Fast and Accurate Enumeration of Combinatorial Libraries

Keith A Harrington, Julian Hayward
& Roger Upton
Pre-Accelrys CombiChem History

• Synopsys
  – Accord Chemistry Software
    • Components, Toolkits, Applications
      – Enumeration capabilities
    – Reaction Databases
      • Selective, Thematic
        – Solid-Phase Synthesis Database

• MSI
  – Enumeration within MedChem explorer as critical part of library design
CombiChem

- The chemists answer to the challenges posed by the HTS revolution
- The purpose being to increase productivity by several orders of magnitude
- Chemists now 100 times more productive
- Millions of new compounds are created per year

- Has software kept pace with requirements?
Scope of Talk

• ‘CombiChem’ encompasses:
  – Molecular Diversity
  – Library Design* and Library Analysis
  – Library Synthesis and Robotics
  – Library Registration*
  – Quality Control
  – High Throughput Screening

• Focus today on ‘Enumeration’ of discrete structures and libraries
Fast and Accurate?

- Accuracy Vital
  - Over enumeration e.g. symmetry
  - Structures simply wrong e.g. stereochemistry
  - Presentation issues e.g. bond overlaps

- Fast
  - Enumeration of small, medium and large libraries
    - What is fast?
  - Use the same enumeration engine in either case
    - Advantages all around
Software Requirements

• Difficult to imagine CombiChem without software
  – Would it be possible?
• Software needs to grow to meet the needs
• Chemical representation to the fore
  – Get the chemical structure correct first time, every time
• Integration standards critical
  – Oracle, Microsoft, third-party software
• Simple to use applications for the end-user chemist
• Component code base highly desirable for more complex, bespoke applications
Enumeration: Principle

- Compound generation, based on user input and common features
End-user Interaction

• How does the end-user interact with the system?
  – Applications must be intuitive
    • Wizard approach rather than rules?
  – Must utilize existing data
    • Reading and writing of standard file formats

• To what extent does the software need to be ‘intelligent’
  – Does it need to ‘understand’ chemistry?
  – or simply ‘do as it is told’?
### Reaction-based Enumeration

<table>
<thead>
<tr>
<th>R1\text{COOH}</th>
<th>+</th>
<th>R2\text{NH}_2</th>
<th>→</th>
<th>R1\text{CONH}R2</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{R1=OH})</td>
<td></td>
<td>(\text{R2=NH}_2)</td>
<td></td>
<td>(\text{R1\text{CONH}R2})</td>
</tr>
<tr>
<td>(\text{R1=OH})</td>
<td>(\text{R2=NH}_2)</td>
<td>(\text{R1=OH})</td>
<td>(\text{R2=NH}_2)</td>
<td></td>
</tr>
<tr>
<td>(\text{R1=OH})</td>
<td>(\text{R2=NH}_2)</td>
<td>(\text{R1=OH})</td>
<td>(\text{R2=NH}_2)</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>8 amines</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>(12 acids)</td>
<td>(8 amines)</td>
<td>(96 amides)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Reaction-based Enumeration

• Advantages
  – Obviates the need to pre-process (clip) the reactants
    • more intuitive to the chemist
  – Flexibility to enumerate the individual reactions as well as molecules
    if desired
  – Complete ‘experiment’ may be subsequently stored generically within
    a database

• Designed into all Accord solutions
More Complex Issues (1)

- “Intramolecular” examples

\[
R_1\text{-NH}_2 + R_2\text{-CO}_2\text{H} \rightarrow R_2\text{-NH}_2\text{-CO}_2R_1
\]

R1-NH2 =

- \[
\text{OH} \quad \text{NH}_2
\]
- \[
\text{NH}_2\text{-C}_6\text{H}_4\text{-NH}_2
\]
- \[
\text{HO-CH}_2\text{-CH}_2\text{NH}_2
\]
- \[
\text{CH}_2\text{-NH}_2
\]
- \[
\text{HO-CH}_2\text{-CH}_2\text{-NH}_2
\]
- \[
\text{C}_6\text{H}_4\text{-CO}_2\text{H}
\]
- \[
\text{H}_2\text{N}-\text{C}_6\text{H}_4\text{-NH}_2
\]
More Complex Issues (2)

- Multifunctional substrates

\[
R_1\text{-NH}_2 \quad + \quad R_2\text{COOH} \quad \rightarrow \quad R_2\text{NH}_2\text{R}_1
\]
More Complex Issues (3)

- Competing functionality

\[ R_1\text{-NH}_2 + R_2\text{-CO}_2\text{H} \rightarrow R_2\text{-CONHR}_1 \]

R1-NH2 =
‘Diverse’ Reactions

\[ \text{R}_1\text{SH} + \text{R}_2\text{Cl} \rightarrow \text{R}_2\text{S}\text{R}_1 \]

\[ \text{R}_1\text{NH}_2 + \text{R}_2\text{SOCl} \rightarrow \text{R}_2\text{SO}_2\text{N}_2\text{R}_1 \]

\[ \text{R}_1\text{NH}_2 + \text{R}_2\text{N}=\text{O} \rightarrow \text{R}_2\text{NO}_2\text{R}_1 \]
Chemical Integrity

- Enumeration must be chemically reliable:
  - No ‘over-enumeration’
    - no ‘impossible’ solutions
    - No multiple, identical solutions for symmetrical molecules
  - Stereochemistry:
    - Absolute & Relative configurations dealt with correctly
    - Diastereomers
  - Double-bond Geometry
    - Olefins, Allenes etc.
- The software must adequately deal with these issues
‘Impossible’ Products

R1 + R2 → R1R2

Br
Cl
F

Cl
Br
F

Br
Cl
F

X
Symmetry Perception (1)

\[
\begin{align*}
\text{R1} & \quad \text{R1} \\
\text{O} & \quad \text{OH} \\
\text{O} & \quad \text{O} \\
\text{O} & \quad \text{OH} \\
\text{O} & \quad \text{OH} \\
\text{OH} & \quad \text{OH} \\
\text{O} & \quad \text{OH} \\
\text{OH} & \quad \text{OH} \\
\end{align*}
\]
Symmetry Perception (2)

18-fold 'symmetry'
Stereochemistry

- Absolute & Relative configurations supported at the atom & bond level:
Control of Inversions
Control of Diastereomers

\[ \text{R1} \text{CO}_2\text{H} + \text{H}_2\text{N} - \text{R2} \rightarrow \text{R1} - \text{CONH}_2 \text{R2} \]
Presentation

• Look & Layout of Library
  – Suitable for visual inspection/comparison?
  – Adherence to corporate registration standards?

• ‘Redraw’ programs
  – Often fail for bridged systems etc.
  – Will fail to maintain a ‘standard’ orientation
    • Could this be template-based?
Swivel & Rotate Features

R1\text{Cl}\quad + \quad N^{R2} \quad \rightarrow \quad R1\text{Cl}_2^{N^{R2}}
Chemical Intelligence

- e.g. Presentation
  - “But steroids are always this way up!”
Scalability

• How do we differentiate between ‘project level’ and ‘large’ libraries
  – Desktop software?
    • Single PC
    • Up to 10,000 compounds?
  – Server software?
    • Parallel processing, Multithreading…
    • Millions of compounds?

• The ‘chemistry engine’ should be based upon the same components
Scalability and Flexibility

- Multi platform, component based
  - Windows, Sun/Solaris, SGI/Irix, NT, Linux
Example Application

- Accord for Excel Combichem add-on
  - Desktop enumeration tool
  - Project-level libraries
  - Utilizes workbook concept to adopt a workflow approach
  - Enumerates compounds or reactions directly within Excel
  - CLogP, MW calculation environment
<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>Generic Reaction Specification.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td><strong>Help</strong></td>
<td><strong>Options...</strong></td>
<td><strong>&lt; Back</strong></td>
<td><strong>Next &gt;</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[Reaction Builder...]

[Set Byproducts...]
**Generic Reaction Specification.**

```
\[ R_1 COOH + H_2N^+ R_2 \rightarrow R_2 R_1 CO \]
```

**CombiChem Options**

- **Reaction Options**
  - [ ] Allow Incomplete Enumeration
  - [ ] Show Intermediate Build Fragments

- **Reactant Options**
  - Output Results as Molecules
  - Output Results as Reactions

- **Enumeration Options**
  - Enumerated results naming
    - Prefix
    - Separator
    - Suffix

**Options...**

**Help**
<table>
<thead>
<tr>
<th>INDEX</th>
<th>CHEMISTRY</th>
<th>CLIPPED</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>G</td>
<td></td>
</tr>
</tbody>
</table>

**Reactant A**

![Chemical Structure]

- Enter Reactants...
- Add Match Column...
- Apply Filter
- Add Data Column...
- Clip Reactants
- Clear Sheet
<table>
<thead>
<tr>
<th>INDEX</th>
<th>CHEMISTRY</th>
<th>CLIPPED</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>5</td>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>7</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>8</td>
<td>H</td>
<td>H(1)</td>
</tr>
<tr>
<td>9</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>10</td>
<td>J</td>
<td>J</td>
</tr>
</tbody>
</table>
Select Clipped Reactants

The reactant below contains more than one occurrence of the functional group to be clipped. Clipped reagent combinations can be selected below.

Original Reactant

Clipped Reactants Selected
Name: H
Reactants Selected: 3
Clipped Reactant Options

Clipped Reactant 1 of 3

Name: H
Index 8

Options:
- Previous
- Next
- Remove Choice from Worksheet
- Add Choice to Worksheet

Help
Close
Select Clipped Reactants

The reactant below contains more than one occurrence of the functional group to be clipped. Clipped reactant combinations can be selected below.

Original Reactant

Clipped Reactants Selected

Clipped Reactant Options

Name: S
Reactants Selected: 2
Clipped Reactant 1 of 2

Index 8
<table>
<thead>
<tr>
<th>CHEMISTRY</th>
<th>INDEX A</th>
<th>INDEX B</th>
</tr>
</thead>
<tbody>
<tr>
<td>AZ</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>AY</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>AX</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>AW</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>AV</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>AU</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Enumeration</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[\text{Chemistry}]</td>
<td>[\text{Composition}] (Cell)</td>
</tr>
<tr>
<td>Computes the elemental composition of a chemistry object.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHEMISTRY</th>
<th>INDEX A</th>
</tr>
</thead>
<tbody>
<tr>
<td>AZ</td>
<td>1</td>
</tr>
<tr>
<td>AY</td>
<td>1</td>
</tr>
<tr>
<td>AX</td>
<td>1</td>
</tr>
<tr>
<td>AW</td>
<td>1</td>
</tr>
<tr>
<td>AV</td>
<td>1</td>
</tr>
<tr>
<td>AU</td>
<td>1</td>
</tr>
<tr>
<td>AT</td>
<td>1</td>
</tr>
<tr>
<td>AS</td>
<td>1</td>
</tr>
<tr>
<td>BZ</td>
<td>2</td>
</tr>
<tr>
<td>BY</td>
<td>2</td>
</tr>
<tr>
<td>BX</td>
<td>2</td>
</tr>
<tr>
<td>RW</td>
<td>2</td>
</tr>
<tr>
<td>CHEMISTRY</td>
<td>INDEX A</td>
</tr>
<tr>
<td>-----------</td>
<td>---------</td>
</tr>
<tr>
<td>AZ</td>
<td>1</td>
</tr>
<tr>
<td>AY</td>
<td>1</td>
</tr>
<tr>
<td>AX</td>
<td>1</td>
</tr>
<tr>
<td>AW</td>
<td>1</td>
</tr>
<tr>
<td>AV</td>
<td>1</td>
</tr>
<tr>
<td>AU</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Enumeration</td>
<td><img src="" alt="Diagram" /></td>
</tr>
<tr>
<td><img src="" alt="Diagram" /></td>
<td><img src="" alt="Diagram" /></td>
</tr>
<tr>
<td><img src="" alt="Diagram" /></td>
<td><img src="" alt="Diagram" /></td>
</tr>
<tr>
<td><img src="" alt="Diagram" /></td>
<td><img src="" alt="Diagram" /></td>
</tr>
<tr>
<td><img src="" alt="Diagram" /></td>
<td><img src="" alt="Diagram" /></td>
</tr>
<tr>
<td><img src="" alt="Diagram" /></td>
<td><img src="" alt="Diagram" /></td>
</tr>
<tr>
<td>CHEMISTRY</td>
<td><img src="" alt="Diagram" /></td>
</tr>
<tr>
<td>WJ</td>
<td><img src="" alt="Diagram" /></td>
</tr>
</tbody>
</table>
Accelrys Accord Solutions

- Manual pre-clipping step can be avoided
- Simple wizard type interface in Excel
- No over-enumeration problems
  - no duplicate or ‘impossible’ structures
- No random inversions of stereochemistry or double-bond geometry
- Correctly enumerates all valid diastereomers
- Presentation greatly enhanced by ‘swivel & rotate’ functions
Accelrys Accord Solutions

- Able to integrate with other ‘best of breed’ applications
  - Inherent in Accord’s component architecture
- Do not place unreasonable constraints on the chemist
  - User interface, accuracy & speed
- Are flexible enough to adapt to rapidly changing requirements
- Are scalable from client (small/medium libraries) to server (large libraries)
Acknowledgements

- Osman Guner and ACS Committee
- The audience today
- Tony Cook
- Dan Thomas
- Jim Clarke
- For demonstrations and further discussions
  - Booths 829 & 837
  - Keith@accelrys.com
  - www.accelrys.com
Science, faster

Keith@accelrys.com