Battling the Data Avalanche

A Chemical Data Management Solution for the Smallcap Company

Antony J William
Kevin Turnbull

Advanced Chemistry Development

Spectroscopy • Chromatography • PhysChem • Naming
Drawing and Databasing • Enterprise Solutions
What do you really need?

▸ A heavy, central registration system?
  – Probably not yet:
    • Expensive
    • Requires sophisticated IT infrastructure
    • Hard to maintain
    • Long time to get started

▸ Isolated files of data?
  – It’s time to get beyond that.
    • Cannot share data
    • Spend too much time managing data sets, rather than thinking about the data
What do you really need?

- A friendly, sophisticated tool that
  - Helps you standardize your data
  - Represents all of your chemistry
  - Supports all important chemical query types
  - Provides access to your data throughout the organization
  - Gives high quality, standardized reports for printing and sharing
  - Integrates your chemical structure data to other systems
  - Solves real problems on the same day you open the box
What do you really need?

- **Data Interchange**
  - Exchange information through standards – Molfiles and ChemDraw files

- **Extensible**
  - If you have IT staff available let them extend it
  - Add-on components as necessary as the requirements are defined and the organization grows
Integrated Software Solutions

Proven Structure Representation

Spectroscopy • Chromatography • PhysChem • Naming
Drawing and Databasing • Enterprise Solutions
All of your chemistry

Spectroscopy • Chromatography • PhysChem • Naming
Drawing and Databasing • Enterprise Solutions
All of your chemistry

\[
\left[ \text{CH}_2\text{CH}==\text{CH}==\text{CH}_2 \right]_n \left[ \text{CH}_2\text{CH}==\text{CH}==\text{CH}_2 \right]_m \quad \left[ \text{CH}_2==\text{CH}==\text{CH}_2 \right]_n
\]

\[
\text{H}_3\text{C} - \text{C}_\text{H}_3 \xrightarrow{\text{Catalyst}} \text{H}_2\text{C}==\text{CH}==\text{CH}_2 + \text{H}_2\text{H}_2
\]

\[600 \, ^\circ \text{C}\]
What you might use it for

- **Chemical Inventories**
  - Create a shareable database of reagents

- **Planning your syntheses**
  - And store this know-how in a searchable database

- **Molecular Design studies**
  - Compute molecular properties

- **Structure Property Relationships**
  - Examine structures alongside result data

- **Reports**
  - Prepare standardized reports

- **Connect all this information**
  - Find the data you need, no matter where it is
Standardize your data

Spectroscopy • Chromatography • PhysChem • Naming
Drawing and Databasing • Enterprise Solutions
Plan your syntheses

To a well-stirred solution of 9.22 g (50 mmol) of cyanuric chloride in 100 mL of ether cooled to -20°C was added, over 45 min, a solution of 5.6 mL (82 mmol) of isopropylamine and 6.72 g (52 mmol) of N,N-diisopropylethylamine in 50 mL of ether. The mixture was filtered, and the filtrate was washed sequentially.
Design New Molecules
Structure-Property Data

<table>
<thead>
<tr>
<th>Structure</th>
<th>Sigma(norm)</th>
<th>Sigma(res)</th>
<th>Sigma(nrel)</th>
<th>Sigma(norm)</th>
<th>Sigma(res)</th>
<th>Sigma(nrel)</th>
<th>Sigma(norm)</th>
<th>Sigma(res)</th>
<th>Sigma(nrel)</th>
<th>Sigma(norm)</th>
<th>Sigma(res)</th>
<th>Sigma(nrel)</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>0.210</td>
<td>-0.010</td>
<td>0.000</td>
<td>0.210</td>
<td>-0.010</td>
<td>0.000</td>
<td>0.210</td>
<td>-0.010</td>
<td>0.000</td>
<td>0.210</td>
<td>-0.010</td>
<td>0.000</td>
</tr>
<tr>
<td>21</td>
<td>0.620</td>
<td>-0.030</td>
<td>0.110</td>
<td>0.620</td>
<td>-0.030</td>
<td>0.110</td>
<td>0.620</td>
<td>-0.030</td>
<td>0.110</td>
<td>0.620</td>
<td>-0.030</td>
<td>0.110</td>
</tr>
<tr>
<td>8</td>
<td>0.430</td>
<td>0.150</td>
<td>0.600</td>
<td>0.430</td>
<td>0.150</td>
<td>0.600</td>
<td>0.430</td>
<td>0.150</td>
<td>0.600</td>
<td>0.430</td>
<td>0.150</td>
<td>0.600</td>
</tr>
<tr>
<td>14</td>
<td>-0.170</td>
<td>-0.000</td>
<td>0.110</td>
<td>-0.170</td>
<td>-0.000</td>
<td>0.110</td>
<td>-0.170</td>
<td>-0.000</td>
<td>0.110</td>
<td>-0.170</td>
<td>-0.000</td>
<td>0.110</td>
</tr>
</tbody>
</table>

Spectroscopy • Chromatography • PhysChem • Naming
Drawing and Databasing • Enterprise Solutions
### Chemical Structure (ID=1)

| No. | 157 | 158 | 159 | 160 | 161 | 162 | 163 | 164 | 165 | 166 | 167 | 168 | 169 | 170 | 171 | 172 | 173 | 174 | 175 | 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 | 184 | 185 | 186 | 187 | 188 | 189 | 190 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |

### Spectrum Parameters (ID=1)

- **m/z**: 0.400, 0.860, 1.160, 79.000, 0.860, 67.000, 100.000, 0.460, 0.630, 1.330, 2.200, 0.540, 0.340, 68.000, 2.080, 0.310, 0.490, 0.710, 0.350, 2.560, 0.670, 56.000, 79.000, 0.840, 12.620, 0.840, 4.310, 87.000, 45.000, 0.500, 2.560, 139.000, 0.730, 91.000, 0.890, 1.330, 104.000, 0.690, 92.000, 3.500, 2.560, 12.620, 3.500, 8.180, 0.370, 5.180, 0.430, 126.000, 0.500, 0.540, 0.640, 0.740, 78.000, 0.670, 86.000, ...

### Spectrum (ID=1)

- **RI (%):** 55.000, DI
- **RI (%):** 122.000, DI
- **RI (%):** 0.910, 123.000
- **RI (%):** 0.990, 63.000
- **RI (%):** 0.640, 118.000
- **RI (%):** 0.860, 0.310
- **RI (%):** 99.000, 85.000
- **RI (%):** 0.410, 0.640
- **RI (%):** 1.250, 0.520
- **RI (%):** 118.000, 63.000
- **RI (%):** 0.840, 4.310
- **RI (%):** 87.000, 45.000
- **RI (%):** 0.500, 2.560
- **RI (%):** 139.000, 0.730
- **RI (%):** 91.000, 0.890
- **RI (%):** 1.330, 104.000
- **RI (%):** 0.690, 92.000
- **RI (%):** 3.500, 2.560
- **RI (%):** 12.620, 3.500
- **RI (%):** 8.180, 0.370
- **RI (%):** 5.180, 0.430
- **RI (%):** 126.000, 0.500
- **RI (%):** 0.540, 0.640
- **RI (%):** 0.740, 78.000
- **RI (%):** 0.670, 86.000
- **RI (%):** 0.840, 120.000
- **RI (%):** 97.000, 59.000
- **RI (%):** 0.800, 107.000
- **RI (%):** 1.700, 115.000
- **RI (%):** 0.690, 2.200
- **RI (%):** 1.670, 0.740
- **RI (%):** 1.270, 96.000
- **RI (%):** 1.060, 1.360
- **RI (%):** 95.000, 103.000
- **RI (%):** 0.500, 1.180
- **RI (%):** 121.000, 4.310
- **RI (%):** 1.760, 0.990
- **RI (%):** 114.000, 83.000
- **RI (%):** 1.360, 1.670
- **RI (%):** 0.640, 0.560
- **RI (%):** 1.700, 0.730
- **RI (%):** 0.910, 0.490
- **RI (%):** 132.000, 1.890
- **RI (%):** 1.190, 1.180
- **RI (%):** 121.000, 4.310
- **RI (%):** 1.250, 54.000
- **RI (%):** 0.400, 1.750
- **RI (%):** 4.220, 0.790
- **RI (%):** 61.000, 0.500
- **RI (%):** 69.000, 1.330
- **RI (%):** 0.630, 0.500
- **RI (%):** 1.600, 0.400
- **RI (%):** 0.550, 4.220
- **RI (%):** 0.890, 0.840
- **RI (%):** 0.690, 1.190
- **RI (%):** 0.710

### RI Table (ID=1)

| No. | 22 | 24 | 27 | 31 | 32 | 34 | 38 | 41 | 44 | 46 | 50 | 51 | 54 | 55 | 56 | 57 | 59 | 60 | 62 | 63 | 67 | 68 | 69 | 70 | 74 | 75 | 76 | 77 | 78 | 79 | 80 | 81 | 84 | 85 | 87 | 90 | 92 | 95 | 97 | 107 | 115 | 120 | 121 | 129 | 132 | 137 | 140 | 141 | 143 | 144 | 145 | 147 | 150 | 152 | 153 | 154 | 155 | 156 | 157 | 158 | 159 | 160 | 161 | 162 | 163 | 164 | 165 | 166 | 167 | 168 | 169 | 170 | 171 | 172 | 173 | 174 | 175 | 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 | 184 | 185 | 186 | 187 | 188 | 189 | 190 |
|-----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|

### Notes

- This document contains a mass spectrum with detailed parameters and a chemical structure diagram.
- The data includes m/z values, DI, and RI percentages for various components.
- The spectrum is well annotated with peak integrations and software solutions highlighted.
- The layout is typical of scientific data presentation, with structured tables and visual aids.
The lining of your stomach contains millions of special cells that produce acid via "acid pumps." It is the job of these pumps to produce the acid to help digest food. For people with acid reflux disease (GERD) this acid backs up into the esophagus where it doesn't belong. NEXIUM works by deactivating (turning off) some of the pumps. By reducing acid production in the stomach, NEXIUM reduces the chance of acid backing up into the esophagus and causing heartburn symptoms.
And reporting....

- Standard desktop products
And reporting....

- Web pages

Spectroscopy ● Chromatography ● PhysChem ● Naming
Drawing and Databasing ● Enterprise Solutions
And reporting.....

- PDF files

Spectroscopy • Chromatography • PhysChem • Naming
Drawing and Databasing • Enterprise Solutions
Beta-lapachone, a quinone, is derived from lapachol (a naphthoquinone) which can be isolated from the lapacho tree (Tabebuia avellanedae), a member of the catalpa family (Bignoniaceae). Like camptotheca and topotecan, 8-lapachone inhibits DNA topoisomerase I. Researchers have found that this compound has promising anticancer and antiviral properties.

Topoisomerase inhibitors, including beta-lapachone, seem to be effective against several types of cancer, including lung, breast, colon and prostate cancers and malignant melanoma. The use of beta-lapachone in humans has been limited due to its toxicity. However, 3-allyl-beta-lapachone (a close chemical relative) has been found to have lower toxicity in cell culture tests, and therefore may prove to be more useful than beta-lapachone.

Beta-lapachone works by disrupting DNA replication. Topoisomerase I is an enzyme that unwinds the DNA that makes up the chromosomes. The chromosomes must be unwound in order for the cell to use the genetic information to synthesize proteins; beta-lapachone keeps the chromosomes wound tight, and so the cell can’t make proteins. As a result, the cell stops growing. Because cancer cells grow and reproduce at a much faster rate than normal cells, they are more vulnerable to topoisomerase inhibition than normal cells. Beta-lapachone also interferes with the replication of HIV-1, a virus that causes AIDS, thereby slowing the advancement of the disease.

**Spectroscopy** • **Chromatography** • **PhysChem** • **Naming**

**Drawing and Databasing** • **Enterprise Solutions**
Searching Across the Organization

- Organizational details
  - Independence
  - Interdependence
- Structure, substructure and similarity structure searching
- Up to 16,000 user definable data fields
- Template construction
- Text searching
Structure search queries
Searching Across the Organization
Substructure Searching

Spectroscopy • Chromatography • PhysChem • Naming
Drawing and Databasing • Enterprise Solutions
Substructure Searching

Multiple Databases Search Result

<table>
<thead>
<tr>
<th>Database Path</th>
<th>Total Structures</th>
<th>Number of Hits in DB</th>
</tr>
</thead>
<tbody>
<tr>
<td>C:\6.0\E...FOLDER\marinlit.CFD</td>
<td>7097</td>
<td>4</td>
</tr>
<tr>
<td>C:\6...ER\Journal Formats.cfd</td>
<td>74</td>
<td>0</td>
</tr>
<tr>
<td>C:\6.0\E...FOLDER\reaction.cfd</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>C:\6.0\EX...FOLDER\WILEYdb.CFD</td>
<td>1000</td>
<td>1</td>
</tr>
<tr>
<td>C:\6.0...\LDER\rarelibrary.CFD</td>
<td>4026</td>
<td>86</td>
</tr>
<tr>
<td>C:\6.0\EX...FOLDER\ACDDemo.cfd</td>
<td>22056</td>
<td>0</td>
</tr>
<tr>
<td>C:\6.0\EX...FOLDER\acddict.cfd</td>
<td>9765</td>
<td>19</td>
</tr>
</tbody>
</table>

Query

Substructure

*Spectroscopy • Chromatography • PhysChem • Naming
Drawing and Databasing • Enterprise Solutions*
Accessing the ACD World

- Systematic nomenclature
  - CAS Index
  - IUPAC
- PhysChem prediction
  - LogP/pKa/LogD
- NMR prediction
  - H1,C13, N15, F19, P31
  - 2D NMR

- Chromatography prediction
- Analytical Data Management
  - Processing – NMR, MS, UVIR, Curve, Chrom.
  - Databasing
Add-on Capability

- 3D Viewer
- Aldrich Library (Pro)
- CNMR DB
- CNMR Predictor
- ChemFolder
- ChemSketch (Loaded)
- Combi NMR
- GC Simulator
- Gene Manager
- Gene Manager Pro
- HNMR
- HNMR DB
- LC Simulator
- LogD
- LogP
- Name
- SolDB
- Spec Viewer
- SpecManager
- Spectrum Database
- Structure Elucidator
- XNMR
- pKa
- pKa
- pKa
- pKa
- pKa
Add-on Capability

- Spectroscopy
- Chromatography
- PhysChem
- Naming

Drawing and Databasing

Enterprise Solutions
ACD Integration

Spectroscopy • Chromatography • PhysChem • Naming
Drawing and Databasing • Enterprise Solutions
All experimental spectra may now be assigned, and updated to a searchable structure based archive.

**Spectroscopy** • **Chromatography** • **PhysChem** • **Naming**

**Drawing and Databasing** • **Enterprise Solutions**
Structure Searchable Images

- Include images AND structures in one record. Images are structure and substructure searchable
  - Microscopy images
  - Pharmaceuticals – “the purple pill”
  - Sea slugs
Flexibility of Database

Spectroscopy • Chromatography • PhysChem • Naming
Drawing and Databasing • Enterprise Solutions
Conclusion

- No need to make dramatic investments in IT early on
- A heavy registration system is NOT necessary to solve your chemical data management needs
- ChemFolder databases can be used to
  - Manage Chemical Inventory
  - Plan your syntheses
  - Design new molecules
  - Examine Structure Property Relationships
  - Generate Reports
- These can be queried from within, and across multiple databases to find what you need.
ACD/ChemSketch

- Available as Freeware from www.acdlabs.com
- >188,000 copies downloaded to date
- The structure drawing interface for ChemFolder
Free ACD/ChemSketch with IUPAC Naming
<table>
<thead>
<tr>
<th>Country</th>
<th>Downloads</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afghanistan</td>
<td>964</td>
</tr>
<tr>
<td>Albania</td>
<td>582</td>
</tr>
<tr>
<td>Algeria</td>
<td>941</td>
</tr>
<tr>
<td>American Samoa</td>
<td>333</td>
</tr>
<tr>
<td>Andorra</td>
<td>531</td>
</tr>
<tr>
<td>Angola</td>
<td>436</td>
</tr>
<tr>
<td>Anguilla</td>
<td>316</td>
</tr>
<tr>
<td>Antarctica</td>
<td>464</td>
</tr>
<tr>
<td>Antigua and Barbuda</td>
<td>89</td>
</tr>
<tr>
<td>Argentina</td>
<td>2749</td>
</tr>
<tr>
<td>Armenia</td>
<td>148</td>
</tr>
<tr>
<td>Aruba</td>
<td>147</td>
</tr>
<tr>
<td>Australia</td>
<td>5675</td>
</tr>
<tr>
<td>Austria</td>
<td>2524</td>
</tr>
<tr>
<td>Azerbaijian</td>
<td>191</td>
</tr>
<tr>
<td>Uruguay</td>
<td>300</td>
</tr>
<tr>
<td>Uzbekistan</td>
<td>15</td>
</tr>
<tr>
<td>Vanuatu</td>
<td>3</td>
</tr>
<tr>
<td>Vatican City State</td>
<td>4</td>
</tr>
<tr>
<td>Venezuela</td>
<td>1338</td>
</tr>
<tr>
<td>Vietnam</td>
<td>331</td>
</tr>
<tr>
<td>Virgin Isl. (British)</td>
<td>2</td>
</tr>
<tr>
<td>Virgin Isl. (U.S.)</td>
<td>14</td>
</tr>
<tr>
<td>Wallis and Futuna Is.</td>
<td>24</td>
</tr>
<tr>
<td>Western Sahara</td>
<td>23</td>
</tr>
<tr>
<td>Yemen</td>
<td>28</td>
</tr>
<tr>
<td>Yugoslavia</td>
<td>660</td>
</tr>
<tr>
<td>Zaire</td>
<td>20</td>
</tr>
<tr>
<td>Zambia</td>
<td>12</td>
</tr>
<tr>
<td>Zimbabwe</td>
<td>90</td>
</tr>
</tbody>
</table>
180,000 Downloads
200/Day Worldwide A-Z

<table>
<thead>
<tr>
<th>Country</th>
<th>Downloads</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afghanistan</td>
<td>964</td>
</tr>
<tr>
<td>Albania</td>
<td>582</td>
</tr>
<tr>
<td>Algeria</td>
<td>941</td>
</tr>
<tr>
<td>American Samoa</td>
<td>333</td>
</tr>
<tr>
<td>Andorra</td>
<td>531</td>
</tr>
<tr>
<td>Angola</td>
<td>436</td>
</tr>
<tr>
<td>Anguilla</td>
<td>316</td>
</tr>
<tr>
<td>Antarctica</td>
<td>464</td>
</tr>
<tr>
<td>Antigua and Barbuda</td>
<td>89</td>
</tr>
<tr>
<td>Argentina</td>
<td>2749</td>
</tr>
<tr>
<td>Armenia</td>
<td>148</td>
</tr>
<tr>
<td>Aruba</td>
<td>147</td>
</tr>
<tr>
<td>Australia</td>
<td>5675</td>
</tr>
<tr>
<td>Austria</td>
<td>2524</td>
</tr>
<tr>
<td>Azerbaiján</td>
<td>191</td>
</tr>
<tr>
<td>Uruguay</td>
<td>300</td>
</tr>
<tr>
<td>Uzbekistan</td>
<td>15</td>
</tr>
<tr>
<td>Vanuatu</td>
<td>3</td>
</tr>
<tr>
<td>Vatican City State</td>
<td>4</td>
</tr>
<tr>
<td>Venezuela</td>
<td>1338</td>
</tr>
<tr>
<td>Vietnam</td>
<td>331</td>
</tr>
<tr>
<td>Virgin Isl. (British)</td>
<td>2</td>
</tr>
<tr>
<td>Virgin Isl. (U.S.)</td>
<td>14</td>
</tr>
<tr>
<td>Wallis and Futuna Is.</td>
<td>24</td>
</tr>
<tr>
<td>Western Sahara</td>
<td>23</td>
</tr>
<tr>
<td>Yemen</td>
<td>28</td>
</tr>
<tr>
<td>Yugoslavia</td>
<td>660</td>
</tr>
<tr>
<td>Zaire</td>
<td>20</td>
</tr>
<tr>
<td>Zambia</td>
<td>12</td>
</tr>
<tr>
<td>Zimbabwe</td>
<td>90</td>
</tr>
</tbody>
</table>

**Spectroscopy • Chromatography • PhysChem • Naming**
**Drawing and Databasing • Enterprise Solutions**
Extending Capability

- If there is IT support – ChemBasic
- A full-featured programming language
- ACD users create applications: “The Goodies” – made available from the ACD website
ChemBasic Examples
IChI in ACD/ChemSketch

Spectroscopy • Chromatography • PhysChem • Naming
Drawing and Databasing • Enterprise Solutions
Source website offers sdf or cfd file for download

GAP-ACD ChemFolder Database
Giveaway ChemFolder
Databases

Spectroscopy • Chromatography • PhysChem • Naming
Drawing and Databasing • Enterprise Solutions