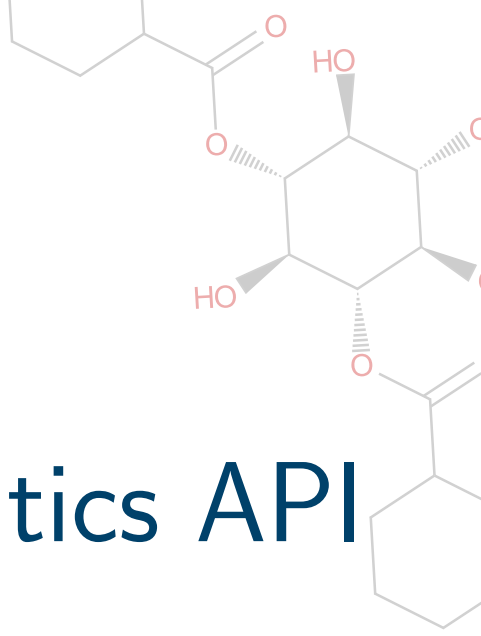
GGA SOFTWARE SERVICES

SciTouch Open-source software



Indigo: Universal Cheminformatics API

Dmitry Pavlov, Mikhail Rybalkin

November 7, 2010

Introduction

GGA SOFTWARE SERVICES

Indigo SDK is an open-source cheminformatics library with several tools.

Goals of Indigo API

- Easy access to the library of GGA cheminformatics algorithms
- Portability across OS and programming languages
- Extensibility with plugins (incl. third-party)

Preceding products

- Bingo — a chemical search engine for Oracle and SQL Server
- Dingo, Cano, Deco — single-purpose libraries

Capabilities

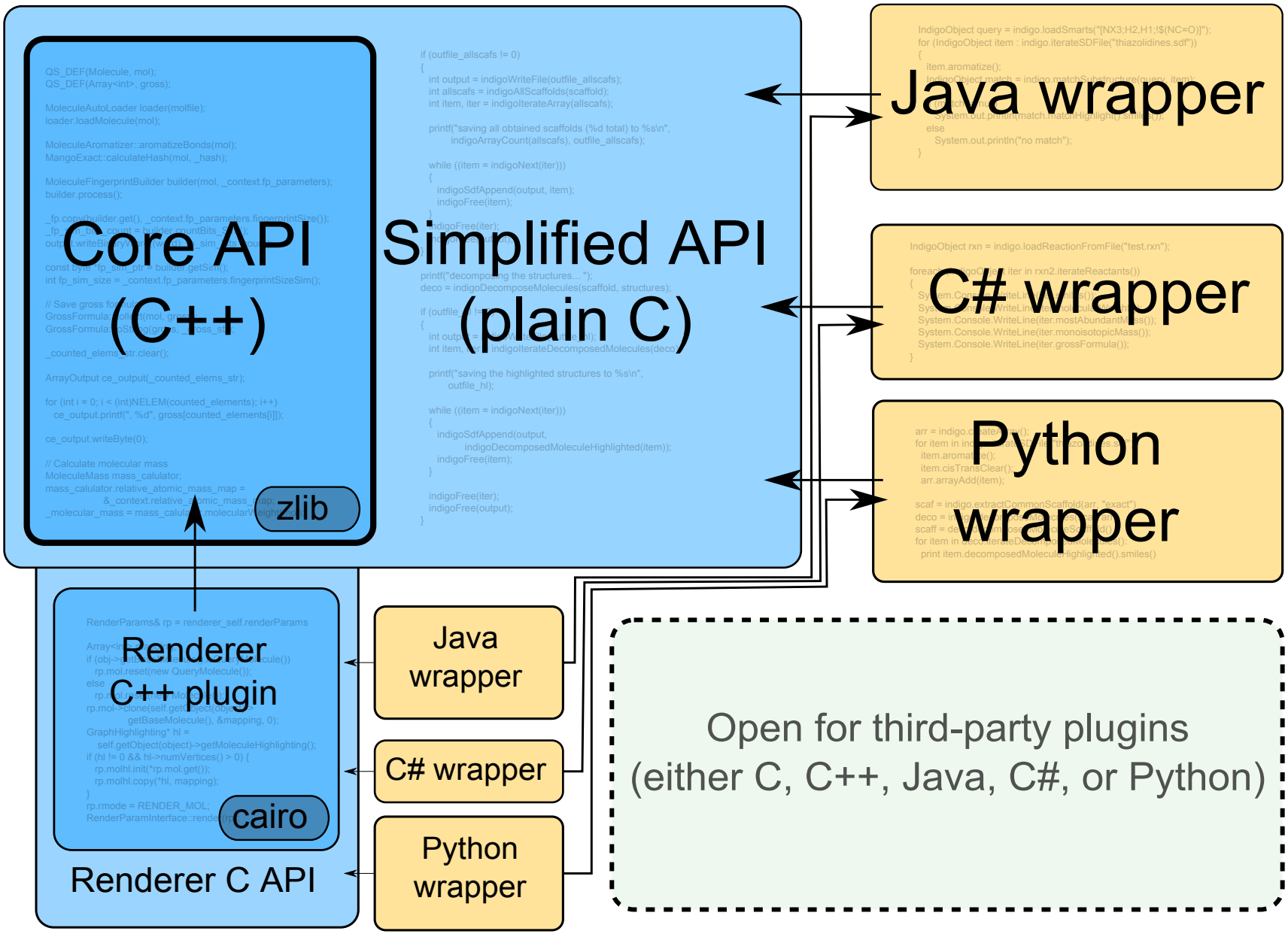
- Support of popular data formats:
  SMILES, SMARTS, Molfile, Rxnfile, SDF, RDF, GZip
- Portability over modern platforms and languages:
  Linux/Windows/Mac OS X, 32/64 bit, Java/Python/C#
- Outstanding performance:
  Original algorithms, fast C++ implementation

Functionality

GGA SOFTWARE SERVICES

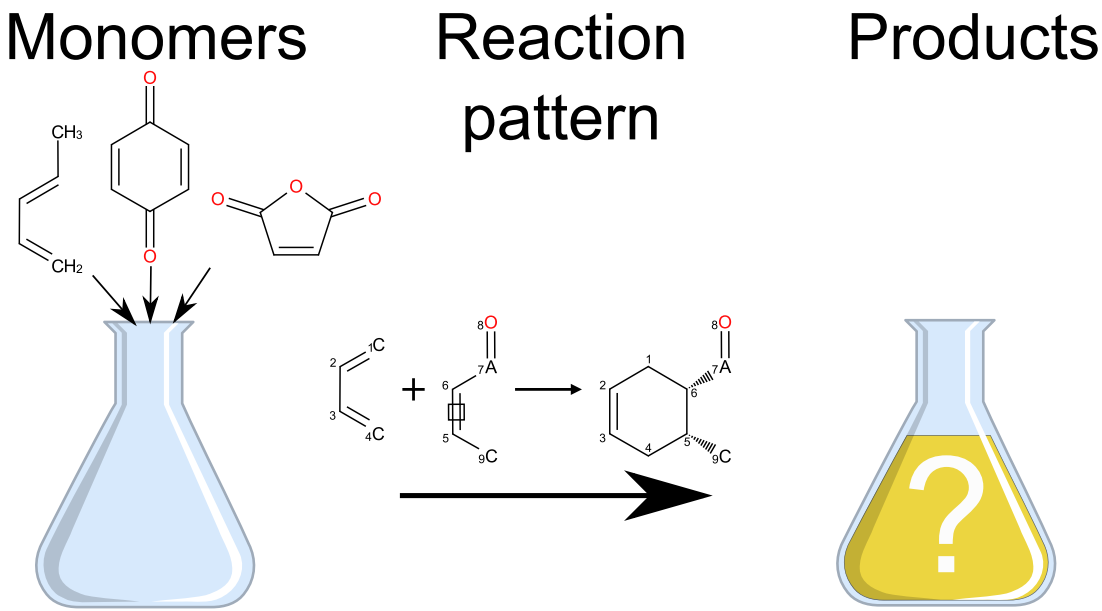
- Calculation of structure properties:
  Canonical SMILES, molecular weight, molecular formula
- Rendering of molecules and reactions:
  SVG, PNG, EMF, PDF, automatic layout, highlighting, ...
- Structure and reaction search:
  Exact, Substructure, Similarity, SMARTS
- Scaffold detection and R-Group decomposition:
  MCS of arbitrary amount of input structures
- Reaction atom-to-atom mapping
- Combinatorial chemistry:
  Stereo transformations, intramolecular and multistep reactions

Design


Core API
(C++)
zlib
Simplified API
(plain C)
Java wrapper
C# wrapper
Python
wrapper
Renderer
C++ plugin
cairo
Renderer C API
Java
wrapper
C# wrapper
Python
wrapper
Open for third-party plugins
(either C, C++, Java, C#, or Python)

Applications: Legio

Combinatorial chemistry GUI tool



Features:

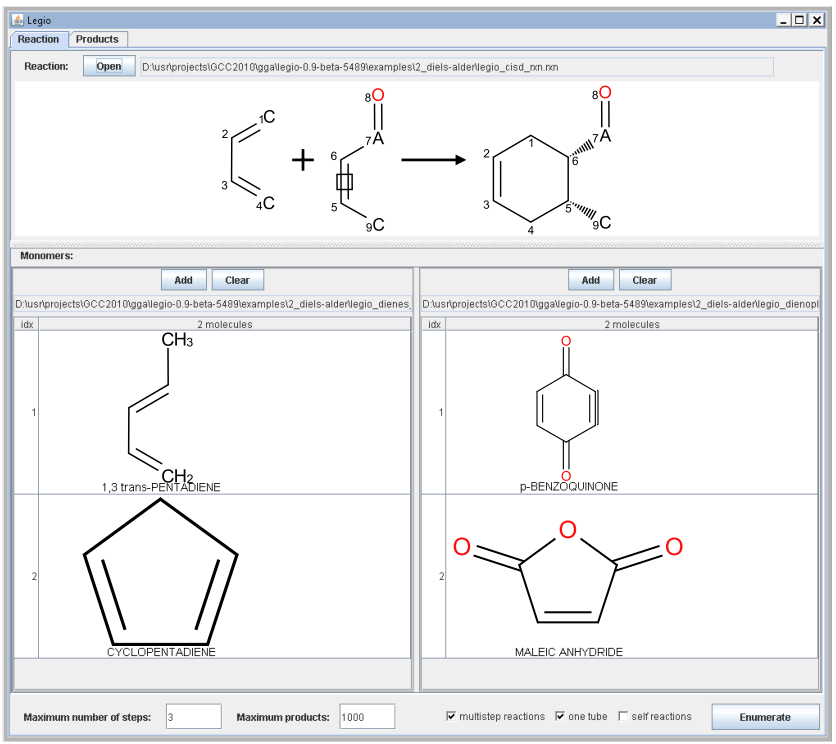
- Support of query features
- Perception of stereochemistry transformation
- Intramolecular and multistep reactions
- Explicit and implicit functional groups

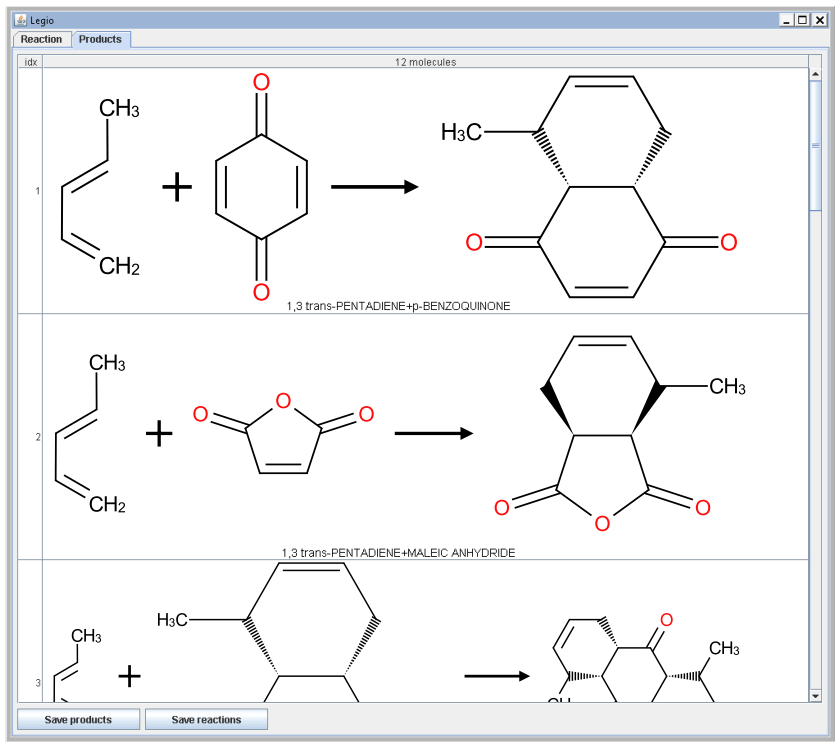
Applications: Legio (2)

Combinatorial chemistry GUI tool

Reaction pattern

Products



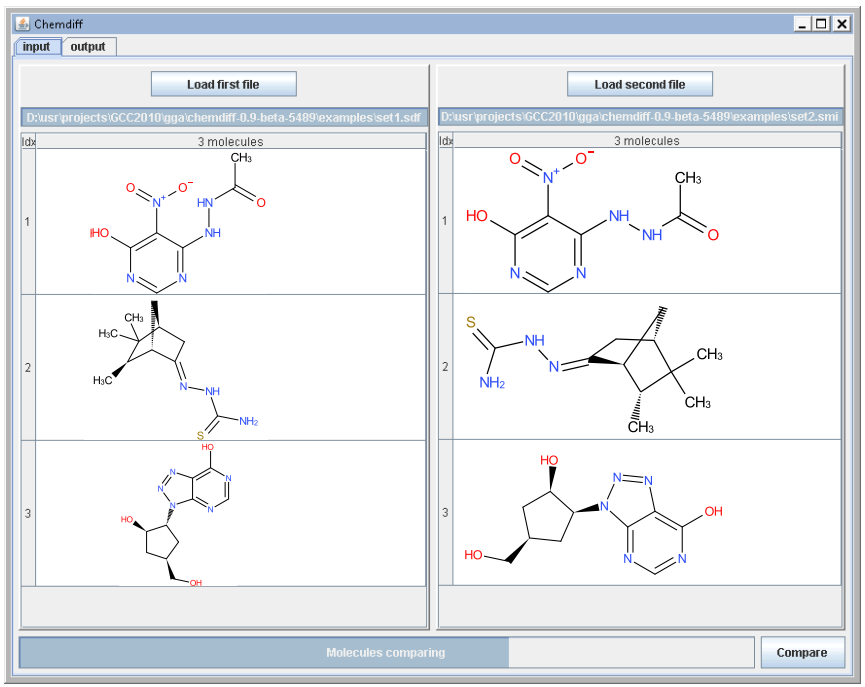


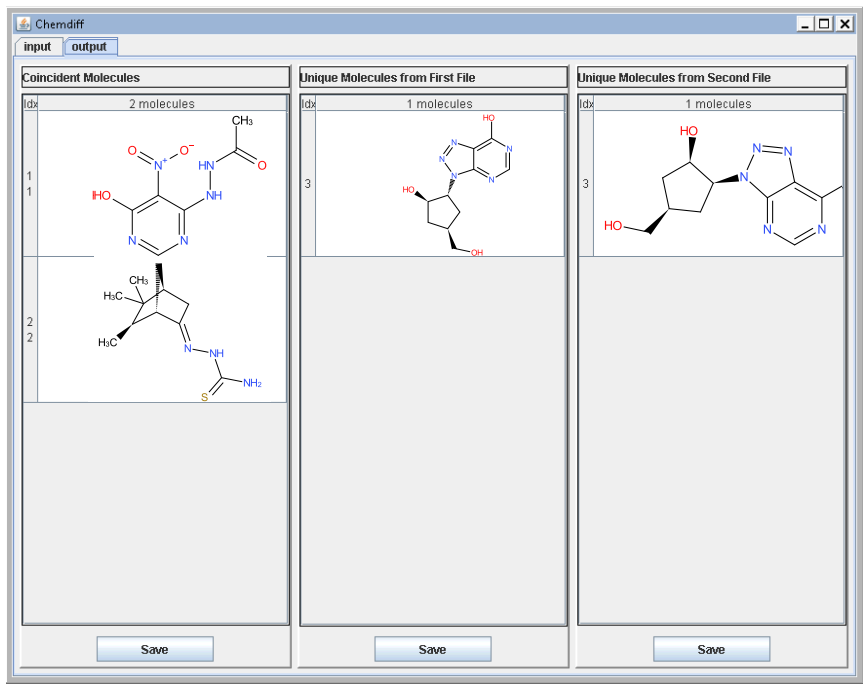
Monomers

Applications: chemdiff

Visual comparison of two SDF or SMILES files

- Find common and unique molecules

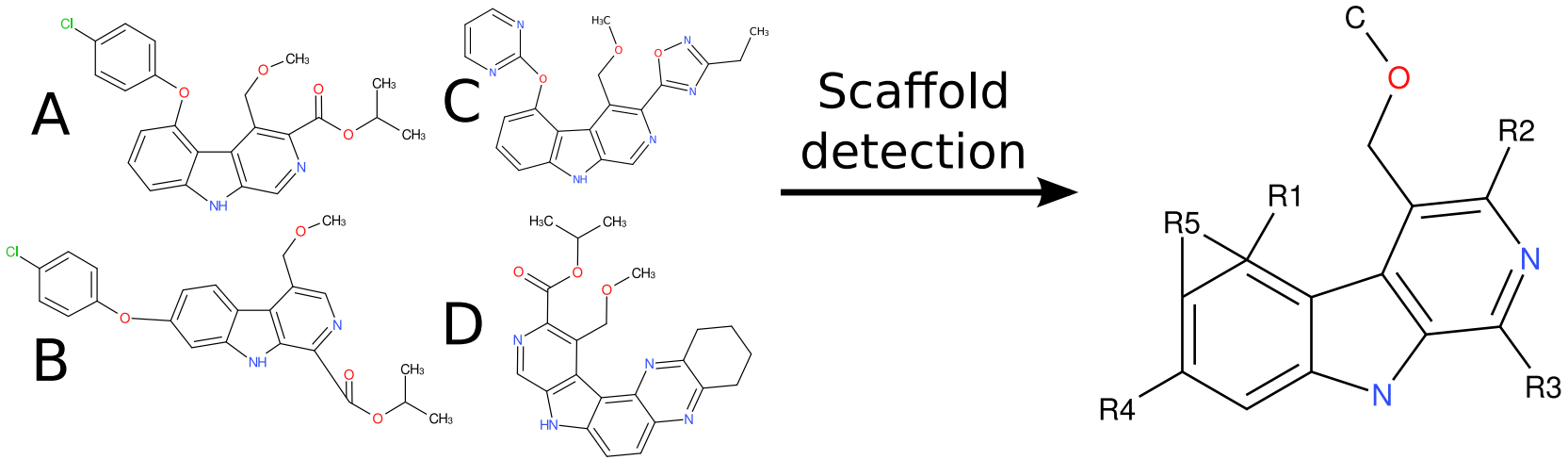


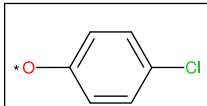
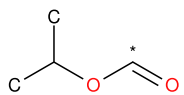
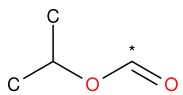
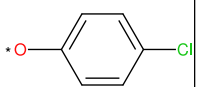
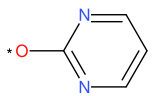
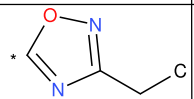
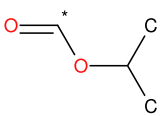
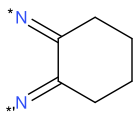


Set 1 Set 2 Common Unique 1 Unique 2

Applications: indigo-deco

R-Group deconvolution utility



	R1	R2	R3	R4	R5
A					
B					
C					
D					

Applications: indigo-depict and indigo-cano

indigo-depict — molecule and reaction rendering utility

Previously known as `dingo-render`, widely accepted across different companies and institutions.

indigo-cano — canonical SMILES generator

Previously known as `cano-utility`.
Generates absolute (isomeric) SMILES.

Community

GGA SOFTWARE SERVICES

- Web Site:
  `http://scitouch.net`
  (will move to `http://ggasoftware.com` soon)
- Google groups:
  `http://groups.google.com/group/indigo-general`
  `http://groups.google.com/group/indigo-dev`
  `http://groups.google.com/group/indigo-bugs`
- Complete source code on GitHub:
  `http://github.com/ggasoftware/indigo`
- BlueObelisk question board:
  `http://blueobelisk.shapado.com`
- Chemistry Toolkit Rosetta (code examples):
  `http://ctr.wikia.com`