Bridging The Gap Between Chemistry And Chemical Engineering

A critical bottleneck in the computational science grand challenge problem of bridging the gap between space and time scales, from individual reactions (length scale of nanometers and time scale of femtoseconds) to reacting flow (length scale of meters and time scale of minutes to hours).
A major problem in constructing a chemical kinetic model is the very large number of possible reactions, products, and reactions intermediates involved.

**Examples:**
Combustion of n-heptane(1) : 3662 reactions
Involving 470 species
Pyrolysis of tetradecane(2) : 479206 reactions and 19052 species

Bridging The Gap Between Chemistry And Chemical Engineering

NetGen (L. Broadbelt et al.):
Matrix representation based system.
Single molecule is represented as a square matrix whereas chemical reaction as matrix transformation.
Example: ethane fission

\[
\begin{pmatrix}
0 & 1 \\
1 & 0
\end{pmatrix} + \begin{pmatrix}
1 & -1 \\
-1 & -1
\end{pmatrix} = C \cdot \begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix}
\]

\[
CH_3 - CH_3 \rightarrow \cdot CH_3 + \cdot CH_3
\]

Literature:
NetGen (L. Broadbelt et al.):

**Advantages:**

✓ Simplicity – coding chemical reactions is easy and generator is very fast

**Disadvantages:**

✓ Only limited set of reaction and molecular species can be coded using this approach, no cyclic compounds handled
✓ The notation is unclear for a-lay audience
✓ Reaction types are hard-coded
Linear notation approach

EXGAS (Come et al.)

Linear chemical notation based system. Each specie is externally represented by a 1D-string. This notation is non-ambiguous, but also non canonical. That means that species are not represented uniquely, an complicated algorithm is required to handle them.

\[ CH_3/CH_3 \rightarrow CH_3 + CH_3 \]

Literature:
Linear notation approach

EXGAS (Come et al.)

Advantages:

✓ Wider (then in matrix approach) spectrum of reaction classes available
✓ Chemical database associated with the program, group additivity approach implemented

Disadvantages:

✓ The canonicity algorithm and the manner of handling species is very complicated and depends on the class of specie (for example different canonicity algorithms are implemented to cyclic and acyclic species)
✓ Reaction types are hard-coded
COMGEN – A New Automatic Mechanism Generator

A new graph theory based automatic reaction mechanism generator COMGEN is presented. COMGEN will be a part of integrated web-based problem solving environment for research in reaction kinetics and engineering.

COMGEN is based on chemical graph theory. Species are naturally represented as molecular graphs. The mechanism is generated using a set of reaction patterns (sub-graphs). They are internal representation for a given class of reactions thus eliminating the possibility of generating unimportant species a priori. Every specie is canonically represented as a set of real numbers (topological indices)
Reaction Design Solving Environment

COMGEN

TheRate
CIMS
Analysis

Graphic User Interface

Computational Reaction Engineering

http://vklab.hec.utah.edu/
COMGEN Procedure

Chemical Graph Theory to match reaction pattern

Matched

Generated
Pattern recognition is accomplished by comparing the tree-like structures starting from C atom in the pattern and every C atom in the molecule:

Pattern

```
  O
 /|
/  |
H CH
 H
```

Molecule

```
H C--C--C--CH
  H   H   H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HH
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HC
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```

Molecule

```
CH HO
 H
HH
 H
```

Pattern

```
C
 /|
/  |
H HO
 H
```
COMGEN Pattern Recognition

Diagram showing molecular structures with labels and connections indicated by arrows.

(1→a 2→b 3→c 4→d)
Molecular isomorphism is a procedure that determines whether newly generated molecule has been already generated or not. A **molecular ID** concept has been implemented – each species (graph) is represented by a set of real numbers (topological indices).

**Indices calculated:**

1. Randic indices of all possible orders (from $0^{th}$ to $(n-1)^{th}$) when $n$ is a number of heavy atoms in molecule
2. Wiener indices
3. Schulz indices (MTI)
4. Determinant of (distance + adjacency) matrix (TI)
5. WID index.
COMGEN Molecular Isomorphism

\[
\begin{align*}
\chi_0 & \quad 3.8284271 & \quad 3.9915638 \\
\chi_1 & \quad 2.2071068 & \quad 2.1927053 \\
\chi_2 & \quad 1.2071068 & \quad 1.1969234 \\
\chi_3 & \quad 0.6035534 & \quad 0.5773503 \\
\chi_4 & \quad 0.2500000 & \quad 0.2886751 \\
\chi_5 & \quad 0.0000000 & \quad 0.0000000 \\
\text{Wiener} & \quad 20.0000000 & \quad 20.0000000 \\
\text{MTI} & \quad 74.0000000 & \quad 74.0000000 \\
\text{WID} & \quad 5.0581531 & \quad 5.0581531
\end{align*}
\]
COMGEN Molecular Isomorphism

\[
\begin{align*}
\chi_0 & = 3.5689141 & \chi_0 & = 3.5689141 \\
\chi_1 & = 1.8569367 & \chi_1 & = 1.8569367 \\
\chi_2 & = 0.8796528 & \chi_2 & = 0.8796528 \\
\chi_3 & = 0.4023689 & \chi_3 & = 0.4023689 \\
\chi_4 & = 0.1666667 & \chi_4 & = 0.1666667 \\
\chi_5 & = 0.0000000 & \chi_5 & = 0.0000000 \\
\text{Wiener} & = 17.0000000 & \text{Wiener} & = 18.0000000 \\
\text{MTI} & = 62.5000000 & \text{MTI} & = 67.0000000 \\
\text{WID} & = 5.0743406 & \text{WID} & = 5.0656449
\end{align*}
\]
COMGEN selected reaction types available in literature

Bond fission and beta-scission

\[
\begin{align*}
&\ce{H-CH} \rightarrow \ce{H-CH + H} \\
&\ce{H-CH} \rightarrow \ce{H-CH + H} \\
&\ce{C-C} \rightarrow \ce{C-H + C-H}
\end{align*}
\]

Radical recombination:

\[
\begin{align*}
&\ce{H-H} + \ce{H-H} \rightarrow \ce{H-H} + \ce{H-H} \\
&\ce{H-H} + \ce{H-H} \rightarrow \ce{H-H} + \ce{H-H}
\end{align*}
\]

Shifts

\[
\begin{align*}
&\ce{H-C-X-CH} \rightarrow \ce{C-X-CH} \\
&\ce{H-COO} \rightarrow \ce{C-OOH}
\end{align*}
\]
COMGEN selected reaction types available in literature

Meta - thesis

\[
\begin{align*}
&\text{H} \quad \text{CH} \quad H \\
&\text{H} + \cdot X \quad \rightarrow \quad \cdot \text{CH} \quad H + \text{XH}
\end{align*}
\]

Cyclic ether formation

\[
\begin{align*}
&\text{H} \quad \text{O} \\
&\text{O} \quad \text{O} \\
&\text{C} \quad \cdot \quad \text{C} \quad \text{H} \\
&\text{H} \quad \text{H} \quad \rightarrow \quad \text{C} \quad \text{C} \\
&\text{H} \quad \text{H}
\end{align*}
\]

Branching agent formation

\[
\begin{align*}
&\cdot \quad \text{O} \quad \text{O} \\
&\text{O} \quad \text{O} \\
&\text{C} \quad \text{C} \quad \text{H} \quad \text{H} \\
&\text{H} \quad \text{H} \\
&\rightarrow \quad \cdot \quad \text{O} \\
&\text{C} \quad \text{C} \quad \text{H} \quad \text{H} \\
&\text{H} \quad \text{H} \\
&+ \quad \text{H}_2\text{O}_2
\end{align*}
\]
COMGEN selected reaction types not available in literature

Diradicals are automatically converted into alkenes:

\[
\begin{align*}
\text{HC} & \text{C} \text{CH} \\
\text{H} & \text{H} \text{H} & \text{H} & \text{H}
\end{align*}
\rightarrow
\begin{align*}
\text{HC} & \text{C} \text{CH} \\
\text{H} & \text{H} \text{H} & \text{H} & \text{H}
\end{align*} + H
\]

The program is able to recognize resonance structures:
Read patterns and input molecules, initial set up reaction and species counters

Group additivity analysis

End of the program

Last molecule?

Select first (next) molecule

Select first (next) pattern

Matches?

Last pattern?

New molecule generated?

Add new molecule

New reaction generated?

Add new reaction

Increase reaction multiplicity
A Benson group additivity approach is used. Each compound is treated as a sum of groups, each of them having been defined as an atom (ligancy >1) and its ligands. Values of thermochemical parameters are generated “on the fly”. An appropriate ring and symmetry corrections are applied, the program is able to determine symmetry number of each molecule.

The mean accuracy claimed by Benson for molecules is $\pm 0.5$ kcal/mol for $H_f^0$ and $\pm 1$ e.u./mol for $S^0$ and $C_p^0$, for free radicals it is claimed to be $\pm 1$ kcal/mol for $H_f^0$ and $\pm 2$ e.u./mol for $S^0$ and $C_p^0$. 
n – propane oxidation up to C3 species – 244 species and 4952 reactions with thermochemical parameters
n – heptane oxidation up to C7 species – 14584 species and 44956 reactions with thermochemical parameters
Acknowledgments

The authors are grateful for financial support from the National Science Foundation by the NSF-NATO Postdoctoral Fellowship in Science and Engineering (grant DGE-000798).