Chemical Markup Language

Introduction

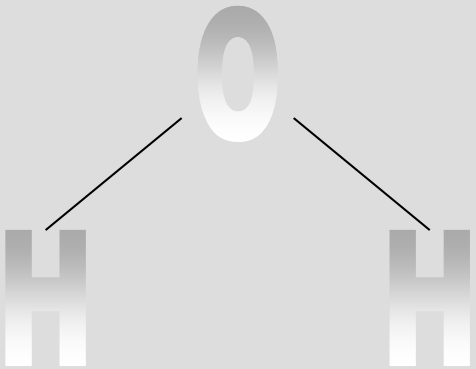
- **Markup for "managing molecular information"**
  - capable of holding extremely complex information structures and so acting as an interchange mechanism or for archival
  - used to manage documents and information in:
    - *macromolecular sequence and structure*
    - *spectra*
    - *organic molecules*
    - *quantum chemistry*
    - *inorganic crystallography*
  - can hold 2D and 3D molecular models
  - more properly described as an SGML-based system than an XML-based one:
    - "CML has used an XML-like philosophy for many months and so there is no problem in making it follow the emerging 'standard.'"

More Introduction

- **CML claims major advantages in the following areas:**
  - input and output of database entries
  - authoring, editing, merging of chemical documents
  - managing legacy data, including future-proofing
  - supporting structured documents
  - direct capture of data from instruments
  - robotic control of experiments
  - analysis of complex program output (e.g., theoretical chemistry)
  - publication by individuals and small groups
  - adding meta-data to chemical documents
  - multidisciplinary areas
  - capture of chemical knowledge in organisations
  - support for terminologies

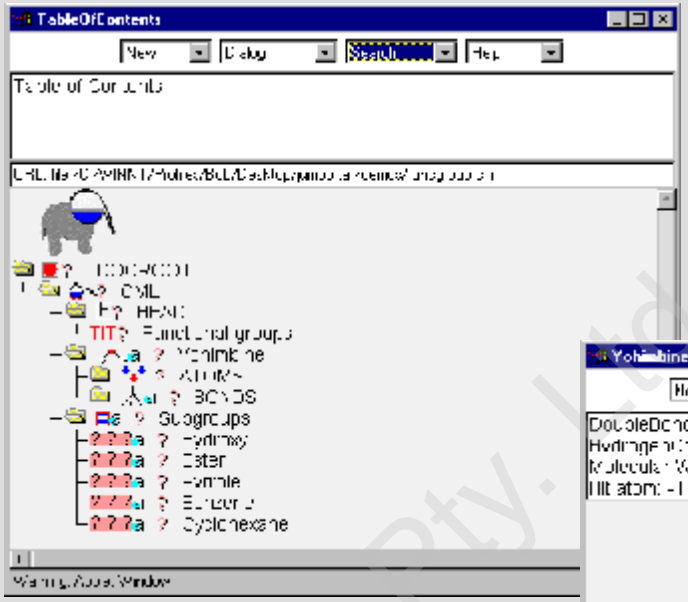
CML Example: H_2O

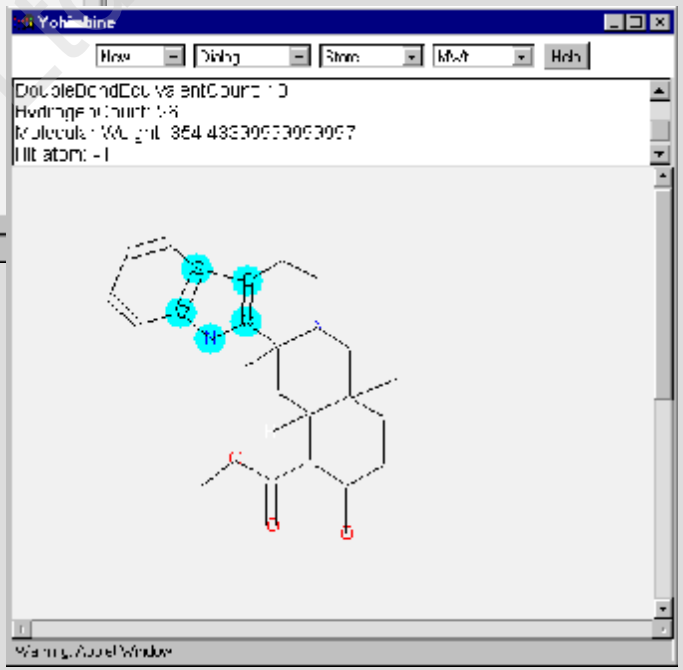
```
<?xml version="1.0 "?>
<CML>
  <MOL TITLE="Water ">
    <ATOMS>
      <ARRAY BUILTIN="ELSYM ">H O H</ARRAY>
    </ATOMS>
    <BONDS>
      <ARRAY BUILTIN="ATID1 ">1 2</ARRAY>
      <ARRAY BUILTIN="ATID2 ">2 3</ARRAY>
      <ARRAY BUILTIN="ORDER ">1 1</ARRAY>
    </BONDS>
  </MOL>
</CML>
```

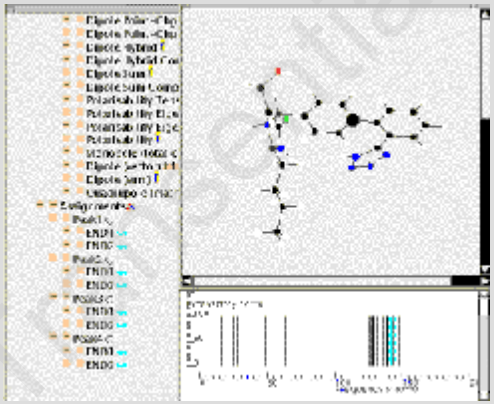


Pictures

- **From the JUMBO system**
  - Java Universal Markup Browser for Objects
    - *note how everything is live and linked…*

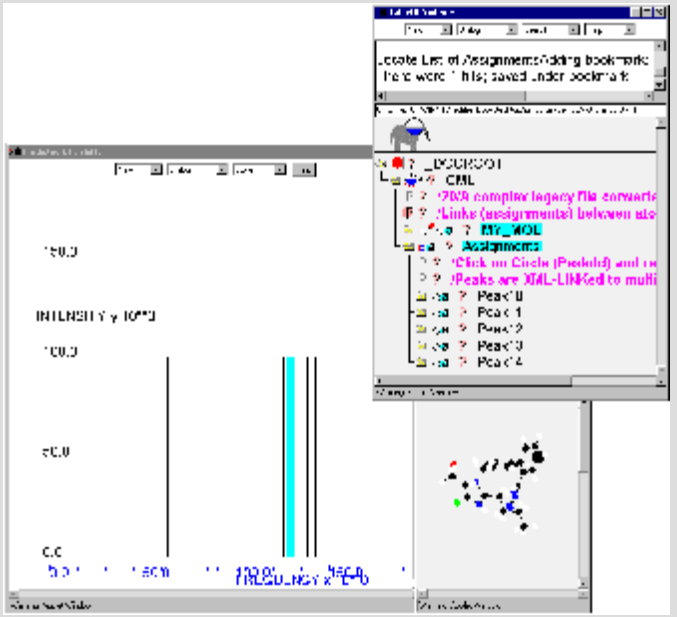






Complex Markup

```
<?JUMBO FONTSIZE="20" MESSAGE="A complex legacy file converted into CML with XML-LINKs" ?>
<?JUMBO MESSAGE="Links (assignments) between atoms and spectral peaks" TEISEARCH="DESCENDANT (1 MOL)"
SELECTED="CLOSE" MESSAGE="Locate List of Assignments" TEISEARCH="D (1 XLIST TITLE Assignments)" SELECTED="OPEN" ?>
<MOL NAME="MY_MOL">
<ATOMS NUMBER="55" >
<ARRAY BUILTIN="ELSYM">
 C C C C C C C C C C C C C C N N N N N C C C C N C O C C C H H H H H H H H H H H H Cl H H H H H H H H H H H H
</ARRAY>
<ARRAY BUILTIN="X3">
 -0.7891 -2.1776 -0.2634 0.1111 -3.0270 -1.1151 0.4139 0.6756 -2.5035 1.2622 -0.1323 1.5201 -3.4151 1.8112 -0.3883 -
0.4933 -4.0868 -0.8647 -0.9403 -5.2184 -3.2995 -4.9669 -6.3961 -3.8560 -1.8448 -6.2828 -1.4367 0.0072 0.4254 -2.5993
0.8264 -4.1209 -0.6919 0.4407 1.4922 1.9539 -2.8350 -4.2378 2.4737 -0.4824 -4.4518 -5.9303 -7.3075 -6.5487 -3.2931 -
1.6118 -1.2220 -6.2717 -1.5725 -2.1061 0.1429 0.6709 0.3232 1.4904 -0.2106
</ARRAY>
<ARRAY BUILTIN="Y3">
 -0.5479 -0.5678 -0.3784 -0.7133 -0.3896 -0.1939 0.3494 -1.9736 -0.1816 0.1313 1.6862 -2.1812 -0.0012 -1.1291 2.6523
2.2829 1.3412 3.7078 3.5184 1.5395 2.5945 2.6863 0.6538 3.3536 2.6432 -0.4199 4.0403 4.0689 5.4547 -0.7241 -0.3963 -
0.4127 -0.0511 -2.8052 0.9598 -3.1775 -0.1655 -0.7731 -1.2918 1.8837 1.3803 3.2613 1.2686 0.1428 3.9493 1.9190 2.3337
-0.0381 4.7562 4.3845 3.3598 3.7110 6.1710 5.4549 5.8198
</ARRAY>
<ARRAY BUILTIN="Z3">
 -0.1908 -0.0117 -1.4797 0.9552 -1.1020 -2.5652 1.8274 1.1913 -2.3819 2.9243 1.6404 2.2795 -3.5490 3.1470 2.5858
0.4146 -3.6682 1.9373 0.6680 -2.7396 -3.5178 -1.9774 -2.7219 -2.4462 -3.8055 -1.8018 -4.2350 -4.6913 -5.1099 0.9931 -
1.6327 -0.9630 -3.5719 0.5094 3.6140 2.4540 -4.5012 -3.4819 4.0104 -0.4995 -4.6236 -0.7214 -2.4810 -3.7137 -1.8853 -
4.6327 -2.9174 -0.9162 -3.3825 -5.0661 -5.5503 -3.8602
-4.2590 -5.4455 -5.9524
</ARRAY>
<ARRAY BUILTIN="C13">
 142.57 127.09 126.19 138.62 128.43 129.08 128.01 125.57 136.48
 127.00 182.64 130.74 63.22 129.08 "" "" "" "" "" 176.01 147.73
 139.41 49.04 "" 36.04 "" 27.71 25.01 15.82 "" "" "" "" "" "" ""
 "" "" "" "" "" "" "" "" "" "" "" "" "" "" "" "" "" "" ""
</ARRAY>
</ATOMS>
```



Complex Markup (Cont'd)

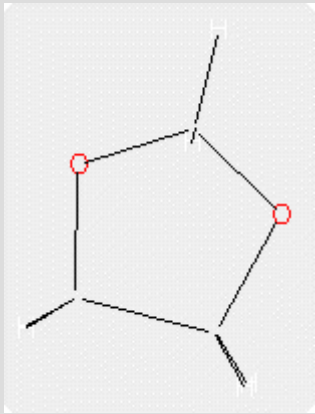
- **A CMLfile can contain many kinds of data**
  - spectra, lab. notes, commands for the control of lab. equipment, etc.

```
<XLIST CONTENT="GRAPH" TITLE="Predicted C13 shifts" NAME="SPECTRUM" DISPLAY="BAR">
<ARRAY DICTNAME="C13SHIFT" SIZE="55" TITLE="FREQUENCY" UNITS="MHZ" TYPE="FLOAT" NAME="C13">
  142.57 127.09 126.19 138.62 128.43 129.08 128.01 125.57 136.48 127.00
  182.64 130.74 63.22  129.08 0.0     0.0   0.0    0.0    0.0    176.01
  147.73 139.41 49.04  0.0    36.04   0.0   27.71  25.01  15.82  0.0
  0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
  0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
  0.0    0.0    0.0    0.0    0.0
</ARRAY>
<ARRAY TITLE="INTENSITY" SIZE="55" TYPE="FLOAT">
 100.01 100.02 100.03 100.04 100.05 100.06 100.07 100.08 100.09 100.10
 100.11 100.12 100.13 100.14  00.15  00.16  00.17  00.18  00.19 100.20
 100.21 100.22 100.23  00.24 100.25  00.26 100.27 100.28 100.29  00.30
  00.31  00.32  00.33  00.34  00.35  00.36  00.37  00.38  00.39  00.40
  00.41  00.42  00.43  00.44  00.45  00.46  00.47  00.48  00.49  00.50
  00.51  00.52  00.53  00.54  00.55
</ARRAY>
</XLIST>
</MOL>
<XLIST TITLE="Assignments">
<?JUMBO MESSAGE="Click on Circle (Peakdd) and resize Molecule and Spectrum windows" ?>
<?JUMBO MESSAGE="Peaks are XML-LINKed to multiple targets.  Click on any circle to show this" ?>
<RELATION TITLE="Peak10"><XVAR BUILTIN="END1" TYPE="ADDRESS">MY_MOL:10</XVAR><XVAR BUILTIN="END2"
TYPE="ADDRESS">SPECTRUM:10</XVAR></RELATION>
<RELATION TITLE="Peak11"><XVAR BUILTIN="END1" TYPE="ADDRESS">MY_MOL:11</XVAR><XVAR BUILTIN="END2"
TYPE="ADDRESS">SPECTRUM:11</XVAR></RELATION>
<RELATION TITLE="Peak12"><XVAR BUILTIN="END1" TYPE="ADDRESS">MY_MOL:12</XVAR><XVAR BUILTIN="END2"
TYPE="ADDRESS">SPECTRUM:12</XVAR></RELATION>
<RELATION TITLE="Peak13"><XVAR BUILTIN="END1" TYPE="ADDRESS">MY_MOL:13</XVAR><XVAR BUILTIN="END2"
TYPE="ADDRESS">SPECTRUM:13</XVAR></RELATION>
<RELATION TITLE="Peak14"><XVAR BUILTIN="END1" TYPE="ADDRESS">MY_MOL:14</XVAR><XVAR BUILTIN="END2"
TYPE="ADDRESS">SPECTRUM:14</XVAR></RELATION>
</XLIST>
</CML>
```

Simple Markup

- **This example shows**
  - display markup
    - *atoms and their positions*
  - structure markup
    - *bonds and bond properties*

```
<?xml:namespace href="jumbo/cml/schema.xml" as="CML"?>
<CML:MOL CONVENTION="MDL-MOLFILE">
<CML:ATOMS>
<CML:ARRAY BUILTIN="X2">-1.342 0.137 0.8 -0.1029 -1.3289 -1.882 -1.8799 0.438 0.464 -0.1409 0.163</CML:ARRAY>
<CML:ARRAY BUILTIN="Y2">-0.527 -0.905 0.335 1.2439 0.879 -0.8419 -0.878 -1.471 -1.441 1.1049 2.28</CML:ARRAY>
<CML:ARRAY BUILTIN="ELSYM">C C O C O H H H H H H</CML:ARRAY>
<CML:ARRAY BUILTIN="FORMCHARGE">0 0 0 0 0 0 0 0 0 0 0</CML:ARRAY>
</CML:ATOMS>
<CML:BONDS>
<CML:ARRAY BUILTIN="ATID1">1 1 1 1 2 2 2 3 4 4 4</CML:ARRAY>
<CML:ARRAY BUILTIN="ATID2">2 5 6 7 3 8 9 4 5 10 11</CML:ARRAY>
<CML:ARRAY BUILTIN="ORDER">1 1 1 1 1 1 1 1 1 1 1</CML:ARRAY>
<CML:ARRAY BUILTIN="STER">0 0 0 0 0 0 0 0 0 0 0</CML:ARRAY>
</CML:BONDS>
</CML:MOL>
```



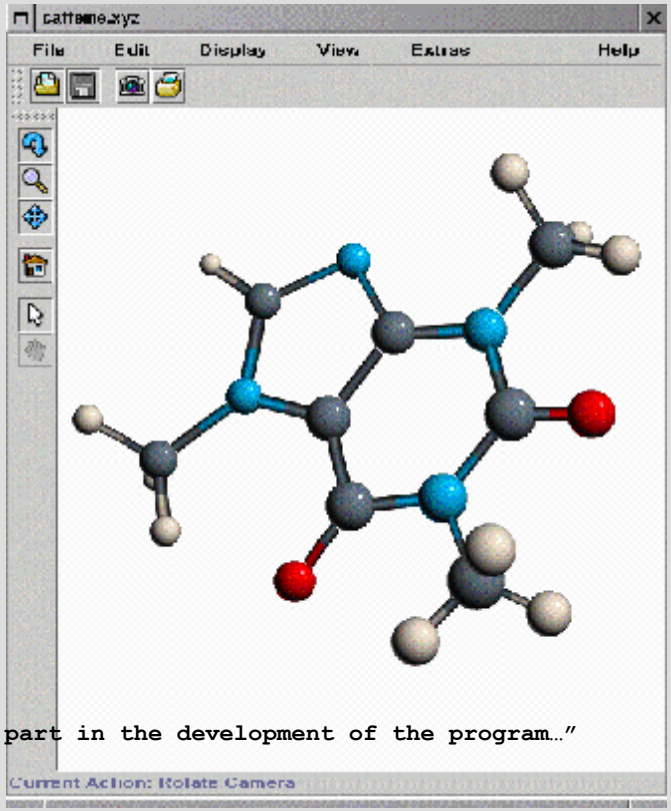
- **A Free, Open Source molecule viewer and editor**
  - understands CML
  - collaboratively developed visualization and measurement tool for chemical scientists
    - *new features being added on a daily basis*

Jmol

More Jmol

– Jmol has a number of extremely useful features for chemical dynamics:

- *can be used to animate the results of simulations that are in the multi-frame XYZ format*
- *can be used to measure inter-atomic distances, bond angles, and dihedral angles from atomic coordinates as a simulation progresses*
- *can animate the computed normal modes*
- *can display vectors (velocity, dipole, etc.), charges, atomic symbols, or atomic indexes during animation*
- *can export high quality images*



"Caffeine, a molecule which has played an integral part in the development of the program…"