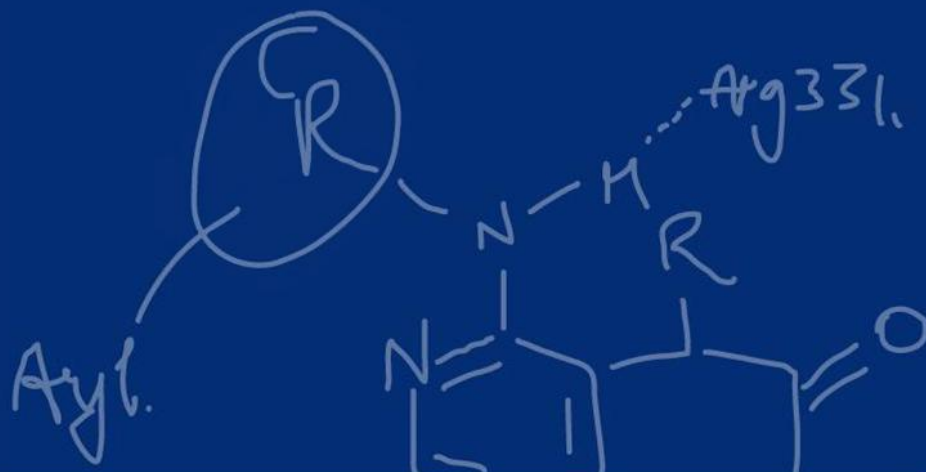
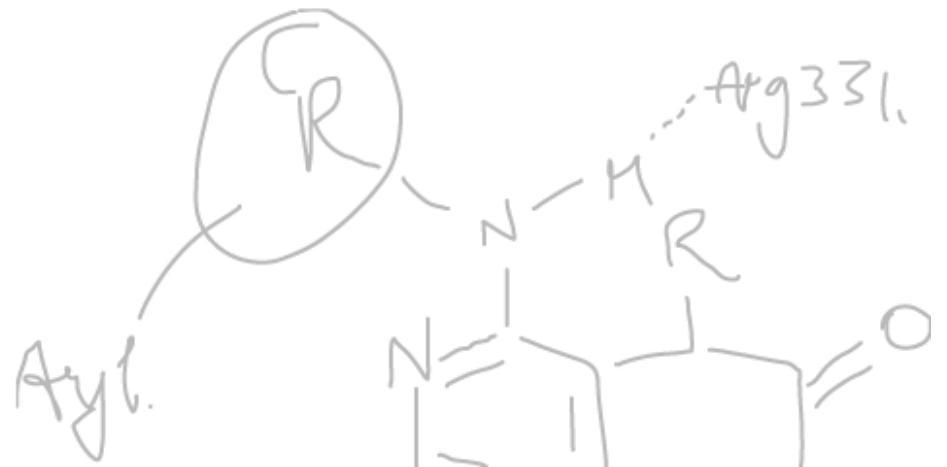

The challenges of loading chemical supplier catalogues

Creating integrated drug discovery
innovation alliances



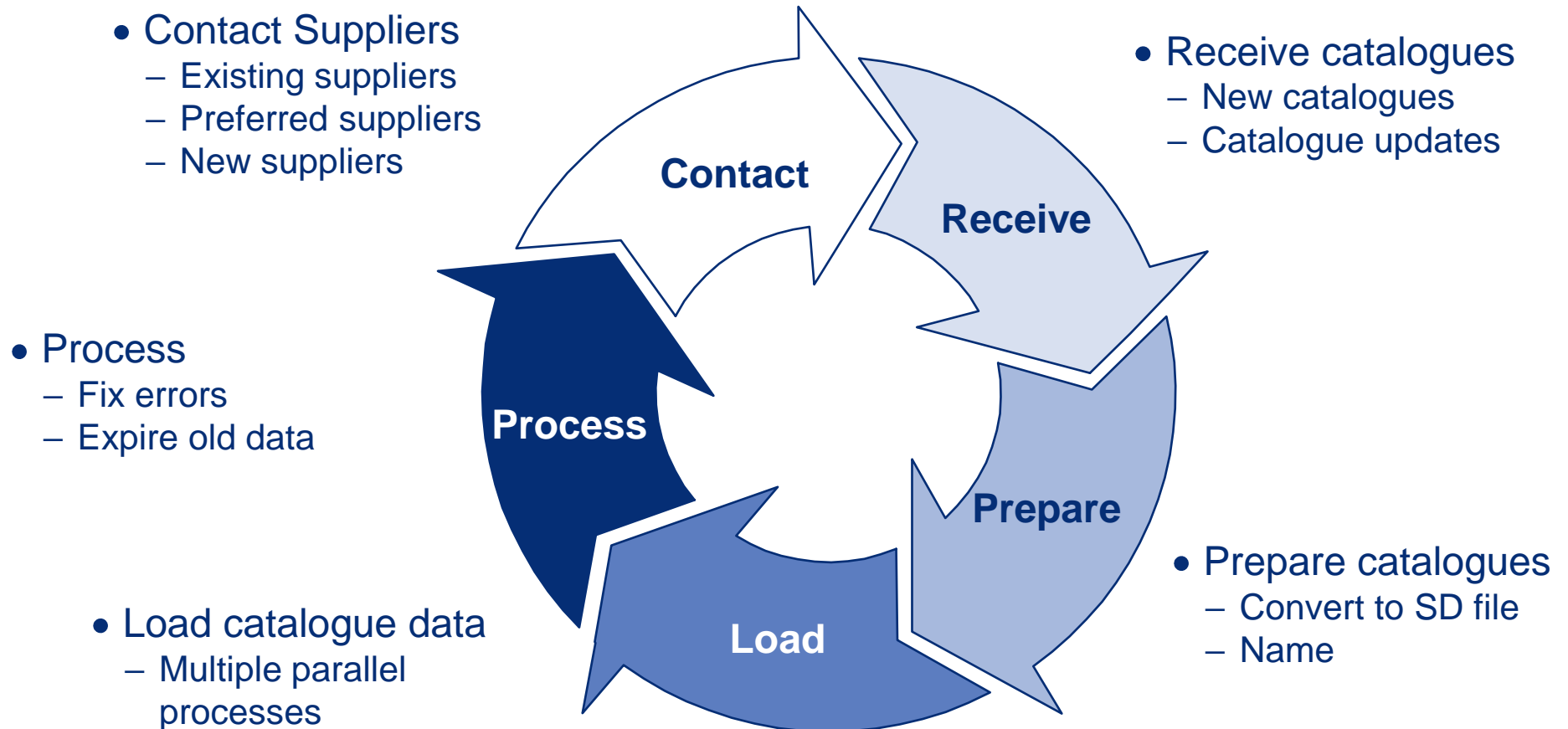
Agenda

- **EVOsource**
- **Suppliers & catalogues**
- **Preparation**
- **Loading**
- **Processing**
- **What would help us**



The challenges of loading supplier catalogues

Integrated cyclic process

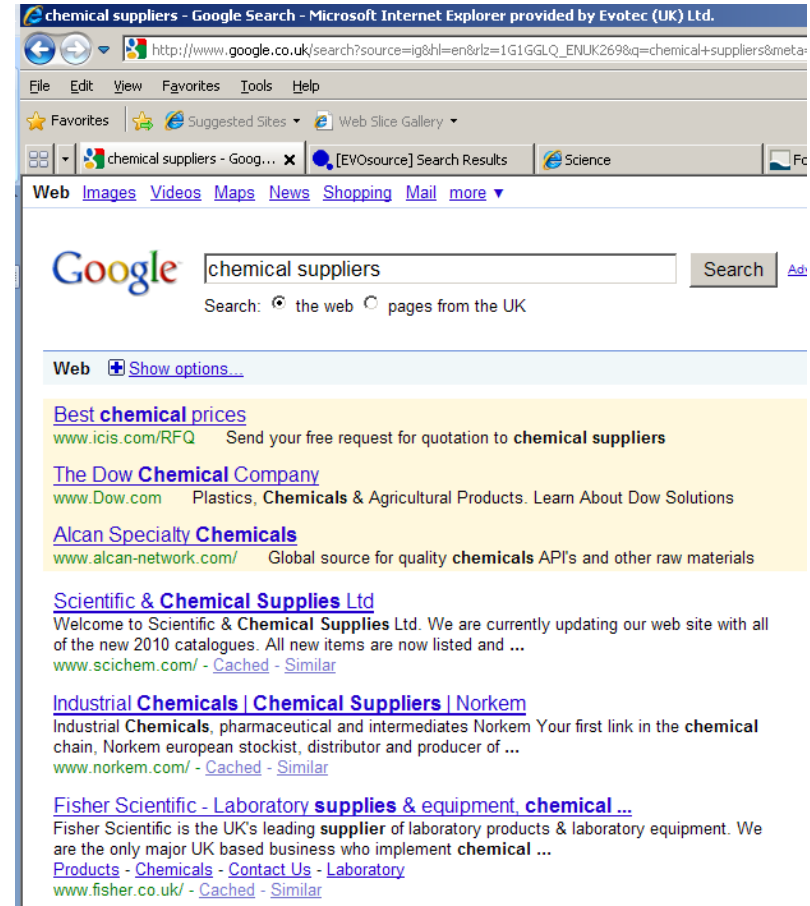


Suppliers

Sources of information

Contact details for suppliers

- Internet
- In house ERP system
- Preferred suppliers
- Buyers
- Chemists contacts issues
- Suppliers no longer exist
- Mergers and purchases
- Over 300 suppliers on original list to around 100 contactable



The screenshot shows a Microsoft Internet Explorer browser window displaying a Google search for "chemical suppliers". The search results are filtered to "the web". The top results include:

- Best chemical prices**: www.icis.com/RFQ Send your free request for quotation to **chemical suppliers**
- The Dow Chemical Company**: www.Dow.com Plastics, **Chemicals** & Agricultural Products. Learn About Dow Solutions
- Alcan Specialty Chemicals**: www.alcan-network.com/ Global source for quality **chemicals** APIs and other raw materials
- Scientific & Chemical Supplies Ltd**: Welcome to Scientific & **Chemical Supplies** Ltd. We are currently updating our web site with all of the new 2010 catalogues. All new items are now listed and ... www.scichem.com/ - [Cached](#) - [Similar](#)
- Industrial Chemicals | Chemical Suppliers | Norkem**: Industrial **Chemicals**, pharmaceutical and intermediates Norkem Your first link in the **chemical** chain, Norkem european stockist, distributor and producer of ... www.norkem.com/ - [Cached](#) - [Similar](#)
- Fisher Scientific - Laboratory supplies & equipment, chemical ...**: Fisher Scientific is the UK's leading **supplier** of laboratory products & laboratory equipment. We are the only major UK based business who implement **chemical** ... [Products - Chemicals](#) - [Contact Us](#) - [Laboratory](#) www.fisher.co.uk/ - [Cached](#) - [Similar](#)

Suppliers

Contacting

Contacts

- Personal contact via email
- General contact via email or website email
- Download files from website
- Files sent on CD

Issues

- No reply
- Long delay in reply
- Unsuitable files
- Files corrupt (download or CD)

Chemical Supplier and Catalogue Links

Useful resources covering chemical companies, products and catalogues.

Related Links

[ChemIndustry.com](#)

A comprehensive directory and search engine for chemical and related industry professionals.

[Chem.com](#)

Contains information about chemical producers and suppliers, MSDS's, structures, chemicals and chemical equipment for sale.

[ChemExper Chemical Directory](#)

Searchable by structure and chemical property, contains information about chemical producers and suppliers, MSDS's, structures, and chemicals for sale.

[ICIS Search](#)

Contains chemical industry news, views, comments, company and product information and chemical profiles.

[ChemConnect: Chemical Suppliers Directory and Exchange](#)

A virtual market place for buyers, suppliers and vendors of chemicals.

[Chemicals Technology](#)

Contains, among other resources, information on current projects - an industry and products and services guide in the field of chemical technology.

[Sigma Aldrich](#)

Produces and sells a broad range of biochemicals, organic and inorganic chemicals and related products. Parent company of Sigma, Aldrich, Fluka and Supelco.

[Chemical Register](#)

Free worldwide supplier database with 15,000 companies and 15,000 products. Database can be browsed and searched by company/chemical name and CAS Registry Number.

Catalogues

Wide variety

Catalogue types - usable

- SD - preferred format
- DB – convert to SD
- Excel with smiles – convert to SD
- Excel with embedded structures – convert to SD
- Excel with names – convert to SD
- Word with names – convert to SD

Catalogue types - unusable

- Excel with pictures and no names
- Word with pictures and no names
- PDF

catalog [Compatibility Mode] - Microsoft Excel

Home Insert Page Layout Formulas Data Review View Developer Add-Ins

Clipboard Font Alignment Number Styles

1 2 3

Times New Rom 10 A A

B I U

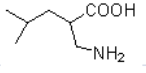
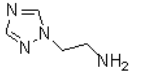
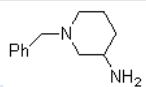
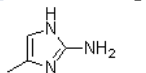
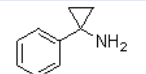
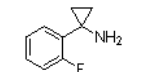
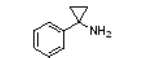
Wrap Text

General

Conditional Formatting as Table

Format

fx 1-(4-chlorophenyl)cyclobutanamine HCl

| | A | B | C | D | E | F | G | |
|----|-------|--|-------------|----------|--------|-----|---|-------|
| 83 | 71111 | 2-(aminomethyl)-4-methylpentanoic acid | 100869-07-6 | C7H15NO2 | 145.2 | 97% |  | 1 g 5 |
| 84 | 63121 | 2-(1H-1,2,4-triazol-1-yl)ethanamine Hydrogenchloride | 51444-31-6 | C4H8N4 | 148.63 | 97% |  | 1 g 5 |
| 85 | 42241 | 1-benzylpiperidin-3-amine | 60407-35-4 | C12H18N2 | 190.28 | 97% |  | 1 g 5 |
| 86 | 65111 | 4-methyl-1H-imidazol-2-amine | 6653-42-5 | C4H7N3 | 97.12 | 97% |  | 1 g 5 |
| 87 | 66111 | 1-phenylcyclopropanamine HCl | 773012-60-5 | C9H11N | 169.69 | 97% |  | 1 g 5 |
| 88 | 66121 | 1-(2-fluorophenyl) cyclopropanamine | 886366-50-3 | C9H10FN | 151.18 | 97% |  | 1 g 5 |
| | 66131 | 1-(3-fluorophenyl) cyclopropanamine | 764647-70-3 | C9H10FN | 152.18 | 97% |  | 1 g 5 |

Viewing catalogue data

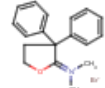
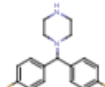
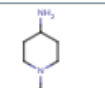
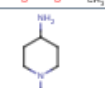
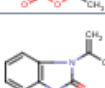
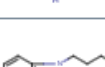
Checking and preparing for loading

MarvinView

- Checking data
- Preparing for loading
- Checking error files

SupplierCatalogue.sdf - MarvinView 5.3.1

File Edit View Table Tools Help

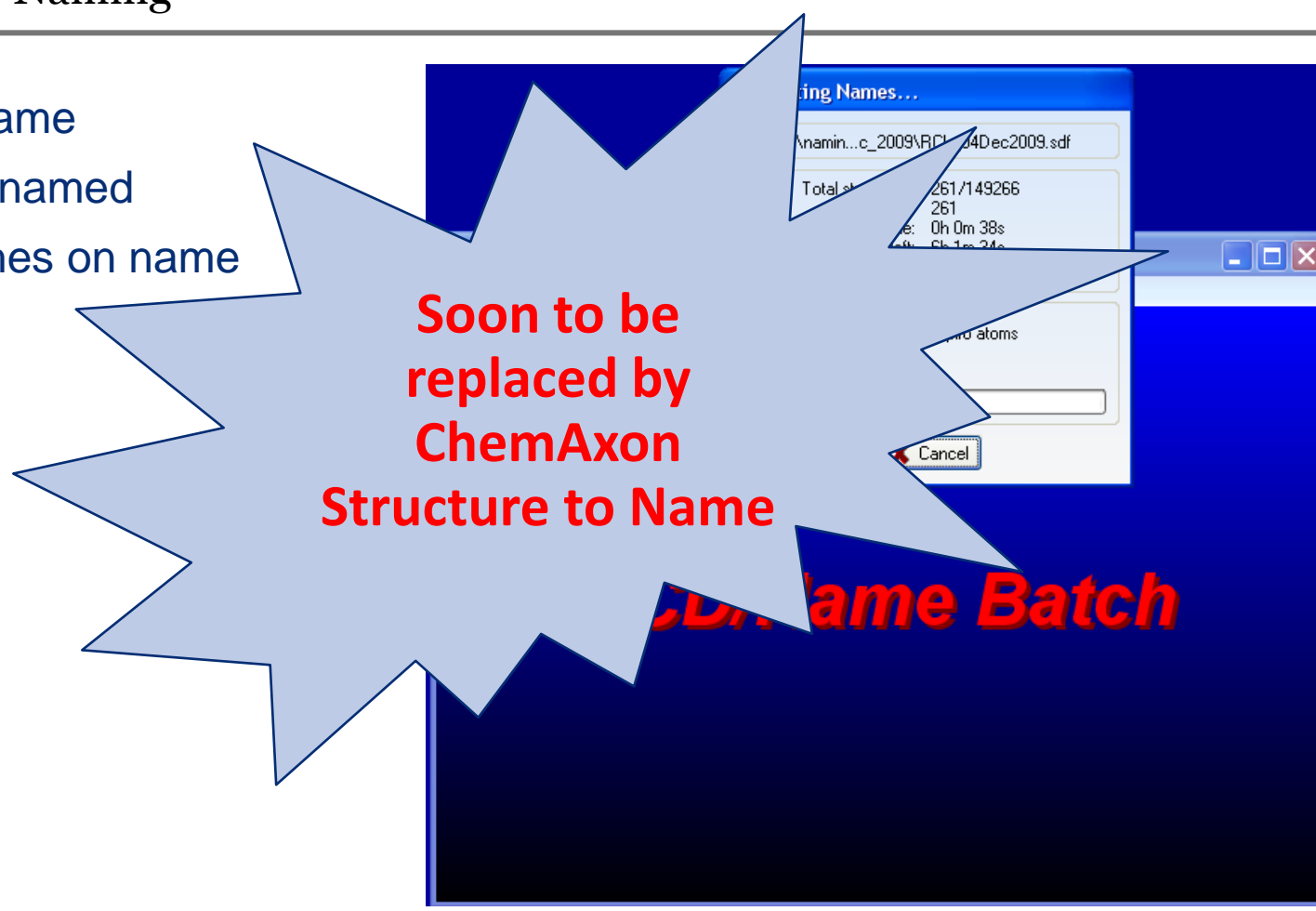
| # | structure | Cdid | Mol Weight | Formula | Reference | Name | Molecular Form... | Molecular Weight | CAS Number | Commercial n |
|---|---|------|------------|---|-----------|--|---|------------------|------------|--|
| 1 |  | 1 | 346.262 | C ₁₈ H ₂₀ BrNO | 10017 | Dihydro-N,N-dimethyl-3,3-diphenyl-2(3H)-furaninium bromide | C ₁₈ H ₂₀ BrNO | 346.2669 | 37743183 | Dihydro-N,N-dimethyl-3,3-diphenyl-2(3H)-furaninium bromide |
| 2 |  | 2 | 288.335 | C ₁₇ H ₁₈ F ₂ N ₂ | 10018 | 1-Bis(4-fluorophenyl)methyl piperazine | C ₁₇ H ₁₈ F ₂ N ₂ | 288.34012 | 27469609 | 1-Bis(4-fluorophenyl)methyl piperazine 97 |
| 3 |  | 3 | 172.2248 | C ₈ H ₁₆ N ₂ O ₂ | 10024 | Ethyl 4-amino-1-piperidinecarboxylate | C ₈ H ₁₆ N ₂ O ₂ | 172.22723 | 58859464 | Ethyl 4-amino-1-piperidinecarboxylate 98% |
| 4 |  | 4 | 172.2248 | C ₈ H ₁₆ N ₂ O ₂ | 18929 | Ethyl 4-amino-1-piperidinecarboxylate | C ₈ H ₁₆ N ₂ O ₂ | 172.22723 | 58859464 | Ethyl 4-amino-1-piperidinecarboxylate 95% |
| 5 |  | 5 | 174.1992 | C ₁₀ H ₁₀ N ₂ O | 10029 | 1-Isopropyl-2-benzimidazolidinone | C ₁₀ H ₁₀ N ₂ O | 174.2022 | 52099726 | 1-Isopropyl-2-benzimidazolidinone 99% |
| 6 |  | 6 | 210.66 | C ₁₀ H ₁₁ ClN ₂ O | 10030 | 1-(3-Chloropropyl)-2-benzimidazole | C ₁₀ H ₁₁ ClN ₂ O | 210.66315 | 62780896 | 1-(3-Chloropropyl)-2-benzimidazole |

Add IUPAC names for all structures

Naming

ACD Structure to name

- All structures are named
- Consistent searches on name
- Catalogue names
- Supplier names



**Soon to be
replaced by
ChemAxon
Structure to Name**

ACD Name Batch

Combining data in Instant JChem

Preparing for loading

Combining data

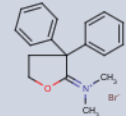
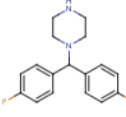
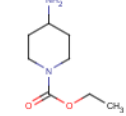
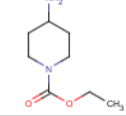
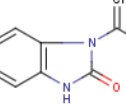
- Instant JChem
- Text Editors
- Excel

Instant JChem 5.3.1

File Edit View Search Data Lists Tools Window Help

Welcome x Grid view for SupplierCatalogue x

Query Browse

| ... | CdId | Structure | Mol Weight | Formula | Reference | Name | Molecular | Molecular | CA5 Number | Commercial | Purity | Package Size | Price GBP |
|-----|------|---|------------|------------|-----------|--|------------|-----------|------------|--|--------|--------------|-------------|
| 1 | 1 |  | 346.26 | C18H20BrNO | 10017 | Dihydro-N,N-dimethyl-3,3-diphenyl-2(3H)-furaniminium bromide | C18H20BrNO | 346.27 | 37,743,183 | Dihydro-N,N-dimethyl-3,3-diphenyl-2(3H)-furaniminium bromide 95% | 95% | 5 GR | 47.7 |
| 2 | 2 |  | 288.33 | C17H18F2N2 | 10018 | 1-Bis(4-fluorophenyl)methyl piperazine | C17H18F2N2 | 288.34 | 27,469,609 | 1-Bis(4-fluorophenyl)methyl piperazine 97% | 97% | 1 GR; 10 GR | 20.2; 107.2 |
| 3 | 3 |  | 172.22 | C8H16N2O2 | 10024 | Ethyl 4-amino-1-piperidinecarboxylate | C8H16N2O2 | 172.23 | 58,859,464 | Ethyl 4-amino-1-piperidinecarboxylate 98% | 98% | 5 ML; 25 ML | 22.6; 73.6 |
| 4 | 4 |  | 172.22 | C8H16N2O2 | 18929 | Ethyl 4-amino-1-piperidinecarboxylate | C8H16N2O2 | 172.23 | 58,859,464 | Ethyl 4-amino-1-piperidinecarboxylate 95% | 95% | 5 GR | 22.6 |
| 5 | 5 |  | 174.20 | C10H10N2O | 10029 | 1-Isopropyl-2-benzimidazolidone | C10H10N2O | 174.20 | 52,099,726 | 1-Isopropyl-2-benzimidazolidone 99% | 99% | 5 GR; 25 GR | 64.2; 206.1 |

Mapping catalogue data to data model

XML file

XML File

- Every catalogue has an XML file
- Maps catalogue data to database
- Catalogue data is read
- Catalogue data mapped
- Catalogue data parsed
- Catalogue data written to database

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<mappingFile xmlns="http://xml.evotec.com/schema/mapping-config">
  <source>Supplier Catalogue</source>
  <compoundMappings>
    <casNumber>CAS Number</casNumber>
    <compoundName>Commercial name</compoundName>
    <compoundNameAlternate>Name</compoundNameAlternate>
    <compoundNameAlternate>Evotec_IUPAC</compoundNameAlternate>
    <compoundNameAlternate>Evotec_IUPAC_1</compoundNameAlternate>
    <hazard>
      <hazards>Hazard</hazards>
      <separator> </separator>
    </hazard>
    <iupacName>Evotec_IUPAC</iupacName>
    <mfdNumber>MDL number</mfdNumber>
  </compoundMappings>
  <partMappings>
    <supplierName>SupplierName</supplierName>
    <catalogueName>CatalogueName</catalogueName>
    <catalogueType>Building Blocks</catalogueType>
    <meltingPoint>melting point</meltingPoint>
    <boilingPoint>boiling point</boilingPoint>
    <purity>Purity</purity>
    <density>density</density>
    <pack>
      <site>1</site>
      <catalogueNumber>Catalog Number</catalogueNumber>
      <packProperty>Supplier</packProperty>
      <packProperty>Risk</packProperty>
      <packProperty>Safety</packProperty>
      <packProperty>Refractive index</packProperty>
      <packProperty>Optical activity</packProperty>
      <packSize>Package Size</packSize>
      <price>Price GBP</price>
      <currencyValue>GBP</currencyValue>
      <separator>;</separator>
    </pack>
  </partMappings>
</mappingFile>
```

Catalogue loading

Loader

Loading

- Split into 1000 structure blocks
- Splitter writes .bat file and splits SD file
- Splitter uses ChemAxon API
- Increases loading efficiency
- Eases error reporting
- Eases error processing
- Run from DOS prompt
- Local machine
- Multiple instances running (12)
- 18k loaded/hour

```

C:\WINDOWS\system32\cmd.exe
28/Apr/2010 11:53:22 [main] DEBUG com.evotec.evosource.loader.Loader - Number of
Parts = 1
28/Apr/2010 11:53:22 [main] DEBUG com.evotec.evosource.loader.Loader - part =
Supplier = ChemII Site = Abingdon Catalogue =2009 Stock BB Catalogu
e number = UYBB-00959 Compound instance = 2021554 Compound name = 1-chloro
naphtho[2,1-b]thiophene-2-carboxylic acid Boiling Point = null Melting
Point = null Density = null Hazard = Pack Size = null Pack Siz
e Unit = null Price = null Currency = null Comments = null Properties = []
Active = true
28/Apr/2010 11:53:22 [main] DEBUG com.evotec.evosource.loader.Loader - Number of
Parts Properties= 0
28/Apr/2010 11:53:22 [main] DEBUG com.evotec.evosource.service.PartService - No
result found for the part with Catalogue Number:UYBB-00959
28/Apr/2010 11:53:22 [main] DEBUG com.evotec.evosource.loader.Loader - Created e
voPart 25542276
28/Apr/2010 11:53:22 [main] INFO com.evotec.evosource.loader.Loader - Loaded: 95
9, %Loaded: 96, Failed: 0
28/Apr/2010 11:53:22 [main] DEBUG com.evotec.evosource.loader.Loader - Before St
andardization CC1CCc2sc(cc2C1)C(O)=O
28/Apr/2010 11:53:22 [main] DEBUG com.evotec.evosource.loader.Loader - After Sta
ndardization CC1CCc2sc(cc2C1)C(O)=O
28/Apr/2010 11:53:22 [main] DEBUG com.evotec.evosource.loader.Loader - Number of
Chiral Centers = 1
28/Apr/2010 11:53:22 [main] DEBUG com.evotec.evosource.loader.Loader - No stereoc
bonds were removed CC1CCc2sc(cc2C1)C(O)=O
28/Apr/2010 11:53:25 [main] INFO com.evotec.evosource.loader.Loader - Purity is
null
28/Apr/2010 11:53:25 [main] DEBUG com.evotec.evosource.loader.Loader - Number of
Parts = 1
28/Apr/2010 11:53:25 [main] DEBUG com.evotec.evosource.loader.Loader - part =
Supplier = ChemII Site = Abingdon Catalogue =2009 Stock BB Catalogu
e number = UYBB-00960 Compound instance = 2356915 Compound name = 5-methyl
-4,5,6,7-tetrahydro-1-benzothiophene-2-carboxylic acid Boiling Point = null
Melting Point = null Density = null Hazard = Pack Size = null
Pack Size Unit = null Price = null Currency = null Comments = null Propert
ies = [] Active = true
28/Apr/2010 11:53:25 [main] DEBUG com.evotec.evosource.loader.Loader - Number of
Parts Properties= 0
28/Apr/2010 11:53:25 [main] DEBUG com.evotec.evosource.service.PartService - No
result found for the part with Catalogue Number:UYBB-00960
28/Apr/2010 11:53:25 [main] DEBUG com.evotec.evosource.loader.Loader - Created e
voPart 25542277
28/Apr/2010 11:53:25 [main] INFO com.evotec.evosource.loader.Loader - Loaded: 96
0, %Loaded: 96, Failed: 0
28/Apr/2010 11:53:26 [main] DEBUG com.evotec.evosource.loader.Loader - Before St
andardization CC1CCc2sc(cc2C1)C(O)=O
28/Apr/2010 11:53:26 [main] DEBUG com.evotec.evosource.loader.Loader - After Sta
ndardization CC1CCc2sc(cc2C1)C(O)=O
28/Apr/2010 11:53:26 [main] DEBUG com.evotec.evosource.loader.Loader - Number of

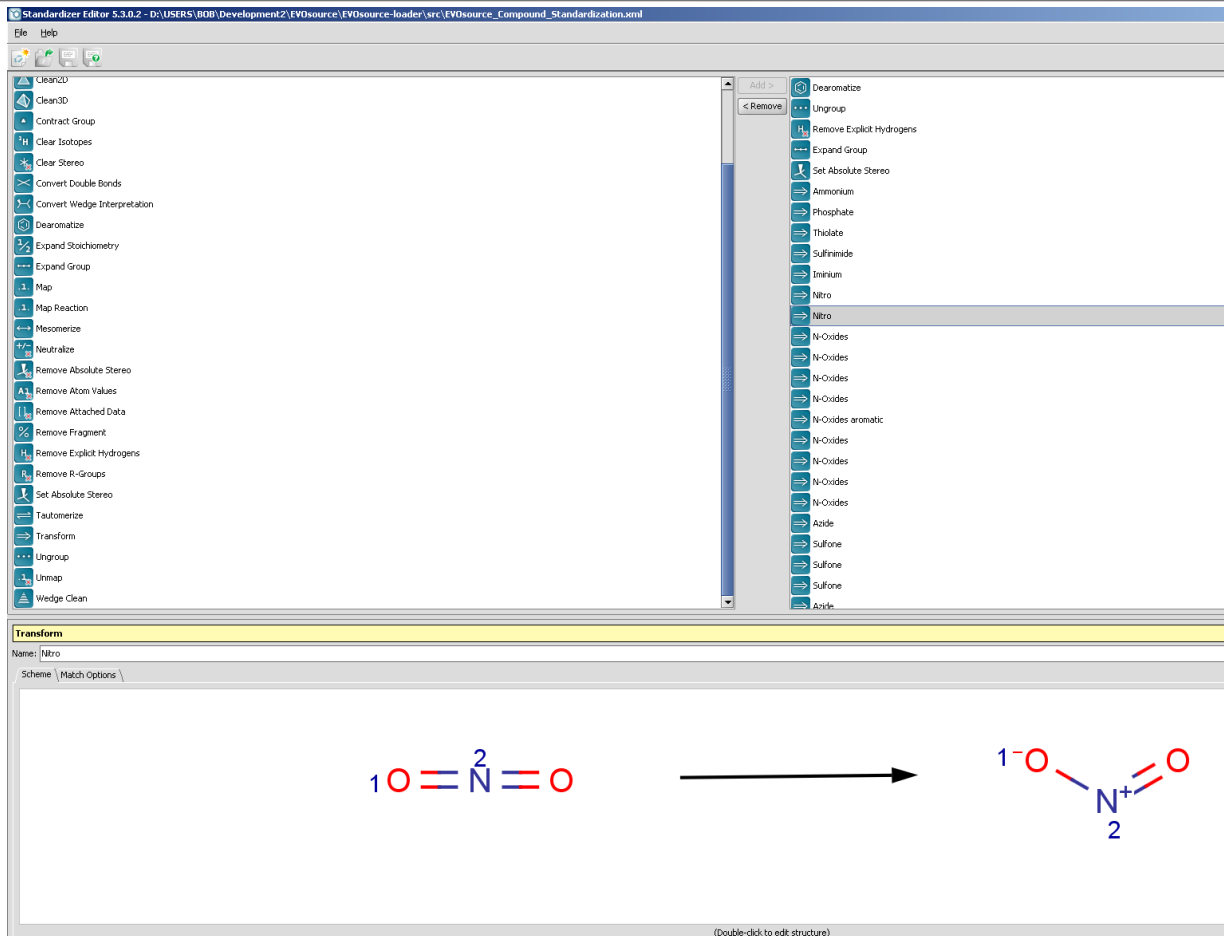
```

Standardization

Cleaning the structures

Standardizer

- Business rules
- Consistent structures
- Carried out by Loader
- List of Transforms
- 65 transforms so far



The screenshot shows the Standardizer Editor interface. The main window displays a list of transforms on the left and a detailed view of the selected 'Nitro' transform on the right. The 'Nitro' transform details show the following chemical structures:

Input structure: $1 \text{O} = \overset{2}{\text{N}} = \text{O}$

Output structure: $1^- \text{O} - \text{N}^+ = \overset{2}{\text{O}}$

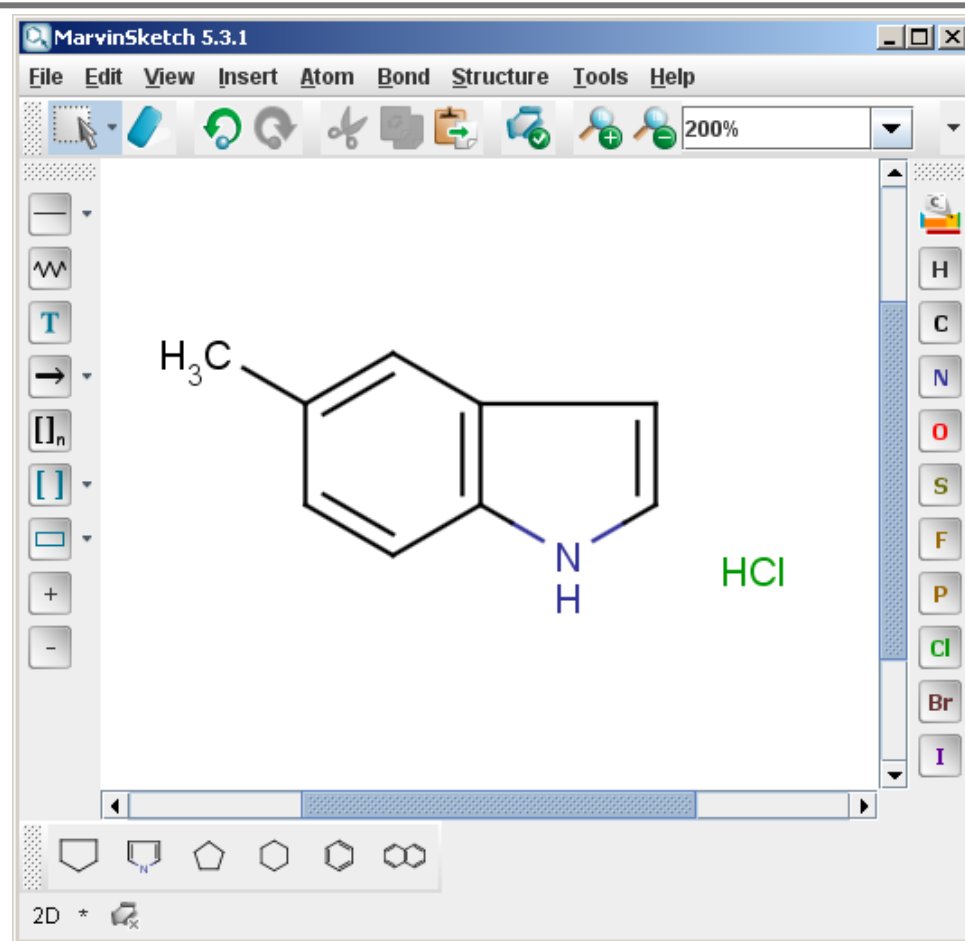
(Double-click to edit structure)

Parent and Version

Finding the salts

Loading

- Compounds are split into components
- Fragments compared to salts table
- Salt ratio stored in salts table
- Compound without salts stored in structure table
- Compound with salts stored in compound table
- Chemists require compound information
- Computational chemists require de-salted compound information

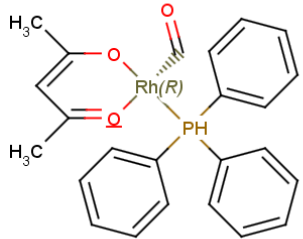
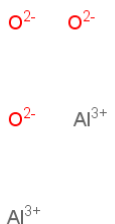
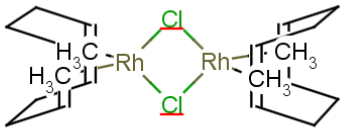


Error and warning files

Processing errors

Processing error files

- Marvin View
- Instant JChem
- Marvin Sketch
- Text editors
- Use DOS to recombine error files

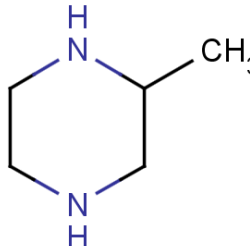
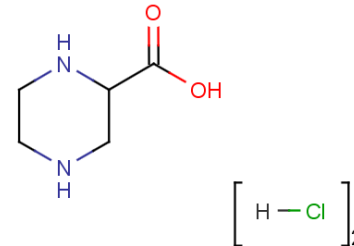
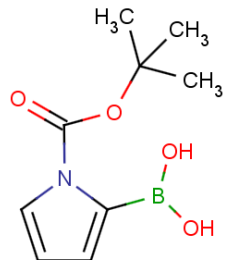
| # | structure | Loader_Error | Exc |
|---|--|---|---|
| 1 |  | The molecule was found to have a VALENCE ERROR when checking <chem>CC1=CC(C)=C(O)Rh@([O-])(C=O)P(c1ccccc1)(c1ccccc1)c1ccccc1 [c:4,t:1]</chem> | formyl[4-(hydroxy-kappaO)pent-3-en-2-onato-k |
| 2 |  | Mixture found for this compound | |
| 3 |  | The molecule was found to have a VALENCE ERROR when checking <chem>C[Rh]1(C)[C]([Rh])C(C)C([C]1.C1C=C=C1C1C=C=C1.C1C=C=C1C1C=C=C1 [c:10,14,18,22]</chem> | di-mu-chloro(tetramethyl)dirhodium - (1Z,5Z)- |

What sort of errors do we get?

Errors

Errors

- Empty structures
- Prices mixed with other data
- Prices and pack sizes together
- Inconsistent separators
- Duplicates with different data
- Incorrect stereochemistry
- Incorrectly drawn structures
- Structures with valence errors
- Incorrect bonding in structures
- Incorrect data – CAS, MFCD
- Incorrect data – purity etc

| SupplierCatalogue2.sdf - MarvinView 5.3.1 | | | |
|---|---|--|------------------------------|
| # | structure | Price | |
| 1 |  | Bulk available | 2-Methylpiperazine |
| 2 |  | 500g: \$1480 1kg: \$1800 Bulk in stock | Piperazine-2-carboxylic acid |
| 3 |  | 1g: \$980 | 1H-Pyrole-1-carboxylic acid |

What do we do with the errors?

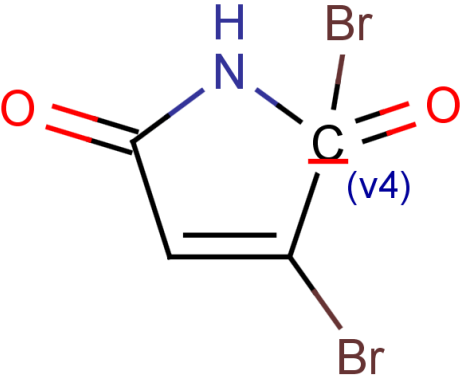
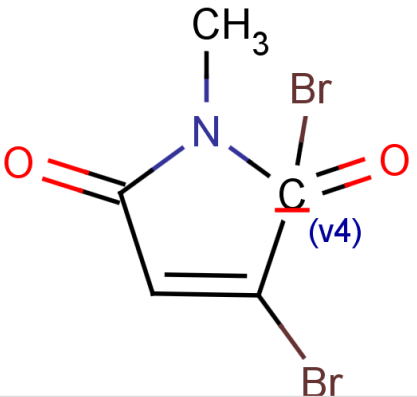
To fix or not?

Fixable

- Structure errors
- Valence errors
- Bonding errors
- Stereochemical errors
- Mixed pricing and sizing

Not fixable

- CAS numbers
- MFCD numbers
- Physical data

| # | structure | Name | Loader_Error |
|---|---|-------------------------------|--|
| 7 |  | 3,4-Dibromomaleimide | The molecule was found to have a VALENCE ERROR when checking BrC1=C(O)N[C]1(Br)O [1:1] |
| 8 |  | 3,4-Dibromo-N-methylmaleimide | The molecule was found to have a VALENCE ERROR when checking CH1C(O)C=C(Br)C1(Br)O [1:4] |



Keeping track

Excel

Keeping track of data

- Loader tracking
- Supplier tracking

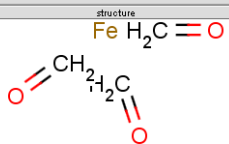
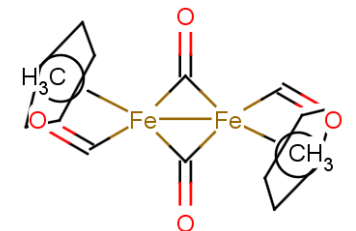
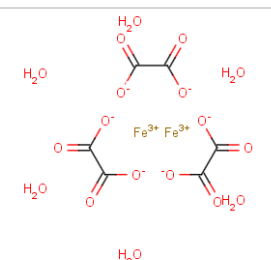
| Supplier | Pref | Active | Logol | Score | Build | Address |
|---------------------------------------|------|--------|-------|-------|-------|--|
| 2 ACB Blocks | 0 | 1 | 1 | | | 1 ACB Blocks Ltd, 121609, P.O. Box 10 Moscow, Russia, +7 (495) 761-4365, +7 (495) 646-0434 |
| 3 Ace Synthesis | 0 | 1 | 0 | | | 1 Ace Synthesis LLC, 21-C Olympia Avenue, Suite 10, Woburn, MA 01801, USA |
| 4 ACROS | 1 | 1 | 1 | | | 1 Thermo Fisher Scientific, Bishop Meadow Road, Loughborough, LE11 3RG |
| 5 Activate Scientific | 0 | 1 | 0 | | | 1 Cadogan House, Dantrées, Wildford Ware, Hertfordshire SG12 8SL, UK, Tel: +44 (0) 1279 / 84 32 44 |
| 6 Adren | 0 | 1 | 0 | | | 1 Adren, Inc., 27 McCullough Drive, New Castle, DE 19720 |
| 7 Advanced Chemical Intermediates Ltd | 0 | 1 | 0 | | | 1 Advanced Chemical Intermediates Ltd, The Old Dairy, Tresplatt, Davidstow, Cornwall, PL32 9YD |
| 8 AF ChemPharm | 0 | 1 | 1 | | | 1 Unit C14 MDC, 39 Allison Crescent, Sheffield, S2 1AS UK, Tel: +44 114 253 0022 |
| 9 Alder Aesar | 1 | 1 | 1 | | | 1 Alder TC, Ltd, Electrodrava St. 2, Bldg. 16, 111524, Moscow, Russia, Tel: +7 (495) 672-78-13, Fax: +7 (495) 672-78-14 |
| 10 Alkermes | 0 | 1 | 0 | | | 1 Alkermes LLC, 11 Melrose Lane, Suite 228, East Hanover, New Jersey 07936, US |
| 11 Alkermes | 0 | 1 | 1 | | | 1 Ambinter SARL, 50 avenue de Versailles, F-75016 Paris, France, Phone: +33 1 4524 4860, Fax: +33 1 4524 6241 |
| 12 AMPRI | 0 | 1 | 1 | | | 1 226 Corporate Circle, Albany, New York 12203, Phone: 518-512-2000, Fax: 518-512-2000 |
| 14 AnalyCon Discovery | 0 | 1 | 1 | | | 1 Hermannswerder Haus 17, 14473 Potsdam, Germany |
| 15 Apin Chemicals | 0 | 1 | 0 | | | 1 Apin Chemicals Ltd, 43d Milton Park, Abingdon, Oxon, OX14 4RU, UK |
| 16 Apollo | 1 | 1 | 1 | | | 1 Apollo Scientific Ltd, Whitefield Rd, Bredbury, Stockport, Cheshire, SK6 2QR, Telephone: +44 (0)161 406 0505, Fax: +44 (0)161 406 0506 |
| 17 Anaspec | 0 | 1 | 0 | | | 1 2149 O'Toole Ave., San Jose, CA 95131, Toll-Free: 800-452-5530, Tel: 408-452-5055, Fax: 408-452-5059 |
| 18 Aray BioPharma | 0 | 1 | 0 | | | 1 Aray BioPharma Inc., 3200 Walnut Street, Boulder, CO 80301 |
| 19 ASCO | 0 | 1 | 0 | | | 1 ASCO Group of Companies, 601 Interchange Blvd., Newark, DE 19711 USA, Telephone: 302-266-6891, Fax: 302-266-6296, Toll Free: 800-452-5530 |
| 20 Astatech | 0 | 1 | 0 | | | 1 Keystone Business Park, 2525 Pearl Buck Road, Bristol, PA 19007, Tel: +1-215-785-3197, +1-215-785-3247, Fax: +1-215-785-3247 |
| 21 Aseanex | 0 | 1 | 1 | | | 1 20 Genevieve Parfittovtzev Str. Bldg 1, Moscow 125480, Russia, +7 495 780 3413 / +7 495 780 3417, +7 495 780 3418 |
| 22 Atlantic Research | 0 | 1 | 0 | | | 1 Atlantic Research Chemicals Ltd, Unit A, Stratton View Business Park, Stratton, Bude, Cornwall, EX23 9WR |
| 23 Aurora fine chemicals | 0 | 1 | 1 | | | 1 Reininghausstr. 49, Graz A-8020, Austria |
| 24 Beta Pharma | 0 | 1 | 0 | | | 1 Bio-Farma Ltd, 44 Alberta St, Greater Manchester, BL3 5SD, UK |
| 25 Bio-Farma | 0 | 1 | 0 | | | 1 Flughafenend 9a P.O. Box 107125 D-28071 Bremen |
| 26 Biology Life Science Institute | 0 | 1 | 1 | | | 1 BIO-SYNTH AG, Rietlistr. 4, 9422 Staad, Switzerland, P: +41 (0)71 858 20 20, F: +41 (0)71 858 20 30 |
| 27 Bioplyth | 0 | 1 | 0 | | | 1 BIOTREND Chemikalien GmbH, Im Technologiezentrum Koehn, Eupener Str. 157, D-50933 Köln |
| 28 BIOTREND Chemicals | 0 | 1 | 0 | | | 1 Borochem, Immeuble Emergence, 7, rue Alfred Kastler |
| 29 Borochem | 0 | 1 | 0 | | | 1 Borochem, Immeuble Emergence, 7, rue Alfred Kastler |
| 30 Boron Molecular | 0 | 1 | 0 | | | 1 Boron Molecular Inc, P.O. Box 12592, Research Triangle Park, NC 27709 |
| 31 Bridge Organics | 0 | 1 | 0 | | | 1 311 W Washington St., Woburn, MA 01897-1200 |
| 32 Butt Park | 0 | 1 | 0 | | | 1 Butt Park Ltd, 21-25 St James St, Okehampton, Devon EX20 1DH |
| 33 Carboxynth | 0 | 1 | 0 | | | 1 Carboxynth Limited, 8 & 9 Old Station Business Park - Compton - Berkshire - RG20 6NE - UK, Tel: +44 (0)1635 578444 - Fax: +44 (0)1635 578444 |
| 34 ChemBridge corporation | 0 | 1 | 1 | | | 1 ChemBridge Corporation, 16981 Via Tazon, Suite G, San Diego, CA 92127 |
| 35 ChemBridge | 0 | 1 | 1 | | | 1 140-144 Leningky pr., 119334 Moscow Russia |
| 36 Chemivate | 0 | 1 | 1 | | | 1 Chemivate Limited, Unit 5, Canal Terrace, Workson, Nottinghamshire, S80 2DF |
| 37 ChemPacific | 0 | 1 | 0 | | | 1 ChemPacific Corporation, 6200 Freeport Centre, Baltimore, Maryland 21224, US |
| 38 Chem-Tek | 0 | 1 | 1 | | | 1 Chemical Technologies & Investigations Ltd, 28 Vavilova Street, office 354, Moscow, Russia 119991 |
| 39 Cheshire Sciences | 0 | 1 | 1 | | | 1 Cheshire Sciences (UK) Ltd Oak Farm Buildings Chester Road Aldford Chester CH3 6HJ |
| 40 Chesa Fine Organics | 0 | 1 | 0 | | | 1 CHESS GmbH, Max-Planck-Str. 1, 58169 Marnheim, Germany |
| 41 Chirac | 0 | 1 | 0 | | | 1 Chirac B.V., 6525 EC Nijmegen, The Netherlands, T: +31 24 365 2194, F: +31 24 365 2259 |
| 42 Chiralvision | 0 | 1 | 0 | | | 1 ChiralVision, Wassenaarseweg 72, 2333 AL Leiden, The Netherlands |
| 43 Chortech | 0 | 1 | 0 | | | 1 Chortech, Inc., 9 Giovanni Drive, Waterford, CT 06385, USA |
| 44 CIP Technologies | 0 | 1 | 0 | | | 1 CIP Technologies, Inc., 104 Heritawh Street, Woburn, MA 01801 |
| 45 Comb-Blocks Inc | 0 | 1 | 0 | | | 1 Comb-Blocks, Inc., 7949 Silverton Avenue, Suite 915, San Diego, CA 92126, USA |
| 46 CombPhos Catalysts | 0 | 1 | 0 | | | 1 CombPhos Catalysts Inc, P.O. Box 220, Princeton, New Jersey 08542-0220, USA |
| 47 CNS Chemicals Co | 0 | 1 | 0 | | | 1 1119 19 St., Shanghai 200001, China, Tel: +86 21 6352 9555, Fax: +86 21 6352 9555 |

| Run | Loader Version | Date Loaded | Re-loading after previous run | SD File Count | Supplier Name | Catalogue Name | Number Compound |
|-----|----------------|-------------|-------------------------------|---------------|------------------------------------|-------------------------------------|-----------------|
| 120 | 119 V2.3 | | | 101881 | VitasLabs | Stock Oct 2009 | |
| 121 | 120 V2.3 | | | 128059 | VitasLabs | Stock Oct 2010 | |
| 122 | 121 V2.3 | | | 98 | 932 Life chemicals | Carbonic Anhydrase Jan2010 | |
| 123 | 122 V2.3 | | | 84 | 722 Life chemicals | Matrix Metalloprotease Jan2010 | |
| 124 | 123 V2.4 | 23/01/2010 | | | 7706 Astatech | Catalogue May 2009 | 7 |
| 125 | 124 V2.4 | 24/01/2010 | | | 100000 Chembridge corporation | Screening Collection June 2009 | 803 |
| 126 | 125 V2.4 | | | | 100000 Chembridge corporation | Screening Collection June 2009 | 803 |
| 127 | 126 V2.4 | 24/01/2010 | | | 100000 Chembridge corporation | Screening Collection June 2009 | 803 |
| 128 | 127 V2.4 | | | | 100000 Chembridge corporation | Screening Collection June 2009 | 803 |
| 129 | 128 V2.4 | | | | 100000 Chembridge corporation | Screening Collection June 2009 | 803 |
| 130 | 129 V2.4 | | | | 100000 Chembridge corporation | Screening Collection June 2009 | 803 |
| 131 | 130 V2.4 | | | | 100000 Chembridge corporation | Screening Collection June 2009 | 803 |
| 132 | 131 V2.4 | | | | 100000 Chembridge corporation | Screening Collection June 2009 | 803 |
| 133 | 132 V2.4 | | | | 5484 Chembridge corporation | Screening Collection June 2009 | 803 |
| 134 | 133 V2.4 | 26/01/2010 | | | 97967 Chembridge corporation | EXP Collection June 2009 | 434 |
| 135 | 134 V2.4 | 26/01/2010 | | | 100000 Chembridge corporation | EXP Collection June 2009 | 434 |
| 136 | 135 V2.4 | 27/01/2010 | | | 100000 Chembridge corporation | EXP Collection June 2009 | 434 |
| 137 | 136 V2.4 | | | | 100000 Chembridge corporation | EXP Collection June 2009 | 434 |
| 138 | 137 V2.4 | | | | 36183 Chembridge corporation | EXP Collection June 2009 | 434 |
| 139 | 138 V2.4 | | | | 14083 Chembridge corporation | GPCR Collection June 2009 | 14 |
| 140 | 139 V2.4 | | | | 3900 Chembridge corporation | KINACORE Collection June 2009 | 3 |
| 141 | 140 V2.4 | | | | 10460 Chembridge corporation | NEBB Collection June 2009 | 10 |
| 142 | 141 V2.4 | 20/01/2010 | | 3 | 17703 Acros | Acros Jul2009 | |
| 143 | 142 V2.4 | 20/01/2010 | | | 26310 Apollo Scientific | Organics Mar2009 | |
| 144 | 143 V2.4 | | | 44 | 1 Key Organics | Agrimedates Dec2009 | |
| 145 | 144 V2.5 | | | 142 182+26 | Apollo Scientific | Organics Mar2009 | |
| 146 | 145 V2.4 | | | | 3014 Activate | Activate Scientific June 2009 | 2 |
| 147 | 146 V2.4 | 29/01/2010 | | | 30196 Florida Centre Het Compounds | ARK Jul 09 | 29 |
| 148 | 147 V2.4 | 29/01/2010 | | | 146 Princeton | Focused Set Collection ADK Jan 2011 | |
| 149 | 148 V2.4 | 29/01/2010 | | | 2267 Princeton | Focused Set Collection CB Jan 2010 | 2 |
| 150 | 149 V2.4 | 29/01/2010 | | | 1770 Princeton | Focused Set Collection CB2 Jan 2 | 1 |
| 151 | 150 V2.4 | 29/01/2010 | | | 1483 Princeton | Focused Set Collection CDK Jan 2 | 1 |
| 152 | 151 V2.4 | 29/01/2010 | | | 1922 Princeton | Focused Set Collection ERB Jan 2 | 1 |
| 153 | 152 V2.4 | 29/01/2010 | | | 554 Princeton | Focused Set Collection GSK3BETA | 1 |
| 154 | 153 V2.4 | 29/01/2010 | | | 1696 Princeton | Focused Set Collection JNK Jan 2010 | 1 |
| 155 | 154 V2.4 | 29/01/2010 | | | 2775 Princeton | Focused Set Collection P38 Jan 2010 | 2 |
| 156 | 155 V2.4 | 30/01/2010 | | | 41782 Princeton | Building Blocks Collection Jan 2010 | 40 |

What can't we load yet?

More challenging data

- PDF catalogues
- Polymers
- Structures with query features (R groups)
- Compounds with no structure (Palm oil)
- Some organometallic compounds
- Mixtures

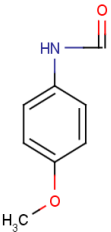
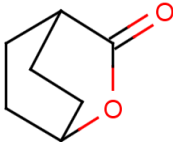
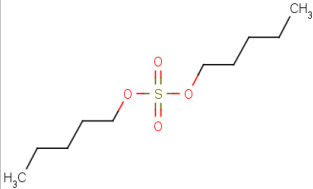
| # | structure | Loader_Error | |
|------|---|---|---|
| |  | | |
| 2420 |  | Mixture found for this compound | Bis(dicarbonylcyclopentadienyl)iron, 99%. |
| 2421 |  | The CAS number 19499-07-9 for O.O.O.O.O.O.[Fe+3].[Fe+3].[O-]C(=O)C([O-])=O.[O-]C(=O)C([O-])=O.[O-]C(=O)C([O-])=O already exists in the database for another molecule, so hasn't been added for this compound... | Iron (II) oxalate |
| 2422 | | The molecule was found to be EMPTY when checking | Carnauba wax |

What would help us?

Data standardization

Suppliers

- Supply catalogues as SD files
- Keep to a consistent layout between different catalogues
- Keep the same layout between issues of catalogues
- Follow guidelines for structure depiction
- Data in separate fields
- Consistent separation of items in fields
- Data checked for consistency

| # | structure | Unit | Price1_GBP | CAS Number | Purity | Product name | Unit2 | Price2_GBP | MP |
|---|---|------|------------|------------|--------|--|-------|------------|-----|
| 1 |  | 50 | 42.85 | 5470-34-0 | 99.0 | 4-Methoxybenzamide | | | |
| 2 |  | 56 | 81.55 | 4250-94-0 | 98.0 | 4-Hydroxy-1-cyclohexanecarboxylic Acid delta-Lactone | 250 | 300.00 | 128 |
| 3 |  | 50 | 95.70 | 5867-98-1 | | Diamyl Sulfate | | | |

Your contact:

Dr Bob Marmon
Senior Scientist

+44.(0).1235 441402

+44.(0).1235 863139

bob.marmon@evotec.com

