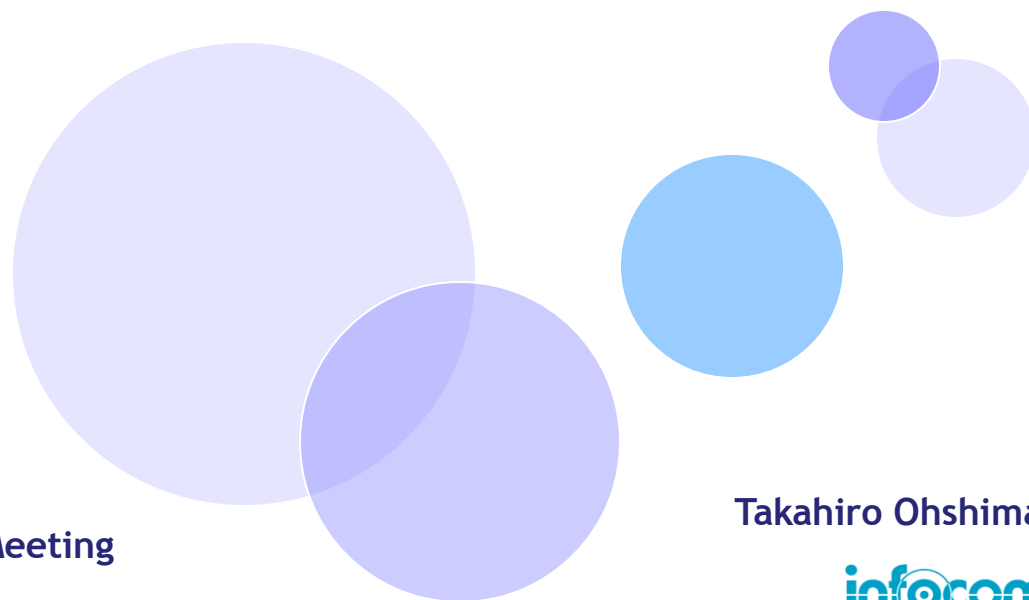


# JChem Extensions for KNIME

## KNIME.com products



ChemAxon 2011 US User Group Meeting  
San Diego, CA

Takahiro Ohshima



# Overview



- INFOCOM
- KNIME
- JChem Extensions
- Marvin Family Nodes
- KNIME.com products
  - ◆ KNIME Enterprise Server
  - ◆ KNIME WebPortal

# INFOCOM

- IT Business company of Teijin Group



- Offering wide varieties of IT services & solutions

- ◆ Head office: Tokyo, JAPAN

- ◆ Capital: 1.59 billion yen (ca. \$15M)

- ◆ No. of employees:

- 683 (stand-alone) / 1,151 (consolidated) , as of March 31, 2010

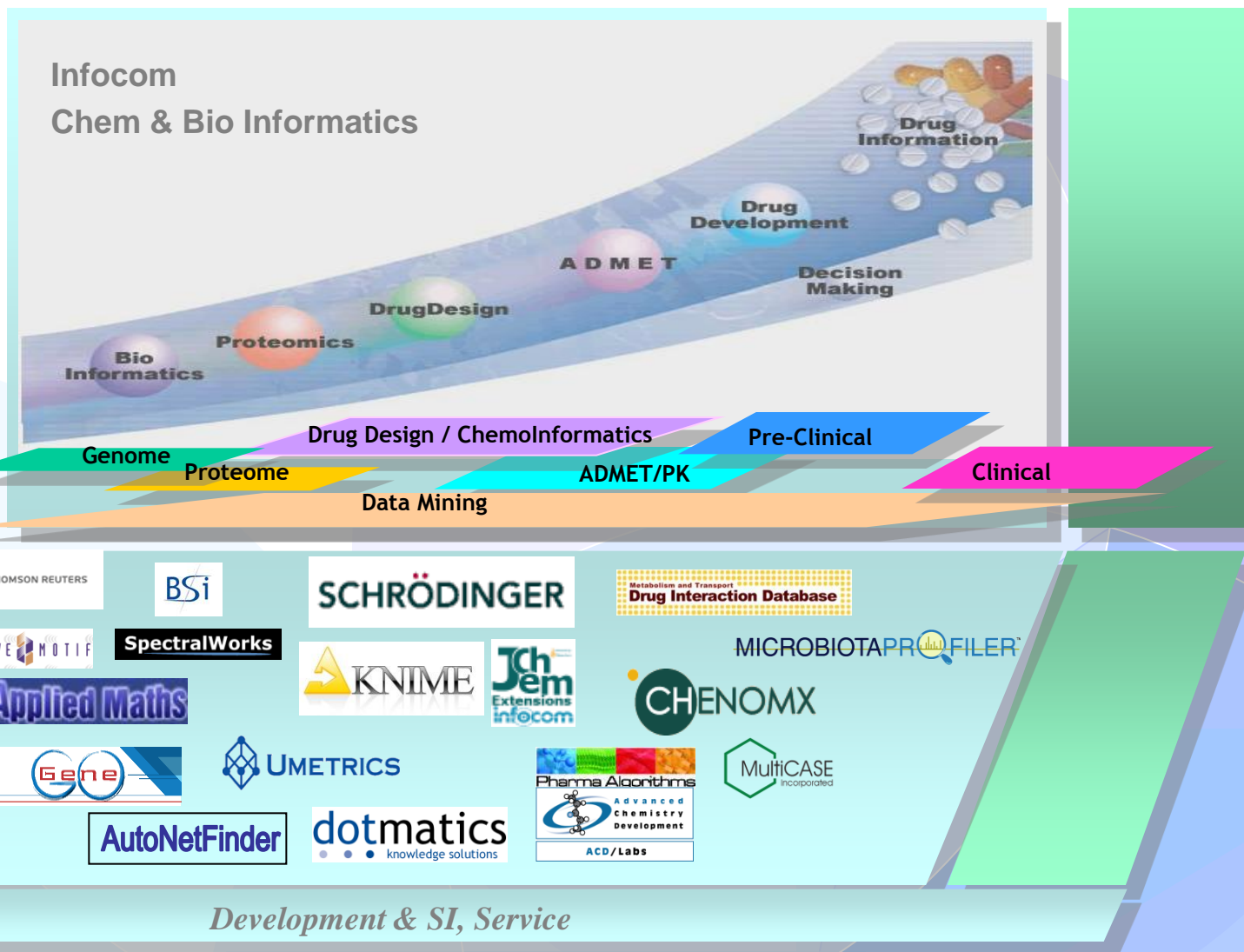
- Chem & Bio Informatics Department

- ◆ for life science market in Japan

- ◆ for more than 25 years



# Our Business (Life science area)



# KNIME (Konstanz Information Miner)

## ■ KNIME [naim]:

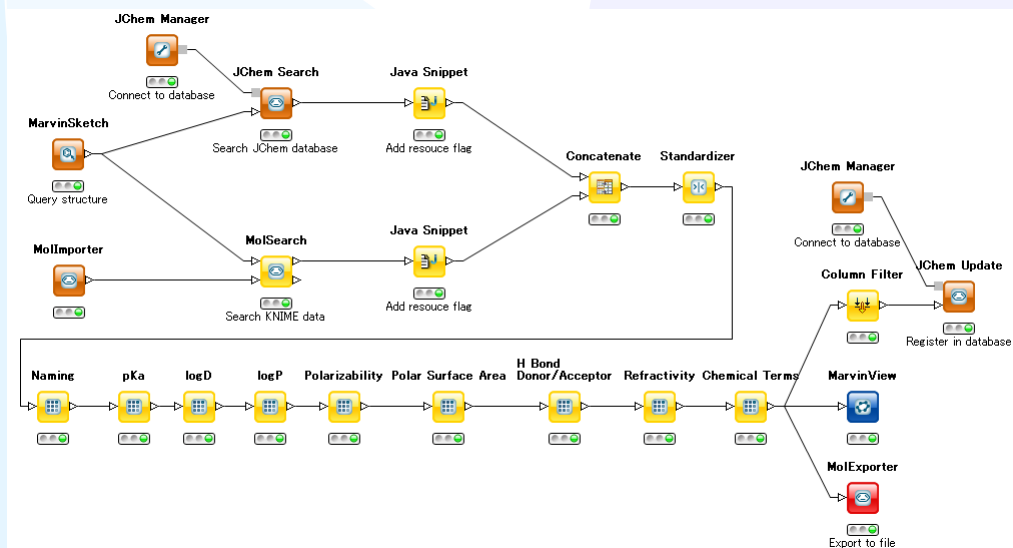
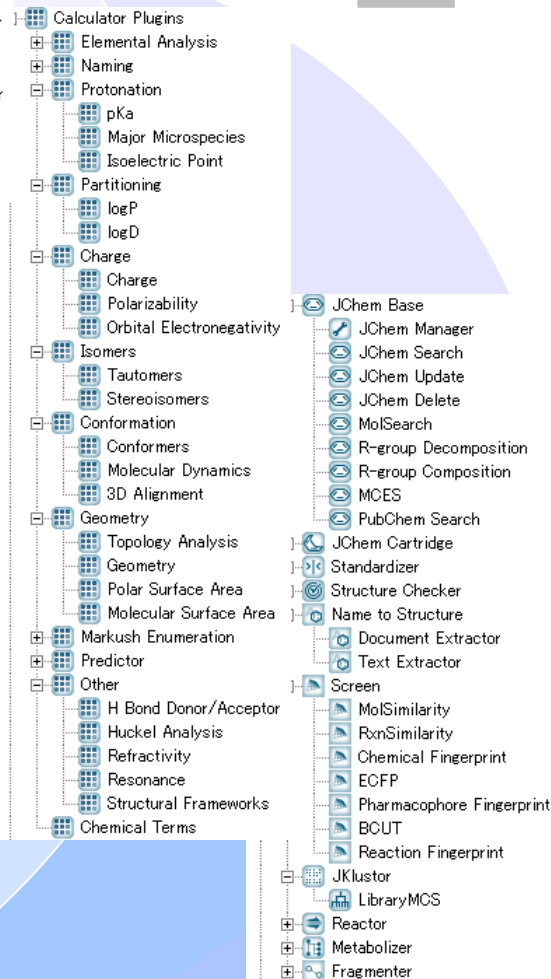
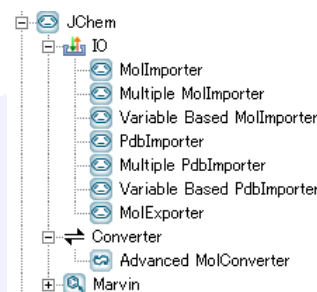
- ◆ Workflow & Data mining platform
- ◆ Developed at the University of Konstanz, Germany
- ◆ Open Source / Eclipse
- ◆ Free of charge !
  - KNIME, CDK, R, Weka ...
- ◆ Commercial suppliers & partners join in
  - ChemAxon, Schrödinger, Tripos, etc



A screenshot of the KNIME software interface. The central workspace displays a workflow with nodes: MarvinSketch, MolSearch, MolImporter, MolExporter, and MarvinView. The MarvinView node is expanded to show a grid of 16 rows of chemical structures. To the right, there are 3D molecular models and a list of chemical structures. The interface includes a menu bar, a toolbar, and a node repository on the left.

# JChem Extensions

- Extensions of ChemAxon's tool for KNIME workflow
- Infocom implement it with the support of ChemAxon
- Contains over 90% of ChemAxon's cheminformatics functionality



# ChemAxon's cheminformatics functionality

## 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...

## Add-ons



### Markush Search

Store & search Markush structures

## 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



### JChem Webservices

Web services integration interface

## 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

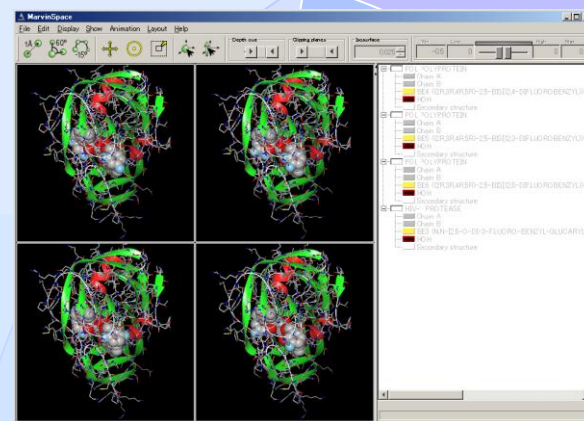
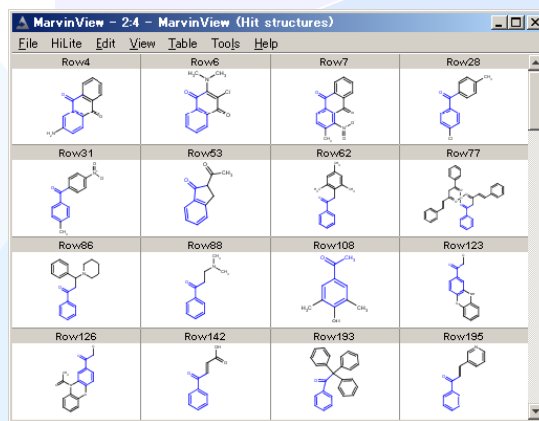
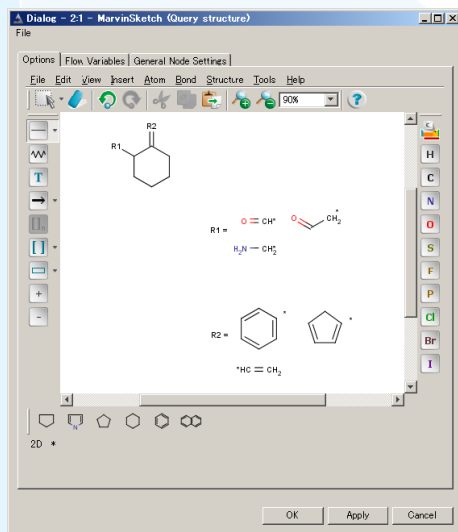
# Marvin Family Nodes and Marvin are FREELY available in KNIME

## ■ What are “Marvin Family Nodes” ?

- ◆ Marvin Sketch node 
- ◆ Marvin View node 
- ◆ Marvin Space node 
- ◆ Converter nodes 

■ Available on KNIME Update Site (Standard Chemistry Tool in KNIME)

■ Free of charge for ALL customers!!

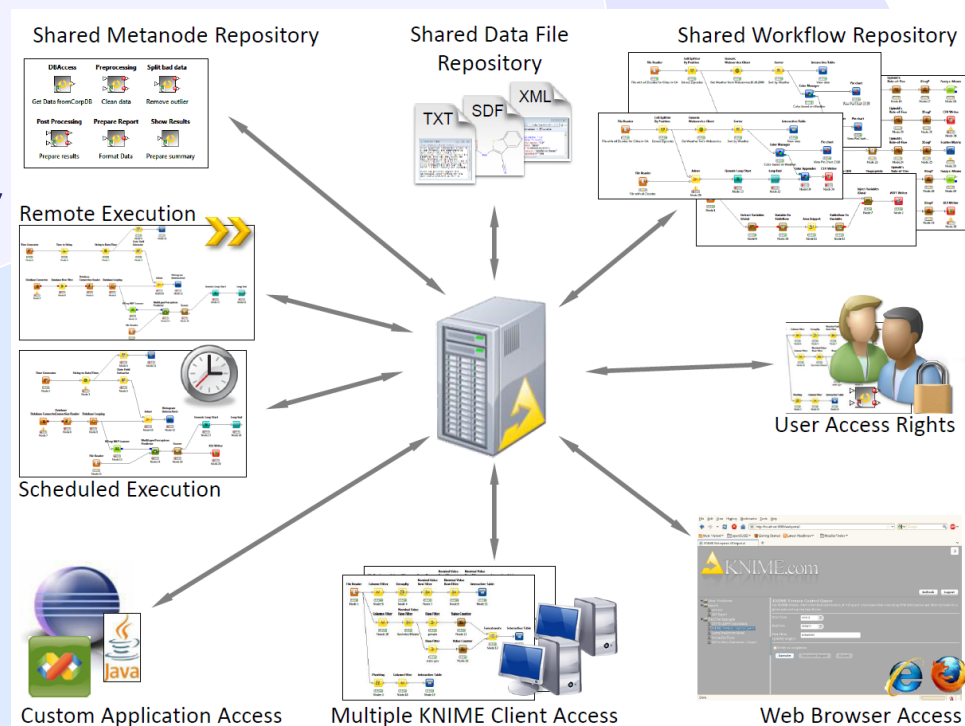


# KNIME.com Product

## ■ KNIME Enterprise Server

- ◆ Hosted repository for workflows
- ◆ User access control
- ◆ Remote and scheduled execution
- ◆ Access to workflows for non-KNIME users
- ◆ Backend data processing engine for SOA integration

KNIME Server



# KNIME WebPortal & QuickForms

## ■ Web Portal

- ◆ Access to workflows through web interface
- ◆ Quickform nodes expose workflow inputs and output(parameters, files, ...)

The image displays the KNIME WebPortal interface. At the top, a 'Node Repository' window shows a list of nodes, including 'Quick Form', 'Date (String) Input', 'Double Input', 'File Download', 'File Upload', 'Integer Input', 'String Input', 'String Radio Buttons', and 'Variable Output'. A red arrow points to the 'File Upload' node icon. Below this, a 'Dialog - 598 - File Upload' window is open, showing configuration options for a file upload node. The dialog includes fields for 'Label', 'Description', 'Weight', 'Variable Name', 'Valid File Extensions', and 'Default File'. The 'Label' field contains the text 'Please select a file for Submission'. The 'Description' field contains 'Please browse to your File for submitting it to the workflow'. The 'Variable Name' field is 'FileNameVar', and the 'Default File' is 'C:\Data\CRMAnalytics\CustomerData.txt'. The 'Valid File Extensions' field is '.txt'. The 'Weight' field is set to 1. The 'Default File' field has a 'Browse...' button. The 'Dialog' window has 'OK', 'Apply', and 'Cancel' buttons.

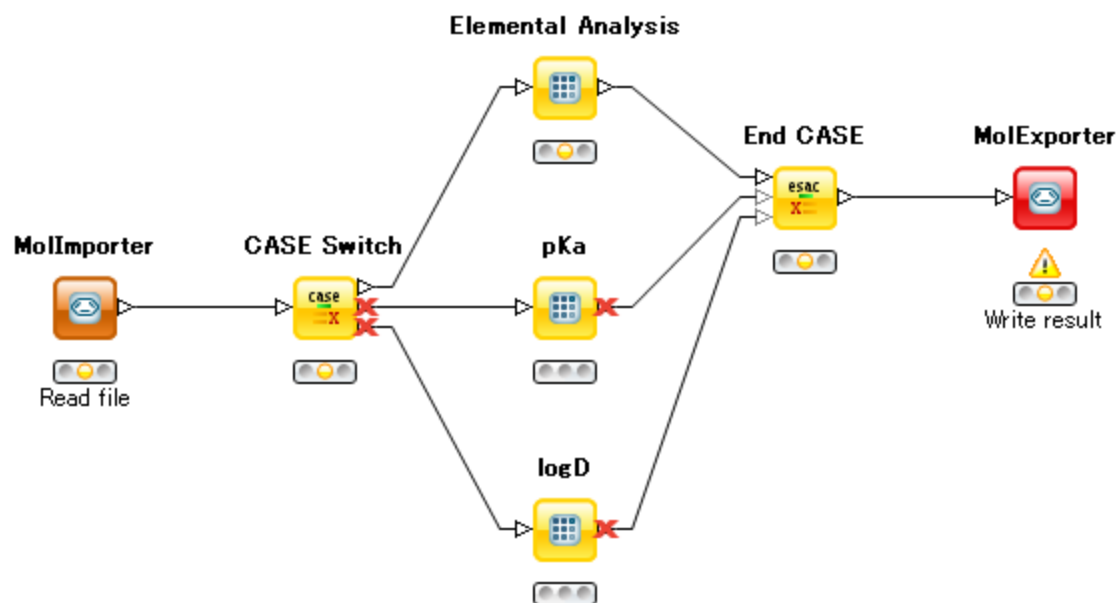
The main web portal interface is shown below. It features the KNIME.com logo and the title 'KNIME Webportal'. There are 'Refresh' and 'Logout' buttons. The 'Workflow Repository' section lists various workflows, including 'CRM Analysis', 'dataMining', 'DataAccess', 'NBO-Calculations', 'Workshop', and 'DataAnalysis'. The 'DataAnalysis' workflow is selected. The 'DataAnalysis' section provides a description of the workflow and a 'Quick Form' for configuration. The 'Quick Form' includes a 'Select File' button, a text input for the number of amines (set to 10), a text input for the number of acids (set to 10), a radio button for the summary view type (set to 'Histogram'), and a checkbox for 'Mail notification on completion'. 'Run' and 'Cancel' buttons are at the bottom.

# JChem Extensions + KNIME Enterprise Server

- Workflow executed via KNIME Quick Forms on a WebPortalUI
  - ◆ User uploads a data file
  - ◆ Enters additional parameters
  - ◆ Executes the workflow on the server via a Web UI
  - ◆ Retrieves the result file via the web browser

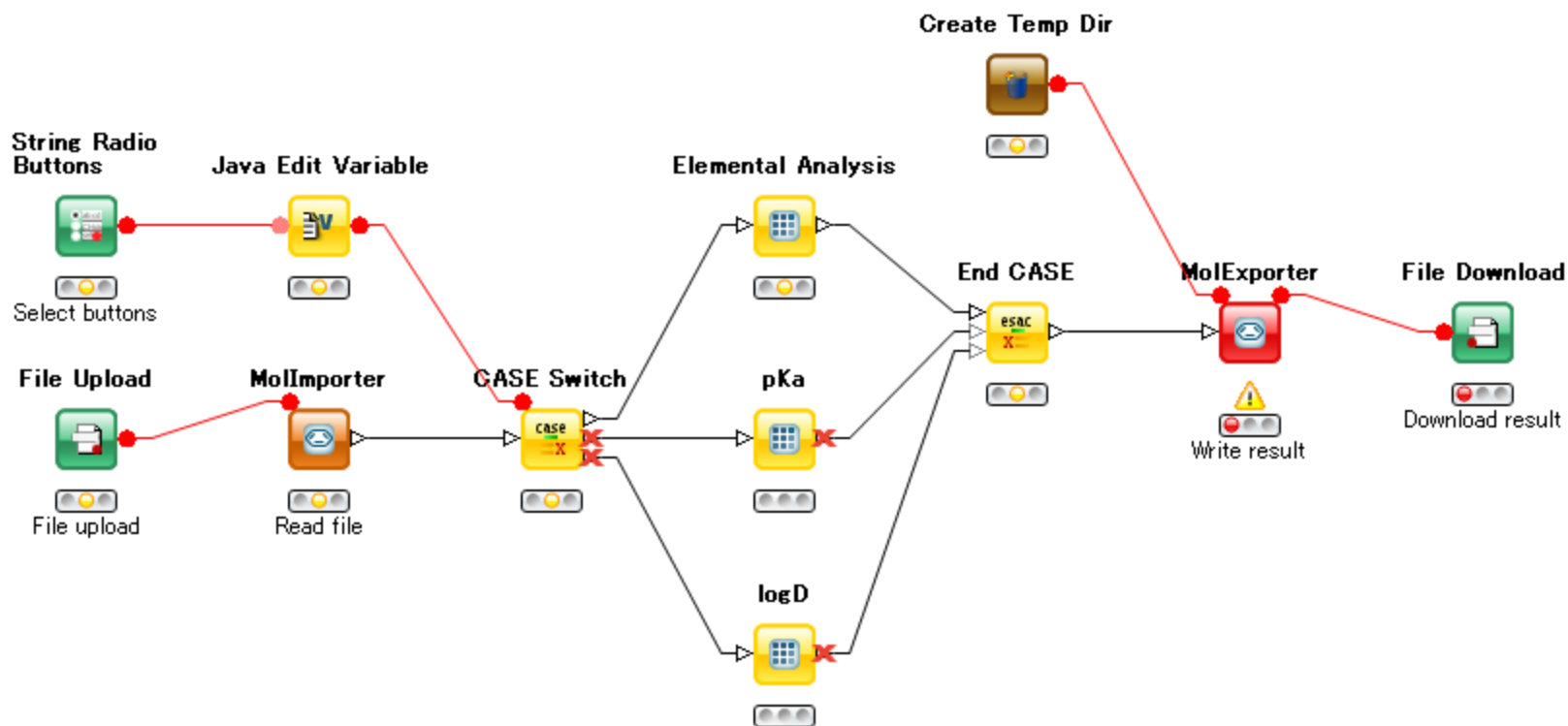
# JChem Extensions + KNIME Enterprise Server

- A general workflow



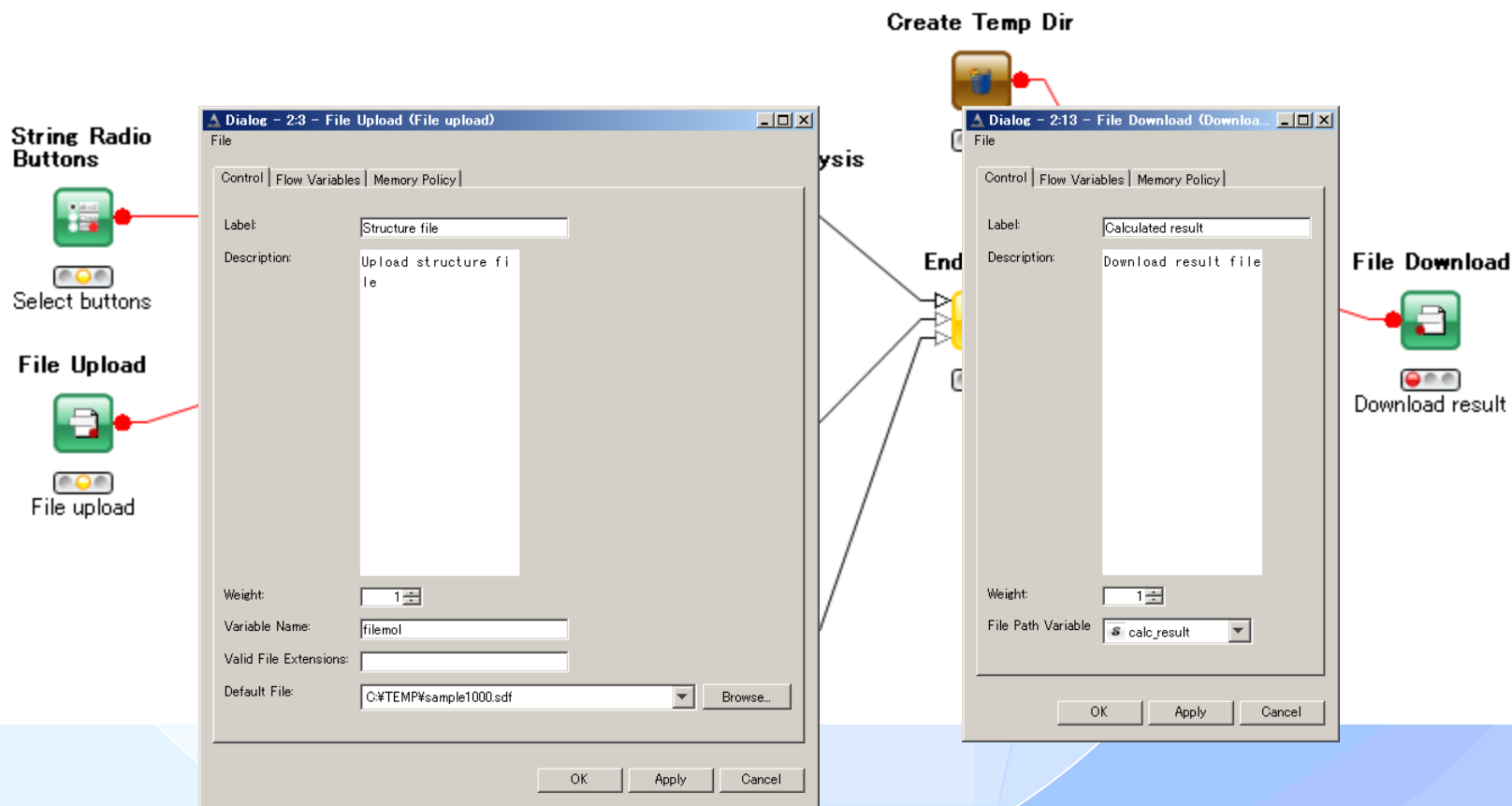
# JChem Extensions + KNIME Enterprise Server

## ■ Workflow + new Quick Form nodes



# JChem Extensions + KNIME Enterprise Server

- The configuration options for the workflow
  - ◆ New Nodes for Input/Output definition
  - ◆ Web UI can be configured through the node configuration



# JChem Extensions + KNIME Enterprise Server

## ■ Upload the workflow to KNIME Server

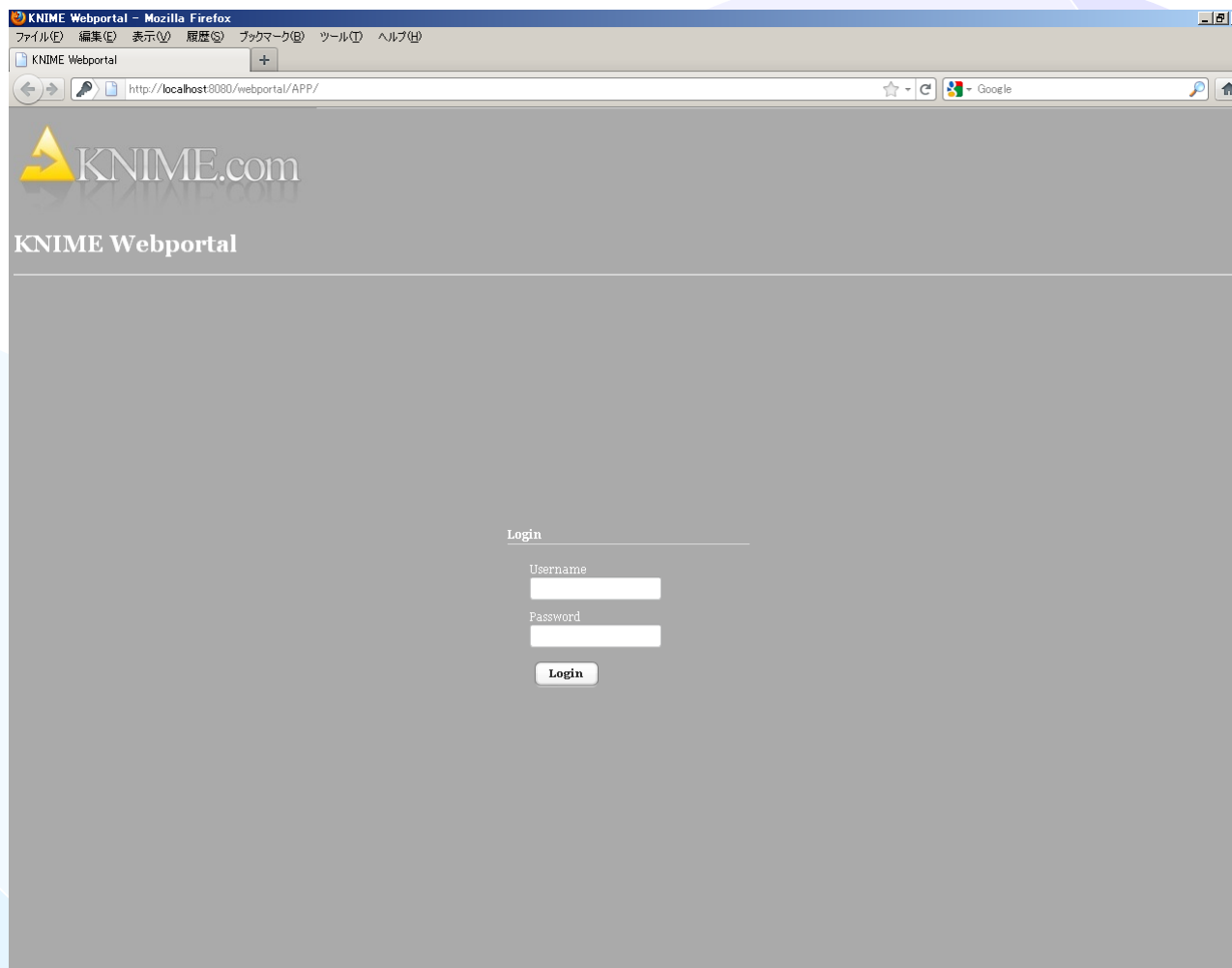
The screenshot displays the KNIME desktop environment. The main workspace shows a workflow titled "0: Calculating\_Properties". The workflow consists of several nodes: "String Radio Buttons" (with "Select buttons" sub-node), "Java Edit Variable", "File Upload" (with "File upload" sub-node), "MolImporter" (with "Read file" sub-node), "GASE Switch", "Elemental Analysis", "pKa", "logD", "End GASE", "Create Temp Dir", "MolExporter" (with "Write result" sub-node), and "Write result". A red dashed arrow originates from the "Calculating\_Properties" project in the "Workflow Projects" pane on the left and points to the "Calculating\_Properties" workflow in the "localhost" pane on the right, illustrating the upload process.

The interface includes several panes:

- Workflow Projects:** Lists the current workflow and other projects like "Chunk\_Flow\_Example" and "Substructure\_Search\_Example".
- Favorite Nodes:** Shows "Personal favorite nodes", "Most frequently used nodes", and "Last used nodes".
- Node Repository:** A tree view of available nodes categorized by function (e.g., IO, Database, Data Manipulation, Chemistry).
- localhost:** Shows the workflow being uploaded to the local KNIME Server. The "Workflow Server" is set to "localhost" and the status is "connect".
- Outline:** A small thumbnail of the workflow.
- Console / KNIME Explorer:** Shows the workspace structure, including "TEMP (Team Space - C:\TEMP)" and "LOCAL (Local Workspace)".

# JChem Extensions + KNIME Enterprise Server

## ■ Accesses to KNIME Webportal



# JChem Extensions + KNIME Enterprise Server

- Executes the workflow on the server via a Web UI

KNIME Webportal - Mozilla Firefox

http://localhost:8080/webportal/APP/

KNIME.com

KNIME Webportal

Refresh Logout

Workflow Repository

Calculating\_Properties

Calculating Properties Example Upload a structural file (SDF, Smiles, etc) Choose a property to calculate Download result file as SDF file

Quick Form

Calculate Property

Elemental

pKa

logP

Structure file

Change File

Uploaded file "sample500.sdf" (750 KB)

Mail notification on completion

toshima@infocom.co.jp

Run Cancel

ファイルのアップロード

ファイルの場所: TEMP

hit\_result2.sdf

hit\_result.sdf

id\_list.txt

markushs.mrv

patent.tif

pubchem\_iupac\_name.txt

reactant1.smiles

reactant2.smiles

reaction.rxn

sample500.sdf

sample1000.sdf

sample10000.smiles

sample10000.cml

sample10000.mrv

ファイル名(N):

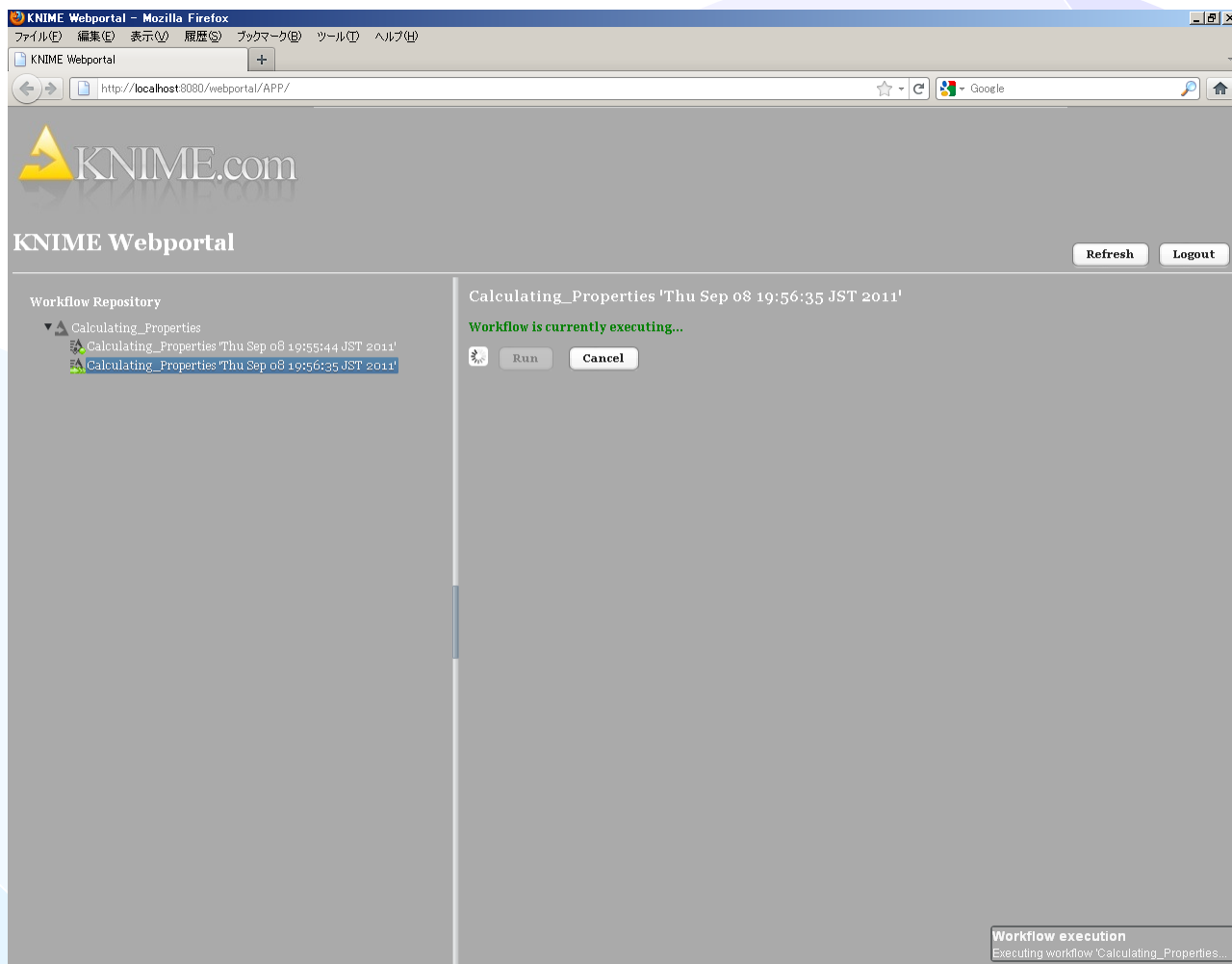
開(O)

ファイルの種類(T): 全てのファイル

キャンセル

# JChem Extensions + KNIME Enterprise Server

- Executes the workflow on the server via a Web UI



# JChem Extensions + KNIME Enterprise Server

- Executes the workflow on the server via a Web UI

KNIME Webportal - Mozilla Firefox  
http://localhost:8080/webportal/APP/

KNIME.com  
KNIME Webportal  
Refresh Logout

Workflow Repository  
Calculating\_Properties  
Calculating\_Properties 'Thu Sep 08 19:55:44 JST 2011'  
Calculating\_Properties 'Thu Sep 08 19:56:35 JST 2011'

Calculating\_Properties 'Thu Sep 08 19:56:35 JST 2011'  
Workflow executed successfully

Quick Form  
Calculated result  
Discard

calc\_result.sdf - MarvinView 5.5.1.0

#	structure	Strongest BASIC pKa 1	Molecule (pKa)	Strongest BASIC pKa 2	Strongest ACIDIC pKa	Strongest ACIDIC pKa
1		-7.606102441714619	CC1=CC(=O)C=C(C1)=O			
2		0.771564077342680 3	S(S(C1=NC2=CC=CC=C2S1)C1=NC2=CC=CC=C1C=C2S1	0.106823690075198	21	
3		-8.561920077865338	OC1=C(C(C)C=C(C=C1N=O)N=O)N=O		2.720660851094925	
4		5.831186176048654	N=C1NC=C(S1)(N=O)=O	-8.173850740366626	7.985620599867345	
5		2.041844180406501	NC1=CC2=C(C(C=C1)C(=O)C1=C(C(=CC=C1)C2=O	-8.50344109960859		
6		-5.427767825761001	OC(=O)C1=C(C(=CC=C1)C1=C2=C(C(=O)C(B)=C2OC2=C1C=CC(O)=C2B1	-8.141092795883687	3.614414022409749	6.248716788344773
7		-0.725665796844792	CN(C)C1=C(C(C)C(=O)C2=C(C=C=C2)C1=O			
8		-7.416119986008756	CC1=C(C2=C(C=C1)C(=O)C1=C(C=C(C=C1)C	-8.26667406185362		

# JChem Extensions + KNIME Enterprise Server

## ■ Web Portal also provides access to reports

The screenshot shows the KNIME Webportal interface in a Mozilla Firefox browser. The page title is 'KNIME Webportal' and the URL is 'http://localhost:8080/webportal/APP/'. The interface includes a navigation menu, a 'Workflow Repository' section, and a main content area displaying a report for the workflow 'Substructure\_Search\_Example'.

**Workflow Repository**

- Calculating\_Properties
- Substructure\_Search\_Example

**Substructure\_Search\_Example 'Thu Sep 08 19:59:28 JST 2011'**

**Workflow executed successfully**

**Quick Form**

[Hit result](#)  
[Non-hit result](#)

**Discard**

**Report Preview for Substructure\_Search\_Example 'Thu Sep 08 19:59:28 JST 2011'**

Molecule (0)	Molecule (1)	Molecule (2)	Molecule (3)

Download Report as  
[PDF](#) [XLS](#) [PPT](#) [DOC](#)

**Workflow execution**  
Executing workflow 'Substructure\_Search\_Example...'

# Summary

## ■ KNIME WebPortal & QuickForms

- ◆ Access to workflows through Web Interface
- ◆ Easy creation of custom Web UI
- ◆ Prompt for workflow variables
- ◆ Upload of data file
- ◆ Download of results
- ◆ Mail notification for long running tasks

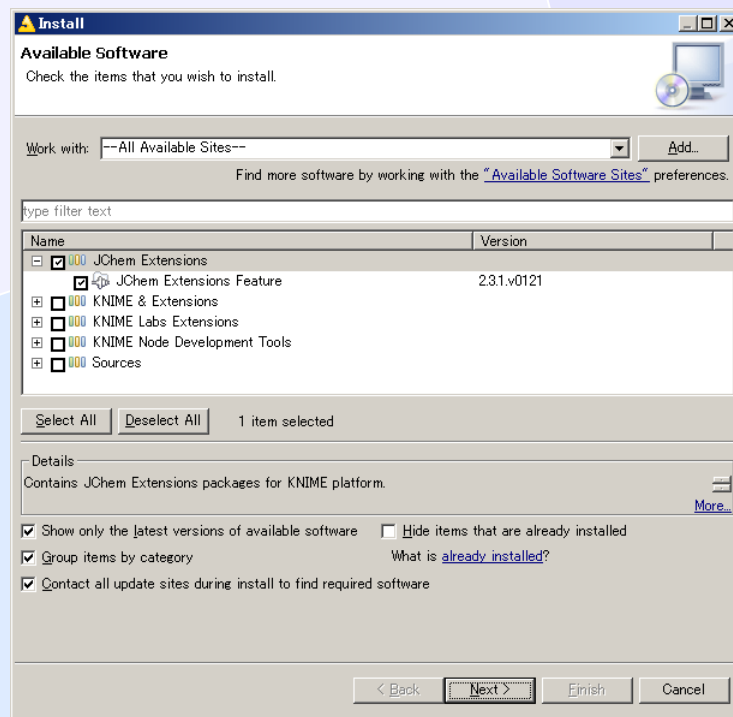
# Future Developments

- Integrate with JChem For Excel (Fall in 2011)
  - ◆ Export to the JC4XL format.
    - Export to XML (supported by Excel 2007 or later)
- Integrate with Instant JChem
  - ◆ Connect to the IJC database
  - ◆ Execute workflow from IJC
    - Integrated with KNIME Enterprise Server
- Marvin Sketch / View QuickForm nodes

# Let's try

## ■ JChem Extensions:

- ◆ [http://www.infocom.co.jp/bio/develop/jchemextension\\_en.html](http://www.infocom.co.jp/bio/develop/jchemextension_en.html)
- ◆ [http://www.infocom.co.jp/bio/develop/jchemextension\\_update\\_en.html](http://www.infocom.co.jp/bio/develop/jchemextension_update_en.html)
- ◆ Download and Install via the update site
  - Update Site : <https://www.infocom.co.jp/bio/knime/update/2.4>



## ■ KNIME:

- ◆ <http://www.knime.org/>
- ◆ <http://tech.knime.org/>

## ■ E-MAIL:

- ◆ [info-science@infocom.co.jp](mailto:info-science@infocom.co.jp)

# Acknowledgements

---

- Infocom development team
- ChemAxon
- KNIME organization

# Thanks

---

Takahiro Ohshima  
[t.ohshima@infocom.co.jp](mailto:t.ohshima@infocom.co.jp)  
Infocom Corporation