Markush structures –
From molecules towards patents

Szabolcs Csepregi

ChemAxon
Solutions for Cheminformatics

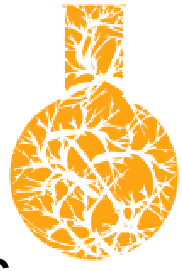
A journey to Markush-land

- Departure

- Markush structures: What are they?
- Getting them,
- Enumeration,
- Storage, search

- Arrival: Recent developments, plans

Departure – ChemAxon

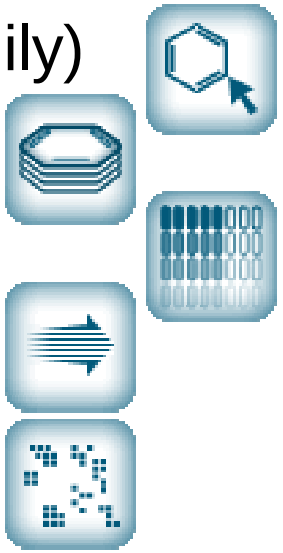


- Cheminformatics toolkits and applications
- HQ: Budapest, Hungary
- Founded: 1998
- Main customers: pharma, biotech, publishing
- 3rd party applications and web sites.
  (e.g. Integrity, Reaxis, PDB ligand search, ELN-s, registration systems, etc)

Departure – ChemAxon

Main products:

- Structure drawing & visualization (Marvin family)
- Chemical DB tools (JChem family)
- Property predictions (Calculator plugins)
- Drug discovery tools (Reactor, JKlustor, etc.)

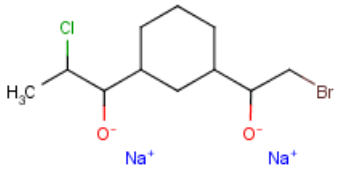


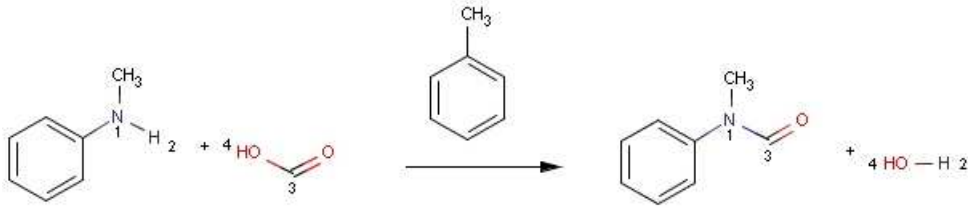
Development strategy: customer-driven

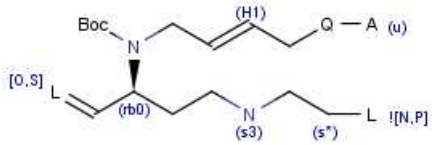
Departure – Initial status

2005

- Chemical Drawing, DB tools
  - molecule, reaction and query structrures







- Customers needed Markush funcionality, especially patents.

What are Markush structures

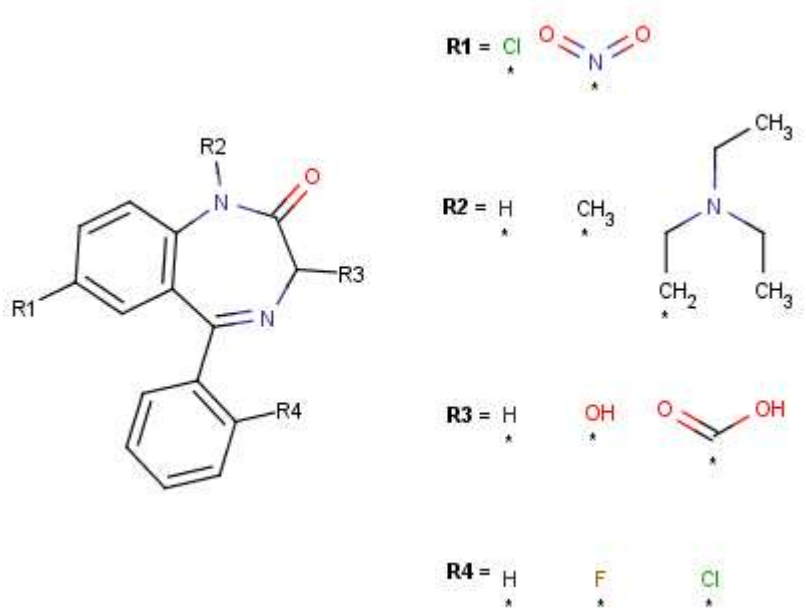
and how to get them?

Markush structures

Generic notation for describing many molecules (= Markush library) in a compact form.

Main usage:

- **Combinatorial chemistry**
- **Chemistry-related patents**



United States Patent [19]
Bitonti et al.

[11] **Patent Number: 5,681,863**
[45] **Date of Patent: Oct. 28, 1997**

[54] **NON-METABOLIZABLE CLOMIPHENE ANALOGS FOR TREATMENT OF TAMOXIFEN-RESISTANT TUMORS**

[75] Inventors: **Alan J. Bitonti**, Maineville; **Russell J. Baumann**, Cincinnati, both of Ohio

[73] Assignee: **Merrell Pharmaceuticals Inc.**, Cincinnati, Ohio

[21] Appl. No.: **350,192**

[22] Filed: **Dec. 5, 1994**

Related U.S. Application Data

[62] Division of Ser. No. 196,817, Feb. 10, 1994, Pat. No. 5,410,080, which is a continuation of Ser. No. 945,305, Sep. 15, 1992, abandoned.

[51] **Int. Cl.**[6] **A61K 31/135**
[52] **U.S. Cl.** **514/648**; 514/649
[58] **Field of Search** 514/648, 649

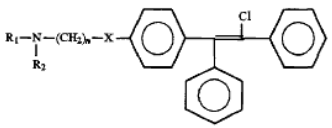
[56] **References Cited**

3,631,109	12/1971	O'Sega et al.	260/570.9
4,696,949	9/1987	Toivola et al.	514/644
4,839,155	6/1989	McCague	514/651
5,114,951	5/1992	King et al.	514/290
5,130,424	7/1992	Weintraub	540/28

Primary Examiner—Jerome D. Goldberg
Attorney, Agent, or Firm—Nelsen L. Lentz

[57] **ABSTRACT**

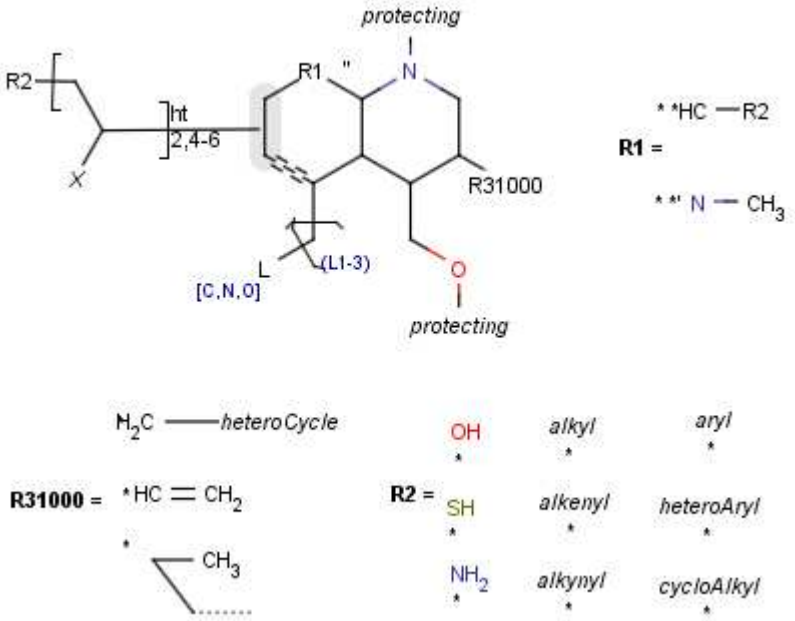
Compounds of the formula:



wherein R_1 and R_2 are each selected from the group consisting of C_1–C_2 lower alkyl; X is NH or S; and n is a whole number within the range of 1–4 inclusive; and when n=0, X is $(CH_2)_3$ and the pharmaceutically acceptable salts thereof

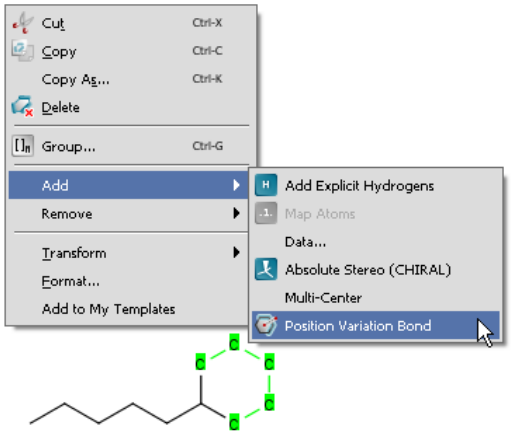
Markush structures

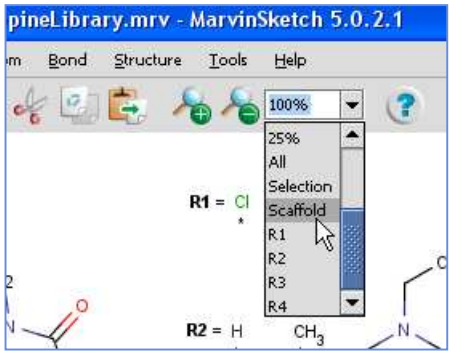
- Current features handled:
  - R-groups
  - Atom lists, bond lists
  - Position variation bond
  - Link nodes
  - Repeating units
  - Homology groups (aryl, alkyl, etc.)

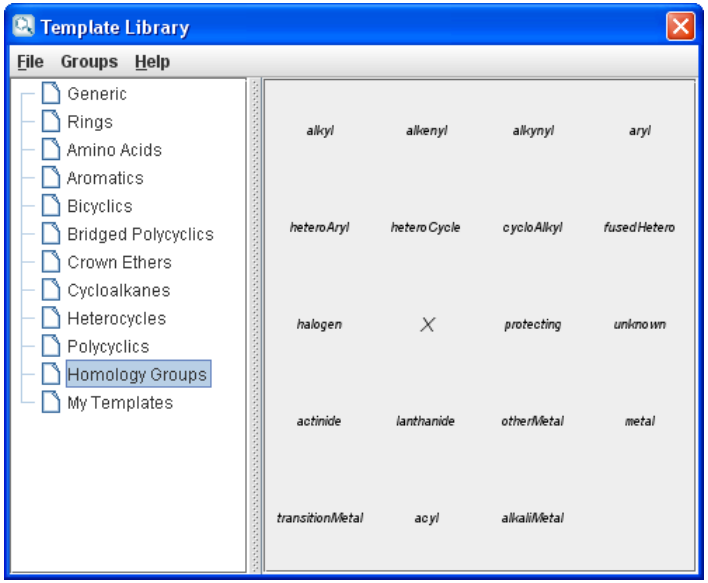


How to get Markush structures?

- Drawing – Marvin Sketch

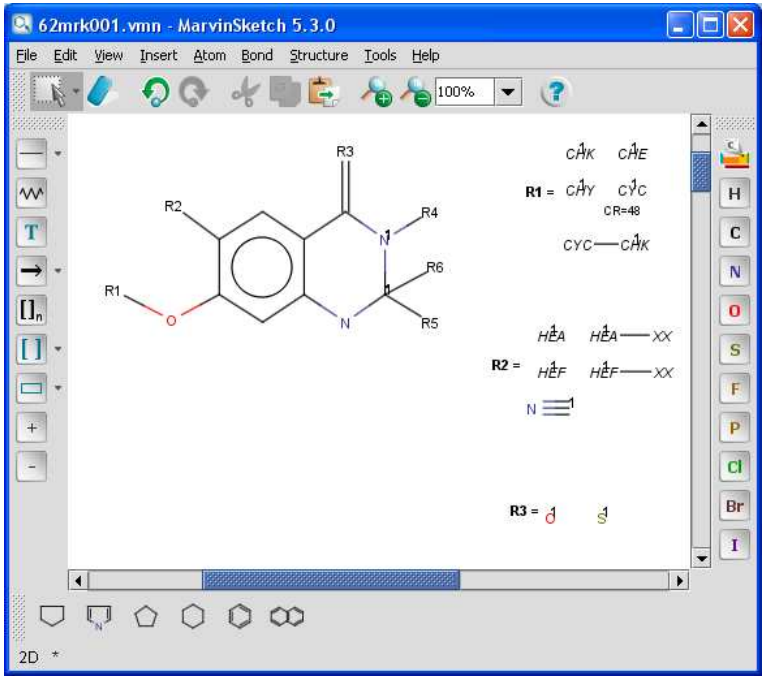






How to get Markush structures?

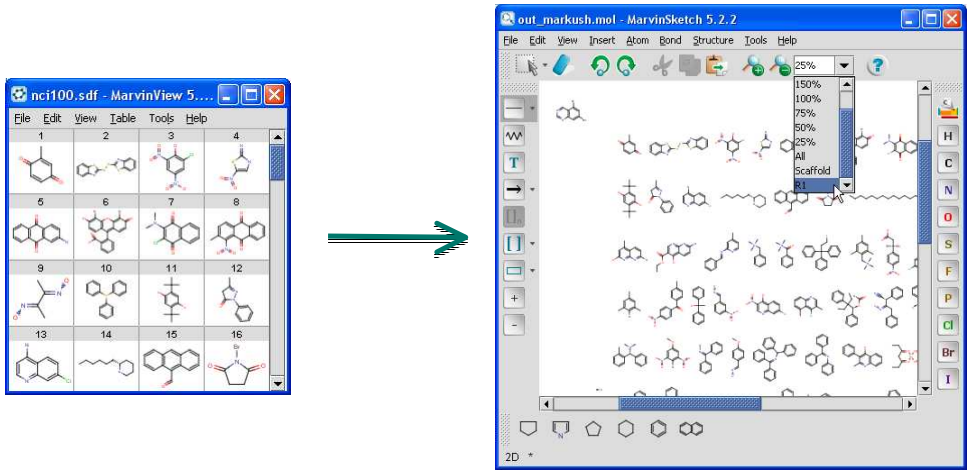
- Patent literature (VMN format coming in 5.3 – Derwent World Patent Index)



How to get Markush structures?

Combinatorial chemistry – **Reagent clipping**

1. Replace reacting group with attachment point (Reactor tool)
2. Turn fragments to R-group definitions (Molconvert tool)
3. Add a scaffold (Molconvert tool)

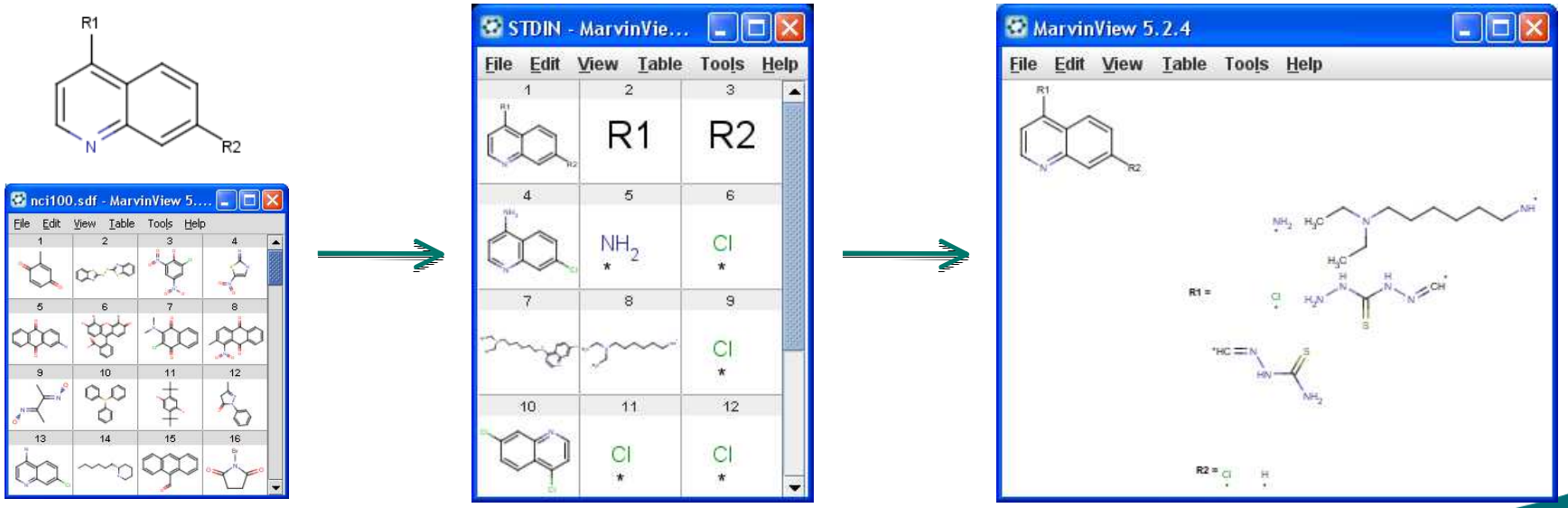


How to get Markush structures?

Combinatorial chemistry – **R-group decomposition**

1. Filter and identify ligands in chemical library
2. Create Markush structure from R-table

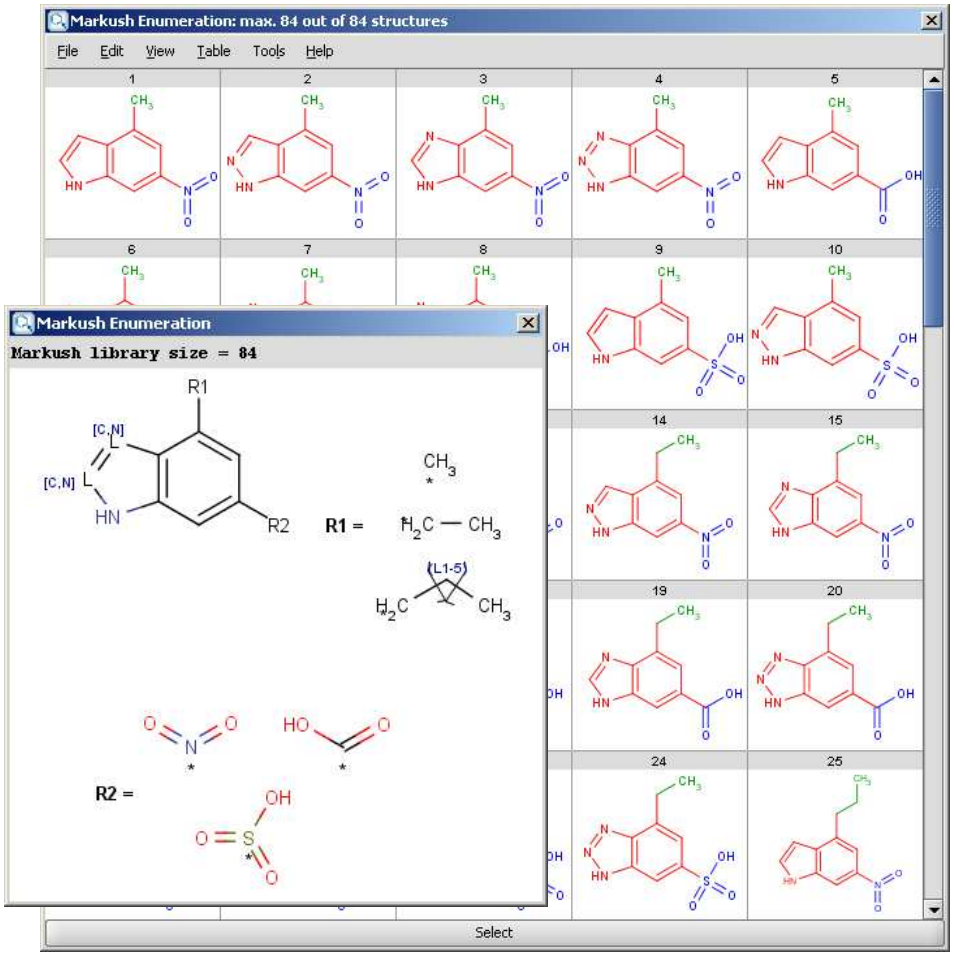
(R-group decomposition tool)



What to do with them?

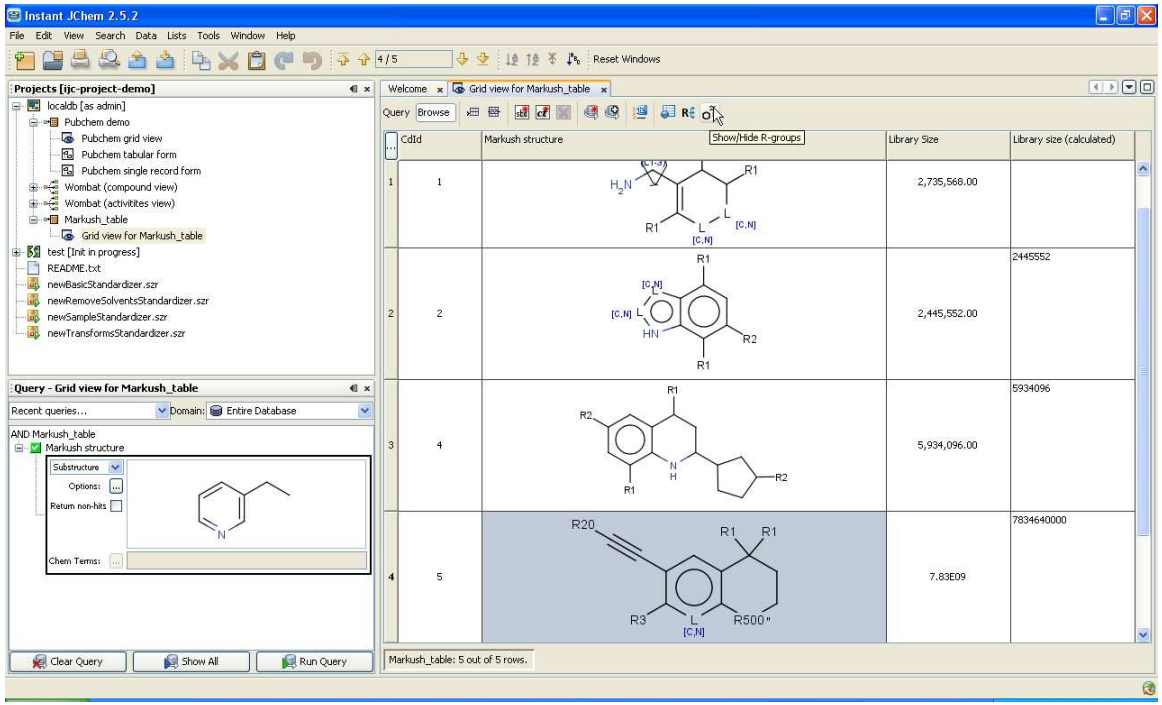
Markush Enumeration

- Markush enumeration plugin
  - Full enumeration
  - Selected parts only
  - Random enumeration
  - Calculate library size
  - Scaffold alignment and coloring
  - Markush code
  - Optional example homology group enumeration



Markush storage & search

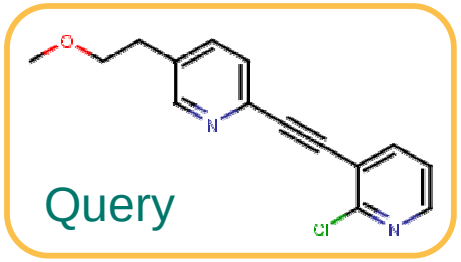
- JChem Base and Instant JChem
- No enumeration involved
- Can handle complex Markush structures (10^{40} or more)

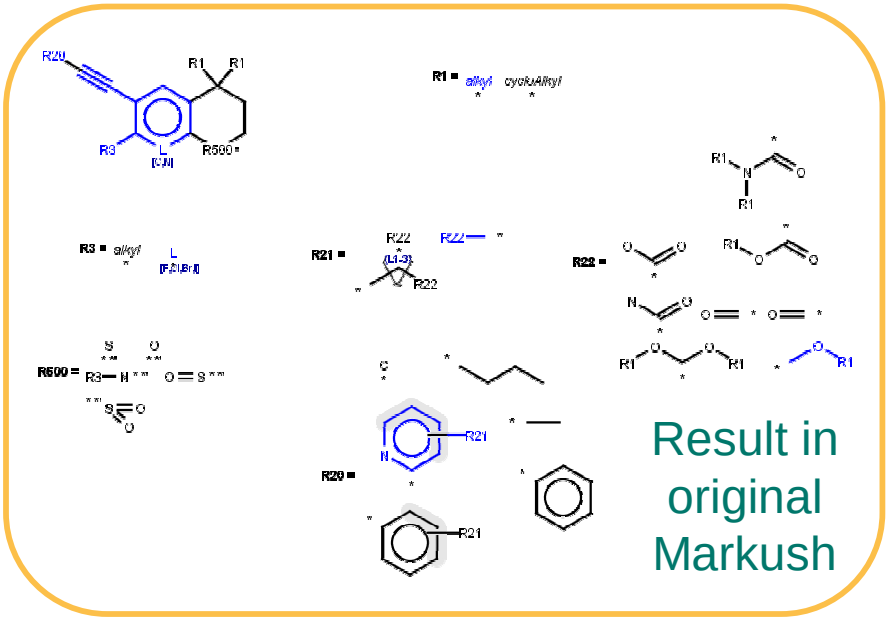


- Substructure and Full structure search
- Basic query features supported

Markush storage & search

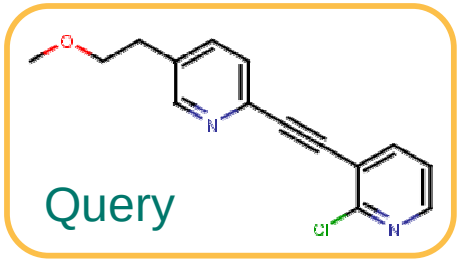
Substructure hit visualization

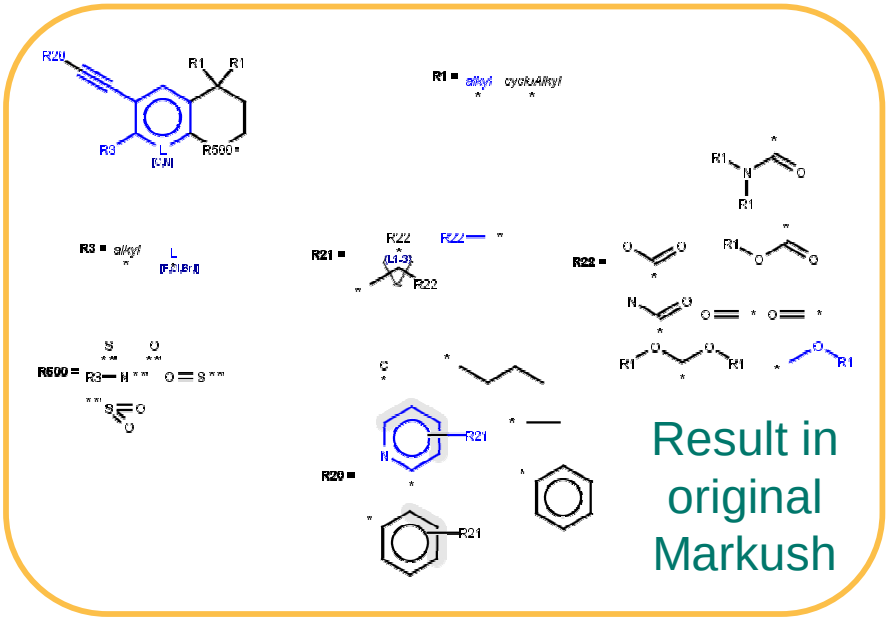


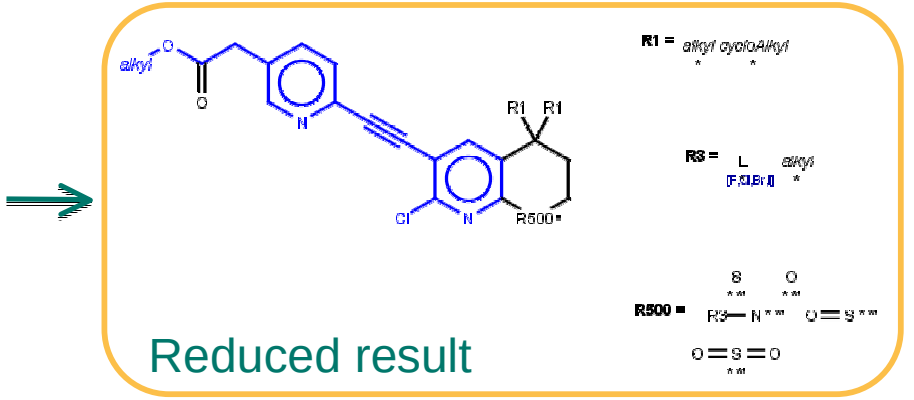


Markush storage & search

Substructure hit visualization:
„Markush structure reduction”





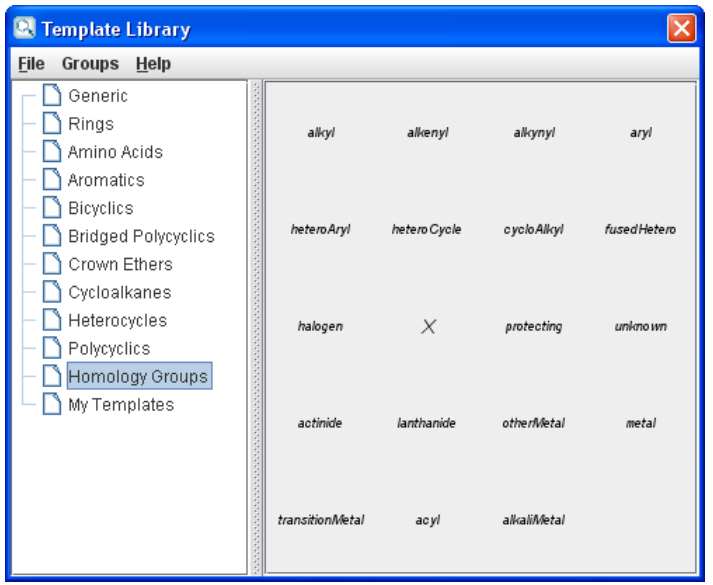


What's new

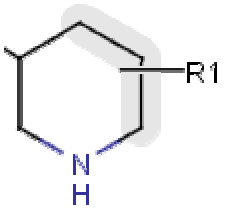
- Homology groups

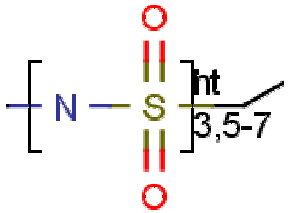
R1 = alkyl * cycloAlkyl *

  - 19 built-in groups
    - Marvin templates for easier sketching
  - Customizable:
    - Examples (for built-in groups),
    - User-defined homology groups

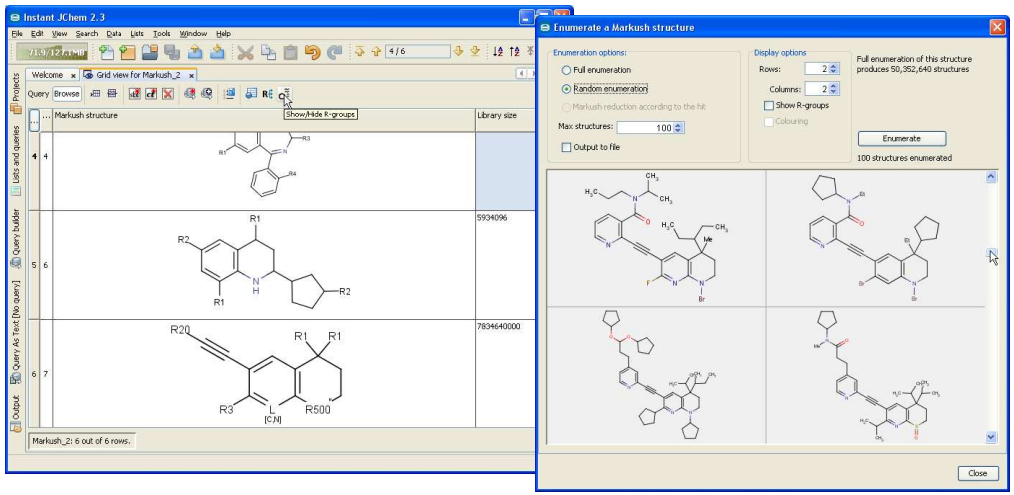


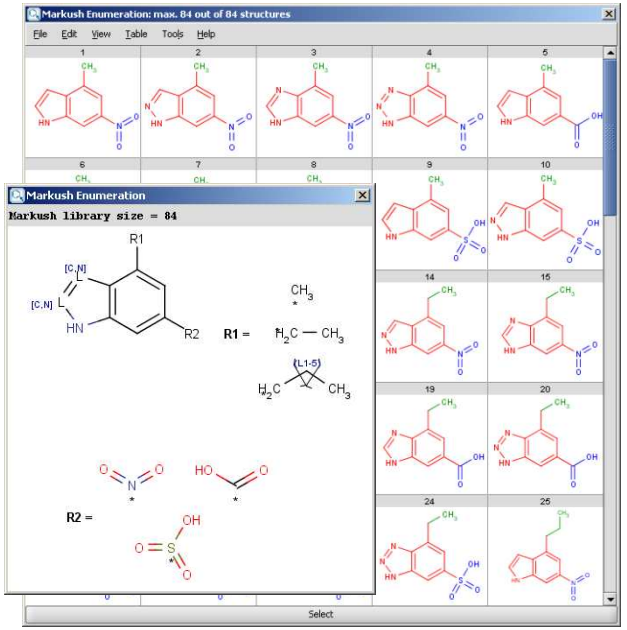
- Import reagent files as R-groups
- Position variation and Repeating units

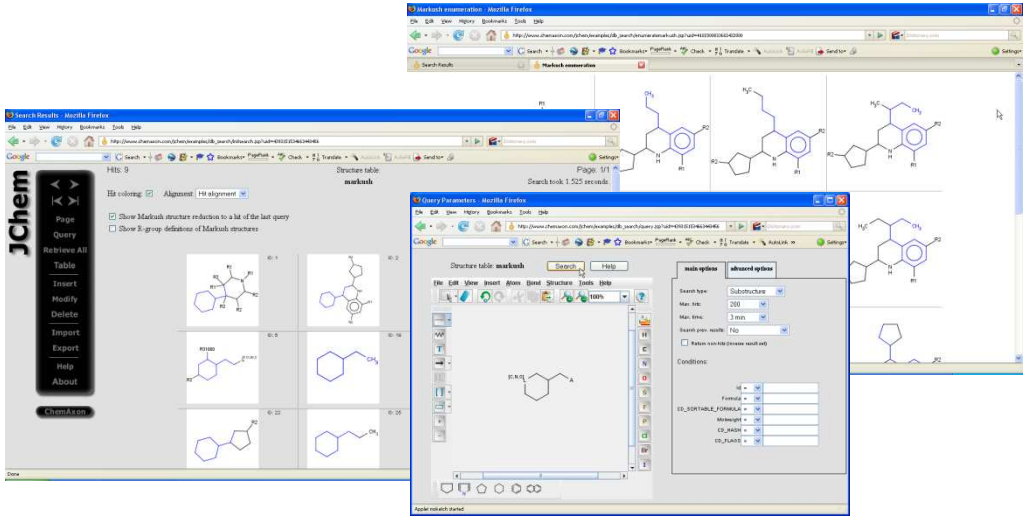




Demo







[Click here...]

Main use cases

- Patent search hits refining,
- White space analysis,
- Markush structure curation,
- In-house storage of small Markush DB,
- etc...

Under development

- .VMN import (Derwent World Patent Index)
  5.3 – this year
- Homology variation queries (narrow translation)
- Maximum common substructure search
- Biased enumeration
- All Markush features of .VMN format
- Overlap analysis of Markush structures
- Conditions for Markush variables

Future work for the community

- Lack of open Markush file format standards.
- Community needs patent Markush data.
- Call for Markush patent content holders to make data accessible.

- Solution?
  - InChI or CML(XML) extensions?
  - Open up existing format specifications?
  - ??

Summary

- Markush structure storage, search and enumeration at ChemAxon now reaching patent coverage
- Continuous development, improvements in the pipeline

Acknowledgements

- Development team: Nóra Máté, Róbert Wágner, Szilárd Dóránt, Tamás Csizmazia, Ferenc Csizmadia, et al.

- Tim Miller and Linda Clark at Thomson Reuters for useful discussions, help and example .VMN files

- Many early adopters and colleagues within the field for suggestions and feedback

Interested?

- We are looking for further early adopters
- Currently running individual projects with pharma companies to test and enhance functionality.
- If you are interested, please contact us.