

2010 – a year of JChem

May 2010, European User Group Meeting

Szabolcs Csepregi



ChemAxon
Solutions for Cheminformatics

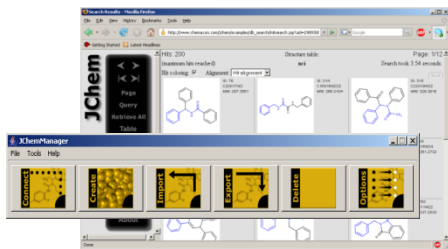
Contents

- ChemAxon chemical database tools
- Main features of JChem Base, Cartridge
- Example interfaces: JSP, ASP, AJAX examples
- Integration with other CXN products
- Recent developments, plans

The JChem DB family

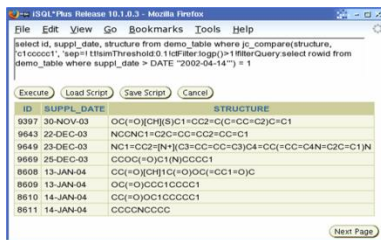
JChem Base

The chemical engine



JChem Cartridge

Oracle SQL integration

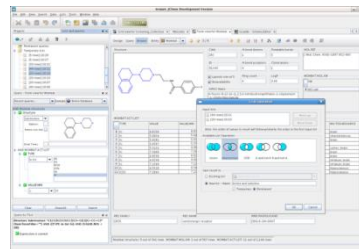


```
select id, suppl_date, structure from demo_table where jc_compare(structure, 'C1CCCC1', 'sup=1 |thamThreshold=0.1 |SF |filter |log(p)=1 |filterQuery select rowid from demo_table where suppl_date > DATE '2002-04-14') = 1
```

ID	SUPPL_DATE	STRUCTURE
9397	30-NOV-03	CC(O)(CH)S(C)1=CC2=C(C(=CC=C2)C=C1
9643	22-DEC-03	NC(=O)1C2=CC=CC=C2C(=O)C1
9649	23-DEC-03	NC1=CC2=NC(=C3=CC=CC=C3)O4=CC(=CC=C4N=C3C=C1)N
9689	25-DEC-03	CCOC(=O)C1(N)CCCC1
8608	13-JAN-04	CC(O)(CH)1C(O)OC(=CC1=O)C
8609	13-JAN-04	CC(O)COC1CCCC1
8610	14-JAN-04	CC(O)COC1CCCC1
8611	14-JAN-04	CCCCCCCC

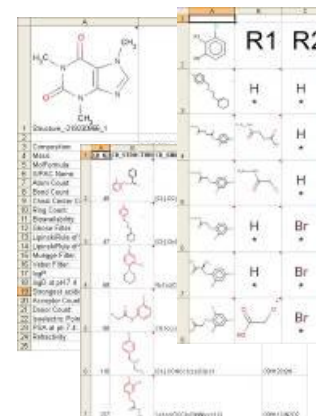
Instant JChem

Desktop application for scientist



JChem 4 Excel

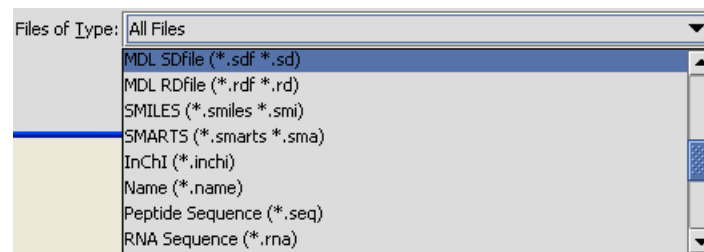
MS Excel integration



Compatibility and interfaces

Chemical file formats:

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

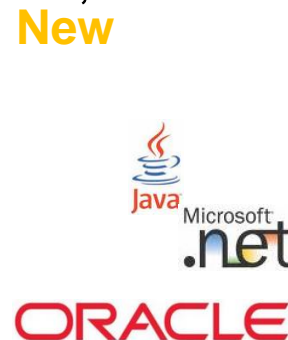


Database engines:

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

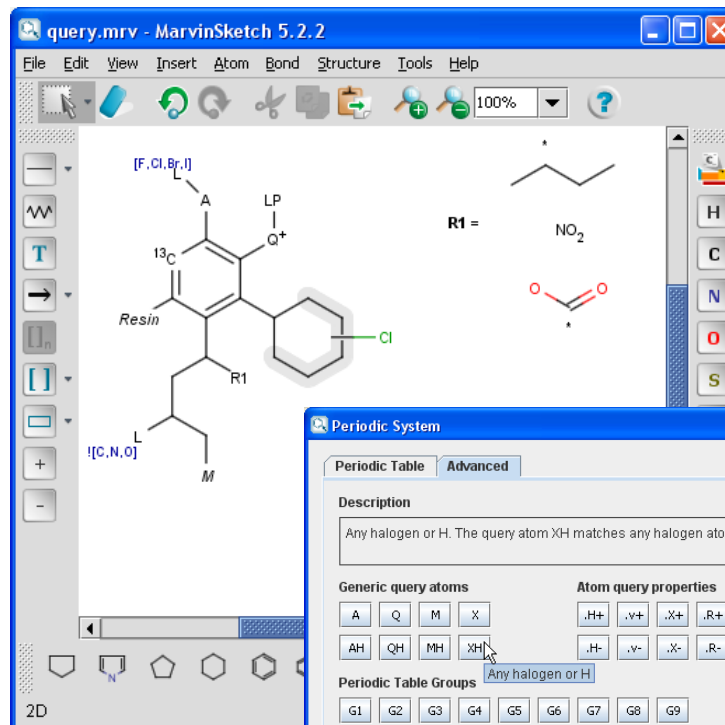
All operating systems through:

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



Structure searching: features

- Substructure, Similarity, Full, etc. search types
- Wide range of query atoms
- Query properties
- R-group queries
- Full SMARTS support
- Pseudo atoms, Lone pairs
- Relative stereo
- Reaction search features
- Polymers
- Position variation
- Hit coloring
- ...



The screenshot shows the 'Periodic System' dialog box in MarvinSketch. The dialog is titled 'Periodic System' and has a 'Close' button. It is divided into several sections:

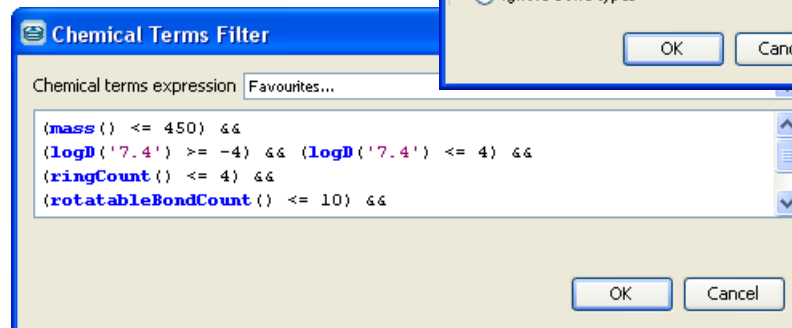
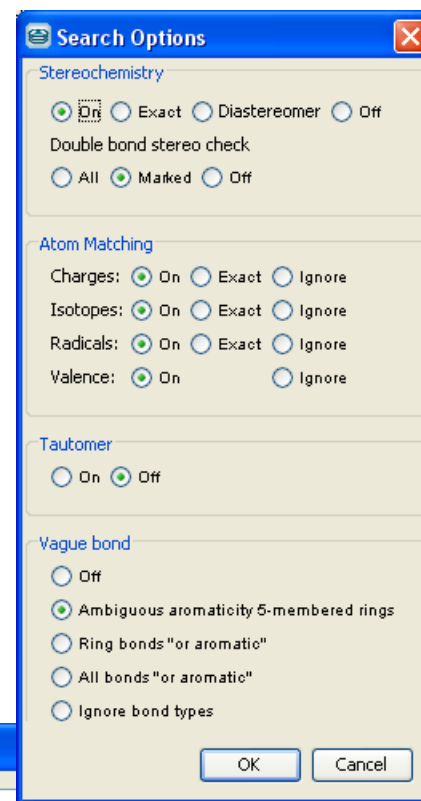
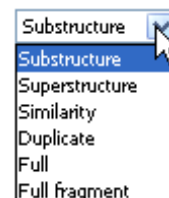
- Periodic Table:** Includes 'Advanced' and 'Periodic Table' tabs. The 'Advanced' tab is selected.
- Description:** A text box containing the description: 'Any halogen or H. The query atom XH matches any halogen atom or hydrogen.'
- Generic query atoms:** A grid of buttons for 'A', 'Q', 'M', 'X', 'AH', 'QH', 'MH', and 'XH'. The 'XH' button is highlighted with a tooltip that says 'Any halogen or H'.
- Atom query properties:** A grid of buttons for various properties: '.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 groups G1 through G18.
- Special nodes:** Buttons for 'LP', 'Pol', and '*'.
- R-groups:** A grid of buttons for R1 through R32.
- Custom Property:** A section with 'Type:' and 'Value:' fields. The 'Type:' field has buttons for 'R-group', 'Alias', 'Pseudo', 'SMARTS', and 'Value'. The 'Value:' field is empty.

www.chemaxon.com/conf/Structural_Search.ppt

Structure searching: 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.

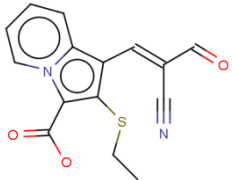
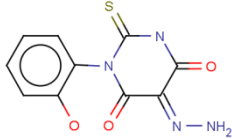
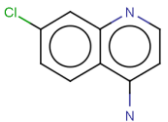
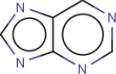


Structure search: performance

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):

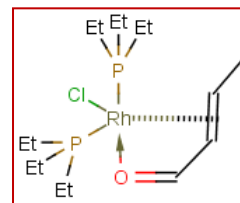
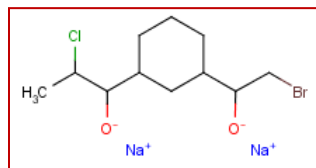
Query	Number of hits	Search time
	2	0.91 s
	93	0.98 s
	6,001	1.30 s
	146,256	5.66 s

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

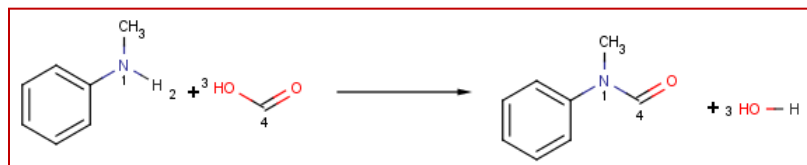
Table types

Control allowed chemical structures and available operations

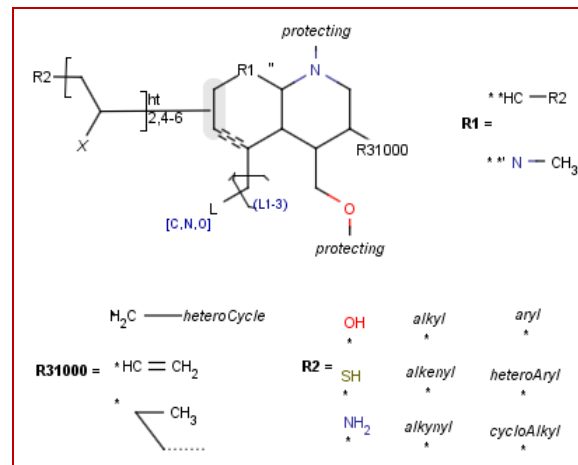
- Molecule



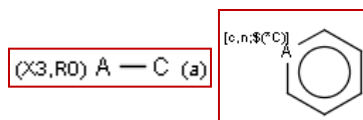
- Reaction



- 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, 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 IDs and molecular weights. The 'Ajax interface to ChemAxon web services' window is also visible, showing a table of search 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 Cl ₂ O ₅	218.551
4		4	C ₇ H ₅ N ₃ O ₂ S	145.14

AJAX example

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

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 and back-end.

Integration – Cartridge extras

- JChem index (arbitrary table structure)
- Communication with Oracle optimizer
- Reaction based enumeration (Reactor)
- Format conversions – image generation also
- Markush enumeration (Calculator plugins)
- Property predictions through Chemical Terms (Calculator plugins)
- Indexing of user fingerprints **New**

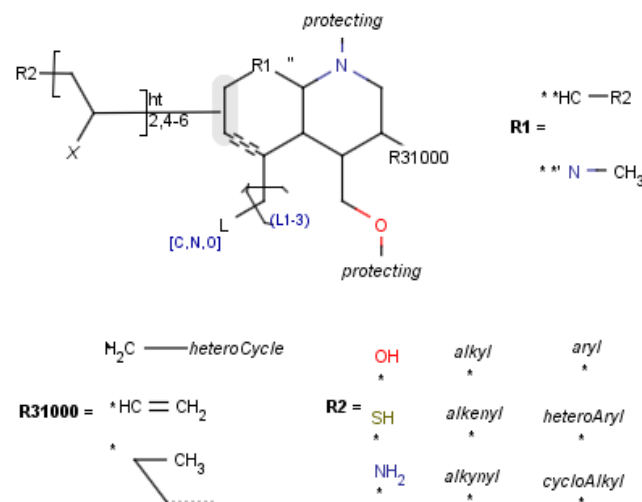
Markush structures

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

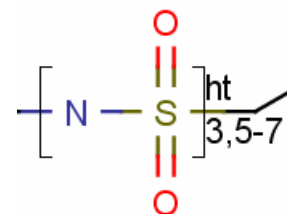
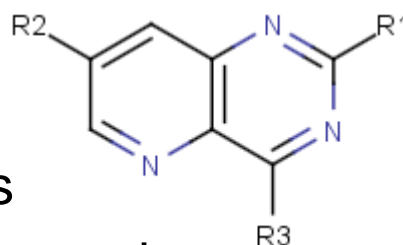


What's new

What's new: JChem Base

5.2.X

- .NET API
- JChem Web Services
- Polymer storage and search
- New query options and features including searching of attached data, group matching of undefined R-atoms, repeating units.
- Improved substructure search performance
- New metrics for similarity search (Tversky, etc.)

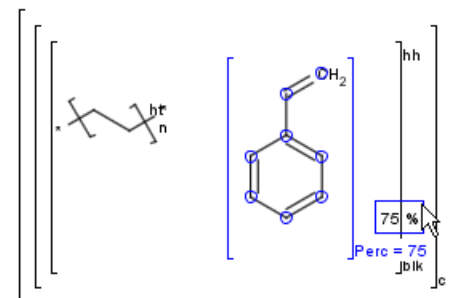


What's new: JChem Base

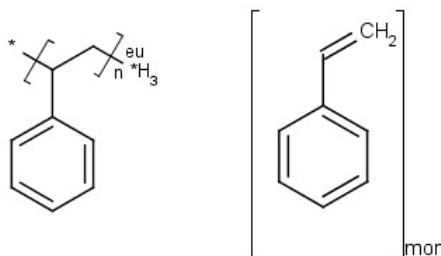
Polymer support details

- Polymer brackets and properties (type, connectivity, etc.)

- Attached data search



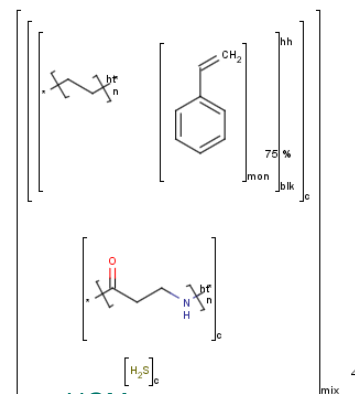
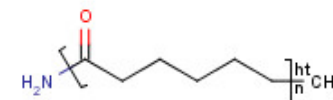
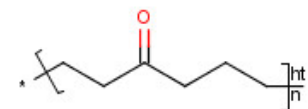
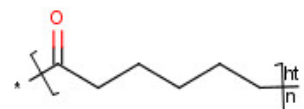
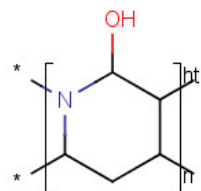
- Source- and structure-based representation



What's new: JChem Base

Polymer support details (cont.)

- Ladder type polymers
- Phase-shifting (for ht SRU)
- End group matching:
 - * atoms: unspecified end groups
 - Search option to ignore end groups
- Copolymer types: co, alt, rnd, blk, grf, xl, mer, mod
- Polymer mixtures
- New search options



What's new: JChem Base

5.3.X

- Homology variation in queries (alkyl, aryl, etc.)
- R-group decomposition API integrated with JChemSearch
- Markush output of R-group decomposition
- Improved undefined R-atom matching feature
- New vague bond level options (improvements in handling fused aromatic ring systems)
- Reduced memory need for ABAS

- Pre-regeneration for less downtime at version updates
- JChem Manager GUI improvements
- New method of handling log tables and register caches, to improve cache update mechanism on batch loading.

What's new: Cartridge-specific

5.2.X

- Interactive installer
- New metrics for similarity search (Tversky, etc.)
- jcf.exec_c to execute native programs

5.3.X

- Markush tables and indexes
- User-defined fingerprints for similarity search
- Screen molecular descriptors for similarity search
- Increased security: encrypted passwords and usernames

Under development

Plans: JChem Base & Cartridge

JChem Base

- Similarity search results visualization: MCS highlighting
- Maximum Common Substructure search type
- R-group decomposition on GUI-s
- Symmetry handling in R-group decomposition
- Multi-threading of fingerprint operations (screening, similarity)
- Arbitrary table structure (JChem index table)
- JChem Server, JChem grid

Cartridge

- R-group decomposition
- Further search speed improvements (SSS, similarity)

Example web applications

Query drawing

Hit alignment,
coloring

Search types,
options

New features

The image displays two screenshots of the ChemAxon web application interface. The top screenshot, titled "Search Results - Mozilla Firefox", shows a search results page for the query "nci". It displays 200 hits (maximum reached) and a structure table with columns for ID, Formula, and MW. The table lists several chemical structures, including those with IDs 76, 314, 315, 318, 325, 334, and 2753. The bottom screenshot, titled "Query Parameters - Mozilla Firefox", shows the query parameters for the search. It includes a "main options" section with fields for "Search type" (Substructure), "Similarity type" (Chemical Hashed Fingerprint), "Screening config" (Default), "Disimilarity threshold" (0.1), "Max. hits" (200), and "Max. time" (3 min.). The "advanced options" section includes a "Conditions" table with fields for ID, Formula, Molweight, CD_HASH, CD_FLAGS, and CD_SORTABLE_FORMULA. A "Chemical Terms filter" section at the bottom contains a list of conditions: `(mass() <= 500) && (logP() <= 5) && (donorCount() <= 5) && (acceptorCount() <= 10)`.

Summary

- JChem Base and Cartridge are comprehensive and efficient
- Good team players – open to integration and extensions
- Continuous development, improvements in the pipeline

Acknowledgements

- Development team: Péter Kovács, Szilárd Dóránt, Ferenc Csizmadia, Edvárd Büki, Róbert Wágner, Tamás Csizmazia, Roland Molnár, Jon Lee, György Pirok, Nóra Máté, Ali Baharev and others at ChemAxon
- The countless suggestions, feedback from you

Find out more

- **Product descriptions & links**
www.chemaxon.com/products.html
- **Forum**
www.chemaxon.com/forum
- **Presentations and posters**
www.chemaxon.com/conf
- **Download**
www.chemaxon.com/download.html

The collage consists of four screenshots from the ChemAxon website:

- Top-left:** The main website page with navigation tabs (Home, Products, Documentation, Download, Forum, About us, Contact us) and a 'Recent News' section listing updates from 2007.
- Top-right:** A screenshot of the 'Insta' software interface, showing a chemical structure and a search bar.
- Bottom-left:** A screenshot of the 'Marvin and Calculator Plugin Demo' page, featuring a list of features such as 'Rich editing', 'wide range of file types supported', and '3D editing'.
- Bottom-right:** A screenshot of the 'Technical Support Forum' page, displaying a table of support topics.

Support	Topics	Posts	Last Post
Structure editing, viewing and file formats	827	4081	Fri Feb 02, 2008 9:08 pm Closed
Support for MarvinSketch, MarvinView and MarvinCenter	60	306	Thu Jan 31, 2008 9:16 pm Closed
MarvinSpace	60	374	Fri Feb 01, 2008 10:02 am Closed
Development discussion area for OpenGL 3D rendering main/3D molecule viewer	351	174	Fri Feb 01, 2008 10:02 am Closed
Structure based prediction and Chemical Terms	251	174	Fri Feb 01, 2008 10:02 am Closed
Support for Calculator Plugin operation through cracks, API, Marvin, InstaXChem and Chemical Terms	426	1952	Fri Feb 01, 2008 10:02 am Closed
Structure search and chemical database	426	1952	Fri Feb 01, 2008 10:02 am Closed
Support for XChem Base and XChem Cartridge	426	1952	Fri Feb 01, 2008 10:02 am Closed
InstaXChem	174	416	Fri Feb 01, 2008 9:18 pm Closed
Discussion area for InstaXChem (Structure database GUI, both Calculator Plugin processor etc)	174	416	Fri Feb 01, 2008 9:18 pm Closed
Structure conversion algorithm / standardization	174	416	Fri Feb 01, 2008 9:18 pm Closed
Support for Standardizer	174	416	Fri Feb 01, 2008 9:18 pm Closed
Library enumeration, virtual synthesis and metabolite generation	116	577	Thu Jan 31, 2008 9:18 pm Closed
Support for Reactor and Fragmenter	116	577	Thu Jan 31, 2008 9:18 pm Closed
Virtual screening, clustering and molecular descriptors	108	471	Thu Jan 31, 2008 9:18 pm Closed
Support for Screen and Shutter	108	471	Thu Jan 31, 2008 9:18 pm Closed
License Issues	4	1	Thu Jan 24, 2008 10:02 am Closed
Support for the technical questions related to license handling	4	1	Thu Jan 24, 2008 10:02 am Closed