

Cresset Nodes for KNIME v2.5.0 Release Notes

Cresset Nodes for KNIME v2.5.0 Release Notes	1
About Cresset Nodes for KNIME	3
Forge™	3
Forge Align	3
FieldTemplater™	3
Activity Miner™	3
'Models' group	4
Forge Build Field QSAR	4
Forge Score Field QSAR	4
Forge Build Activity Atlas	4
Forge Score Activity Atlas	4
Forge Build Machine Learning	4
Forge Score Machine Learning	5
Forge Model Info	6
'Project' group	6
Forge Project Reader	6
Forge Project Writer	6
Forge Project Viewer	6
Forge Project Roles Extractor	6
Forge Project Molecules Extractor	6
Surface Writer	6
Forge Molecule Viewer	6
Spark™	6
Spark Database Search	7
Spark Fragment Selector	7
Generate Spark Database	7
XedTools™	7
XedMin	7
XedeX	7
Flare™	8

Flare Viewer	8
pyflare	8
Supported Platforms	9
Minimum specifications	9
Recommended specifications to install all binaries.....	9
Supported Operating Systems	9
Installation	9
KNIME nodes.....	9
Application binaries	10
Cresset Engine Broker	10
Environments.....	10
Licensing.....	12
Upgrading Cresset Nodes for KNIME	12
Upgrading from 2.4.0 to 2.5.0.....	12
Upgrading from 2.3.0 to 2.4.0.....	12
Change Log.....	13
V2.5.0	13
V2.4.0	14
V2.3.0	14
V2.2.0	14
V2.1.0	15
V2.0.0	15
V1.1.0	15
Known Issues and Bugs.....	16
Support	16

About Cresset Nodes for KNIME

There are currently 24 nodes available from Cresset that incorporate key Cresset applications into the KNIME environment. Each node wraps the corresponding application making it available within KNIME.

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 platform (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 electrostatic, shape 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.

'Models' group

This group includes KNIME nodes for building qualitative and Quantitative Structure-Activity Relationships (QSAR) models in Forge, and scoring new molecules against (Q)SAR models.

Forge Build Field QSAR

Generates a [Field QSAR](#) (3D-QSAR) model for activity from a set of aligned molecules. The aligned set of molecules is used to derive a set of 'sample positions' around the molecules based on their field points, which can be used to probe any molecule for the electrostatic potential or for the volume taken up at those positions. The data matrix derived from the sample values 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:

- Average of Actives: shows you what the average active molecule looks like, by making an analysis of what the active molecules in the data set have in common.
- Activity Cliff Summary: shows you the critical regions of the SAR, based on activity cliffs.
- Regions Explored analysis: makes an assessment of what regions of the aligned molecules have been fully explored and calculates a novelty score for each molecule in the data set. The Regions Explored analysis can be used within the 'Forge Score Activity Atlas' node to compute a 'Novelty' score for new molecules.

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 Machine Learning

Generates a Machine Learning regression or classification model from a set of molecules.

Regression methods are useful when we have real values as the biological activity data (for example, pIC50, pKi), and we aim to calculate a function that can be used to predict further activity values. Classification methods are useful to analyze categorical data, i.e. when the biological data takes the

form of qualitative labels (for example, 'high', 'medium', 'low), or of activity ranges (for example, ranges of % of inhibition) which can be appropriately converted into categories.

The molecules must be pre-aligned for all model types except kNN using 2D descriptors (see below).

The models which can be generated are:

Auto - All the model types will be generated and the best model will be picked for the output. For regression models, this means keeping the model with the highest q^2 ; for classification models, the model with the highest Informedness value.

k Nearest Neighbor (kNN) - The kNN methodology is a well-known and robust distance learning approach where the activity for each new compound is predicted as the weighted average activity of its k nearest neighbors in the training set.

The similarity between the molecules is calculated using Cresset's field/shape similarity method or by using the 2D circular fingerprint methods ECFP4, ECFP6, FCFP4, or FCFP6.

Random Forest (RF) - The aligned set of molecules is used to derive a set of sampling points around the dataset that can be used to probe any molecule for the electrostatic potential or for the volume taken up by molecules. The data matrix derived from the sample values is then processed to generate a RF model for activity. Optionally the data matrix can be exported into the KNIME workspace for use with alternative data analysis methods.

Only Random Forest regression models can be generated.

Relevance Vector Machine (RVM) - The aligned set of molecules is used to derive a set of sampling points around the dataset that can be used to probe any molecule for the electrostatic potential or for the volume taken up by molecules. The data matrix derived from the sample values is then processed to generate a RVM model for activity. Optionally the data matrix can be exported into the KNIME workspace for use with alternative data analysis methods.

Regression or classification models can be generated by setting the activity unit type in the node configuration dialog.

Support Vector Machines (SVM) - The aligned set of molecules used to derive a set of sampling points around the dataset that can be used to probe any molecule for the electrostatic potential or for the volume taken up by molecules. The data matrix derived from the sample values is then processed to generate a SVM model for activity. Optionally the data matrix can be exported into the KNIME workspace for use with alternative data analysis methods.

Regression or classification models can be generated by setting the activity unit type in the node configuration dialog.

Forge Score Machine Learning

ML

Calculates predicted activities for molecules against a Machine Learning model. Except for 2D kNN models, the molecules must be pre-aligned in a manner consistent with the original model to make sensible predictions.

Forge Model Info



Extracts information on a Field QSAR or kNN 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/neighbors to calculate with the Forge Score nodes or in collating information about multiple Forge Build experiments.

'Project' group

This group includes KNIME nodes for reading, writing, viewing and extracting information from Forge projects.

Forge Project Reader



Reads 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



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 to view the input Forge project.

Forge Project Roles Extractor



Extracts the role information out of a Forge project.

Forge Project Molecules Extractor



Extracts molecules, conformers or alignments from a Forge project.

Surface Writer



Writes the Positive, Negative, Van der Waals and Hydrophobic surfaces for molecules to a directory. The surfaces can be written in Cube, CCP4, Insight or MOE format.

Forge Molecule Viewer



The Forge Molecule Viewer launches Forge for viewing results.

Spark™



[Spark](#) finds biologically equivalent replacements for key moieties in your molecule and guides you through the process of growing and linking ligands. 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 can 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™

The [XedTools](#) provide access to XED, one of the most innovative molecular mechanics force fields in existence. XED uses off atom charges on electronegative atoms to give a more accurate representation of the charge density surrounding an atom, resulting in an improved description of molecular interactions.

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.

Flare™



[Flare](#) is a structure-based design tool for the visualization and analysis of protein-ligand complexes.

Current capabilities in Flare include:

- Smart import and handling of protein-ligand complexes
- Accurate and reliable protein preparation
- Sequence alignment and superimposition
- Protein interaction potentials
- Electrostatic complementarity™
- Accurate docking using Lead Finder™
- Minimization of protein ligand complexes
- 3D-RISM water analysis (using Cresset XED and AMBER FFs)
- WaterSwap analysis for analyzing ligand and binding site energetics
- The Flare Python™ API.

In KNIME the Flare nodes provide access to all Flare functionality through the Flare Python API. An additional viewer node is included to enable visualization of ligand, proteins and results.

Flare Viewer



The Flare Viewer node is used to launch Flare for viewing ligands, proteins and results.

pyflare



The pyflare node to use the Flare Python API from within KNIME. This enables access to all Flare functionality including protein preparation, docking and Electrostatic Complementarity.

The full documentation for the Flare Python API can be found in the Flare installation directory.

Supported Platforms

Cresset KNIME nodes are supported under KNIME 3.1.2 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 6GB of disk space up to a maximum of 40GB.

Minimum specifications

Hardware	Specifications
RAM	2GB
Processor	Intel i5 processor or equivalent
Disk space	1GB

Recommended specifications to install all binaries

Hardware	Specifications
RAM	8GB
Processor	Intel i7 processor or equivalent
Disk space	16GB

Supported Operating Systems

Forge KNIME nodes are supplied as 64bit binaries for Windows® and Linux®.

Windows®	Linux®
Windows 7	Red Hat Enterprise Linux 6.10+ CentOS 6.10+
Windows 10	Red Hat Enterprise Linux 7

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 requires a valid license file. Please contact Cresset (enquiries@cresset-group.com) for more information.

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

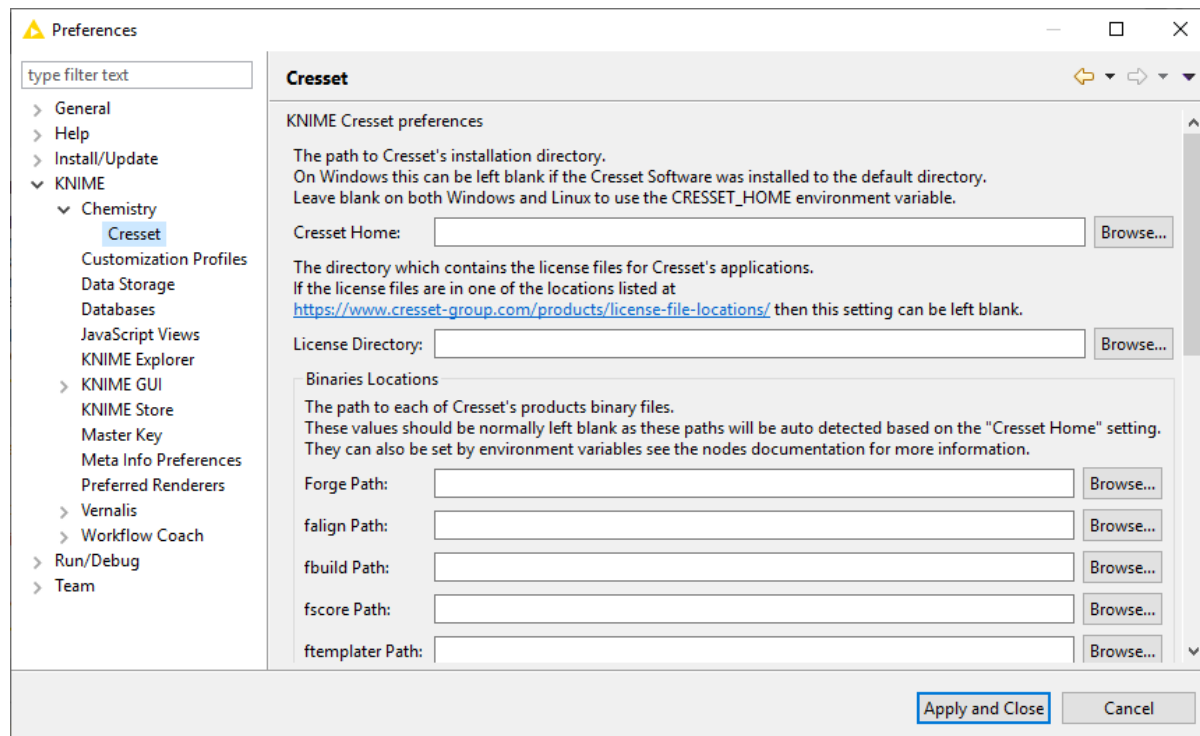
Installation of the Cresset application binaries is described in detail in the release notes for the individual binaries or in general on [our website](#). Note that the binaries require a license file or license server to operate correctly: please contact Cresset (enquiries@cresset-group.com) for more information. 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 several 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.

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 Cresset binaries. Default values are shown in the table below.

Operating System and Install level	Default Location
Windows Global	%PROGRAMFILES%\Cresset-BMD
Windows Local install	%APPDATA%\Cresset-BMD
Linux global install from rpm	/opt/cresset
Linux local install from tgz	\$HOME/cresset

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

Node	Environment variable
Forge Align	CRESSET_FORGEALIGN_EXE
Forge Build	CRESSET_FORGEBUILD_EXE
Forge Score	CRESSET_FORGESCORE_EXE
Activity Miner	CRESSET_ACTIVITYMINER_EXE
FieldTemplater	CRESSET_FIELDTEMPLATER_EXE
Surface Writer	CRESSET_FORGESURFACE_EXE
Spark Fragment Selector	CRESSET_SPARK_EXE
Spark Database Search	CRESSET_SPARKCL_EXE
Generate Spark Database	CRESSET_SPARKDB_EXE
XedeX	CRESSET_XEDEX_EXE
XedMin	CRESSET_XEDMIN_EXE
Forge Molecule Viewer	CRESSET_FORGE_EXE
Flare Viewer	CRESSET_FLARE_EXE
pyflare	CRESSET_PYFLARE_EXE

Cresset Engine Broker	CRESSET_BROKER
------------------------------	----------------

The licensing for the XedMin and XedeX binaries requires that either the 'cressetbmd_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) for more information.

Upgrading Cresset Nodes for KNIME

Upgrading from 2.4.0 to 2.5.0

- Cresset KNIME nodes 2.5.0 now require KNIME 3.1.2 or newer.
- The 'Forge Build k Nearest Neighbor' has been deprecated, use the 'Forge Build Machine Learning' node instead.
- The 'Forge Score k Nearest Neighbor' has been deprecated, use the 'Forge Score Machine Learning' node instead.

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

V2.5.0

- Dropped support for older versions of KNIME. KNIME version 3.1.2 or newer is now required.
- New Flare Viewer node to view ligands, proteins and results in Flare.
- New pyflare node to use the pyflare Python interpreter with ligands and proteins in a KNIME workflow.
- New Surface Writer node to write molecule surfaces to a directory.
- New Forge Build Machine Learning to create QSAR models using various Machine Learning methods.
- New Forge Score Machine Learning nodes to calculate predicted activities for molecules using a Machine Learning model.
- Deprecated the Forge Build k Nearest Neighbor (kNN) and Forge Score k Nearest Neighbor (kNN) nodes. The Forge Build Machine Learning and Forge Score Machine Learning nodes should be used instead.
- The new Weighted Sum Activity Cliffs Summary algorithm is now the default in the Forge Build Activity Atlas node.
- The XedMin, Forge Align, Flare Viewer and pyflare nodes now accept proteins in PDB format.
- New option to set pharmacophore constraints to the Forge Align node.
- New option to set field and pharmacophore constraints to the Spark Database Search node.
- New option to remove boats conformations before performing the alignment in the Forge Align node.
- In the Spark Database Search node, multiple reference molecules can now be used to help guide the search.
- In the Spark Database Search node, users can now specify a database to search in addition to the databases it autodetects.
- The Forge Project Roles Extractor node now also outputs additional data for each role.
- Forge Build Activity Atlas node now also writes the Activity Atlas surfaces to a directory.
- Added Weighting Scheme option to the Forge Build Machine Learning node in support of kNN models.
- The Forge Build Activity Atlas node can now read similarity data from a column in the input table.
- The Forge Align node and Spark Database Search node now support the Tanimoto and Tversky similarity metric in addition to the default Dice metric.
- The nodes Forge Align, Forge Build Machine Learning, Activity Miner, Forge Molecule Viewer and Generate Spark Database now accept molecules in SMILES format.
- The 'Torch/Forge Molecule Viewer' node has been renamed to 'Forge Molecule Viewer'.

- Fixed Field Engines not being stopped when a node is cancelled.
- Fixed bug which prevented 2D kNN models from being built.

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 $\log(K_i)$.

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

There are no known bugs or issues.

Support

Cresset
New Cambridge House
Bassingbourn Road
Litlington
Cambridgeshire, SG8 0SS, UK
SG8 0SS

+44 (0)1223 858890
support@cresset-group.com
www.cresset-group.com
Twitter: [@cressetgroup](https://twitter.com/cressetgroup)
LinkedIn: [Cresset group](https://www.linkedin.com/company/cresset-group)
+44 1223 85 88 90