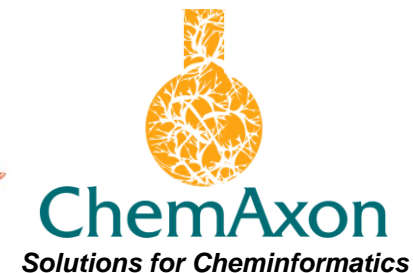




The year of chemistry with JChem - Formula search, unique structures, Markush, speedups and more

Szabolcs Csepregi

Sept 2011, US User Group Meeting



Contents

- ChemAxon chemical database tools
- Quick intro to JChem back end
 - JChem Base
 - Cartridge
 - Web Services
 - Markush search
- Recent developments, plans

The JChem family – back & front end

Visualization



Marvin

Structure, query & reaction editor, viewer & visualization

Property Prediction



Calculator Plugins

Structure property prediction & calculation

Selected calculations listing

- pKa, Major microspecies
- logP, logD
- Charge
- Tautomerization
- Stereoisomer
- Conformation and 3D alignment
- Topology Analysis
- Molecular Surface Area
- Markush Enumeration
- Hydrogen bond donor/ acceptor
- Structural Frameworks
- Structure to Name
- etc...

Chemical DB – toolkit



JChem Cartridge

JChem/Oracle integration



JChem Base

Structure searching & db access



Standardizer

Chemical business rules processing



Structure checker

Batch structure file validation and correction

Chemical DB – desktop



Instant JChem

Structure db management, search & prediction



JChem for Excel

Enabling chemistry in Excel

Nomenclature



Name to Structure

Import & search chemical names

Enumeration



Reactor

Enumeration via reaction modelling

Library analysis



JKluster

Clustering & diversity analysis



Fragmenter

Decomposition to fragments and R-groups



Screen

HT pharmacophore screening

Add-ons



Markush Search

Store & search Markush structures



JChem Webservice

Web services integration interface

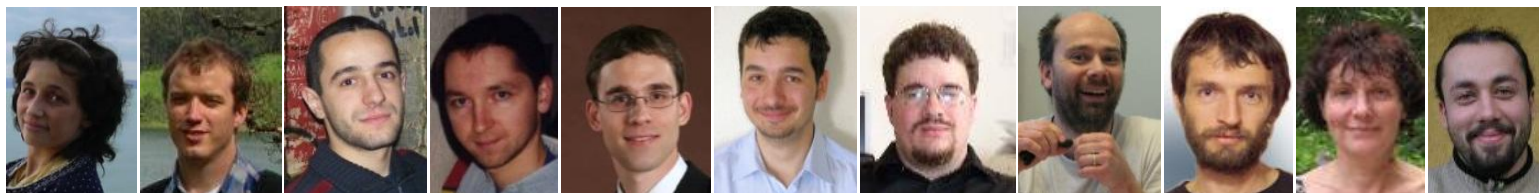
JChem for SharePoint

Live chemistry, search and structure based properties

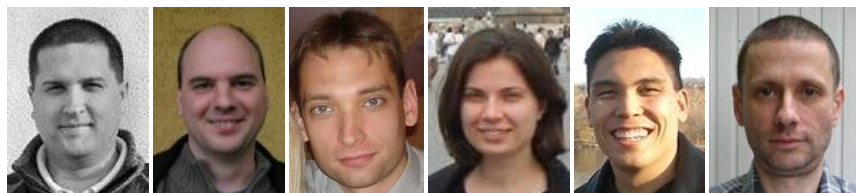
New

Development team – back end

JChem Base and Markush



JChem Web Services and Cartridge



Helpers

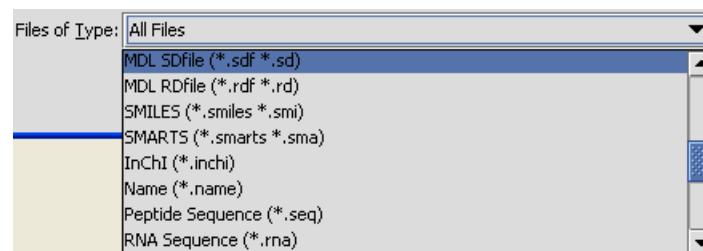


Compatibility and interfaces

Chemical file formats:

- MDL mol/rxn/sdf/rdf (v2.0 and v3.0), SKC, CDX, CDXML
- Smiles
- CML, MRV (Marvin)
- IUPAC and traditional names
- InChI, mol2, PDB, etc.

Markush DARC



Database engines:

- **Oracle**, **MySQL**, MS SQL Server, MS Access, PostgreSQL, IBM DB2, **Derby**, **Composite**, etc.

All operating systems:

- Java API (JChem Base)
- .NET API (JChem Base)
- SQL (Cartridge)
- JChem Web Services

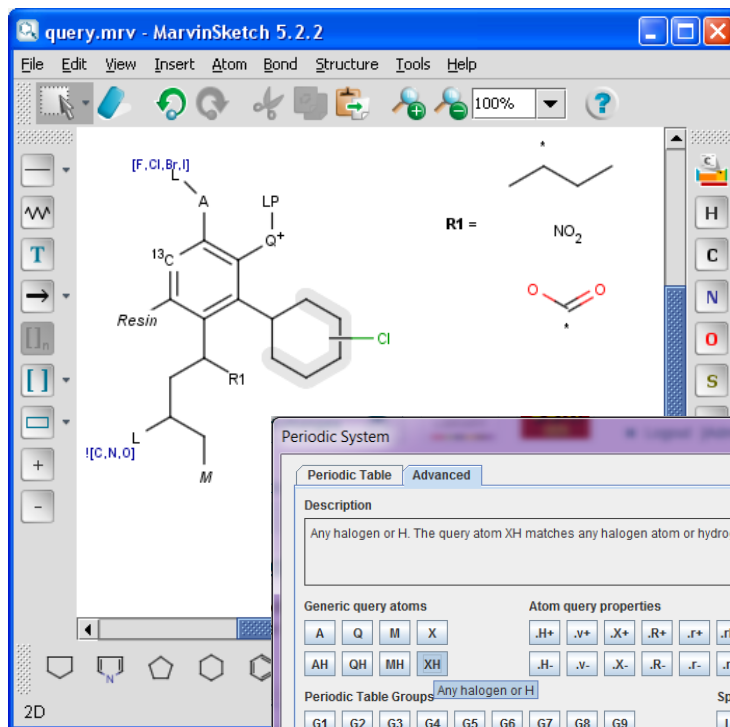


ORACLE



Structure search: features

- Substructure, Similarity, Full, etc. search types
- Wide range of query atoms
- Reaction search features
- Polymers
- Position variation
- Hit coloring
- R-groups
- Homology groups with properties **New**
- Tautomers
- SMARTS
- ...



The screenshot shows the Periodic System dialog box in MarvinSketch. The dialog has two tabs: Periodic Table and Advanced. The Advanced tab is selected. The Description field contains the text: "Any halogen or H. The query atom XH matches any halogen atom or hydrogen." The dialog includes several sections for defining query properties:

- Generic query atoms:** A grid of buttons for A, Q, M, X, AH, QH, MH, and XH. The XH button is highlighted.
- Atom query properties:** A grid of buttons for .H+, .v+, .X+, .R+, .r+, .rb+, .s+, .h+, .D+, .u, .H-, .v-, .X-, .R-, .r-, .rb-, .s-, .h-, .D-, and .a/A.
- Periodic Table Groups:** A grid of buttons for G1 through G18. The text "Any halogen or H" is displayed above the grid.
- Special nodes:** Buttons for LP, Pol, and a dropdown menu currently showing "alkyl".
- R-groups:** A grid of buttons for R1 through R32.
- Custom Property:** Buttons for Type: R-group, Alias, Pseudo, SMARTS, and Value. The Value field is empty.

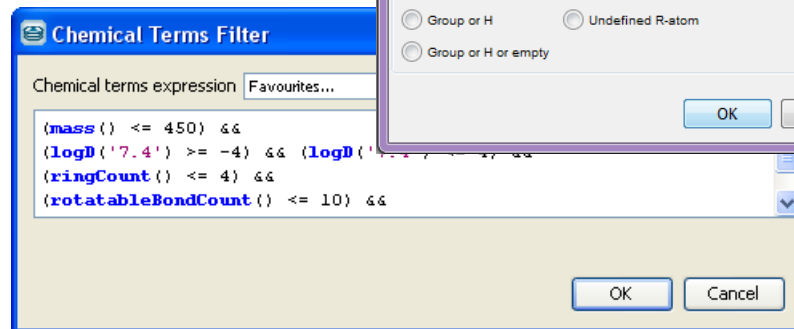
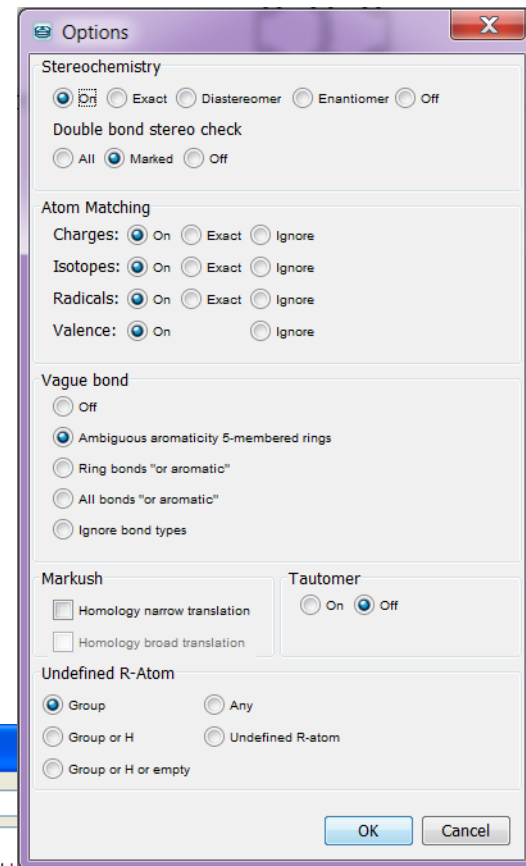
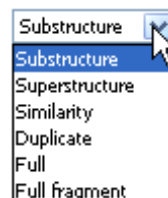
A Close button is located at the bottom right of the dialog.

www.chemaxon.com/conf/Structural_Search.ppt

Structure search: options

Some selected structure search options:

- Stereo on/off
- Ignore charge/isotope/radical/valence/polymers, etc.
- Vague bond matching options
- Chemical Terms filter
- Tautomer search
- Inverse hit list
- Maximum search time / number of hits
- Combine with non-structure conditions
- Ordering of results
- etc.

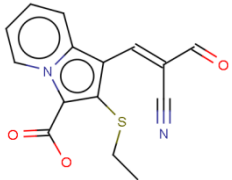
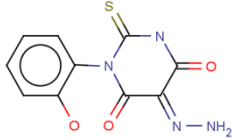
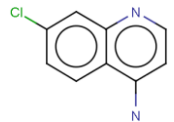
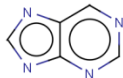


Structure search: performance

Compound registration:

Number of compounds	Elapsed time	
	Duplicates not checked	Duplicates checked
10,000	21 s	25 s
100,000	2 min 4 s	2 min 34 s
200,000	4 min 24 s	5 min 13 s

Substructure search in PubChem (19.5 million compounds):

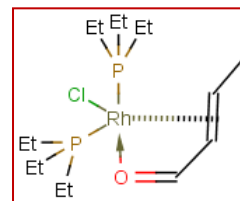
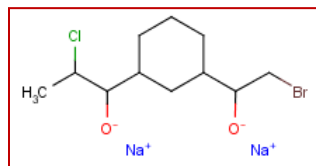
Query	Number of hits	Search time
	2	0.91 s
	93	0.98 s
	6,001	1.30 s
	146,256	5.66 s

JChem Base 5.2.2,
Intel Quad Q6600 2.4GHz,
8GB RAM; Oracle 10.2.0.3

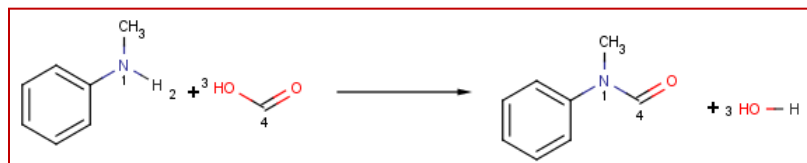
Table types

Control allowed chemical structures and available operations

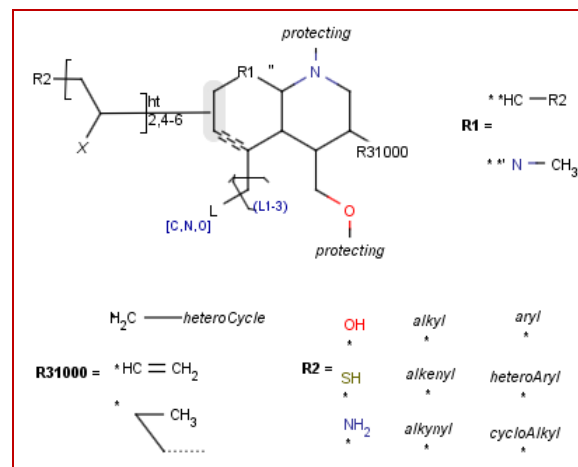
- Molecule



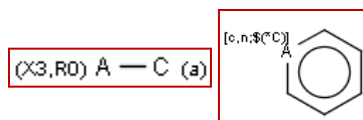
- Reaction



- Markush



- Query



- Any structure

Example web applications

Java Server Pages (JSP) example

- Marvin applets are used for query drawing and structure visualization

The screenshot displays the ChemAxon web services interface. On the left, the 'Query Parameters' window shows search settings: Search type: Substructure, Similarity type: Chemical Hashed Fingerprint, Screening config: Default, Dissimilarity threshold: 0.1, Max. hits: 200, Max. time: 3 min, and Return prev. results: No. The main window shows search results for 'nci' with a grid of chemical structures and their properties (ID, Formula, Molweight). A table below the grid lists the results:

No.	Structure	ID	Formula	Molweight
1		1	C ₇ H ₆ O ₂	122.1213
2		2	C ₁₄ H ₈ N ₂ S ₄	332.487
3		3	C ₆ H ₃ ClN ₂ O ₅	218.551
4		4	C ₃ H ₃ N ₃ O ₂ S	145.14

AJAX example

- Back end is JChem Web Services
- No Java is needed for browsing

The screenshot shows the 'Ajax interface to ChemAxon web services' window. It features a navigation sidebar with options: Query, Export, Print, and About. The main content area displays search results for 'nci' with a table of results:


No.	Structure	ID	Formula	Molweight
1		1	C ₇ H ₆ O ₂	122.1213
2		2	C ₁₄ H ₈ N ₂ S ₄	332.487
3		3	C ₆ H ₃ ClN ₂ O ₅	218.551
4		4	C ₃ H ₃ N ₃ O ₂ S	145.14

Integration – other ChemAxon tools

»« Standardizer — customizable chemical representation

 Calculator plugins — properties by Chemical Terms
Calculated columns

 Screen — alternative similarity types and metrics

 Tautomer support:

- Tautomer search – duplicate or SSS
- Tautomer duplicate filter option





 Marvin — Query drawing and structure visualization
Provides the most consistent interface to back end

Integration – Cartridge & WS extras

JChem Cartridge:

- JChem index (arbitrary table structure)
- Indexing of user fingerprints
- Communication with Oracle optimizer

JChem Cartridge & Web Services:

-  Reaction based enumeration (Reactor)
-  Format conversions – image generation also
-  Markush enumeration (Calculator plugins)
-  Property predictions through Chemical Terms (Calculator plugins)

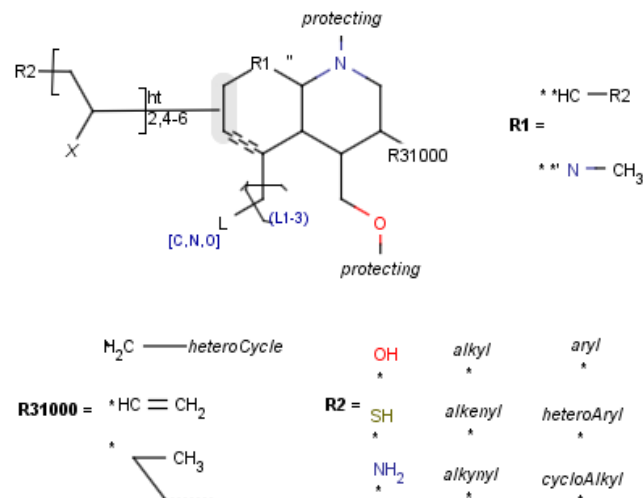
Markush structures

Markush structure registration and search

- Markush features

- R-groups
- Atom lists, bond lists
- Position variation bond
- Link nodes and repeating units
- Homology groups

- Compatible enumeration plugin

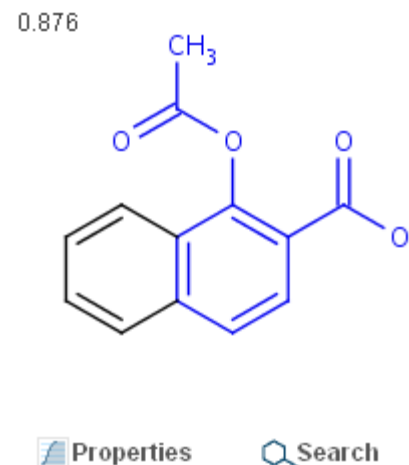
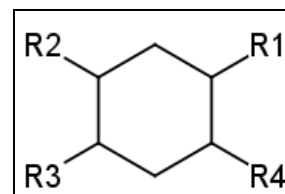


What's new

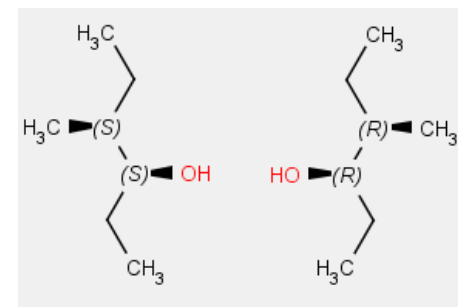
since last year – JChem versions 5.4-5.6

What's new: Search engine

- Hit visualization of similarity search results using MCS
- More consistent R-tables for symmetrical scaffolds.

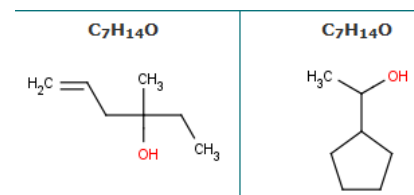


- Multi-threading enhancements:
faster first results and similarity search.
- Non-tautomer duplicate search on tautomer duplicate tables.
- Enantiomer stereo search option
- ECFP & FCFP in similarity search
(with JChem Screen)

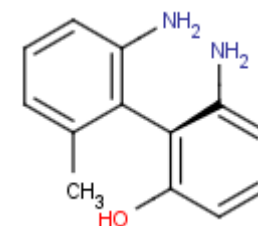
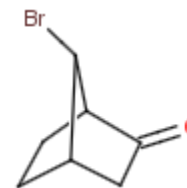


What's new: Search engine

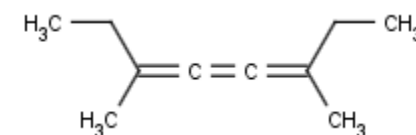
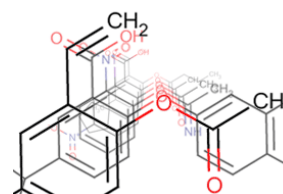
- Sophisticated formula search
 - intervals, excluded atom types,
 - components, coefficients,
 - polymers,
 - isotopes,
 - combinatorial groups, etc.
 - Subformula, exact formula search, etc.



- New stereo types support:
 - Syn, anti
 - Cis/trans of cumulenes (in memory only)
 - Axial: atrop, allenes and cumulenes (in memory only)



- Faster R-group search



What's new: JChem Base, WS

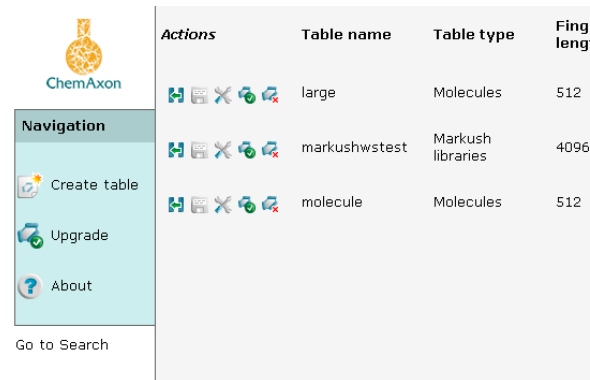
- Duplicate filtering table option
- Jcunique – duplicate filtering of a stream in command-line

Filter out duplicate structures

- JChem Manager in AJAX & JChem Web Services

- New Web Services:

- Molecule search in lists
- Retrieve or export related table data
- Markush search and enumeration
- Batch insert & delete in JChem table
- Batch Chemical Terms evaluations



The screenshot displays the JChem Manager interface. On the left is a navigation sidebar with the ChemAxon logo and options: 'Create table', 'Upgrade', and 'About'. Below the sidebar is a 'Go to Search' button. The main area shows a table with columns: 'Actions', 'Table name', 'Table type', and 'Fing leng'. The table contains three rows of data.

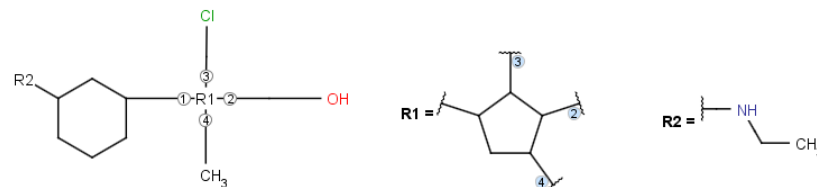
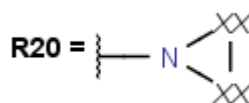
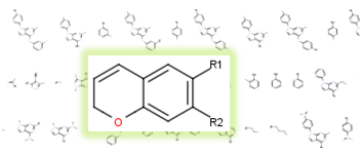
Actions	Table name	Table type	Fing leng
	large	Molecules	512
	markushwstest	Markush libraries	4096
	molecule	Molecules	512

What's new: Markush I.

Improved:

- Chemical representation

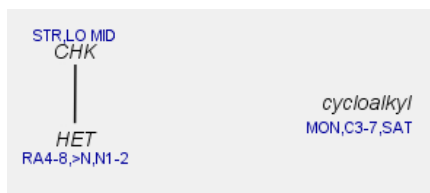
- New homology groups
- Homology properties
- Large combinatorial Markush structures



Edit properties of atom (cycloalkyl)

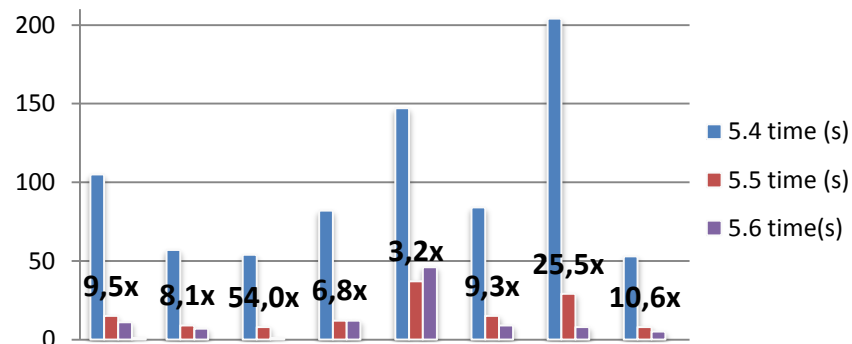
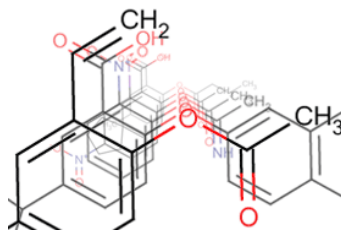
Property key	Property value
Deuterium or Tritium Count	
Ring Type	Monocyclic
Saturation	Not Specified
Additional Text Notes	
<Type new property name here>	

Ok Cancel



- Performance

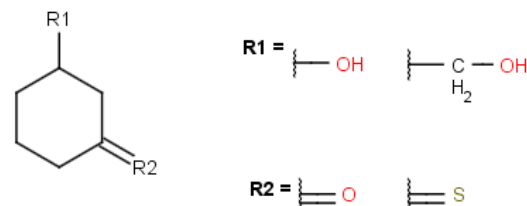
- SSS: **24x**
- FSS: **72x**
- Enum: **>100x**



What's new: Markush II.

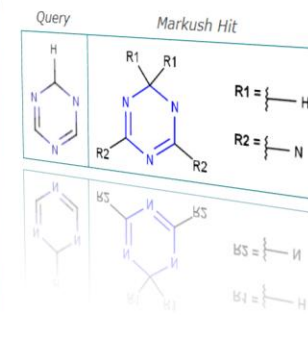
Improved:

- Search features
 - Simple R-group queries
 - Atomic query properties
 - Full H-matching support
 - Homology translation option on/off



- Robustness

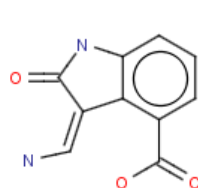
Property	Query Structure	Markush Target Hit
a aromatic		
A aliphatic		
R part of a ring		
R0 not part of a ring		
R₂ number of rings the atom is member of		
Rⁿ substitution as drawn (including its substituents in the query structure)		



What's new: Markush III.

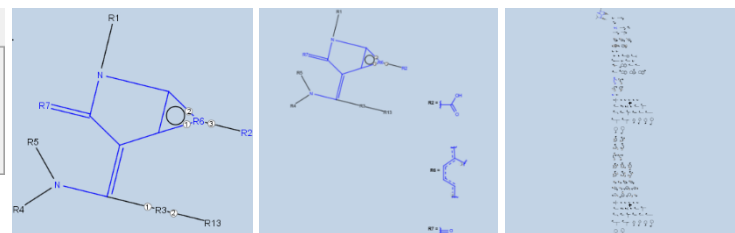
Improved:

- Visualization
- Enumeration



o₂

No R-groups
 Relevant R-groups
 All R-groups



Markush structure

Query structure

Enumerate Display Filter Output

Full enumeration
 Random enumeration
 Markush reduction according to the hit
 Expand homology groups

Max structures: 1,000

Enumerate Generated 184 structures

Full enumeration of this structure produces 39204 structures



Enumerate Display Filter Output

Alignment: Off On Partial clean

Colouring

Show R-groups

Enumerate Display Filter Output

Post-enumeration Chemical Terms filter:

```
(mass() <= 500) &&  
(logP() <= 5) &&  
(donorCount() <= 5) &&  
(acceptorCount() <= 10)
```

Applies only to full or random enumeration

Markush Viewer - D:\work\documentation\chem\markush\demo_markush.vwr

Fragment | Markush |
050611_Demo_Scaffold

050611_Demo_R10

H -CHK

050611_Demo_enumeration view

Previous Next

Under development

Under development

Markush search

- Further speed improvements – scale up full MMS searching efficiently
- Improved query capabilities – according to feedback

JChem Base, WS and Cartridge

- Computational cluster / grid solution
- More usable JChem Manager
- Arbitrary table structure (JChem index table) API support
- Easier installation for Cartridge

Example web applications

Query drawing

Hit alignment,
coloring

Search types,
options

New features

The image displays two screenshots of the JChem web application interface. The top screenshot shows search results for a query, displaying a table of hits with chemical structures and associated IDs and molecular weights. The bottom screenshot shows the query parameters configuration screen, including search type, similarity type, screening configuration, and search options.

Search Results - Mozilla Firefox
URL: http://www.chemaxon.com/jchem/examples/db_search/intsearch.jsp?uid=298930
Hits: 200 (maximum hits reached)
Structure table: nci
Search took 3.54 seconds

ID	Formula	MW	Structure
ID: 76	C20H17NO	MW: 287.2551	
ID: 314	C16H16N2O2	MW: 268.3104	
ID: 315	C22H19NO2	MW: 329.3918	
ID: 318	C11H13NO2		
ID: 325	C13H13NO4		
ID: 334	C14H15NO4	MW: 281.2732	
ID: 2753	C15H11NO2	MW: 237.2533	

Query Parameters - Mozilla Firefox
URL: http://www.chemaxon.com/jchem/examples/db_search/query.jsp?uid=29893

Structure table: nci

main options: Search type: Substructure; Similarity type: Chemical Hashed Fingerprint; Screening config: Default; Dissimilarity threshold: 0.1; Max. hits: 200; Max. time: 3 min.; Search prev. results: No

advanced options: Conditions table with fields for Id, Formula, Molweight, CD_HASH, CD_FLAGS, CD_SORTABLE_FORMULA

Chemical Terms filter:
`(mass() <= 500) &&
(logP() <= 5) &&
(donorCount() <= 5) &&
(acceptorCount() <= 10)`

Summary

- JChem back end is comprehensive and efficient
- Good team player – open to integration and extensions
- Best Markush support
- Continuous development, improvements in the pipeline