

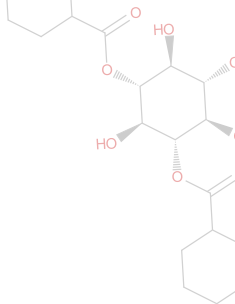


Open-source software

Indigo KNIME nodes

Mikhail Rybalkin

KNIME Open Source Days
October 6, 2011



- ▶ Saint-Petersburg & Boston since 1995

- ▶ 380 employees:

development department: 150

content department: 100

QA department: 70

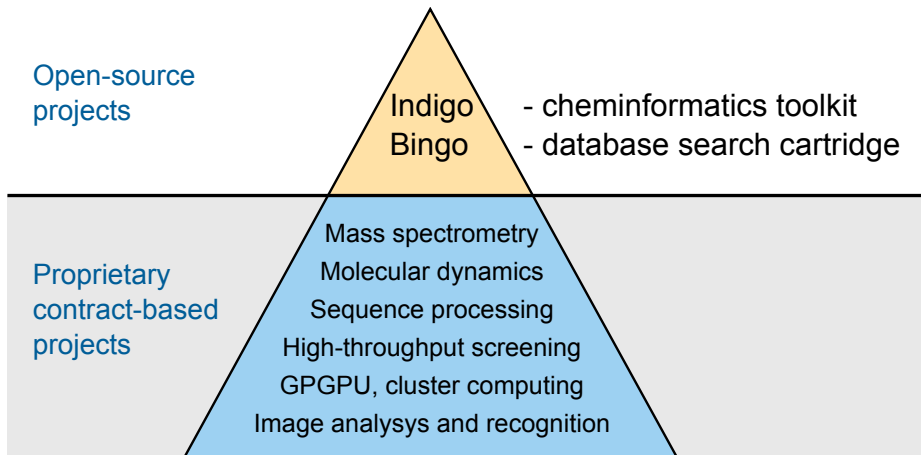
- ▶ Projects with  NOVARTIS



...

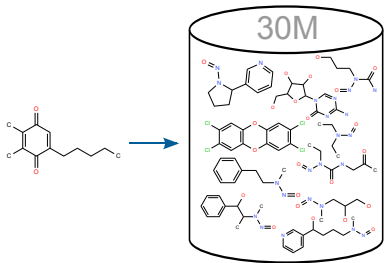
- ▶ <http://ggasoftware.com>

GGA projects



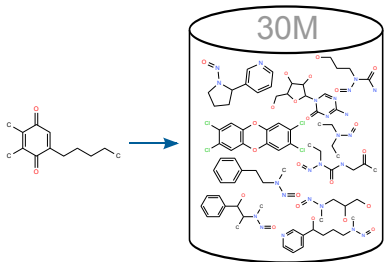
History of Indigo projects

- ▶ Different algorithms
- ▶ Bingo for Oracle



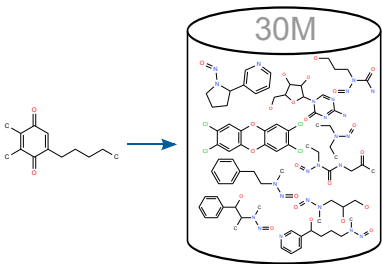
History of Indigo projects

- ▶ Different algorithms
- ▶ Bingo for Oracle
- ▶ Bingo for SQL Server
- ▶ Ketcher - molecule sketcher:
bingo-demo.ggasoftware.com/



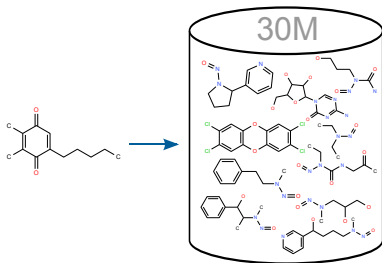
History of Indigo projects

- ▶ Different algorithms
- ▶ Bingo for Oracle
- ▶ Bingo for SQL Server
- ▶ Ketcher - molecule sketcher:
bingo-demo.ggasoftware.com/
- ▶ Indigo toolkit API:
C#, Java, Python



History of Indigo projects

- ▶ Different algorithms
- ▶ Bingo for Oracle
- ▶ Bingo for SQL Server
- ▶ Ketcher - molecule sketcher:
bingo-demo.ggasoftware.com/
- ▶ Indigo toolkit API:
C#, Java, Python
- ▶ Bingo for PostgreSQL



Why Open Source?

- ▶ Bussiness model

Why Open Source?

- ▶ Bussiness model
- ▶ Relationship with the scientific commnity
- ▶ Feedback from the commnity
 - ▶ Suggestions
 - ▶ Testing

Why Open Source?

- ▶ Business model
- ▶ Relationship with the scientific community
- ▶ Feedback from the community
 - ▶ Suggestions
 - ▶ Testing
- ▶ Indigo modules are used in commercial projects

Indigo SDK is an open-source cheminformatics library.

Goals of Indigo

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

- ▶ 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

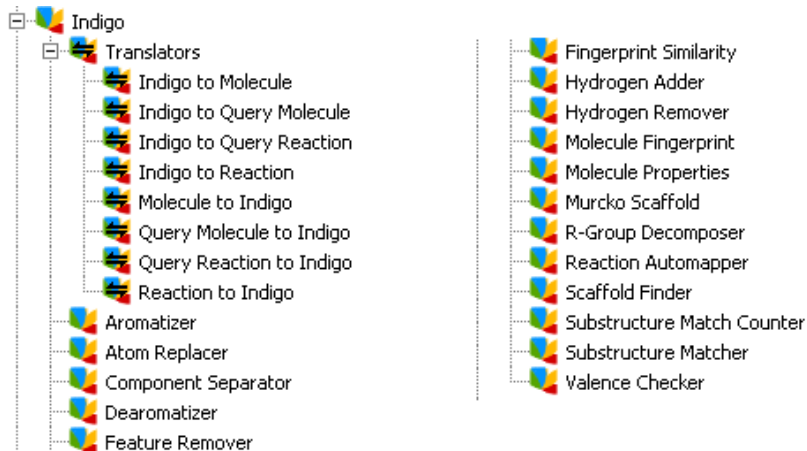
Functionality

- ▶ 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, fingerprints

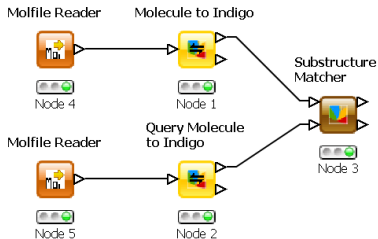
- ▶ Calculation of structure properties:
Canonical SMILES, molecular weight, molecular formula
- ▶ Rendering of molecules and reactions:
SVG, PNG, EMF, PDF, automatic layout, highlighting, ...
- ▶ Structure and reaction search:
Exact, Substructure, Similarity, SMARTS, fingerprints
- ▶ Scaffold detection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

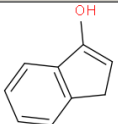
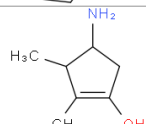
Indigo Nodes for KNIME


- ▶ First release: May 2011
- ▶ Intensive communication with the community

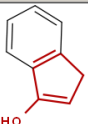
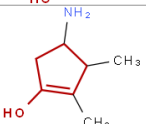


Substructure search



Row ID	Molecules
t1	
t2	

Row ID	Molecules
q	

Row ID	Molecules
t1	
t2	

Atom-to-atom mapping

Chemical Reactions
File Reader

Reaction to Indigo

Reaction Automapper

Indigo to Reaction



Node 1



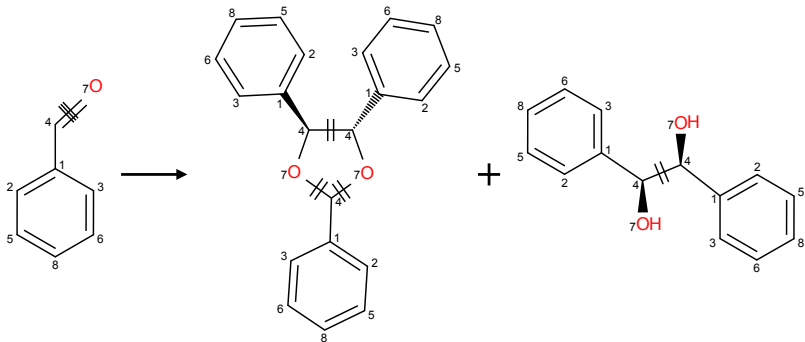
Node 2



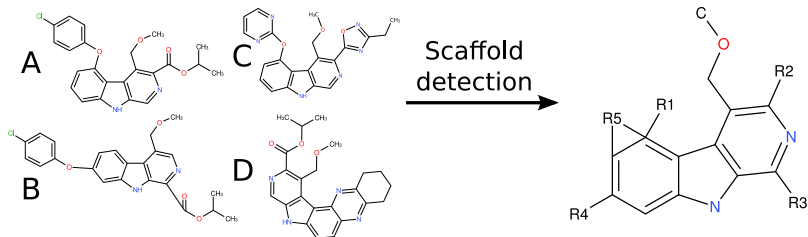
Node 3











Node 4



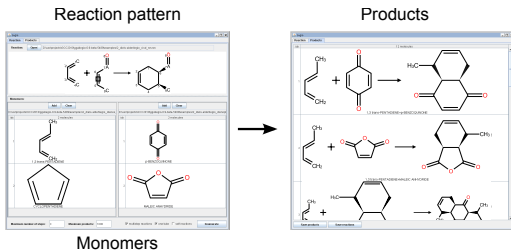
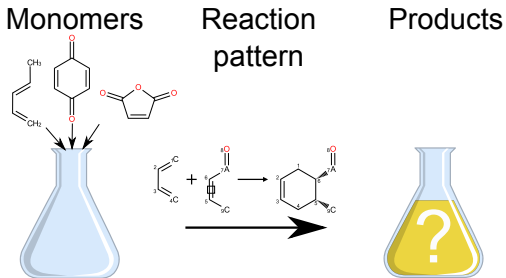
R-Group decomposition



	R1	R2	R3	R4	R5
A					
B					
C					
D					

Combinatorial chemistry

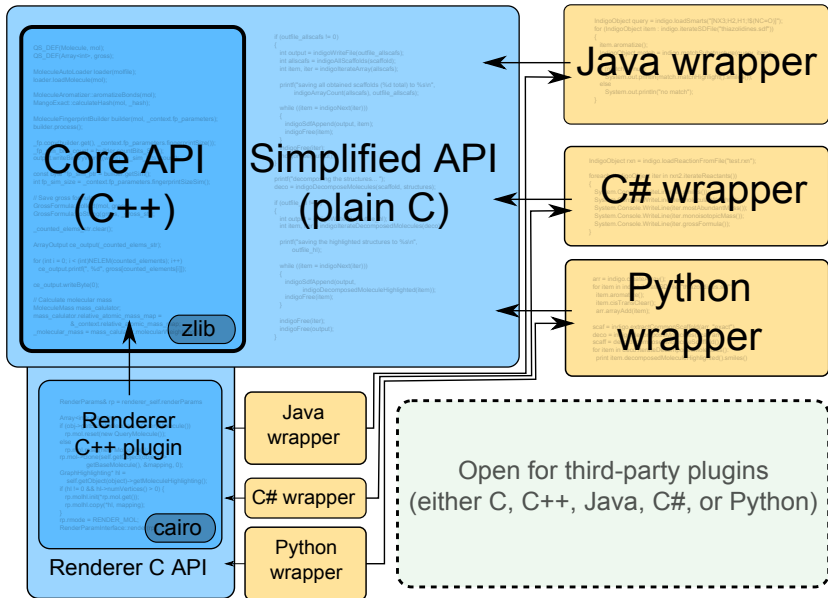
- ▶ Integrate Legio tool into KNIME



- ▶ Cross-platform:
 - ▶ Linux/Windows/Mac OS X
(Solaris is also supported by request)
 - ▶ 32/64 bit
- ▶ Testing:
 - ▶ Jenkins
 - ▶ Python/IronPython/Jython

- ▶ Cross-platform:
 - ▶ Linux/Windows/Mac OS X
(Solaris is also supported by request)
 - ▶ 32/64 bit
- ▶ Testing:
 - ▶ Jenkins
 - ▶ Python/IronPython/Jython
- ▶ C++ core code, plain C API
- ▶ Indigo API for Java/C#/Python
SWIG ?

- ▶ Cross-platform:
 - ▶ Linux/Windows/Mac OS X
(Solaris is also supported by request)
 - ▶ 32/64 bit
- ▶ Testing:
 - ▶ Jenkins
 - ▶ Python/IronPython/Jython
- ▶ C++ core code, plain C API
- ▶ Indigo API for Java/C#/Python
SWIG ?
- ▶ Hydrogens support.
SMARTS: h notation is not supported on purpose.



```
Indigo indigo = new Indigo ();
```

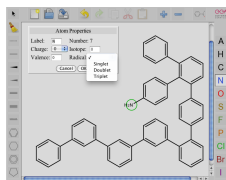
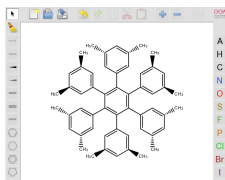
```
IndigoObject rxn = indigo.createReaction ();  
rxn.addReactant (mol1 );  
rxn.addReactant (mol2 );  
rxn.addProduct (indigo.loadMolecule (" C1C1CCCCC1" ));  
rxn.automap (" discard" );
```

```
System.out.println (rxn.smiles ());  
for (IndigoObject item : rxn.iterateReactants ())  
    System.out.println (item.molfile ());
```

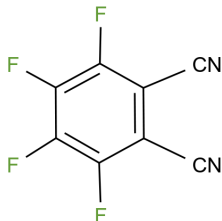
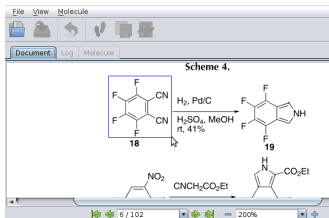
1. Add reaction support into KNIME:
 - ▶ automatic atom-to-atom mapping (already available)
 - ▶ substructure and exact search
 - ▶ combinatorial chemistry
 - ▶ reaction-based transformations
2. Add reaction readers into KNIME core
3. Add other additional nodes (community requests)
4. Multithreading
5. Testing workflows

Other open-source projects

Ketcher — web-based chemical structure editor



Imago — 2D chemical structure image recognition toolkit



Acknowledgments

- ▶ Indigo community
- ▶ KNIME community
- ▶ KNIME Open Source Days organizers
- ▶ Dmitry Pavlov

- ▶ Web Site:

<http://ggasoftware.com/opensource>

- ▶ 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>

- ▶ Chemistry Toolkit Rosetta (code examples):

<http://ctr.wikia.com>

- ▶ KNIME Community:

<http://tech.knime.org/forum/indigo>

Suggestions to KNIME

- ▶ Pipes between nodes to avoid intermediate data
- ▶ Automatic node conversion to a newer version
- ▶ Universal domain-specific language for table processing