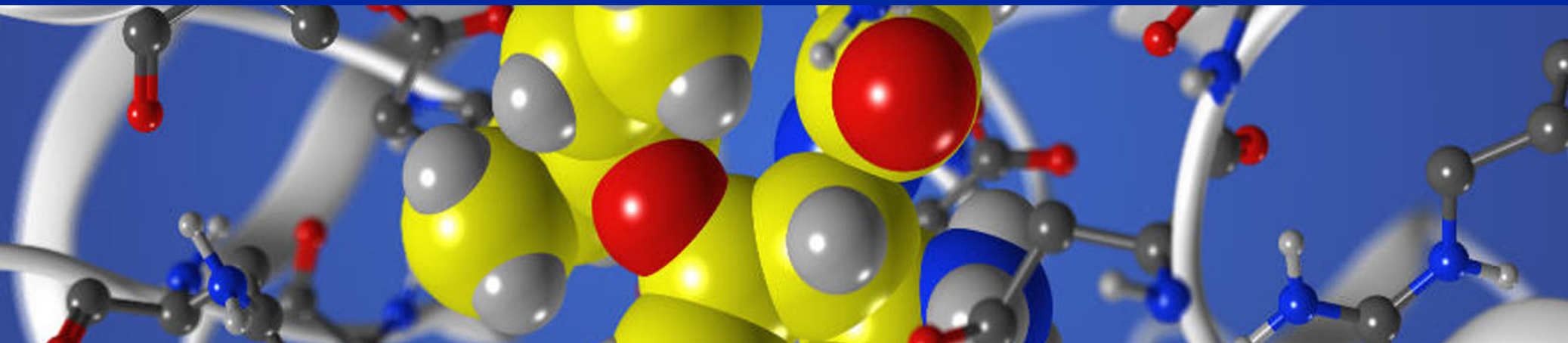

Analyzing Search Hits with ChemAxon's Markush Enumeration Tool

Dr. Guy de Weck - pRED Informatics



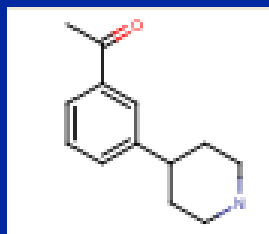
General workflow for a Novelty Search

ChemAxon's Markush Enumeration Tool

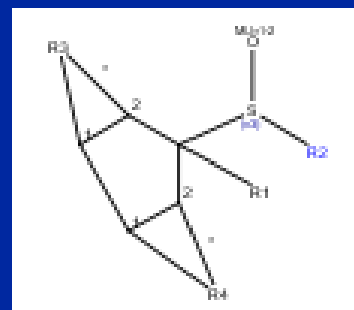
Suggested Improvements

General Workflow for a Novelty Search

- Idea for potentially new chemical structures
- Verifying patentability in various databases:
 - MMS, MARPAT, REGISTRY, Reaxys, ...
- Analyzing search hits
 - Tedious work: determining structural proximity of a searched substructure compared to the claimed structures



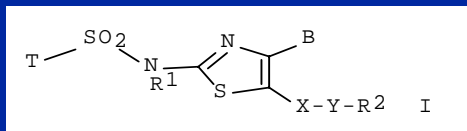
Searched substructure



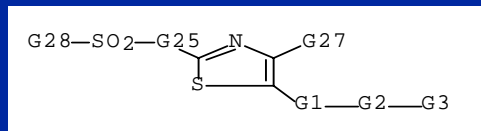
Markush structure

General Workflow for a Novelty Search

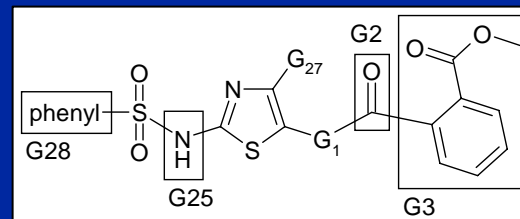
Graphic Image



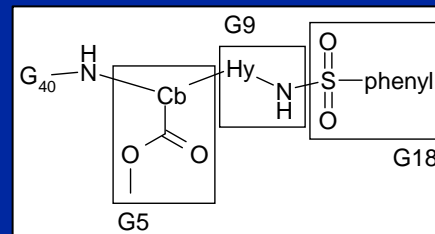
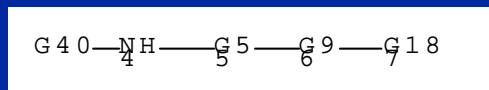
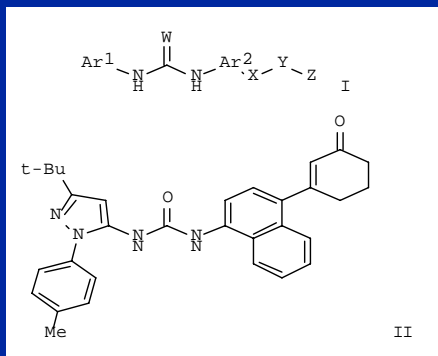
Markush structure



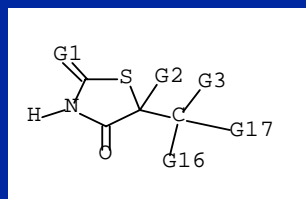
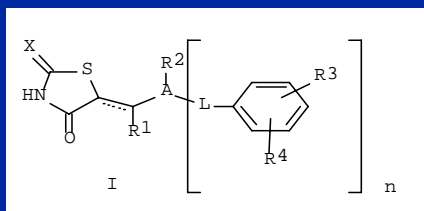
Structural proximity



close



medium

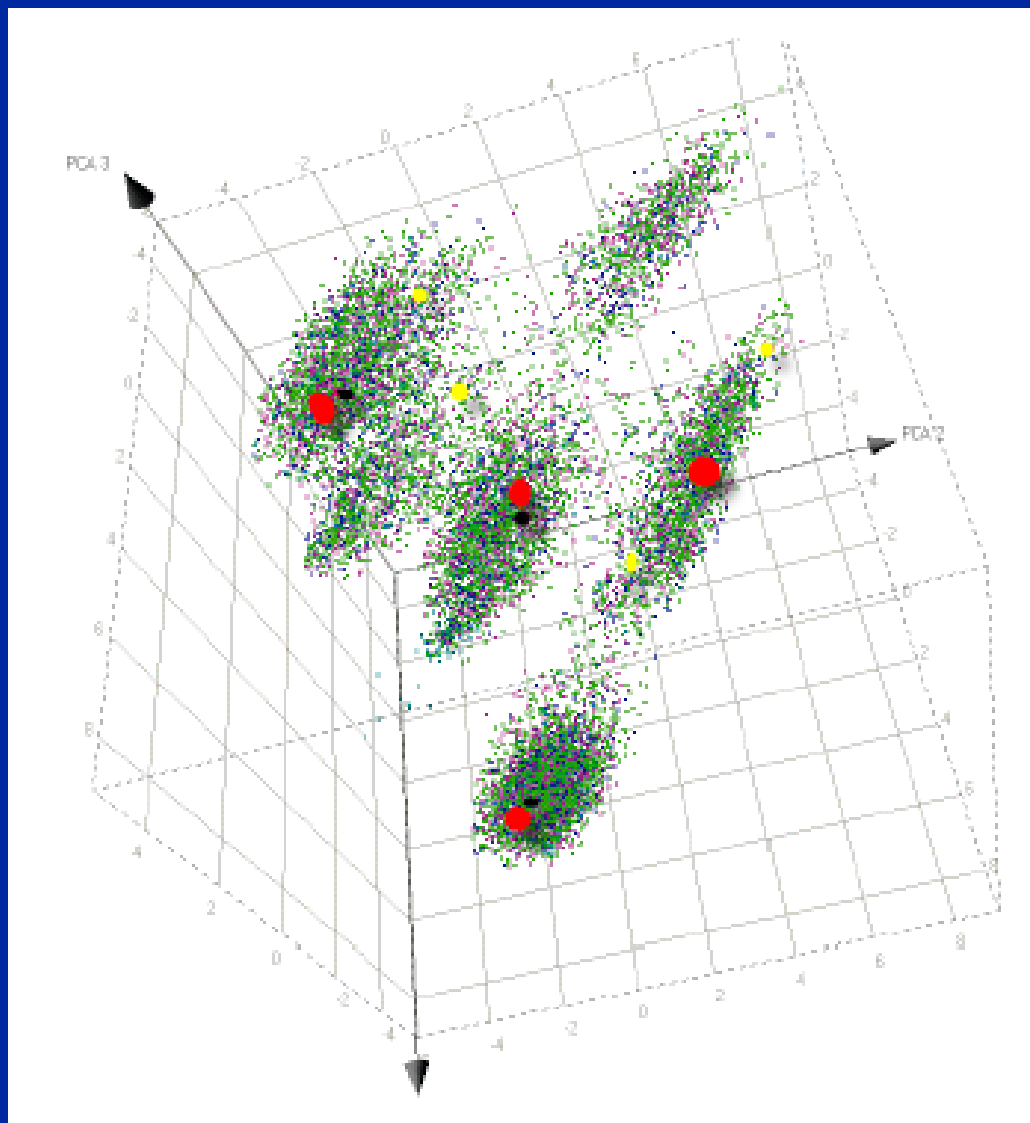


G₁₇ contains the searched substructure

distant

General Workflow for a Novelty Search

- A graphical representation of structural proximity:
 - Markush space density of a patent
 - Red: exemplified and prophetic structures
 - Yellow: distant structural proximity
 - Black: close structural proximity
- Markush space from decriPt, PCA of normalized fingerprints



General Workflow for a Novelty Search

ChemAxon's Markush Enumeration Tool

Suggested Improvements

ChemAxon's Markush Enumeration Tool

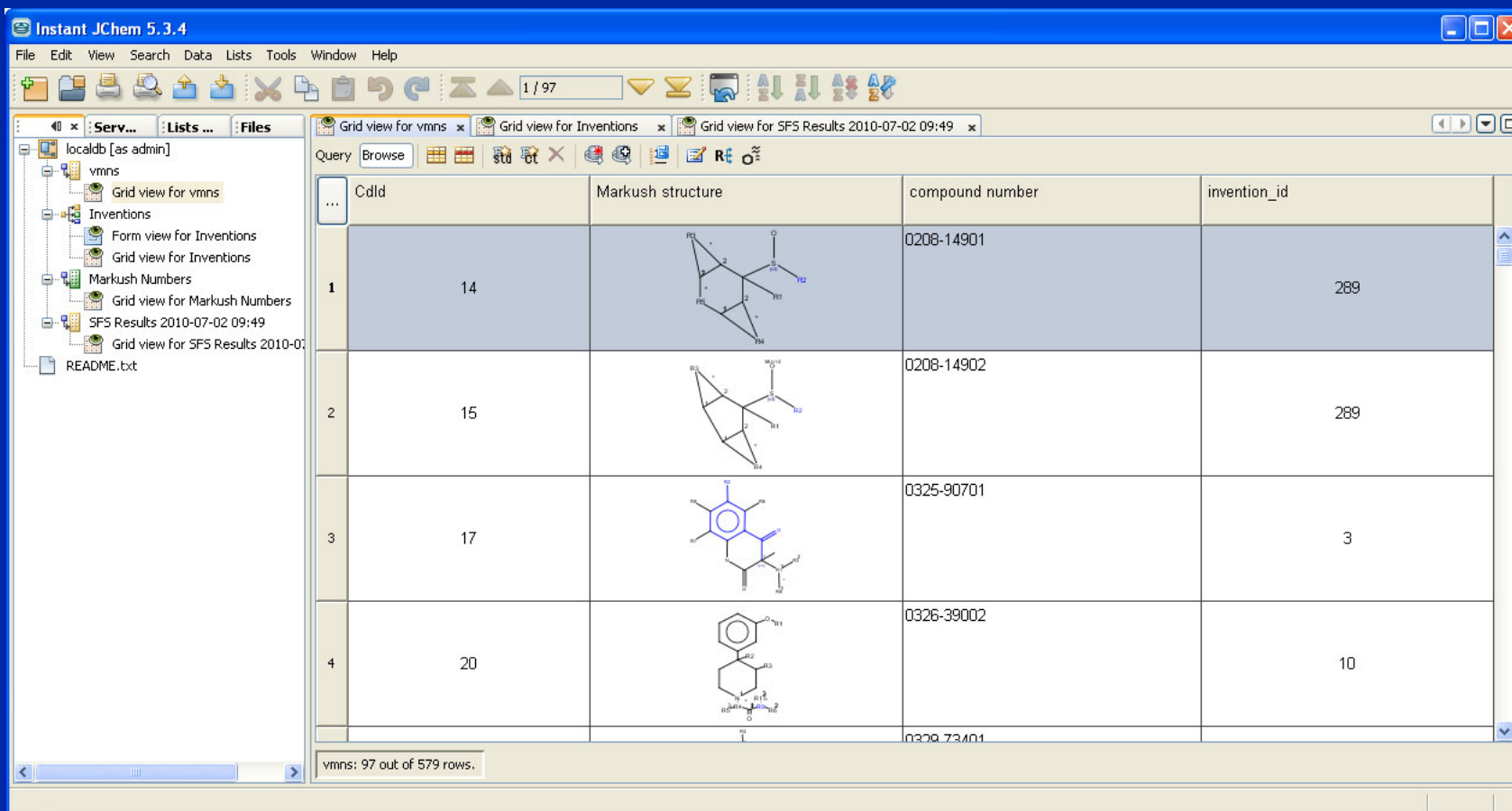
- How can ChemAxon's Markush enumeration tool help defining structural proximity?
- The 300 patents and their corresponding 579 Markush structures were loaded in Instant JChem and searched by substructure.

The screenshot displays the Instant JChem 5.3.4 application window. The interface includes a menu bar (File, Edit, View, Search, Data, Lists, Tools, Window, Help), a toolbar with various icons, and a status bar showing '1 / 579'. The main workspace is divided into several panes:

- Left Pane:** A file explorer showing a tree structure under 'localdb [as admin]'. The tree includes folders for 'vmns', 'Inventions', 'Markush Numbers', and 'SFS Results 2010-07-02 09:49', each with its own 'Grid view' icon. A 'README.txt' file is also visible.
- Top Right Pane:** A search interface with buttons for 'Query', 'Browse', 'Clear Query', 'Show All', and 'Run Query'. Below these buttons is a table with the following columns: 'CdId', 'Markush structure', and 'compound num'. The 'Markush structure' column contains a chemical structure of a benzamide derivative with a piperidine ring attached to the para position. Below the structure, the word 'Substructure' is displayed.

ChemAxon's Markush Enumeration Tool

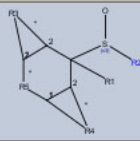
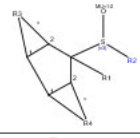
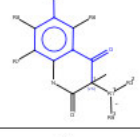
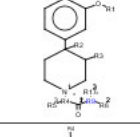
- This substructure search retrieved 97 Markush structures.



Instant JChem 5.3.4

File Edit View Search Data Lists Tools Window Help

Grid view for vmns x Grid view for Inventions x Grid view for SFS Results 2010-07-02 09:49 x

Cdid	Markush structure	compound number	invention_id
1		0208-14901	289
2		0208-14902	289
3		0325-90701	3
4		0326-39002	10

vmns: 97 out of 579 rows.

ChemAxon's Markush Enumeration Tool

- An example of close structural proximity using Markush reduction according to the hit:
- Full enumeration: 41310 structures.
- Markush reduction according to the hit: 8 structures.
- The highlighted part suggests close structural proximity of the searched substructures with the exemplified/prophetic structures.

Grid view for vmns x Grid view for Inventions x Grid view for SFS Results 2010-07-02 09:49 x

Query Browse

Cdid	Markush structure	compound number
...		
12	130	0327-06501

Enumerate a Markush structure

Enumeration options:

Full enumeration
 Random enumeration
 Markush reduction according to the hit

Max structures: 100

Output to file

Display options

Rows: 3

Columns: 3

Show R-groups
 Colouring

Enumerate

8 structures enumerated

Full enumeration of this structure produces 41310 structures

ChemAxon's Markush Enumeration Tool

- An example of distant structural proximity using Markush reduction according to the hit:
- Full enumeration: 10^{42} structures.
- Markush reduction according to the hit: 6 structures.
- The highlighted part suggests distant structural proximity of the searched substructures with the exemplified/prophetic structures.

The screenshot displays the ChemAxon Markush Enumeration Tool interface. At the top, there are three browser tabs: "Grid view for vmns", "Grid view for Inventions", and "Grid view for SFS Results 2010-07-02 09:49". Below the tabs is a toolbar with various icons, including "Query", "Browse", and "R".

The main interface is divided into two main sections. The top section is a table with the following columns: "Cdid", "Markush structure", and "compound number". The table contains one row with the following data:

Cdid	Markush structure	compound number
4		0326-39002

The bottom section is titled "Enumerate a Markush structure" and contains several options and controls:

- Enumeration options:**
 - Full enumeration
 - Random enumeration
 - Markush reduction according to the hit
 - Max structures:
 - Output to file
- Display options:**
 - Rows:
 - Columns:
 - Show R-groups
 - Colouring
- Enumerate button:** A button labeled "Enumerate" with the text "6 structures enumerated" below it.
- Results grid:** A 2x3 grid of chemical structures showing different variations of the original Markush structure, with various substituents highlighted in blue and red.

ChemAxon's Markush Enumeration Tool

- An example of medium structural proximity using Markush reduction according to the hit:
- Full enumeration:
 10^{16} structures.
- Markush reduction according to the hit:
2 structures.
- In this case subsequent partial enumeration steps are needed to define the structural proximity.

The screenshot shows the 'Enumerate a Markush structure' window. It features two main sections: 'Enumeration options' and 'Display options'. In the 'Enumeration options' section, 'Markush reduction according to the hit' is selected, and 'Max structures' is set to 100. In the 'Display options' section, 'Rows' is set to 3, 'Columns' is set to 4, and 'Colouring' is checked. A 'Full enumeration of this structure produces 4698509682954150 (~ 10¹⁶) structures' message is displayed. An 'Enumerate' button is present, and below it, it says '2 structures enumerated'. The bottom of the window shows a grid of chemical structures. The first two cells contain chemical structures with various R-groups (R1-R3, R6, R7, R8, R9) and a 'HET' label. The remaining cells are empty.

ChemAxon's Markush Enumeration Tool

- Summary:
 - Patent records having a distant structural proximity relative to the searched substructure are easily identified with ChemAxon's Markush enumeration tool. This is saving time in the novelty search workflow.
 - Patent records with a medium to close structural proximity relative to the searched substructure need to be analyzed in more detail:
 - Repeated partial enumeration by one or more R groups within the Markush enumeration tool is needed.

Novelty Search Workflow

ChemAxon's Markush Enumeration Tool

Suggested Improvements

Suggested Improvements

- Repeated partial enumeration by one or more R groups within the Markush enumeration tool.

Enumerate a Markush structure

Enumeration options:

- Full enumeration
- Random enumeration
- Markush reduction according to the hit

Max structures: 100

Output to file

Display options:

Rows: 3

Columns: 4

Show R-groups

Colouring

Enumerate

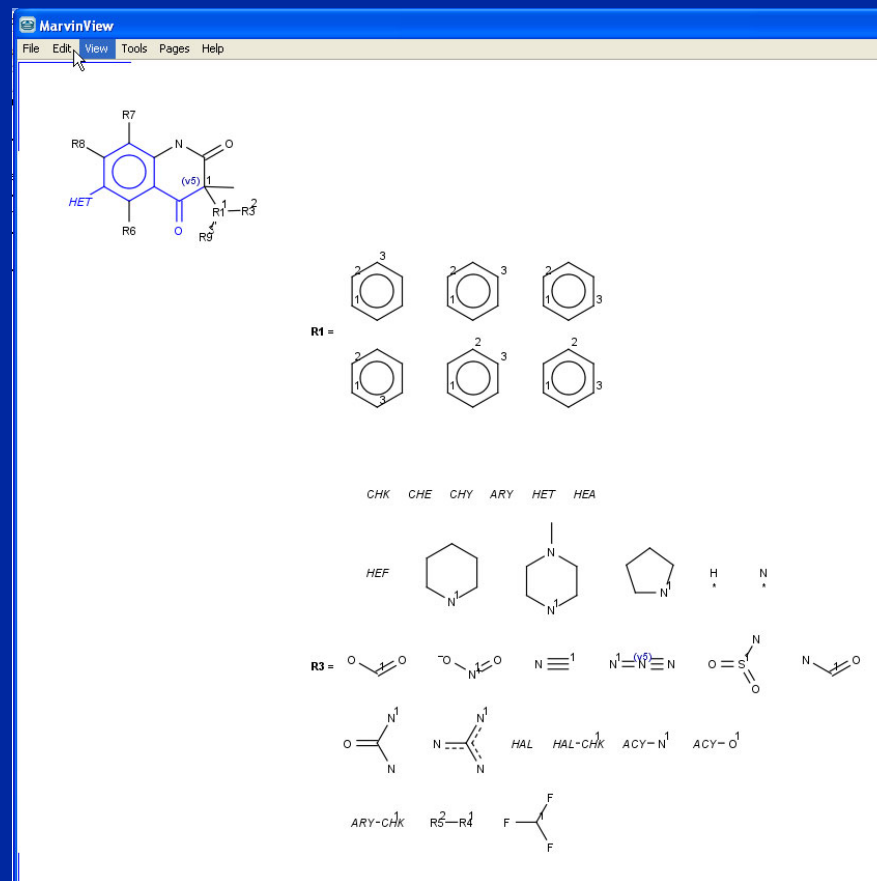
2 structures enumerated

Full enumeration of this structure produces 4698509682954150 ($\sim 10^{16}$) structures

- Possibility of selecting one or more R-groups and enumerating these R-groups.

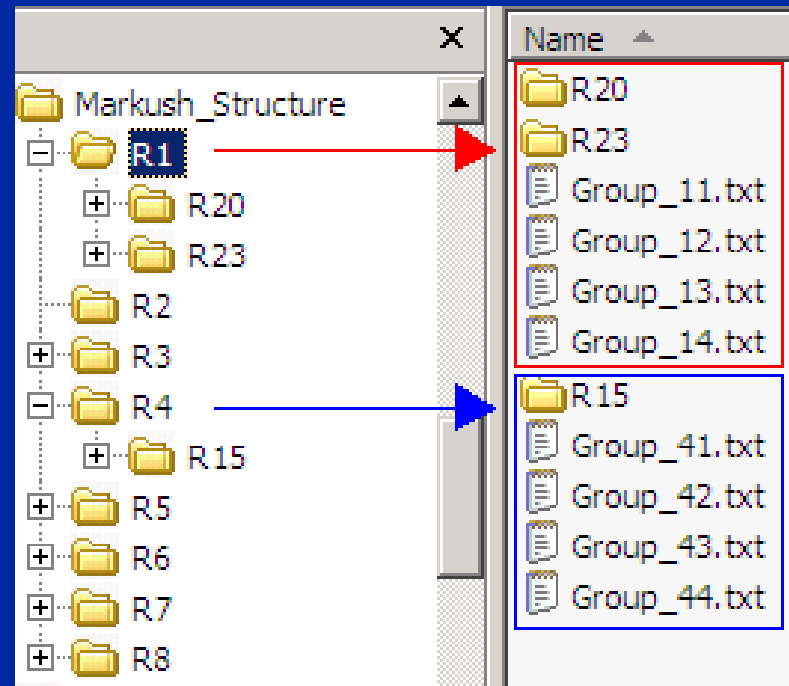
Suggested Improvements

- Display of enumerated structures when “Show R-groups” is selected:
 - The (partially) enumerated structure must be visible at all times
 - Only selected R-groups should be enumerated.



Suggested Improvements

- Explorer like display of enumerated structures: (similar to the Markush display in Torus from digital Chemistry)
- Multiple selection of R-groups
- Additional suggestions:
 - Displaying search and filtering steps: action thread on top and results below.



Suggested Improvements

- Display of exemplified / prophetic structures.
 - Substructure search on the set of exemplified/prophetic structures of the corresponding patent.
 - Highlighting of the searched substructure

The screenshot shows the Instant JChem 5.3.7 interface. The main window displays a 'Markush structure' of a piperazine derivative with various substituents labeled R1 through R9. Below this, there is a 'Patents' list and a 'Description' of the patent. The 'examples' table is as follows:

Structure	Mol Weight	Formula	PREF_NAME	SYSTEMATIC	SYNONYM	RIN	CHEMISTRY
	466.21	C20H22Br2N2O		1-((1R,2R)-2-(3,5-Dibromo-benzyloxy)-indan-1-yl) piperazine		01391	
	482.21	C20H22Br2N2O2		1-((3R,4S)-3-(3,5-Dibromo-benzyloxy)-1-benzopyran-4-yl) piperazine			
		C21H24Br2N2		1-((1R,2R)-2-(3,5-			

(Screen shot provided by ChemAxon)

Suggested Improvements

- Running novelty searches in MMS out of ChemAxon's Instant JChem.
- Federated novelty search over all relevant sources?

Acknowledgements

Thomson Reuters: Steve Hajkowski

ChemAxon: Szabolcs Csepregi, Alex Allardyce, Douglas Drake

Roche: Torsten Schindler



We Innovate Healthcare