

Markush project — Introducing Markush DARC support

Szabolcs Csepregi

May 2010, European UGM

- What are Markush structures?
- How to get them?
- What can be done with them?
 - Enumeration
 - Storage, search
- Recent developments, plans

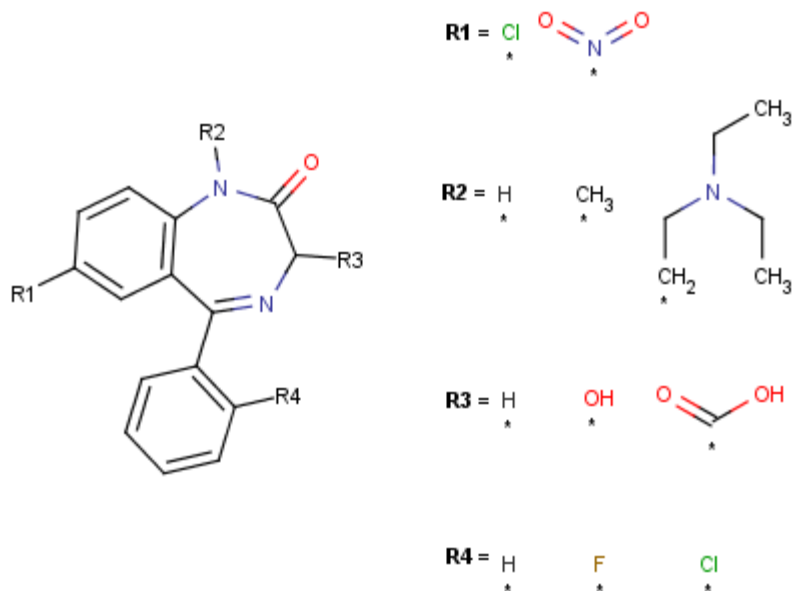
**What are Markush structures
and how to get them?**

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

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

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

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

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

[57] ABSTRACT

[21] Appl. No.: 350,192

Compounds of the formula:

[22] Filed: Dec. 5, 1994

Related U.S. Application Data

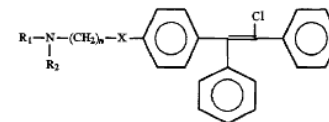
[62] Division of Ser. No. 196,817, Feb. 10, 1994, Pat. No. 5,410,080, which is a continuation of Ser. No. 945,305, Sep. 15, 1992, abandoned.

[51] Int. Cl.⁶ A61K 31/135

[52] U.S. Cl. 514/648; 514/649

[58] Field of Search 514/648, 649

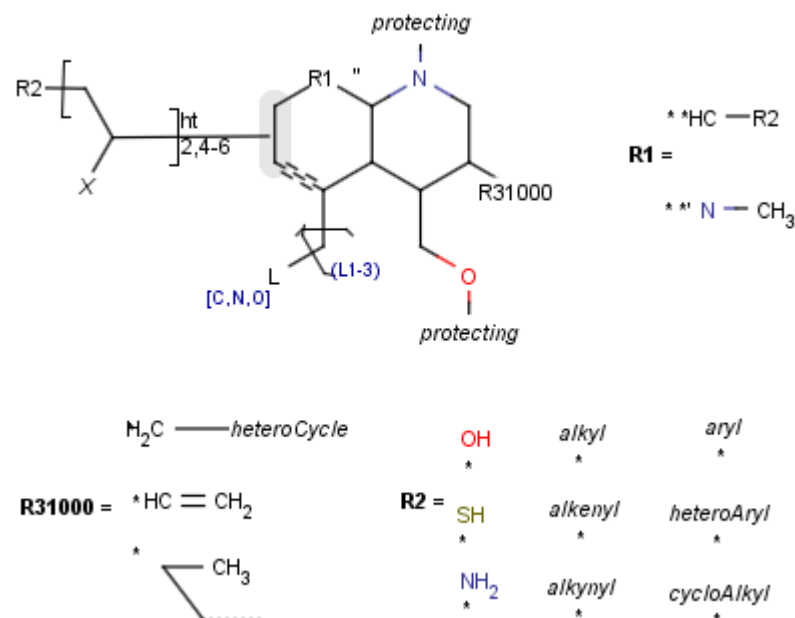
[56] References Cited



wherein R₁ and R₂ are each selected from the group consisting of C₁-C₂ 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 (CH₂)₂ 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.)



ChemAxon Markush project

Goals:

- Extend structural search capabilities to combinatorial Markush structures
- Markush enumeration

Complications:

- Practical examples may be very complex, methods using explicit enumeration may be impossible
- Extension of current molecular formats (generic features)

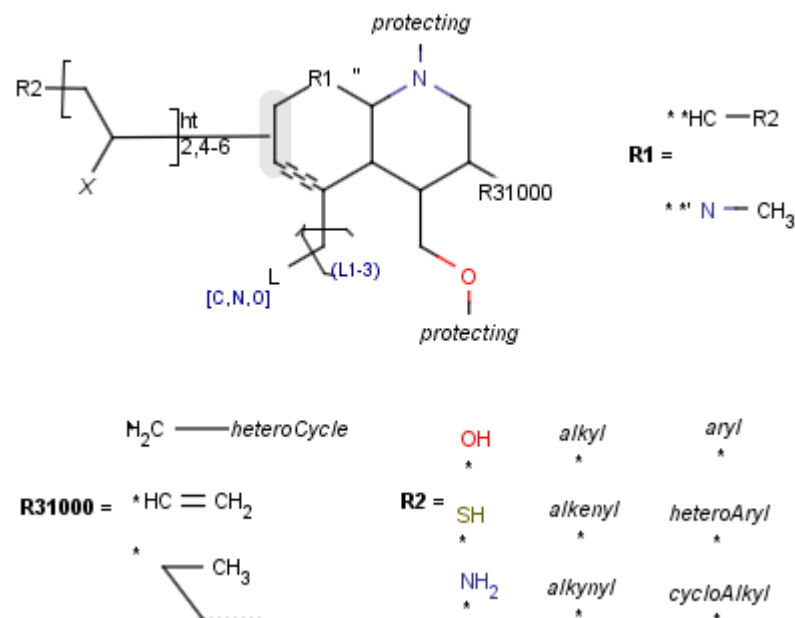
Timeline

- Pilot study started in 2005 Q4,
- First prototype shown at UGM, 2006 June
- Released in JChem 5.0, 2008
- Markush DARC format support 5.3.0 2010

Markush structures

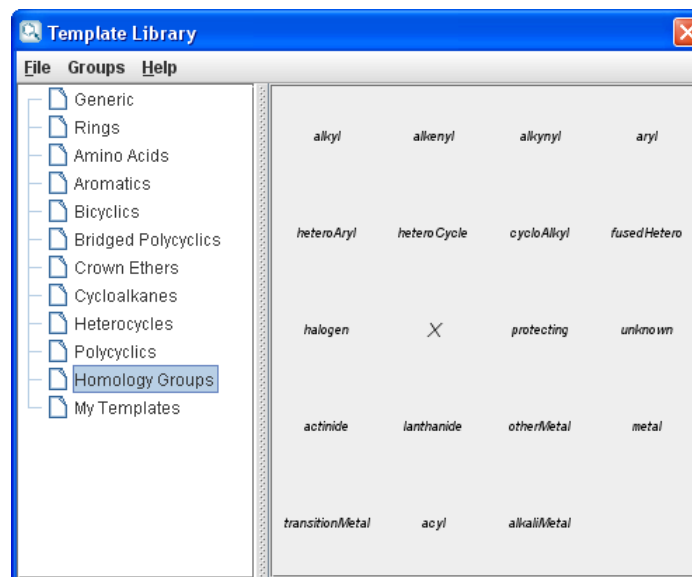
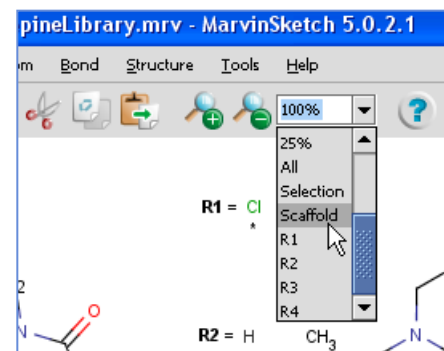
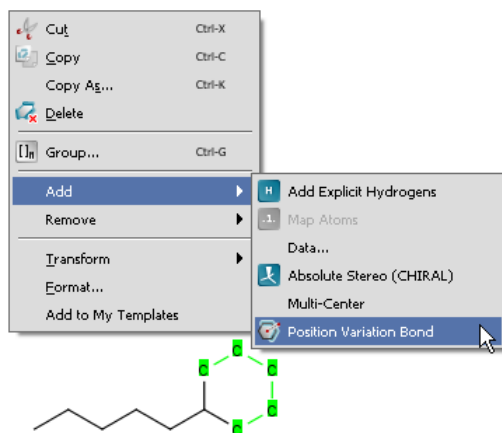
Current features handled:

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



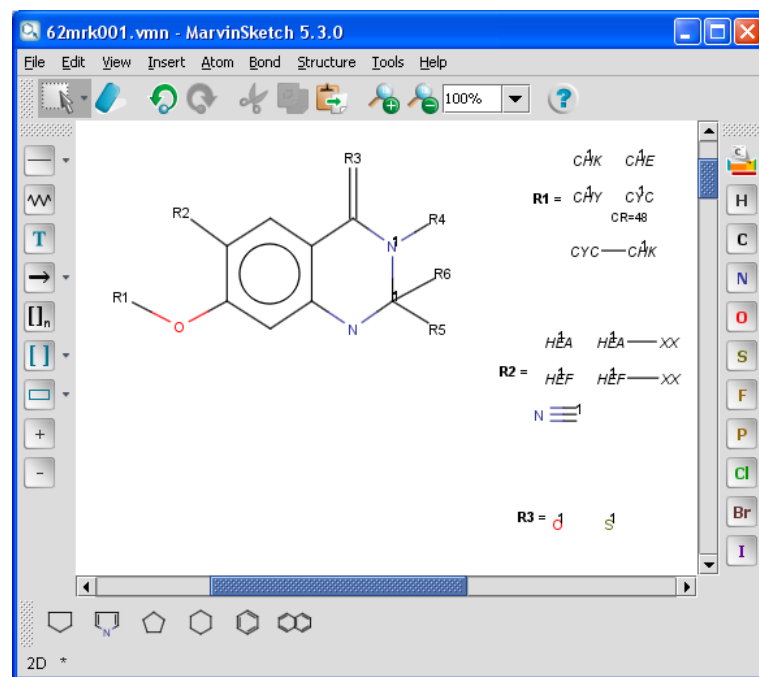
How to get Markush structures?

- Drawing – Marvin Sketch



How to get Markush structures?

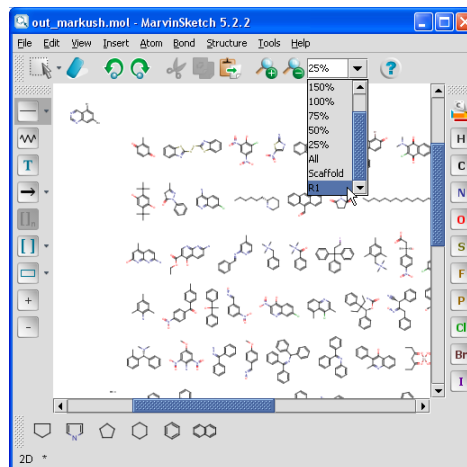
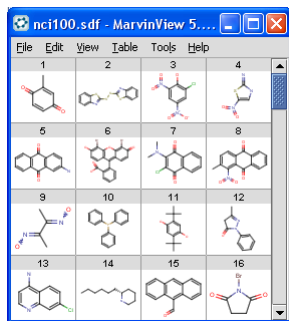
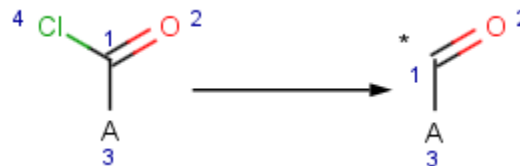
- Patent literature – Markush DARC format (*.vmn)
- Compatible with Thomson Reuters MMS patent Markush database (Test set available.)



How to get Markush structures?

Combinatorial chemistry – Reagent clipping

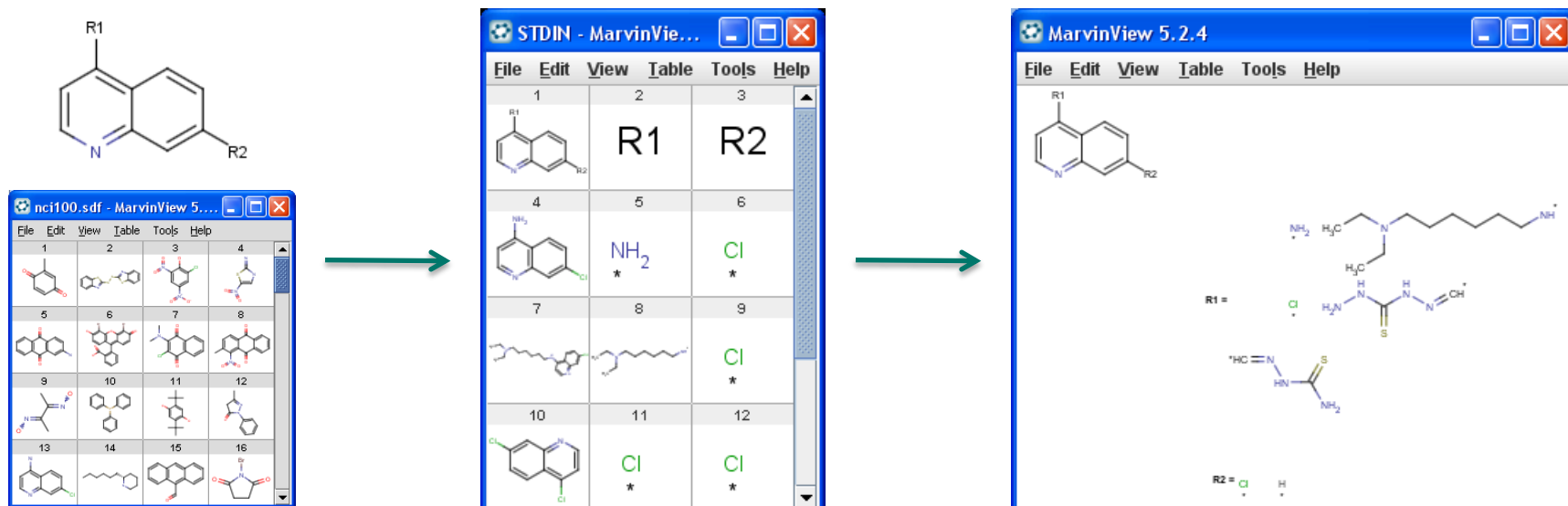
1. Replace reacting group with attachment point (Reactor tool)
2. Turn fragments to R-group definitions (Molconvert tool)
3. Add a scaffold (Molconvert tool)



How to get Markush structures?

Combinatorial chemistry – R-group decomposition

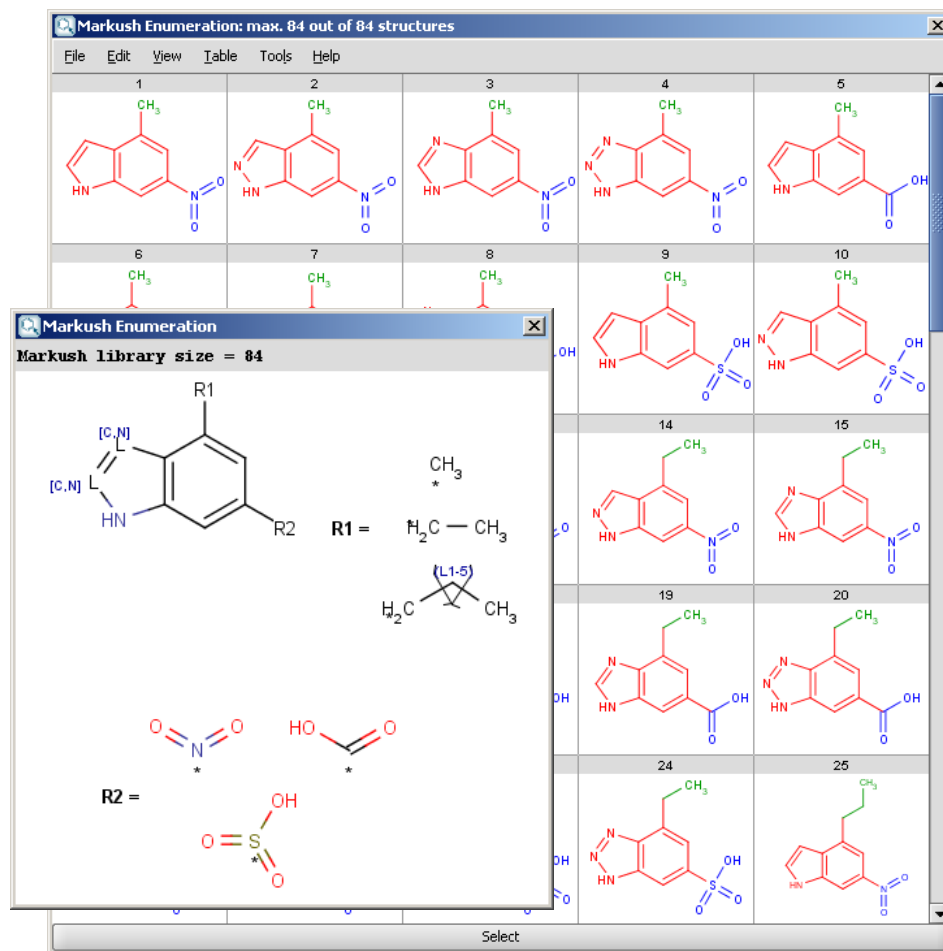
1. Filter and identify ligands in chemical library
2. Create Markush structure from R-table
(R-group decomposition tool)



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)

The screenshot displays the Instant JChem 2.5.2 interface. The main window shows a 'Grid view for Markush_table' with a table of search results. The table has columns for 'Cdid', 'Markush structure', 'Library Size', and 'Library size (calculated)'. The results are as follows:

Cdid	Markush structure	Library Size	Library size (calculated)
1		2,735,568.00	
2		2,445,552.00	2445552
3		5,934,096.00	5934096
4		7.63E09	7834640000

The interface also shows a 'Query - Grid view for Markush_table' window with a 'Substructure' search option and a chemical structure of a pyridine ring with an ethyl substituent. The status bar at the bottom indicates 'Markush_table: 5 out of 5 rows'.

- Substructure and Full structure search
- Broad translation of homology groups is supported. (Homology in DB, specific in query.)

Example web application

- Uses Marvin applets
- Markush search, insert, import, enumeration, coloring, Markush reduction integrated

The screenshot displays the JChem web application interface, which is a web-based chemical search tool. The main window shows search results for a Markush query. The search results are displayed in a grid format, with each cell containing a chemical structure and its ID. The structures are color-coded to highlight the Markush groups (R1 and R2). The search results are sorted by relevance, and the search took 1.525 seconds.

The interface includes a sidebar with navigation options: Page, Query, Retrieve All, Table, Insert, Modify, Delete, Import, Export, Help, and About. The ChemAxon logo is visible in the bottom left corner.

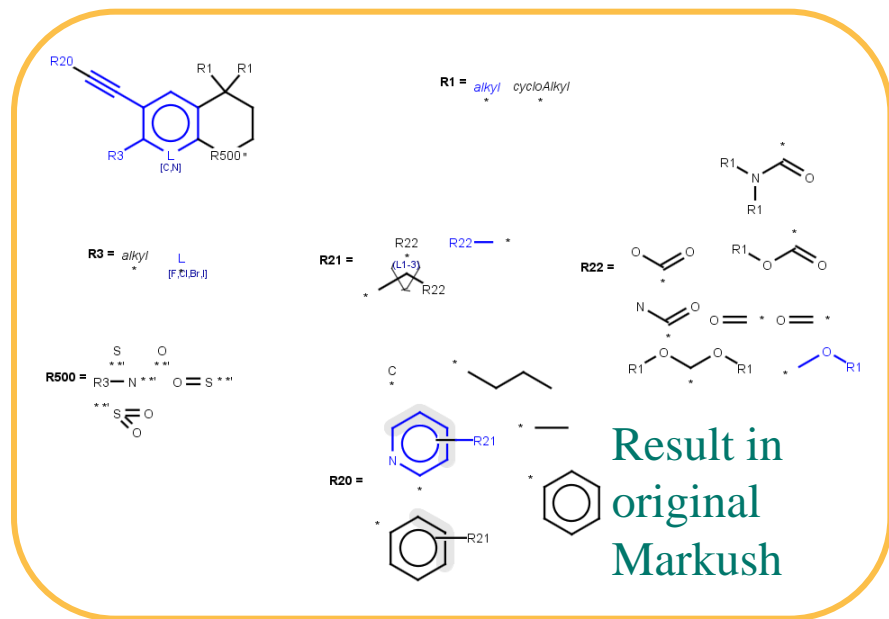
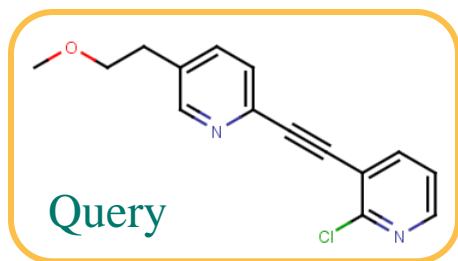
The search results are displayed in a table format with the following columns: ID, Structure, and Markush. The structures are color-coded to highlight the Markush groups (R1 and R2). The search results are sorted by relevance, and the search took 1.525 seconds.

The search parameters are displayed in a separate window, showing the query structure and the search options. The search type is set to Substructure, and the maximum number of hits is 200. The search time is 3 minutes. The search results are sorted by relevance, and the search took 1.525 seconds.

The search results are displayed in a table format with the following columns: ID, Structure, and Markush. The structures are color-coded to highlight the Markush groups (R1 and R2). The search results are sorted by relevance, and the search took 1.525 seconds.

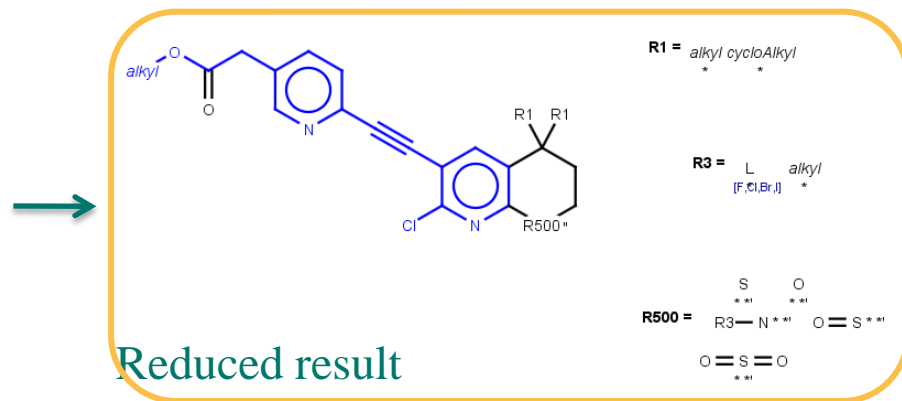
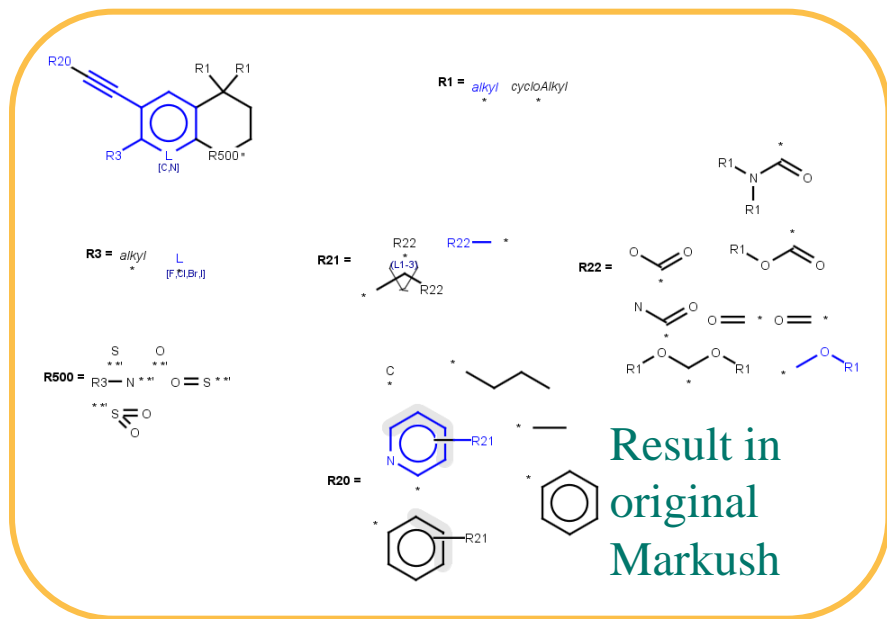
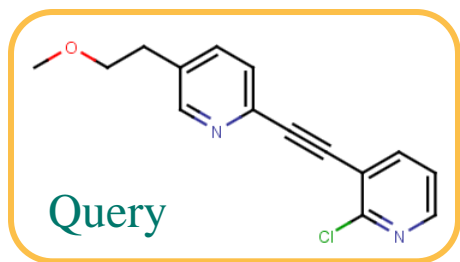
Markush storage & search

Substructure hit visualization



Markush storage & search

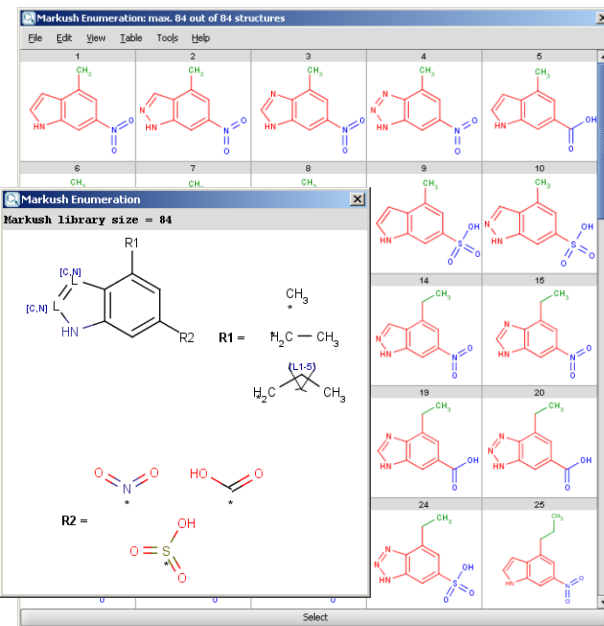
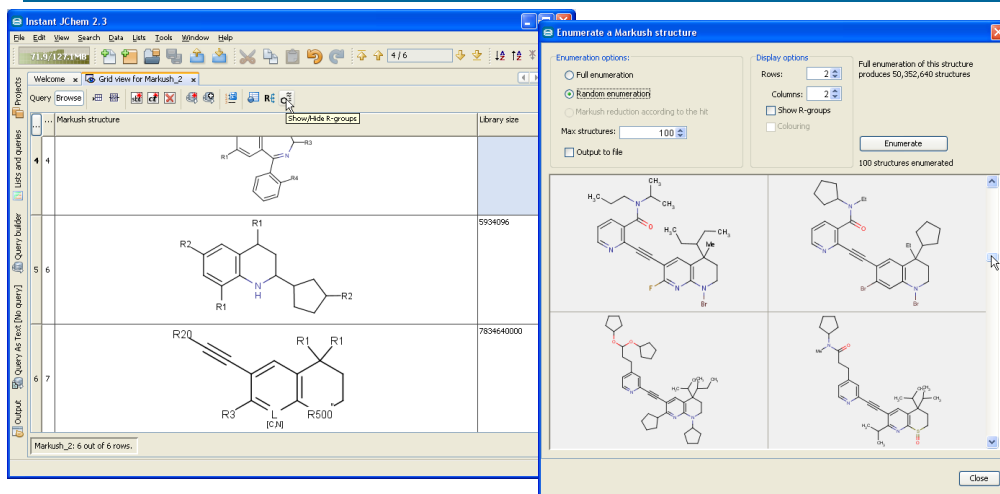
Substructure hit visualization:
„Markush structure reduction”



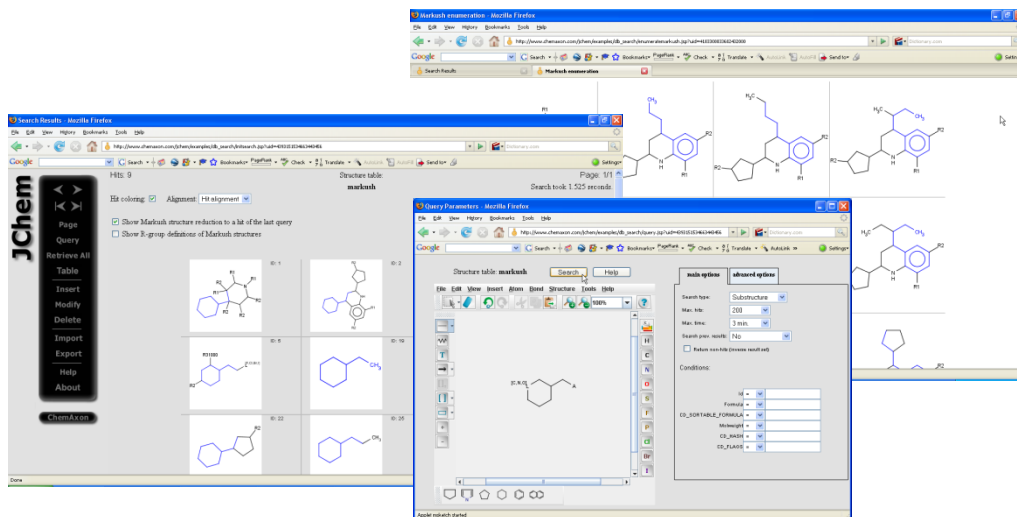
What's new

- **Support of Markush DARC (Thomson Reuters MMS patent Markush database) file format**
- Handling of multiple attachment points (>2)
- New homology definitions obtained from statistical analysis of a large chemical database.
- Instant JChem: batch converter vmn -> mrv

Demo



[Click here...](#)



Main use cases

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

Under development

- Further improvements in Markush DARC support:
 - Interpretation of homology groups – use Thomson Reuters definitions
 - Ring segment groups (XX form a ring)
 - New, more robust representation for attachment points
 - Homology properties (low alkyl, fused aryl, C1-3, N2-5, etc)
- Maximum common substructure search
- Biased enumeration and covering Markush – based on examples in patent.
- Improve search speed to handle larger Markush sets.
- Overlap analysis of Markush structures
- Conditions for Markush variables

Summary

- Markush structure storage, search and enumeration at ChemAxon now patent coverage
- Compatible patent data is available from Thomson Reuters
- Continuous development, improvements in the pipeline

Acknowledgements

- Development team: Nóra Máté, Róbert Wágner, Szilárd Dóránt, Tamás Csizmazia, Tim Dudgeon, Ali Baharev, Ferenc Csizmadia, et al.
- Tim Miller, Steve Hajkowski, Gez Cross and Linda Clark at Thomson Reuters for useful discussions, help and example Markush DARC 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.

Thank you for your attention!