

## Cresset Nodes for KNIME v2.4.0 Release Notes

About Cresset Nodes for KNIME .....	2
Forge .....	2
Spark .....	4
XedTools Nodes .....	5
Torch/Forge Viewer .....	5
Supported Platforms .....	5
Minimum specifications .....	5
Recommended specifications to install all binaries .....	6
Installation .....	6
KNIME nodes .....	6
Application binaries .....	6
Cresset Engine Broker .....	7
Environments .....	7
Licensing .....	9
Upgrading from 2.3.0 to 2.4.0 .....	9
Change Log .....	9
V2.4.0 .....	9
V2.3.0 .....	9
V2.2.0 .....	10
V2.1.0 .....	11
V2.0.0 .....	11
V1.1.0 .....	11
Known Issues and Bugs .....	11
The Cresset Support Team .....	11

## About Cresset Nodes for KNIME

There are currently 21 nodes available from Cresset that incorporate key Cresset applications into the KNIME environment. Each node wraps the corresponding application making it available within KNIME.

All Cresset's KNIME nodes use our proprietary XED molecular mechanics force field and give realistic, low energy conformations and accurate molecular fields. However, the XED force field is optimised for molecules that exist in biological systems and hence is sub-optimal for tasks that lie away from the study of bioactivity. In particular support for non-organic elements is limited and transition states are not supported at all.

### Forge

Forge is Cresset's powerful computational suite to understand SAR and design. It is available with both a graphical user interface and as a series of command line binaries which can be operated through KNIME nodes. Additional nodes are included to enable reading and writing of Forge project files.

### Forge Align



The Forge Align node gives access to Cresset's unique ligand alignment method. It uses one or more reference molecule(s), supplied in a specified 3D conformation, as a template for the alignment of congeneric or diverse ligands. When used on a congeneric series the tool can help in library design and give a rationale for the prioritization of compounds for synthesis. Using Forge Align on a diverse set of active molecules can help define the requirements of the protein of interest, aiding the synthetic chemist in the design of new actives. Additionally the scores from the alignment routine can be used to select molecules for screening in the same way as using the dedicated [Blaze](#) virtual screening system (albeit with a lower throughput). Optionally the Forge Align node can load a protein structure that can be used as an excluded volume for the alignment.

### FieldTemplater



[FieldTemplater](#) is a tool for comparing molecules using their shape and electrostatic and hydrophobic fields in order to find common patterns. When applied to several structurally-distinct molecules with a common activity, FieldTemplater can determine the bioactive conformations and relative alignments of these molecules without requiring any protein information. Other pharmacophore generation packages simply attempt to generate a very crude idea of what the protein wants in terms of donor points, acceptor points and the like: FieldTemplater attempts to provide a full picture of how the active molecules bind, which features they use, what shape they are, and how different series can be compared. The output of FieldTemplater can be used as reference structure in the Forge Align node or saved to a project for subsequent visual inspection.

### Activity Miner



[Activity Miner](#) is a tool for generating similarity, distance or disparity matrices from a set of aligned molecules using a selection of 2D similarity methods or using Cresset's 3D field and shape similarity. Using these matrices on compounds of known activity indicates regions where the SAR is acute or has 'cliffs'. The output can be viewed in multiple third party nodes or used as the distance matrix for clustering nodes for example. The molecules must be pre-aligned if you wish to use Cresset's field and shape similarity - the Forge Align node is ideal for this.

### **Forge Build Field QSAR**



Generates a [Field QSAR](#) (3D-QSAR) model for activity from a set of aligned molecules. It determines a set of 'sample positions' around the molecules based on their field points, and then calculates the value of the molecular fields at those positions. The data matrix derived from this is then processed by partial least squares (PLS) to generate a linear model for activity. Optionally the data matrix can be exported into the KNIME workspace for use with alternative data analysis methods.

### **Forge Score Field QSAR**



Calculates predicted activities for molecules against a Field QSAR model. The molecules must be pre-aligned in a manner consistent with the alignment used to create the original Field QSAR model.

### **Forge Build Activity Atlas**



Generates an [Activity Atlas](#) model for activity from a set of aligned molecules. Activity Atlas is a new method for summarizing the SAR for a series into a visual 3D model that can be used to inform new molecule design, particularly useful for those project teams where there is not enough SAR for a traditional 3D-QSAR approach.

Activity Atlas calculates and displays as 3D visualizations:

- 'Activity Cliff Summary': what do the activity cliffs tell us about the SAR?
- 'Average of actives': what do active molecules have in common?
- 'Regions explored': where have I been? For a new molecule, would making it increase our understanding? This analysis also calculates a novelty score for each molecule.

The result is a global picture of activity, highlighting those regions around the series where the SAR is clear.

### **Forge Score Activity Atlas**



Calculates a novelty score for molecules against an Activity Atlas model. The molecules must be pre-aligned in a manner consistent with the alignment used to create the original Activity Atlas model.

### **Forge Build *k* Nearest Neighbour (kNN)**



Generates a [k Nearest Neighbour \(kNN\)](#) regression model using similarity and activity data for a set of molecules. The similarity between the molecules can be calculated using Cresset's field and shape similarity method or by using 2D descriptors such as ECFP4, ECFP6, FCFP4, or FCFP6 fingerprints.

### **Forge Score *k* Nearest Neighbour (kNN)**



Calculates predicted activities for molecules against a kNN model. If the model was generated using Cresset's field and shape similarity, the molecules must be pre-aligned in a manner consistent with the alignment used to create the original kNN model.

### **Forge Model Info**



Extracts information on a (Q)SAR model saved within a Forge project. The Forge project may have been created by the Forge Build node or read in by the Forge Project Reader node. The information is useful for choosing the number of PLS components to calculate with the Forge Score nodes or in collating information about multiple Forge Build experiments.

### **Forge Project Reader**



The 'Forge Project Reader' node is used to read a Forge project file which can be used as the input for Forge Score. The Forge project can be viewed using the Forge Project Viewer node.

### **Forge Project Writer**



The 'Forge Project Writer' node writes a Forge project generated with Forge Build to a file. The Forge application may then be used to view the Forge project outside of KNIME or the project can be used as an input to the Forge Score node.

### **Forge Project Viewer**



Launches Forge, Torch or TorchLite (depending on which application is installed on the executing machine) to view the input Forge project.

### **Forge Project Roles Extractor**



Extracts the role data out of a Forge project.

### **Forge Project Molecules Extractor**



Extracts molecules, conformers or alignments from a Forge project.

## **Spark**

[Spark](#) enables you to move to new series and non-obvious IP by swapping scaffolds and exploring accessible R-groups. Given a 'starter molecule', you can select a portion to replace and Spark will suggest a list of new molecules containing replacement moieties with similar electrostatic and steric properties. Spark comes with several databases of moieties generated from fragmentation of commercially-available compounds, active molecules reported in the literature, and available reagents. In addition you are able to generate your own databases of replacement moieties using the database generator functionality.

In KNIME the Spark nodes provide access to the scaffold hopping and R-group exploring capabilities with a specialised node to enable graphical specification of the portion of the starter molecule to be replaced. Lastly a database creation node is included.

### **Spark Database Search**



This node is the main Spark node and is used to search databases of replacement moieties for novel replacements for a portion of the starter molecule. If the starter molecule is supplied in 2D then it is first converted to the nearest 3D conformation before the search is performed. Unlike other fragment replacement methodologies, Spark works in product space so all potential replacements are merged into the original starter molecule before scoring, leading to more diverse yet realistic results.

### **Spark Fragment Selector**



The Spark Fragment Selector enables the graphical selection of the portion to be replaced by Spark in a database search experiment. It takes the first molecule in the input and displays a dialog which allows the user to select the portion of the molecule to replace. The output of this node should be connected to the input of the Spark Database Search node.

### Generate Spark Database



Generate Spark Database is a tool for generating or updating Spark databases. It reads a list of molecules and breaks them into fragments, and stores the fragments into a database file for use with the Spark Database Search node.

### XedTools Nodes

#### XedMin



XedMin is a ligand minimizer that uses Cresset's proprietary XED force field. As well as free minimization a second input to the KNIME node enables a context for the optimization to be provided. Most commonly the context for the minimization will be the active site of a protein.

#### XedeX



XedeX uses Cresset's XED force field to generate conformations for an input ligand. It uses a torsion randomisation and subsequent minimization algorithm to generate diverse, low energy conformations. In independent testing XedeX performs as well as the best available conformation hunting algorithms at reproducing bioactive conformations.

#### Torch/Forge Viewer



The Torch/Forge Viewer node is used to launch one of Cresset applications for viewing results. It will start Forge, Torch or TorchLite depending on what is installed on the executing machine.

The powerful molecular viewer, editor and sketcher shows your molecules in 3D overlaid with field patterns together with 2D structure and physicochemical properties. It was originally designed for use in evaluating and choosing results from Cresset's unique virtual screening platform 'Blaze' but has developed into a general molecule viewer and editor. Every molecule that is loaded into the viewing application is presented in the 2D molecule table together with its calculated physicochemical properties such as logP and TPSA and any data that was present in the original file (e.g. sdf file tags). The molecule table can be filtered on any of the columns present to focus the view towards molecules with specific values. Molecules that are highlighted in the 2D table are also displayed in the 3D window enabling comparison of molecules in specific alignments either side by side or overlaid with each other. Any molecule can be marked as a 'favorite' or as a 'search molecule' enabling greater control over these entries.

### Supported Platforms

Cresset KNIME nodes are supported under KNIME 2.7.4 and later releases. The tables below show the supported platforms for the KNIME nodes and the underlying command line binaries. Note that Spark requires access to Cresset supplied or internal Spark databases. The Cresset supplied databases require a minimum of 4Gb of disk space up to a maximum of 33Gb.

#### Minimum specifications

Hardware	Specifications
RAM	2GB
Processor	Intel i3 processor or equivalent
Disk space	100MB
Operating Systems	Windows 7 (32 bit) Red Hat Enterprise Linux 6 (32 bit)

## Recommended specifications to install all binaries

Hardware	Specifications
RAM	8GB
Processor	Intel i7 processor or equivalent
Disk space	16Gb
Operating Systems	Windows 7 (64 bit) Red Hat Enterprise Linux 6 (64 bit)

## Installation

The Cresset KNIME nodes wrap functionality of key binaries and hence both the application binary and the nodes must be installed. Operation of the binaries is subject to a license agreement and license file which is available from our website or directly from [Cresset Sales](#).

### KNIME nodes

To install the Cresset KNIME nodes add the Cresset KNIME update site to your software sources. To do this, start KNIME, navigate to the Help menu, choose 'Install new software'. Now click 'Add' then in the Location box enter <http://www.cresset-group.co.uk/knime/update>. Select the nodes that you wish to install and agree to the license agreement. Upon restarting KNIME, the nodes will appear under a top-level 'Cresset' directory.

If you are unable to access the Cresset hosted update site then we are happy to provide a zip file of the site for you to use internally. Please contact [Cresset support](#) for the current url.

### Application binaries

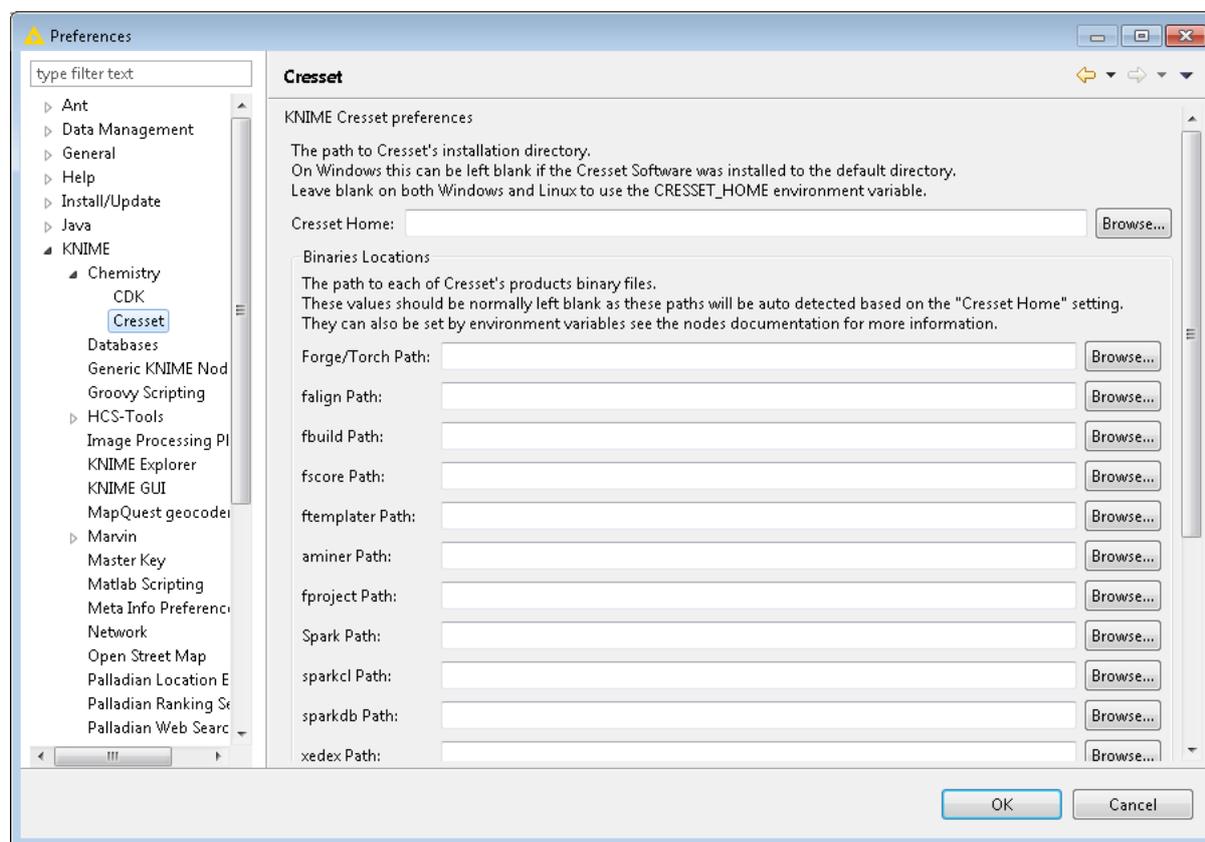
Installation of the Cresset application binaries is described in detail in the release notes for the individual binaries or in general on [our website](#). Note that the binaries require a license file or license server to operate correctly. Licenses can be obtained from Cresset ([enquiries@cresset-group.com](mailto:enquiries@cresset-group.com)). License files should be placed in one of the locations detailed [here](#) or in the KNIME application directory.

## Cresset Engine Broker

The Cresset Engine Broker can enable a number of nodes to use the power of your computing cluster to reduce the time it takes for the node to run. To use the Cresset Engine Broker set the CRESSET\_BROKER environment variable to 'hostname:port' where 'hostname' is the name or IP address of the server running the Cresset Engine Broker and 'port' is the port the Cresset Engine Broker is running on. For more information see <http://www.cresset-group.com/2014/06/cresset-engine-broker-release/>.

## Environments

To execute correctly, each protocol requires the location of the corresponding binary to be set within KNIME. This can be configured in the KNIME .preferences under the KNIME → Chemistry → Cresset menu.



Alternatively this can be achieved by setting the CRESSET\_HOME environment variable to point to the installation directory for the Creset binaries. Default values are shown in the table below.

Operating System and Install level	Default Location
<b>Windows Global</b>	%PROGRAMFILES%\Creset-BMD
<b>Windows Local install</b>	%APPDATA%\Creset-BMD
<b>Linux global install from rpm</b>	/opt/creset
<b>Linux local install from tgz</b>	\$HOME/creset

In addition to setting CRESSET\_HOME, the absolute path to each individual executable can be set using the variables shown below.

Node	Environment variable
<b>Forge Align</b>	CRESSET_FORGEALIGN_EXE
<b>Forge Build</b>	CRESSET_FORGEBUILD_EXE
<b>Forge Score</b>	CRESSET_FORGESCORE_EXE
<b>Activity Miner</b>	CRESSET_ACTIVITYMINER_EXE
<b>FieldTemplater</b>	CRESSET_FIELDTEMPLATER_EXE
<b>Spark Fragment Selector</b>	CRESSET_SPARK_EXE
<b>Spark Database Search</b>	CRESSET_SPARKCL_EXE
<b>Generate Spark Database</b>	CRESSET_SPARKDB_EXE
<b>XedeX</b>	CRESSET_XEDEX_EXE
<b>XedMin</b>	CRESSET_XEDMIN_EXE
<b>Torch/Forge Viewer</b>	CRESSET_TORCH_FORGE_EXE
<b>Creset Engine Broker</b>	CRESSET_BROKER

The licensing for the XedMin and XedeX binaries requires that either the 'cresetbmd\_LICENSE' or 'XED\_LICENSE' environment variable be set to the location of your license file.

Note also that the Spark command line binaries require access to database files or location in which to create databases. These are most commonly specified using the environment variables SPARK\_CRESSET\_DB and SPARK\_DB respectively. Full details can be found in the application [installation instructions](#).

## Licensing

The KNIME nodes require you to accept the terms of our license agreement when they are installed but do not require further license files. However, the software used by the KNIME nodes does require a valid license file. Please contact Cresset ([enquiries@cresset-group.com](mailto:enquiries@cresset-group.com)) for more information.

## Upgrading from 2.3.0 to 2.4.0

- When opening old workflows the following error may appear in the console window.

```
ERROR      LoadWorkflowRunnable      Status: Error: Unable to
insert connection "[6(2) -> 8( 1)]"
```

This happens as the type of connection for Forge projects was changed and the 'Torch/Forge Molecule Viewer' node cannot receive this new type of connection. To fix this problem, replace the 'Torch/Forge Molecule Viewer' node with the 'Forge Project Viewer' node.

- When opening old workflows the following warning may appear in the console window.

```
WARN Align molecules using Forge Align Substituted
'com.cressetgroup.knime.nodes.torch.TorchNodeFactory' for unknown factory
'com.cressetgroup.torch.TorchNodeFactory'
```

This warning can be safely ignored.

## Change Log

### V2.4.0

New preferences page under KNIME → Chemistry → Cresset for configuring binary locations and FieldEngine settings.

Added new nodes for Activity Atlas and k Nearest Neighbors.

Added new 'Forge Project Roles Extractor' and 'Forge Project Molecules Extractor' nodes for extracting data out of a Forge project.

Added setting for "Gradient Cut-off" to the 'Spark Database Search' node.

The nodes ports for Forge projects now use a special kind of connection.

Drag-and-drop of Forge and Torch project files onto a workflow automatically creates a 'Forge Project Reader' node.

### V2.3.0

New Forge model info node that reports the statistics and log file associated with a Forge model.

Added support for Cresset's engine broker environment for distribution of calculations to a Linux cluster.

New options added to the Forge Align and FieldTemplater nodes to control conformation hunting.

New option added to Spark database generator node to process molecules using all the available reagent rules.

## **V2.2.0**

New Generate Spark Database node which can generate Spark Databases to be used with the Spark Database Search node.

New FieldTemplater node which can compare molecules using their electrostatic and hydrophobic fields in order to find common patterns.

Changed the output of the Activity Miner node so that the resulting columns are appended to the input columns instead of replacing the input columns.

The Activity Miner node now has an additional output containing a Forge project. The project can be viewed using the Activity Miner GUI by using the Torch/Forge Viewer node with the 'Open Activity Miner' checkbox ticked. The project can also be written to a file using the Forge Project Writer node.

Added Read Settings to the Torch/Forge Viewer node which define how molecules should be read in e.g. as conformations or aligned molecules.

Added 'Only score molecules' option to Forge Align node. If selected then conformations will not be generated and the molecules will not be moved.

Added 'Skip Conformation Hunt' option to Forge Align node.

Added the 'Attachment types' column to the Spark Database Search node output.

Added Cresset Engine Broker Support to Forge Align, Forge Build, Activity Miner and FieldTemplater nodes.

Forge Align now defaults to 50 conformations instead of 100 when 'quick' mode is selected.

Spark Database Search can now list more than 10 databases.

Fixed an error being displayed when changing the speed option in Forge Align.

Fixed the 'Conformers of achiral molecules can be inverted' in Forge Align doing the opposite of what the help text describes.

Fixed the 'Add enantiomeric conformations' in XedeX doing the opposite of what the help text describes.

Fixed the 'ring library' in XedeX doing the opposite of what the help text describes.

Fixed the Fragment SMILES output column appearing blank when a KNIME plugin present which provides a SMILES render.

### V2.1.0

New Activity Miner node (part of the Forge collection) to calculate similarity and disparity matrices.

Forge\_align node now correctly outputs additional alignments when requested.

Spark database search node can now be incorporated into a workflow in the absence of installed databases.

Forge\_build node includes new option to specify activity as log(Ki).

Added option to control the number of processors (FieldEngines) used by Spark search node.

### V2.0.0

New nodes to wrap all Forge functionality and Spark search functionality.

Previous TorchLite node replaced by generic Torch/Forge viewer node that uses either Torch, TorchLite or Forge as the molecule viewer.

XedMin node now capable of minimising ligands within a protein cavity.

Introduced update site to facilitate software updates and smooth corporate installations.

### V1.1.0

Updated nodes to wrap new Forge Align and TorchLite binaries.

Introduced CRESSET\_FORGEALIGN\_EXE and CRESSET\_TORCH\_EXE variables to enable use of custom install locations and binaries.

Introduced automatic recognition of CRESSET\_HOME for standard installations.

### Known Issues and Bugs

Stopping the executing nodes fails to stop the associated FieldEngine processes. These will terminate themselves after a short period of time.

### The Creset Support Team

Creset  
New Cambridge House  
Bassingbourn Rd  
Litlington  
Cambridgeshire  
SG8 OSS

[www.creset-group.com](http://www.creset-group.com)  
[support@creset-group.com](mailto:support@creset-group.com)

skype: cresetgroup

[@cresetgroup](#)

Join [Creset group](#) on LinkedIn

+44 1223 85 88 90