United States Environmental Protection Agency

Supporting Chemical Alternative Assessment via Synthesis Information Based On Chemical Structure Using a Name Reaction Database

7th International Congress on Sustainability Science & Engineering
Cincinnati, Ohio

William M. Barrett[a], Sudhakar Takkellapati[a], Kidus Tadele[a,b], Todd Martin[a], and Michael A. Gonzalez[a]

[a]U.S. Environmental Protection Agency, National Risk Management Research Laboratory
26 W. Martin Luther King Dr., Cincinnati, OH 45268
[b]Oak Ridge Institute for Science and Education (ORISE)
100 ORAU Way, Oak Ridge, Tennessee 37830

United States Environmental Protection Agency

Disclaimer

The views expressed in this presentation are those of the author[s] and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency.

Alternative Assessment of Chemicals

- **Involves comparing the potential human health and environmental effects of an identified alternate to its chemical predecessor throughout its life cycle.**
- **Focus of this work is identifying sustainable chemical synthesis routes that can be used for manufacturing of an identified alternate chemical.**
- **Inputs (reactants, catalysts and other reaction aids), as well as outputs (products and by-products) can be identified based upon the type of reaction used.**
- **The life cycle risk evaluation of a candidate manufacturing process can then be developed based upon the reaction participants throughout its lineage.**

EPA Chemical Manufacturing Ontologies

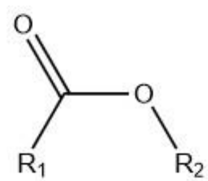
- **Chemical Lineage Ontology:**
  - Traces and identifies synthesis steps for producing a chemical from raw materials (*e.g.* crude oil).
  - Synthesis steps are typically name chemical reactions (*e.g.* Diels–Alder reaction).
  - For each step of the reaction sequence, a chemical has parents (reactants) and children (products).
  - Reaction conditions for each step are part of the ontology.
- **Process Ontology**
  - Unit operations required to carry out each of the synthesis steps.
  - Includes reactors, separation processes, storage, etc.
  - Emissions can be evaluated for each process step.

Mittal, V. K., S.C. Balin, M.A. Gonzalez, D.E. Meyer, W.M. Barrett, R.L. Smith (2018). "Toward Automated Inventory Modeling in Life Cycle Assessment: The Utility of Semantic Data Modeling to Predict Real-World Chemical Production." ACS Sustainable Chemistry & Engineering 6(2): 1961-1976.

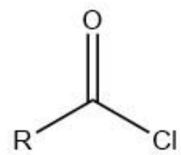
EPA United States Environmental Protection Agency

Functional Group

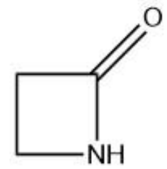
- **A functional group is an arrangement of atoms and bonds within a molecule.**
- **Examples:**



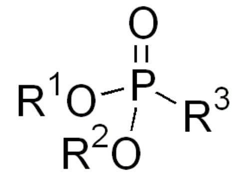
Ester



Acyl halide



Lactam



Phosphonate

- **Functional groups usually undergo the same or similar reactions regardless of the remainder of the molecule.**
- **Allows for systematic identification of reactions associated with a functional group.**

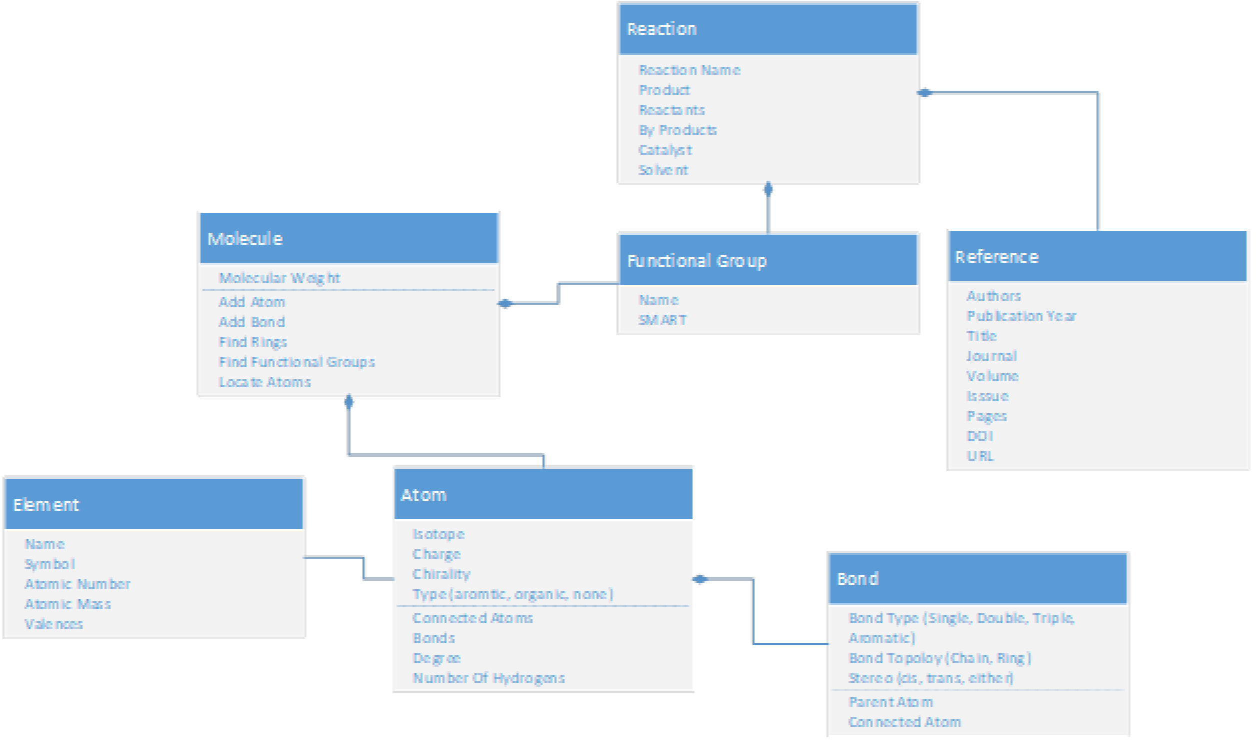
Name Chemical Reaction

- **Are descriptive for a general class of chemical reactions.**
- **Represent the general mechanism that used to produce a class of chemicals from a class of reactants under similar conditions.**
- **Involves formation of one functional group from one or more other functional groups.**
- **Examples:**
  - **Grignard reaction – alkylation of a carbonyl group.**
  - **Wittig reaction – alkene from carbonyl group.**
- **Our current database contains about 200 named reactions.**

Application Overview

- Consists of a chemical informatics class library.
- Class library can be exposed as a web service that allows users to obtain a list of functional groups and named chemical reactions associated with the functional groups within an identified chemical structure.
- Web Service currently written in C#.
- Accessed using a Windows Forms desktop application.
- Input is a structural representation of the molecule
  - Simplified Molecular-Input Line-Entry System (SMILES)
  - Chemical table (MOL) file from TEST.
- Output can either be a JSON serialized set of results or a web page.

Chemical Informatics Class Diagram



Application Overview

- **Atom Class**
  - **Nodes of the molecule graph.**
  - **Contains a collection of connected atom objects and bonds.**
  - **Obtains information about atom from Element Class.**
- **Element Class**
  - **Chemical Element Information from Blue Obelisk Data Repository (https://sourceforge.net/projects/blueobelisk/).**
- **Bond Class**
  - **Edges of the molecule graph.**
  - **Bond type (single, double, triple, aromatic)**
  - **Stereo orientation (cis/trans)**

Application Overview

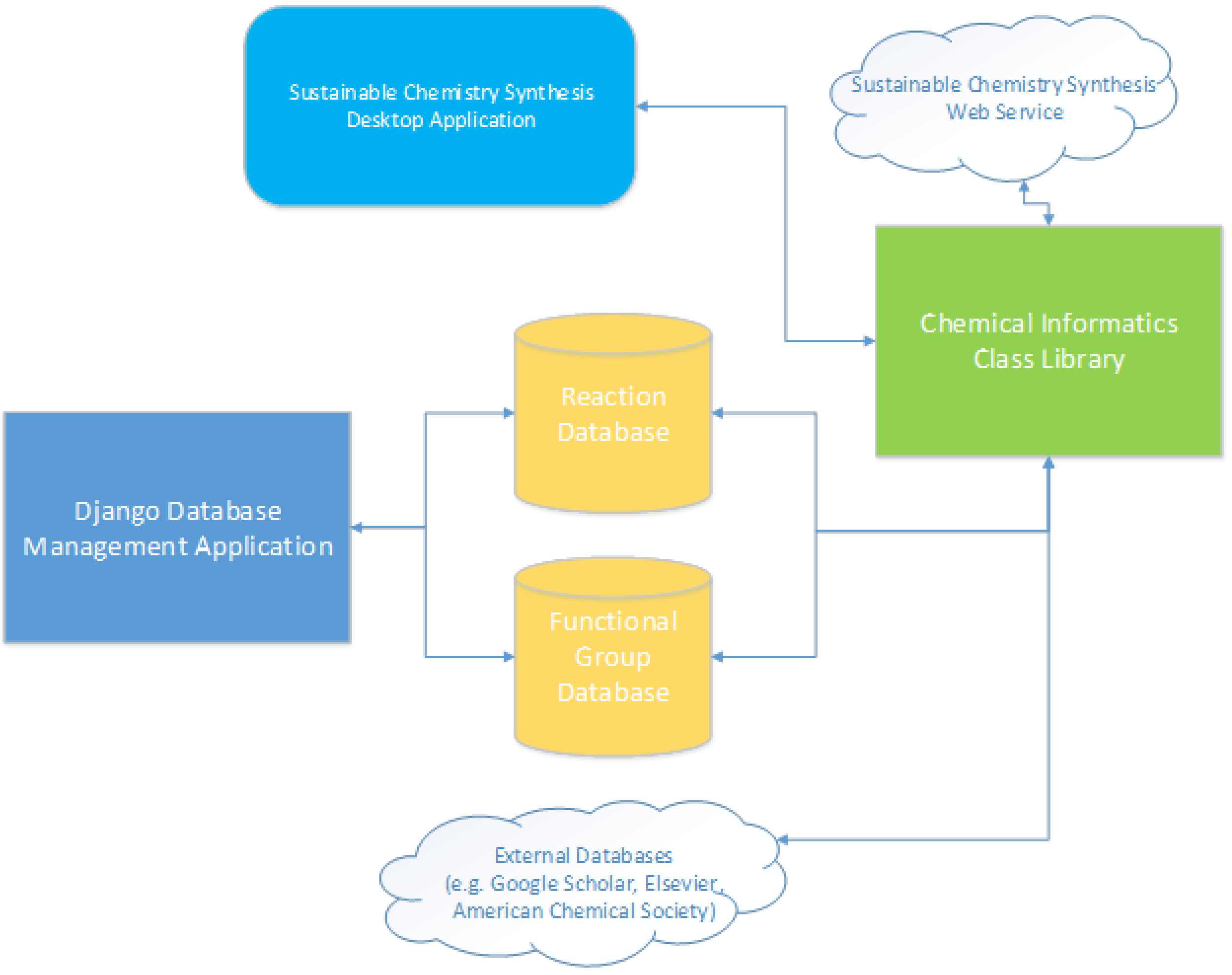
- **Molecule Class**
  - **Contains a collection of Atom objects.**
  - **Ring Identification using Tarjan Depth First Search**
  - **SMILES/SMARTS matching using VF2 Subgraph Isomorphism Algorithm**
  - **Aromaticity detection/Kekulization.**
- **SMILES/SMARTS Parser**
  - **Converts SMILES/SMARTS into an instance of the Molecule Class.**
  - **Grammar based upon OpenSMILES (http://opensmiles.org/)**
  - **Built using ANTLR4 (www.antlr.org).**
- **MOL File Parser**
  - **Used to import TEST output.**

Chemical table file specification is available at http://download.accelrys.com/freeware/ctfile-formats/ctfile-formats.zip (December 2011)
Tarjan, R. (1972). "Depth-First Search and Linear Graph Algorithms." SIAM Journal on Computing 1(2): 146-160.
Cordella, L. P., et al. (2004). "A (Sub)Graph Isomorphism Algorithm for Matching Large Graphs." IEEE Trans. Pattern Anal. Mach. Intell. 26(10): 1367-1372.

Application Overview

- **Functional Group Class**
  - **SMARTS string representing the subgraph of atoms/bonds.**
- **Named Reaction Class**
  - **List of reactant and product functional groups.**
  - **Associated with the desired functional group.**
  - **Keywords to enable online search using Google scholar, etc.**
- **Reference Class**
  - **Contains information for journal article associated with a reaction.**
  - **Input as an Endnote (RIS) file that can be obtained from the journal.**
  - **Has URL/DOI to enable access of the article from journal web site.**

Application Block Diagram

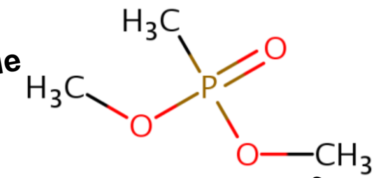


What the application does…

Enter SMILES

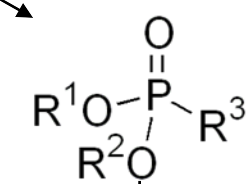
O=P(OC)(OC)C

Step 1: Construct Molecule

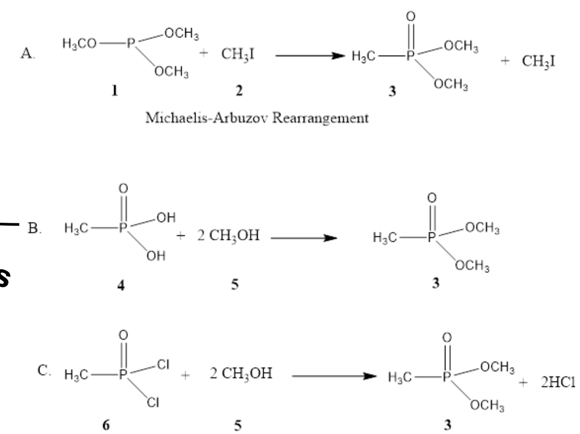


Dimethyl methylphosphonate

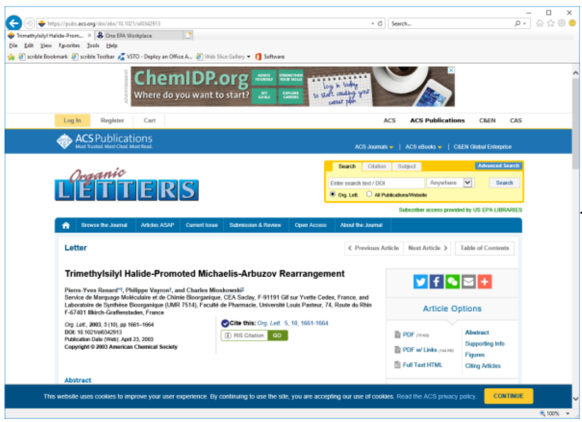
Step 2: Identify Functional Group



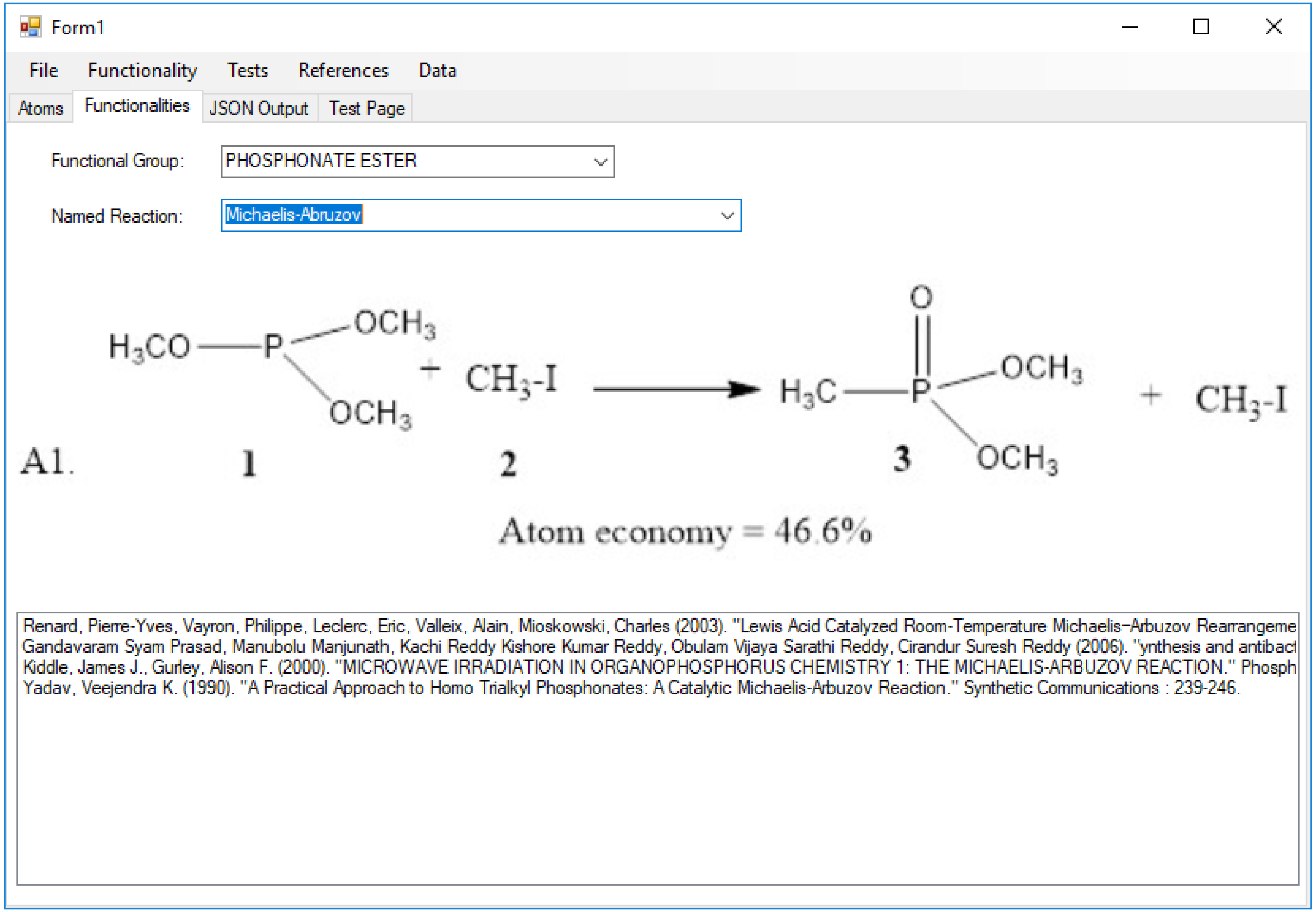
Step 3: Look Up Reactions



Step 4: Obtain References



Application Screen Shot



EPA United States Environmental Protection Agency

Alternative Assessments Dashboard Hazard Profiles for the Synthesis of Dimethyl Methyl Phosphonate

	Acute Mammalian Toxicity Oral	Acute Mammalian Toxicity Inhalation	Acute Mammalian Toxicity Dermal	Carcinogenicity	Genotoxicity Mutagenicity	Endocrine Disruption	Reproductive	Developmental	Neurotoxicity Repeat Exposure	Neurotoxicity Single Exposure	Systemic Toxicity Repeat Exposure	Systemic Toxicity Single Exposure	Skin Sensitization	Skin Irritation	Eye Irritation	Acute Aquatic Toxicity	Chronic Aquatic Toxicity	Persistence	Bioaccumulation
121-45-9 Trimethyl phosphite	M	N/A	M	N/A	L	L	H	H	N/A	N/A	M	N/A	N/A	H	H	L	N/A	L	L
74-88-4 Methyl iodide	H	H	M	VH	H	L	N/A	N/A	N/A	N/A	N/A	M	N/A	H	VH	VH	VH	H	L
2857-97-8 Bromotrimethylsilane	M	L	N/A	N/A	N/A	L	N/A	N/A	N/A	N/A	N/A	N/A	N/A	VH	VH	N/A	N/A	N/A	N/A
993-13-5 methyl phosphonic acid	M	N/A	N/A	N/A	L	L	N/A	H	N/A	N/A	N/A	N/A	N/A	N/A	N/A	M	N/A	N/A	L
67-56-1 Methanol	H	H	H	N/A	H	H	H	H	N/A	N/A	H	H	N/A	N/A	H	L	N/A	H	L
676-97-1 methyl phosphonic dichloride	VH	VH	N/A	N/A	L	L	N/A	H	N/A	N/A	N/A	N/A	N/A	N/A	N/A	L	N/A	N/A	L
67-56-1 Methanol	H	H	H	N/A	H	H	H	H	N/A	N/A	H	H	N/A	N/A	H	L	N/A	H	L

Summary

- Name reactions can be identified as a starting point for the production of chemicals based upon functional groups present in a molecule.
- Chemical lineage information, including required reactants, by-products, catalysts, reaction conditions, etc. can be determined for each of the named reactions.
- Synthesis pathways can be built from the named reactions.
- Alternative assessment of the manufacturing phase of the chemical life cycle can then be performed.