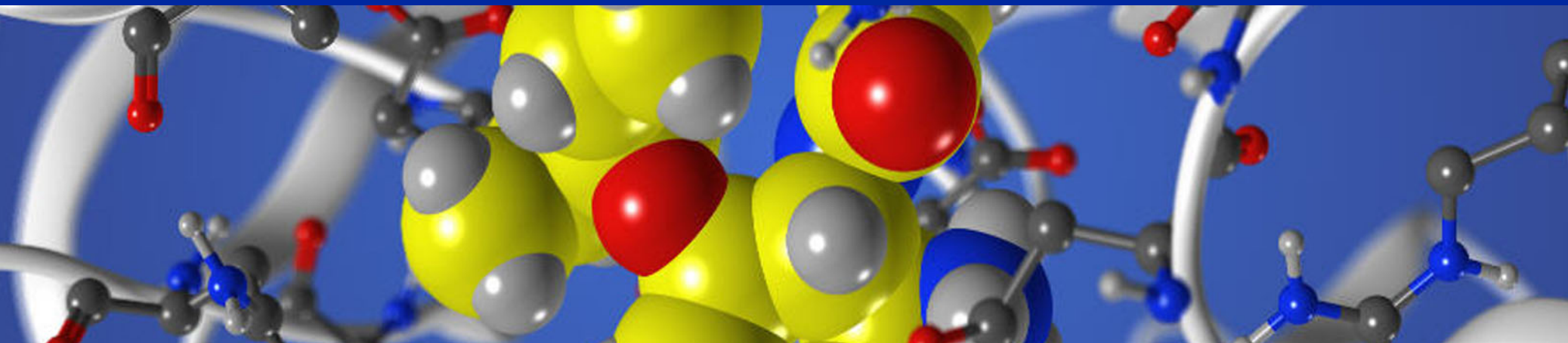


# Analyzing Search Hits with ChemAxon's Markush Enumeration Tool

*Dr. Guy de Weck - pRED Informatics*



---

# General workflow for a Novelty Search

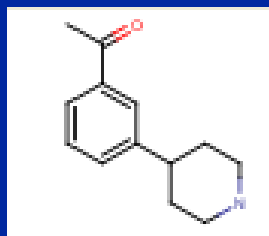
---

ChemAxon's Markush Enumeration Tool

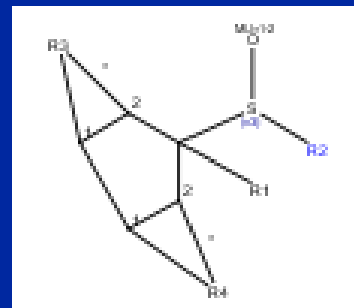
Suggested Improvements

# 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



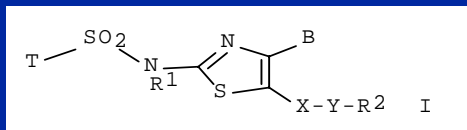
Searched substructure



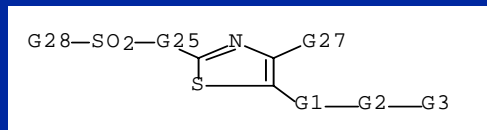
Markush structure

# General Workflow for a Novelty Search

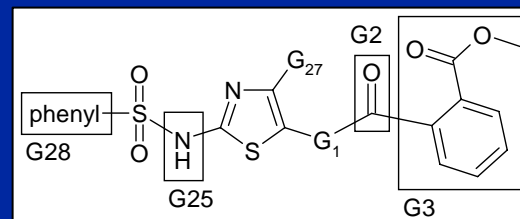
## Graphic Image



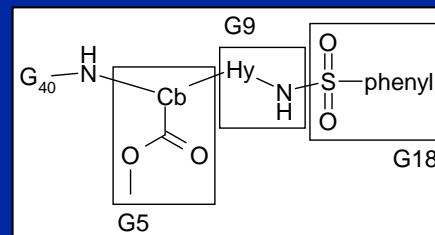
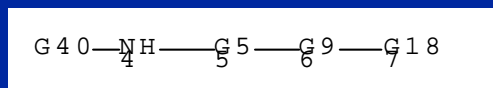
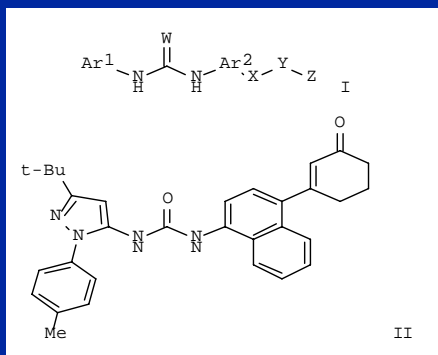
## Markush structure



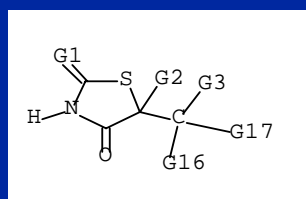
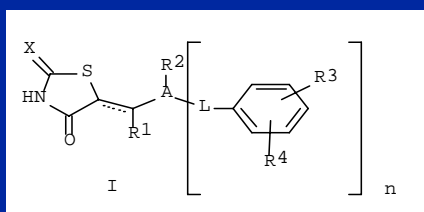
## Structural proximity



close



medium



G<sub>17</sub> contains the searched substructure

distant

# General Workflow for a Novelty Search

**MSTR 1**

$$G1-G3-G15-CH-G16$$

$G22$   $G23$   
 $G15-CH-G16$

G1 = aryl (SO) / alkyl (SR (1-) G2) /  
heteroaryl<EC (0-) N (0-) S (0-) O> (SO) /  
Hy<EC (0-) N (0-) O (0-) S, AR (0)> (SO) / (SC 172)

1924-1935

G2 = aryl (SO) / heteroaryl<EC (0-) N (0-) S (0-) O>  
(SO) / R

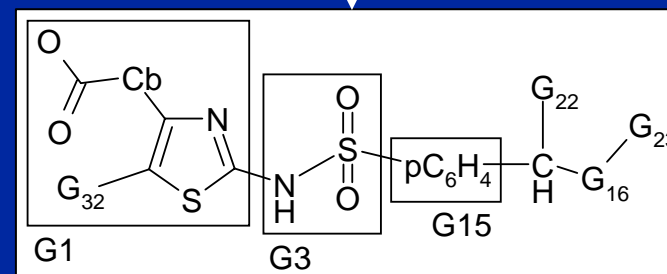
G3 = 5-1 8-3 / 16-1 14-3 / 22-1 20-3 / 24-1 26-3 /  
28-1 31-3 / 38-1 39-3 / 41-1 43-3 / 47-1 49-3 / 54-1 57-3 /  
60-1 65-3 / 66-1 70-3 / 74-1 77-3

$$\begin{matrix} G4 \\ | \\ G5-C(O)-G5 \end{matrix} \quad \begin{matrix} G4 \\ | \\ G7-G8-N-C(O)-G6 \end{matrix} \quad \begin{matrix} G4 \\ | \\ G7-G8-C(O)-N-G6 \end{matrix} \quad \begin{matrix} G4 \\ | \\ G(O)-N-G6 \end{matrix}$$

$$\begin{matrix} G4 \\ | \\ G9-G10-G9 \end{matrix} \quad \begin{matrix} G4 \\ | \\ G3-N-G6 \end{matrix} \quad \begin{matrix} G4 & G11 & G4 \\ | & | & | \\ G4-N & -C- & N-G4 \end{matrix} \quad \begin{matrix} G12 & G4 \\ | & | \\ G4-N & =C- & N-G4 \end{matrix} \quad \begin{matrix} G4 & G4 \\ | & | \\ G4-N & -G14-N-G(O) \end{matrix}$$

Handwritten diagram illustrating the mapping of variables G1 through G36 to a specific chemical structure. The diagram shows the main structure  $G1-G3-G15-CH-G16$  with various substituents and their corresponding variable assignments. Key annotations include:

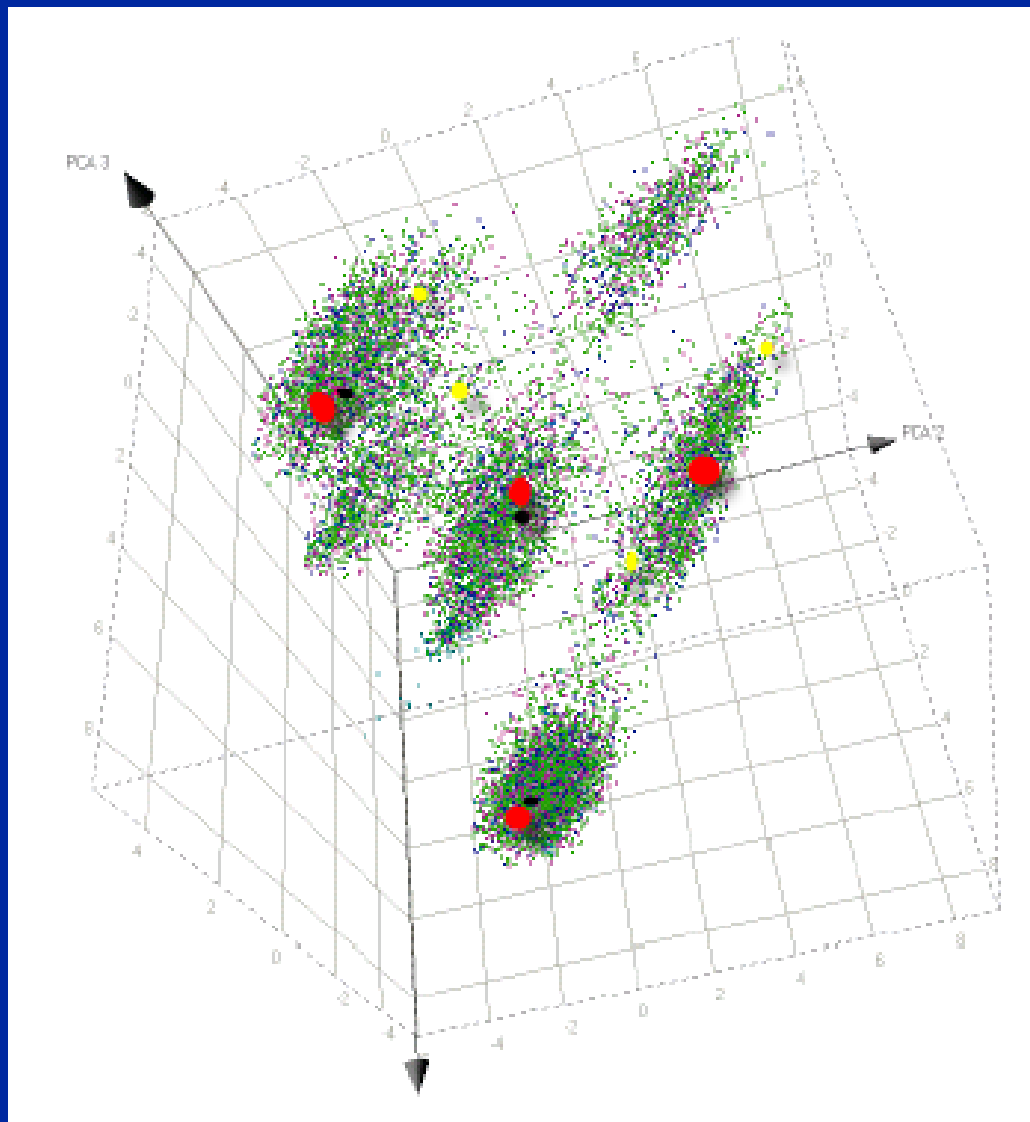
- $G10 = N-SO_2$  (with  $S^1$  and  $O^2$  labels)
- $G25 = C$
- $G34 \rightarrow 180$  (pointing to a carbonyl group)
- $G36$  (pointing to a hydroxyl group)
- $G24 =$  (pointing to a ring system)
- $G32 =$  (pointing to a ring system)
- $G4$  (pointing to a nitrogen atom)
- $P-C_6H_4$  (pointing to a para-substituted benzene ring)



- Defining structural proximity is an extremely time consuming process!

# 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.

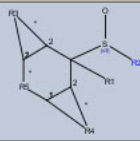
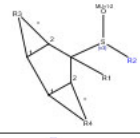
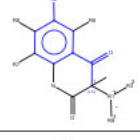
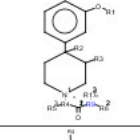
The screenshot displays the Instant JChem 5.3.4 application window. The interface includes a menu bar (File, Edit, View, Search, Data, Lists, Tools, Window, Help), a toolbar with various icons, and a status bar showing '1 / 579'. The main workspace is divided into several panes:

- Left Pane:** A tree view showing the local database structure under 'localdb [as admin]'. It includes folders for 'vmns', 'Inventions', 'Markush Numbers', and 'SFS Results 2010-07-02 09:49', each with its own grid view.
- Top Right Pane:** A search control area with buttons for 'Query', 'Browse', 'Clear Query', 'Show All', and 'Run Query'. Below these are tabs for 'Grid view for vmns', 'Grid view for Inventions', and 'Grid view for SFS Results 2010-07-02 09:49'.
- Main Table:** A table with columns 'CdId', 'Markush structure', and 'compound num'. The first row contains a chemical structure of a benzamide derivative (SMILES: CC(=O)c1ccc(cc1)C2CCNCC2) and the label 'Substructure' below it.

# ChemAxon's Markush Enumeration Tool

- This substructure search retrieved 97 Markush structures.

The screenshot shows the Instant JChem 5.3.4 application window. The main area displays a grid view of Markush structures. The table below summarizes the data shown in the grid:

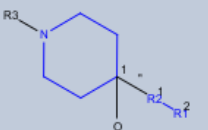
Cdid	Markush structure	compound number	invention_id
1		0208-14901	289
2		0208-14902	289
3		0325-90701	3
4		0326-39002	10

At the bottom of the window, a status bar indicates: 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

...	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

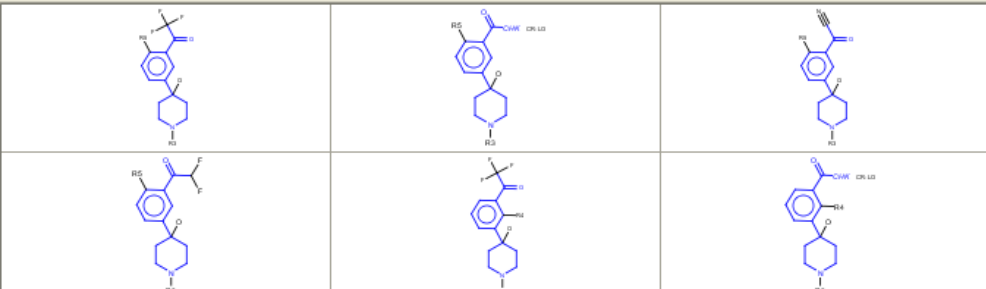
Show R-groups

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 for file operations and search.

The main interface is divided into several sections:

- Table:** A table with columns "Cdid", "Markush structure", and "compound number". The first row shows "4" in the "Cdid" column, a chemical structure in the "Markush structure" column, and "0326-39002" in the "compound number" column.
- Enumerate a Markush structure:** A panel with "Enumeration options" and "Display options".
  - 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
  - Full enumeration of this structure produces  $\sim 10^{42}$  structures
  - 
  - 6 structures enumerated
- Results:** A grid of six chemical structures, each representing a different enumeration result. The structures are variations of the original Markush structure, with different R-groups and functional groups highlighted in blue and red.

# ChemAxon's Markush Enumeration Tool

- An example of medium structural proximity using Markush reduction according to the hit:

Enumerate a Markush structure

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

Full enumeration of this structure produces 4698509682954150 (~ 10<sup>16</sup>) structures

Enumerate

2 structures enumerated

Chemical structures shown in the grid:

- Structure 1: A benzene ring with substituents R7, R8, R6, and R9. A nitrogen atom is attached to the ring, and a carbonyl group is attached to the nitrogen. A central carbon atom is bonded to R1, R2, and R3, and also to the nitrogen and a carbonyl group. A label (vi) is near the central carbon.
- Structure 2: A benzene ring with substituents R2, R8, R6, and R9. A nitrogen atom is attached to the ring, and a carbonyl group is attached to the nitrogen. A central carbon atom is bonded to R1, R2, and R3, and also to the nitrogen and a carbonyl group. A label (vi) is near the central carbon.

- Full enumeration:  
10<sup>16</sup> structures.

- Markush reduction  
according to the hit:  
2 structures.

- In this case subsequent partial enumeration steps are needed to define the structural proximity.

# 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

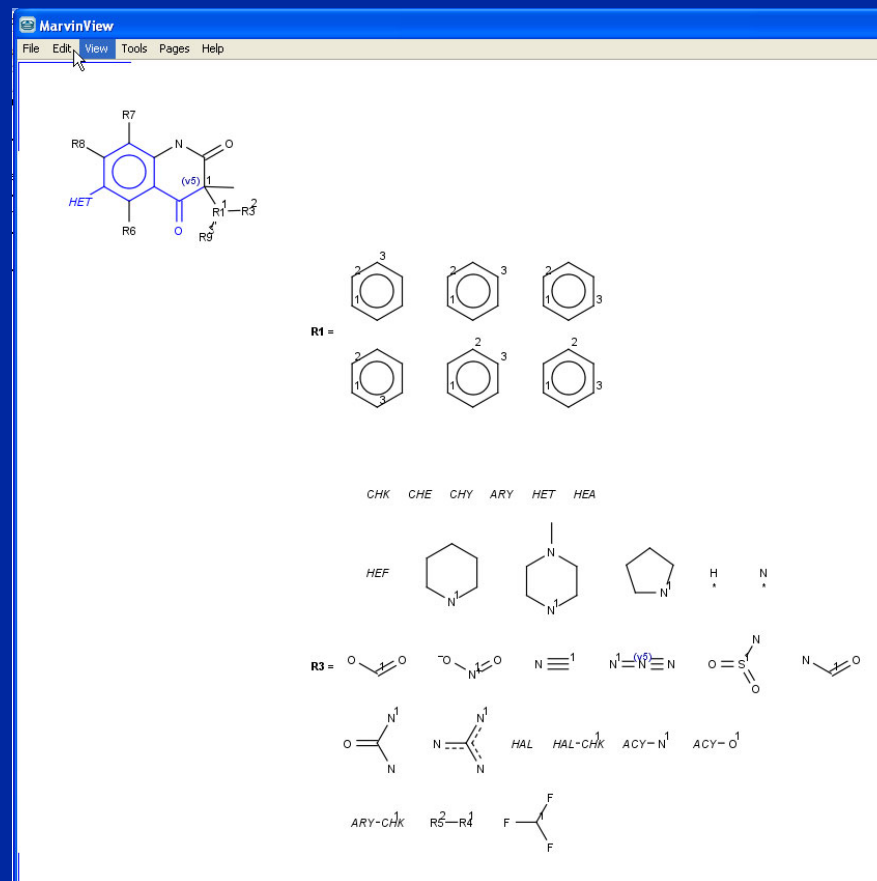
2 structures enumerated

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

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

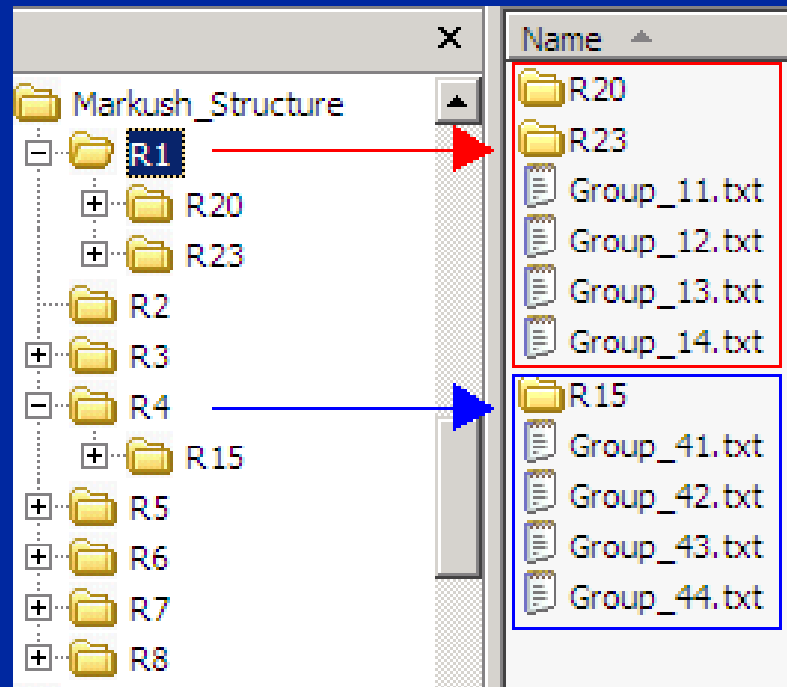
# Suggested Improvements

- 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 chemical structure with various substituents labeled R1 through R9. Below the structure, there is a 'Patents' section listing several patent numbers and a 'Description' section providing a detailed chemical description of the piperazine derivatives. At the bottom, there is a table of 'examples' with the following data:

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