PerkinElmer
For the Better

Lead Discovery 5.1

Installation Guide

Powered by TIBCO Spotfire®

Last Updated: April 10, 2013

Table of Contents

1 Introduction

This guide explains how to install and set up Lead Discovery modules at your site. These modules provide the following extensions to the TIBCO Spotfire® core functionality:

- Ability to open an SDFile and view its fields in the table visualization, including the molecular structures.
- Ability to export data to an SDFile or a ChemDraw for Excel file.
- Ability to use the Structure Viewer to view the structures, create, export, and import ID lists.
- Ability to use the Structure Filter to filter a data table based on structure.
- Perform an R-Group Decomposition on a substructure search.

Note: The Structure Viewer and Structure Filter are not available in the Web Player.

- Ability to search on structures using the similarity and substructure search algorithms of the supported chemistry cartridges.
- For more information about system requirements for Lead Discovery see:

http://www.cambridgesoft.com/services/EnterpriseSupport/KnowledgeBase/SystemRequirements/?fid=230

2 Prerequisites

This Installation Guide explains the steps required to deploy and set up Lead Discovery, based on the assumption that you have previously performed the following tasks:

1. Installed the Spotfire Server at your site.
2. Deployed the core Distribution Package and Installation Documentation to the Spotfire Server.
3. Installed one or more of the supported structure renderers on the client computers. For more information about supported renderers and editors see:

http://www.cambridgesoft.com/services/EnterpriseSupport/KnowledgeBase/SystemRequirements/?fid=230

Note: Lead Discovery 5.1 includes a dedicated ChemDrawActiveX Enterprise chemistry renderer. It will be automatically installed and activated on the client computers during the first use of Lead Discovery. It is therefore not necessary to independently install ChemDraw on the client computers in order to render chemical structures in Lead Discovery. The automatic installation of the ChemDraw renderer should not interfere with pre-existing ChemDraw installations on the client computer. However, Lead Discovery will not make use of pre-existing versions of ChemDraw.

Tip: Tasks 1 and 2 are explained in detail in the *TIBCO Spotfire® Installation and Configuration Manuals* and *TIBCO Spotfire® Deployment and Administration Manual*

3 Installation

This chapter explains the installation procedures for installing Lead Discovery 5.1.

3.1 Installation Overview

The actual features and functionality of Lead Discovery are implemented in software packages. These packages are bundled into a distribution which must be deployed on the TIBCO Spotfire Server and Web Player Server.

When the end users start Spotfire on their computers, they will log into the Spotfire Server and download the deployment locally, automatically downloading any newer or added packages. Depending on the licenses that are enabled for each user, varying features may appear in each user's application.

There are some initial steps you must perform when deploying Lead Discovery on a Spotfire Server. These are:
1. Deploying the distribution package.
2. Setting up licenses.
3. Configuring information links.
4. Verifying the configuration.

Please follow the instructions in this guide carefully.

3.2 Deploying the Distribution Package

Note: For additional details on deploying new packages, refer to the *TIBCO® Spotfire® Deployment and Administration Manual*.

To deploy the extension package to the server:

1. Open the Administration Console located at http://spotserver/spotfire/administration (where spotserver is the name of one of your Spotfire servers) in a web browser.
2. Log in as a Spotfire Admin.
3. Select the Deployment tab.
4. Select the Environment to which you want to deploy Lead Discovery.

   Note: The Environment you select must have a functioning TIBCO Spotfire deployment.

5. Click **Add**. The Add to Deployment dialog appears.
6. Click **Browse**. The File Upload dialog appears.
7. Browse to the LeadDiscovery.spk file you want to include in the deployment.
8. Click **OK** in the Add to Deployment dialog to upload and add the file to the distribution. The list of packages is updated with the contents of the file.
9. Click **Validate** in the lower left corner to make sure the deployment is not damaged or corrupted in any way.
10. Click **Save** in the lower left corner to save and publish the deployment. The Save Deployment dialog appears.
11. Enter a version number and a description for the deployment, for instance, 5.1 and Spotfire with Lead Discovery and click **OK**.

3.3 Setting up Licenses

All Lead Discovery users must have certain license functions enabled in order to use the all new functionalities like viewing structures, opening SDFiles, exporting data to SDFiles and conducting a structure search. If you are using anonymous/preconfigured authentication, then the preconfigured single user that has been set up must have these license functions.

You can configure licenses from the Spotfire Administration Manager.

At present, the following license functions are relevant for Lead Discovery:

- Access to Extensions - this license function is needed in order to use any extension to TIBCO Spotfire.
- Accord - this license function is needed for a user to be able to change the settings for chemical structures in the Spotfire visualizations using the Accord Chemistry Control renderer.
- ChemDraw - this license function is needed for a user to be able to change the settings for chemical structures in the Spotfire visualizations using the ChemDraw renderer and editor.
- ChemIQ - this license function is needed for a user to be able to change the settings for chemical structures in the Spotfire visualizations using the ChemIQ renderer.
- ChemistryViewer - this license function is needed to open the Structure Viewer from the menu and toolbar in new analyses.
- ExportToSDF - this license function is needed for a user to be able to reach the Export to SDFile functionality from a Spotfire table visualization.
- ExportToCDExcel – this license function is needed for a user to be able to reach the Export to ChemDraw for Excel functionality from a Spotfire table visualization.
- Marvin - this license function is needed for a user to be able to change the settings for chemical structures in the Spotfire visualizations using the MarvinSketch renderer and editor.
- SDFParser - this license function is needed to open an SDFile in a new analysis.
- Structure Filter – this license function is needed to open the Structure Filter from the menu and toolbar in new analyses.
- StructureSearch - this license function is needed to reach the structure search functionality.
- Symyx® Draw (MDL) - this license function is needed for a user to be able to change the settings for chemical structures in the Spotfire visualizations using the Symyx® Draw (MDL) renderer and editor.

3.3.1 Configuring License Functions

To configure license functions:

1. Start Spotfire and log in as an administrator.
2. Select Tools > Administration Manager.
3. Select the Groups and Licenses tab.
4. Click on the Licenses tab in the right hand pane.
5. For each group of users that will use Lead Discovery, click the Edit button, select the check boxes for the above mentioned license functions and click **OK**.

3.4 Web Player Considerations

If Lead Discovery is going to be used by Spotfire Web Player users, you also need some preparations to the Spotfire Web Player server.

1. Use the upgrade tool to upgrade the Web Player server with the new packages. See the *TIBCO Spotfire Web Player Installation Manual*, chapter "*Deploying Extensions and Upgrades*" for more information.
2. The authors of Spotfire analysis documents have control over which chemistry renderers are used for each visualization feature or application panel. The specific renderer(s) designated in the analysis document will be used by the Web Player Server to deliver chemical structure images to the end-user's web browser. The Web Player end-user does not need to have a chemistry renderer installed. However, any renderers that may potentially be used in analyses documents must be installed on the Web Player server.

Note: A ChemDrawActiveX Enterprise installer is provided with the Lead Discovery distribution. It should be installed on each Web Player server because the ChemDraw renderer is automatically delivered to all Lead Discovery users. The Web Player will only make use of the provided ChemDrawActiveX control. It is not sufficient to install any other version of ChemDraw on the Web Player.

3. Using Lead Discovery with the Web Player requires the use of a renderer that supports the Operating System the Web Player runs on. Check the system requirements for the renderer before installing. For renderer system requirements please refer to the respective renderer provider.
4. The chemistry renderers run on the Web Player Server in a dedicated 32-bit process. You must therefore install a 32-bit compatible version of the renderers, even if the Web Player Server OS runs on 64-bit processor.
5. For some renderers it may also be important to consider the user account under which the renderer runs. For example a renderer may require running under a designated user account which has been setup with appropriate environment variables and file system access. When installing renderers pay attention to whether they are installed for use by a specific user or by all users of the system. The user account under which the renderers will execute is indirectly controlled by the Identity of the account used to run the Microsoft Internet Information Services Application Pool associated with the TIBCO Spotfire Web Player. Refer to IIS administration documentation for modifying the Application Pool Identity. By default the application pool will run as the built-in NETWORKSERVICE Windows account.
6. It is recommended that you Stop IIS before installing renderers to the Web Player.

4 Configuration

To search for structures and also to connect to large structure databases through the tools of Lead Discovery**, it is necessary to set up information links and filters** with a set of specific properties using the Information Designer. Please follow the instructions below carefully to ensure that all functionality will be available.

Chemical structures loaded into Spotfire in ChemDraw's native file format (CDX, CDMXL) are supported for chemical rendering, chemical searching and exporting. This content in data tables will be displayed natively in renderers that support the format (ChemDraw) or converted to MOLfiles or SMILES for renderers that do not offer support. Query molecules will be extracted from renderers that support the format (ChemDraw) and sent unchanged to cartridges that support the format or converted to MOLfiles or SMILES for cartridges that do not.

When defining Information Links for Structure Query, it is necessary to add a new property (<prefix>.Cartridge) to the Information Link. When this property is set to CSCartridge, the native cdx format is preserved.

Instructions for configuring this property are included in the respective sections below.
Note: Make sure that one or more of the supported structure data sources are installed at your site. The current release of Lead Discovery supports a number of different Oracle cartridges. For more information about the supported cartridges see:

http://www.cambridgesoft.com/services/EnterpriseSupport/KnowledgeBase/SystemRequirements/?fid=230

Tip: For more information about creating and configuring information links, refer to the following sections of the *TIBCO Spotfire - User's Manual*: “Information Links”, “Data Sources”, “Folders”, “Column Elements”, “Filter Elements”.

4.1 Prerequisites

Before you start configuring the information links for Lead Discovery, you must complete the following tasks:

1. **Setting up the data sources.**

Enter the information required to connect to the structure databases through the Data Source tab in the Information Designer.

The Connection URL template for an Oracle JDBC driver is as follows:
jdbc:oracle:thin:@<host>:<port>:databaseName=<database>

For an Oracle DataDirect driver, the template to use is:
jdbc:tibcosoftwareinc:oracle://<host>:<port1521>;SID=<sid>

Add your own host name, port number and database name to the specified locations.
The chemistry cartridges require the following settings on the **Data Source tab**, under **Connection initialization**:

- **For Symyx® Direct (MDL**):
  - **Version 6.0 or later:** Select 1 from dual
  - **Version 5.0 or earlier**: Select cdcaux.ctenvinit('schema.dbname') from dual (Replace 'schema.dbname' with the name of your schema and database).
- **For Accord:** Select 1 from dual
- **For CambridgeSoft CSCartridge:** Select 1 from dual
- **For JCHEM:** Select jchem_core_pkg.set_password('<schema password>') from dual
- **For ChemXtra:** Select 1 from dual
- **For GGA Bingo:** Select 1 from dual

Tip: Do not terminate the above initialization SQL statements with a semicolon. The Information Designer tool may add such a semicolon during the save operation. When editing the information link you may need to remove the added semicolon in order to avoid "invalid character" Oracle errors caused by a double semicolon at the end of the initialization SQL.

Tip: For more details about data sources in Spotfire, refer to section "Data Sources" of the *TIBCO Spotfire - User's Manual.*

Note: To configure the information links for Lead Discovery successfully and to be able to view and search structures, you need to know the columns in the database tables which store the following information:

- Chemical structures
- Unique identifiers of the structures

2. **Creating folders for storing elements and set permissions.**

Information links, and the elements that are used for creating them, are stored in the library. Different groups of users can be given different levels of access to the data by specifying what permission level each user group should have for each folder in the library.
Create a new folder for the elements required for configuring the information links for Lead Discovery, or add these elements to an existing folder.
Tip: For more details about folders in the Spotfire library, refer to section "Folders" of the *TIBCO Spotfire - User's Manual*.

3. **Combining tables by creating joins.**

If you want to work with data from different database tables, you first need to create joins.
Tip: For more details about joins in Spotfire, refer to section “Joins” of the *TIBCO Spotfire - User's Manual*.

4.2 Configuration Overview

To configure Lead Discovery, you must define a coherent set of elements:

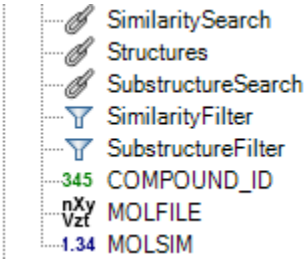
1. **Define column elements** from the columns of the database tables which return:
   - Chemical structures
   - Unique identifiers of the structures
   - Similarity scores for chemical structures
2. **Create filter elements** used for each search algorithm:
   - Substructure search filter
   - Similarity search filter

   These filter elements correspond to the WHERE-statements in the resulting SQL used to retrieve the results from a search.

3. **Create information links:**
   - Information link to use for retrieving structures based on the list of identifiers
   - Information link to use for substructure search
   - Information link to use for similarity search
   - Information link to use for structure hit highlighting

   The information links required for a structure search always come in pairs of one retrieve link and one structure search link. These two information links must have a matching prefix on their defined properties.

The definition of the elements, including specification of certain properties and prompts, is required to make the search and retrieve functionality in Lead Discovery work. Read more about this in the following chapters.
The resulting tree structure in the library should look like the following one:



All these configurations are performed using the Information Designer tool. See the following chapters for details.

To reach the Information Designer:

1. Select **Tools > Information Designer**. The Information Designer opens.

4.3 Defining Column Elements

Define the column elements to use when creating the required information links.

Column Containing	Property Name	Property Value
Chemical structures	ContentType	chemical/x-mdl-molfile chemical/x-mdl-chime chemical/x-daylight-smiles chemical/x-cdx
Unique identifiers of the structures	<prefix>.Column	ID
Similarity scores for chemical structures	<prefix>.Column	SimilarityScore

Note: If the column contains cdxml structures, select chemical/x-cdx for the Content Type value.

Details for defining these column elements follow below.

4.3.1 Defining Chemical Structure Elements

The information below is an example of how to set up the column element containing chemical structures for a Symyx® Direct (MDL) cartridge, creating a molfile structure column. If you are using another cartridge or another MIME type, the entered expressions will be different. See the documentation from your cartridge provider for more information about valid expressions.

To add a column with chemical structures:

1. In the Information Designer, click **New** and select **Column**. The Column Element tab opens.
2. In the Data Sources tree, select the column (leaf node) containing chemical structures.
3. Click **Add >**. The selected element is added to the Source columns list on the Column Element tab.
4. If you are accessing a Symyx® MDL cartridge, type the following in the **Expression** field:
   molfile(%1)

   If you are accessing an IDBS ChemXtra cartridge, type the following in the Expression field:
   CONVERT_MOLECULE(%1, IDBS_MOLECULE.MOLFILE)

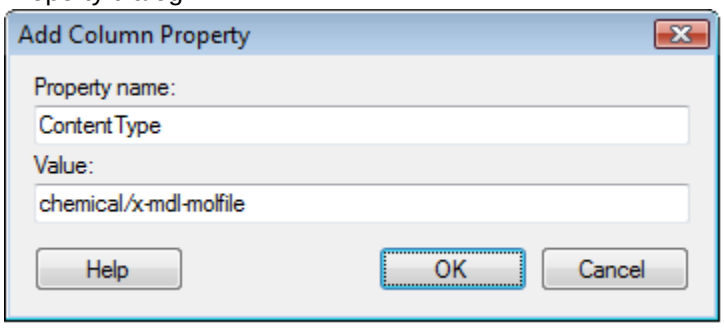
   Comment: For other cartridges, leave %1 in the expression field and proceed to the next step. If you want to convert the column to other MIME types, please see the documentation from the cartridge provider for more information.

5. Select **Clob** as the Data type of the column element.

Comment: For other cartridges this element may be of the Blob data type. See the documentation from your cartridge provider for details on what data type your data source or the expression used returns. The Information Designer automatically displays the returned data type from the data source when you add a single column to the column element tab but it is not updated as you add an expression.

6. In the **Description** field, type your own description of the column.
7. In the **Properties** section, click  to show the column properties controls.
8. Click **Add** to add the column property.

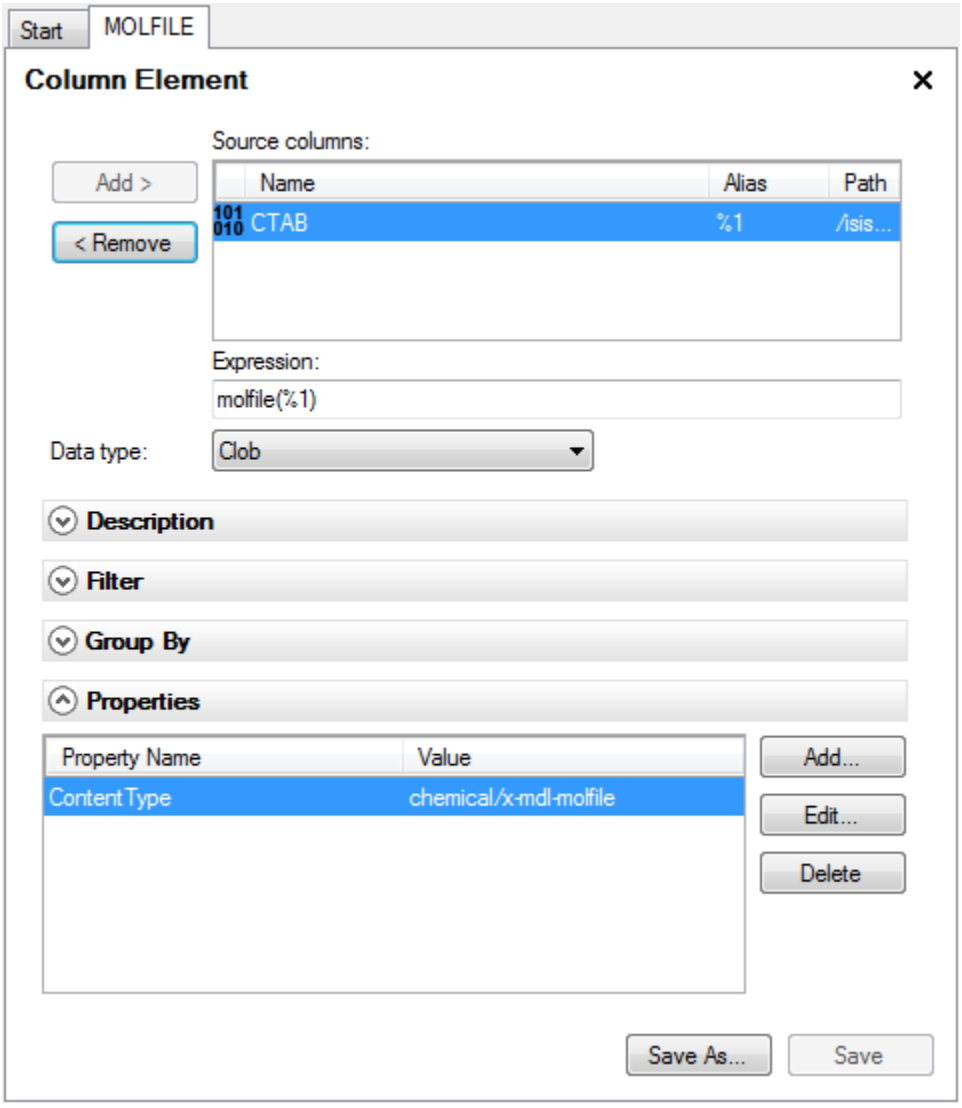
9. Type **ContentType** for the **Property** name and **chemical/x-mdl-molfile** for the **Value** in the Add Column Property dialog:



Note: No prefix is allowed for the property name of the structure column.

Comment: The value must be a MIME type string as defined in the table above. Make sure that the content type matches the output from the expression defined above.

10. Click **OK**. The column property displays in the properties list.



11. Click **Save**.
12. Click to select the folder where you want to save the column element.
13. Type a descriptive column name (for example, **MOLFILE**). The column is saved in the library and is shown with an icon denoting its type in the Elements tree.

To activate structure hit highlighting for a CSCartridge:

When performing a chemical substructure search the resulting structures can be returned with the query molecule highlighted.

To activate substructure highlighting requires the modification to the Chemical Structure Element settings as described below.

Chemical Structure Element is used to retrieve structures. It is called by the search structure, search similarity and it can also be used by the Structure Viewer to retrieve a structure, In certain cases, it is not possible to highlight the substructure because there is not an exact match.

For example, if a similarity search results with a 70% degree of similarity, the resulting structure may not contain the exact substructure used for the search. In this case, the substructure will NOT be highlighted.

Comment: The Structure Hit Highlighting is only supported by CSCartridge.

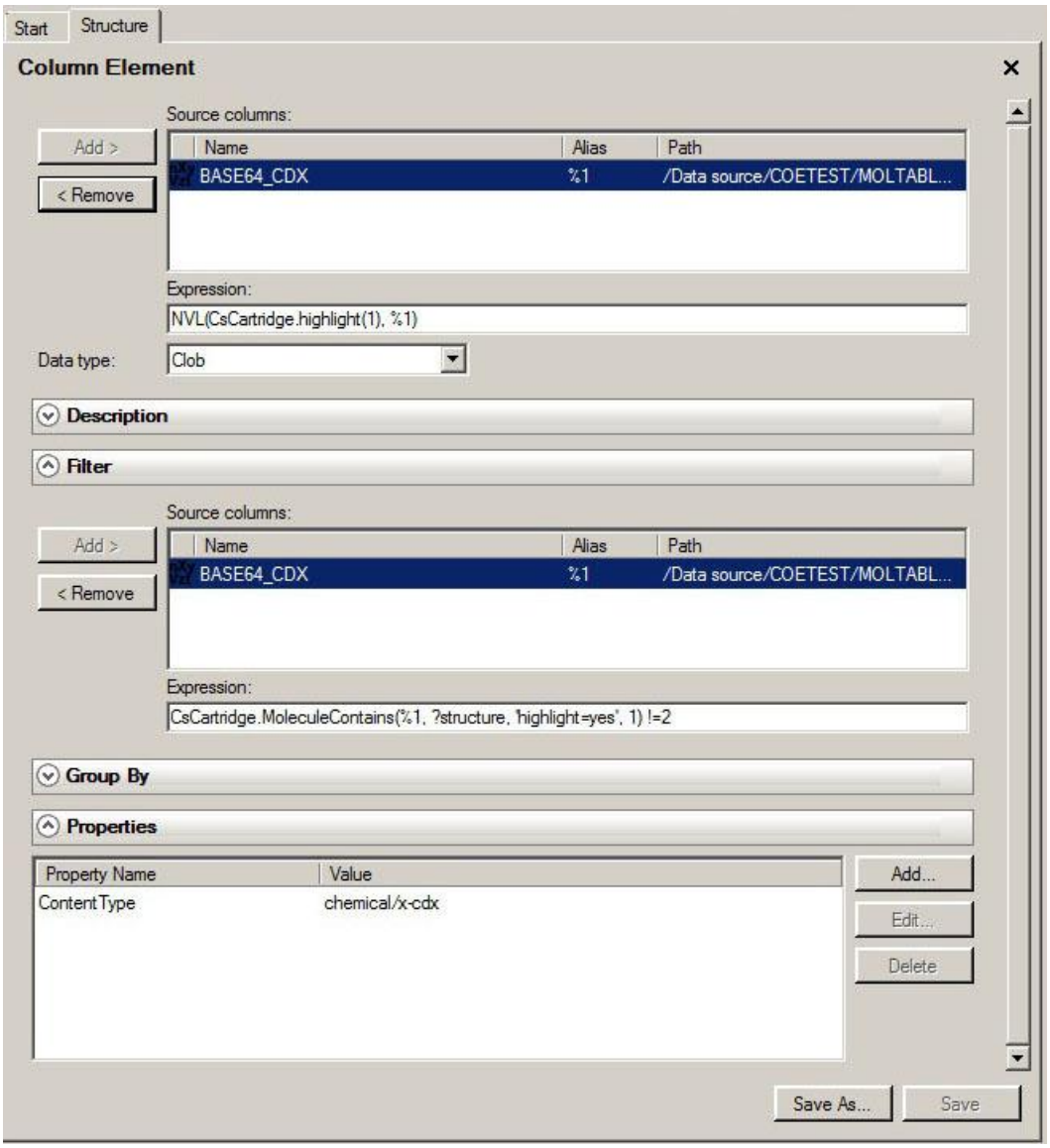
1. In the Information Designer, click **New** and select **Column**. The Column Element tab opens.
2. In the **Data Sources** tree, select the column (leaf node) containing chemical structures indexed by the cartridge.
3. Click **Add >**. The selected element is added to the Source columns list on the Column Element tab.
4. In the **Description** field, type your own description of the column.
5. Enter the following in the Expression field:

NVL(CsCartridge.highlight(1), %1)

6. In the **Filters** section, click  to show the filter controls.
7. In the **Data Sources** tree, select the column (leaf node) containing chemical structures indexed by the cartridge.
8. Click **Add >**. The selected element is added to the Source columns list on the Column Element tab.
9. Enter the following in the Expression field:

CsCartridge.MoleculeContains(%1, ?structure, 'highlight=yes', 1) !=2

10. In the **Properties** section, click  to show the column properties controls.
11. Click **Add >**. Enter the mime type (for the Structure Column) as the Content Type property value. In the example shown below, chemical/x-cdx is entered.
12. Click **'OK'**. The column property displays in the properties list.
13. Click '**Save**'.



To de-activate structure hit highlighting for a CSCartridge:

1. If you are using CSCartridge, and do not want to active the structure hit highlighting feature, the following expression must be entered in the **Expression** field:

   %1

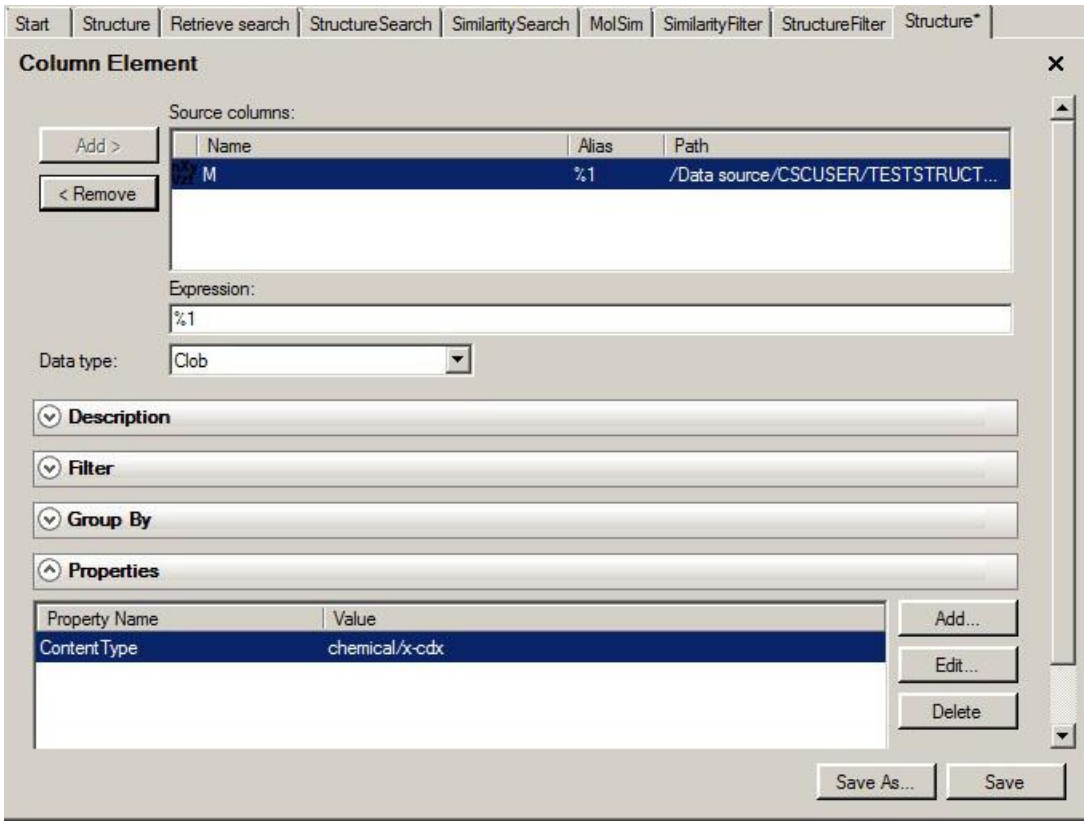
2. Select **Clob** as the Data type of the column element.

Comment: A filter is not required.

3. In the **Properties** section, click  to show the column properties controls.

4. Click **Add >**. Enter the mime type (for the Structure Column) as the Content Type property value.

5. Click **'OK'**. The column property displays in the properties list.
6. Click '**Save**'.

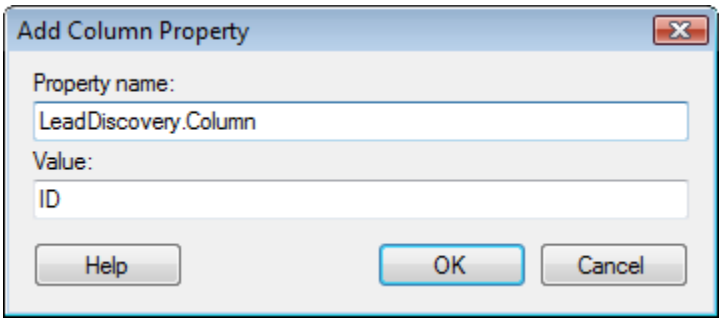


Note: The returned Content type is chemical/x-cdx when using the CSCartridge with no highlighting.

4.3.2 Adding Structure Identifier Elements

To add a column with structure identifiers:

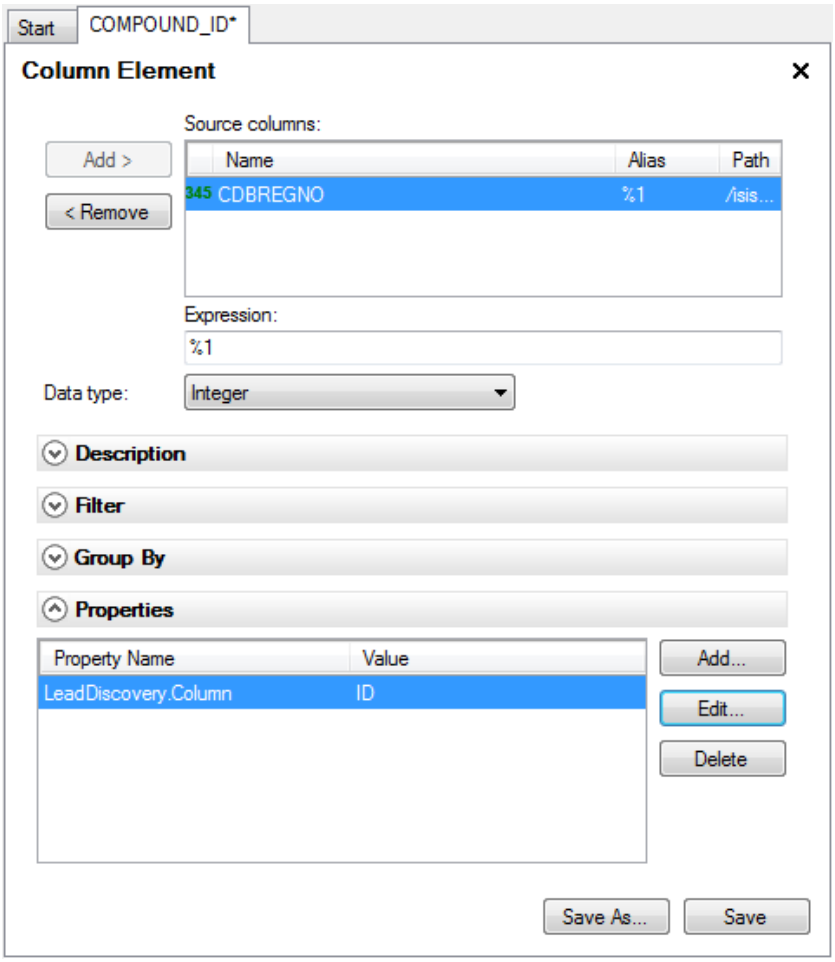
1. In the Information Designer, click **New** and select **Column**. The Column Element tab is opened.
2. In the **Data Sources** tree, select the column (leaf node) containing identifiers of chemical structures.
3. Click **Add >**. The selected element is added to the Source columns list on the Column Element tab.
4. In the **Description** field, type your own description of the column.
5. In the **Properties** section, click to show the column properties controls.
6. Click **Add** to add the column property.
7. Type **<prefix>.Column** for the **Property Name** and ID for the **Value** in the **Add Column Property** dialog.



Note: While the second part of the property name is pre-defined and you must assign it without changes (for the identifier column, this is Column), you can use an arbitrary prefix for the element properties (for example, **LeadDiscovery**).

One set of elements (as shown in the Configuration Overview chapter) should always use one and the same property prefix. If you want to configure alternative information links you can create more than one set of elements (e.g., to set up information links running against different data tables). However, each set of elements should always use its own property prefix.

8. Click **OK**. The column property displays in the properties list.



9. Click **Save**.
10. Click to select the folder where you want to save the column element.
11. Type a descriptive column name (for example, **COMPOUND_ID**). The column is saved in the library and is shown with an icon denoting its type in the Elements tree.

4.3.3 Creating Similarity Score Elements

The information below is an example of how to set up the similarity score column element for a Symyx® Direct (MDL) cartridge. If you are using another cartridge the entered expressions will be different. See the table below this step instruction or the documentation from your cartridge provider for more information about valid expressions.

To add a column with similarity score values:

1. In the Information Designer, click **New** and select **Column**. The Column Element tab opens.
2. In the **Data Sources** tree, select the column (leaf node) containing chemical structures indexed by the cartridge.
3. Click **Add >**. The selected element is added to the Source columns list on the Column Element tab.

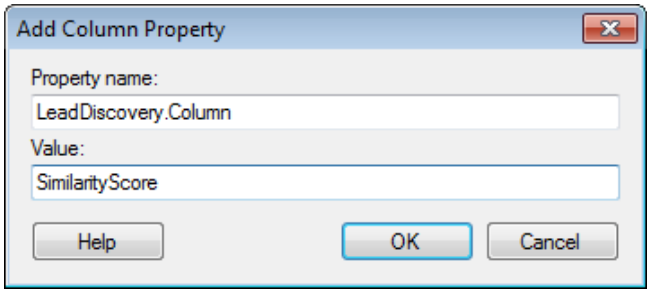
4. In the **Description** field, type your own description of the column.
5. Type the following expression in the **Expression** field:

   molsim(%1, ?structure, ?similarity_type)

   Where **molsim** is the Symyx® Direct (MDL) similarity search operator extracting similarity values for chemical structures, structure is the parameter for the query structure. And **similarity_type** is the parameter for the similarity search type.

Comment: See the table below for a description of the search expressions to use on other cartridges.

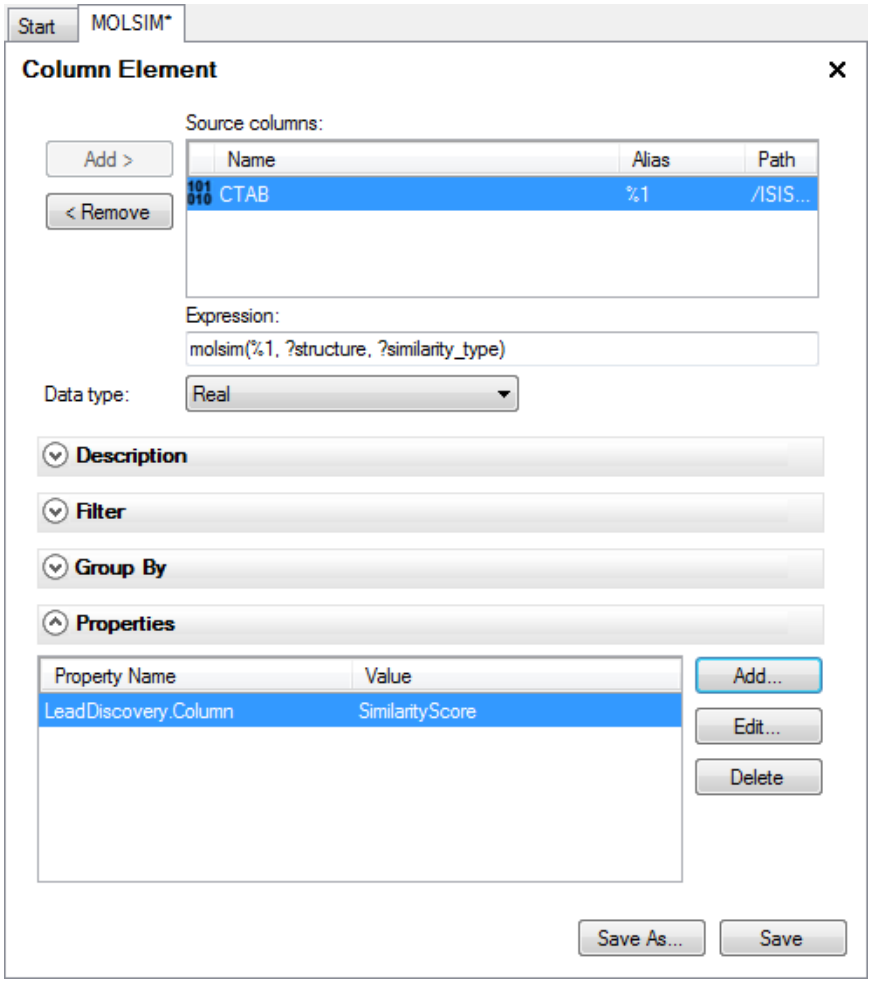
6. Select **Real** as the **Data type** of the column element.
7. In the **Properties** section, click  to show the column properties controls.
8. Click **Add** to add the column property.
9. Type **<prefix>.**Column for the **Property Name** and **SimilarityScore** for the Value in the **Add Column Property** dialog.



Note: While the second part of the property name is pre-defined and you must assign it without changes (for the similarity score column, this is **Column**), you can use an arbitrary prefix for the element properties (for example, **LeadDiscovery**).

One set of elements (as shown in the Configuration Overview chapter) should always use one and the same property prefix. If you want to configure alternative information links you can create more than one set of elements (e.g., to set up information links running against different data tables). However, each set of elements should always use its own property prefix.

10. Click **OK.** The column property displays in the properties list.



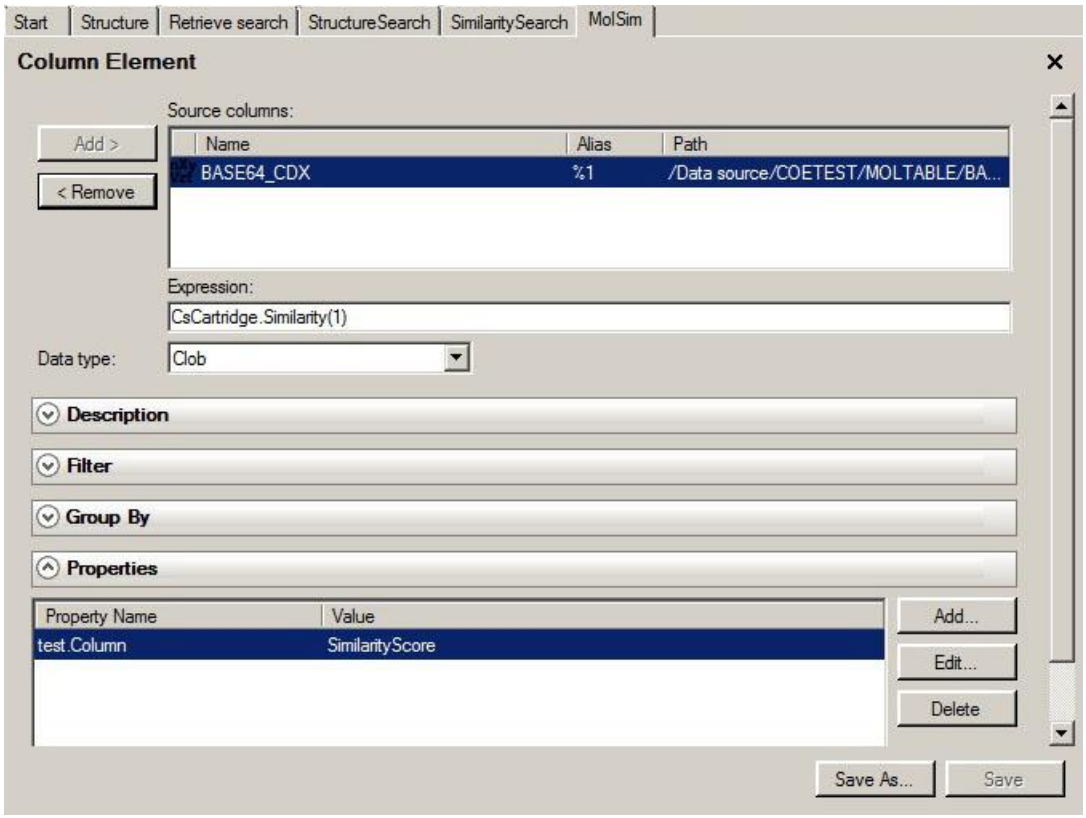
11. Click **Save**.
12. Click to select the folder where you want to save the column element.
13. Type a descriptive column name (for example, **MOLSIM**). The column is saved in the library and is shown with an icon denoting its type in the Elements tree.

Cartridge	Expression
CambridgeSoft CSCartridge	CsCartridge.Similarity(1)
Symyx® Direct Cartridge (MDL)	molsim(%1,?structure,?similarity_type)
Accelrys Accord Chemistry Cartridge	ChemistryMatches(%1, sssConst.MatchTanimoto, ?structure)*100
ChemAxxon JChem Oracle Cartridge	(1- jchem.jc_dissimilarity(%1, jchem.jcf.molconvert(?structure, 'smiles:au')))*100

IDBS ChemXtra Cartridge	SIM(%1, ?structure, IDBS_MOLECULE.TANIMOTO, IDBS_MOLECULE.DELETE_HANDLE) * 100
GGA Bingo	Bingo.Sim(%1, ?structure, 'tanimoto')

If you are using CSCartridge, the following expression must be entered in the **Expression** field:

CsCartridge.Similarity(1)



4.4 Creating Filter Elements

Create filter elements with descriptive names to apply when creating information links:

1. Substructure search filter
2. Similarity search filter

The information below is an example of how to set up the filter elements for a Symyx® Direct (MDL) cartridge. If you are using another cartridge the expressions entered will be different. See the tables below each step instruction or the documentation from your cartridge provider for more information about valid expressions (WHERE-statements).

4.4.1 Creating Substructures Search Filters

To create a substructure search filter:

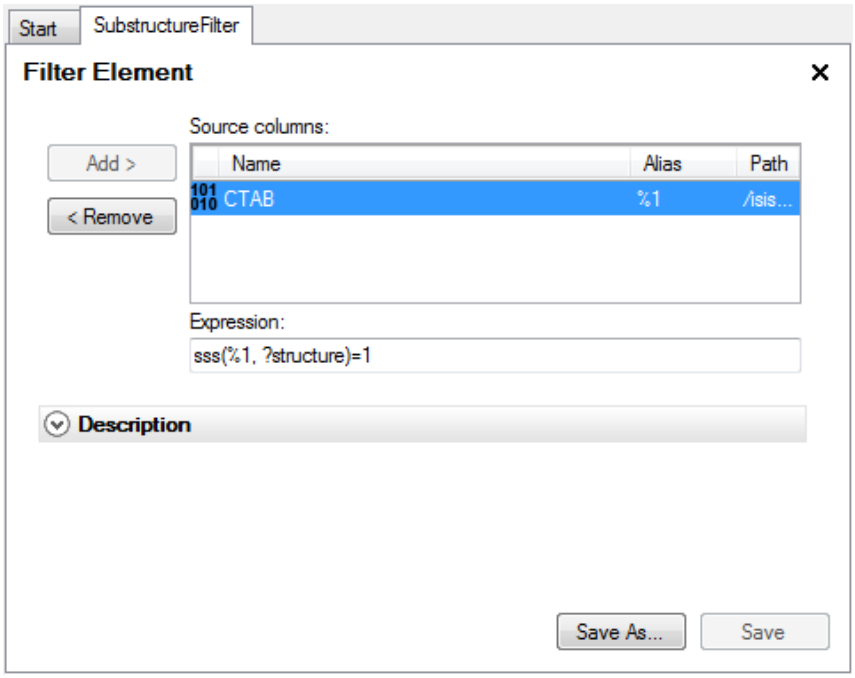
1. In the Information Designer, click **New** and select **Filter**. The Filter Element tab is opened.

2. In the **Data Sources** tree, select the column containing chemical structures indexed by the cartridge, to use in the filter condition.
3. Click **Add >**. The selected column is added to the Source columns list on the Filter Element tab.
4. Define the following filter condition in the Expression field:

   sss(%1, ?structure)=1

   Where **sss** is the Symyx® Direct (MDL) substructure search operator and **structure** is the parameter for the query structure.

Comment: See the table below for a description of the search expressions to use on other cartridges.



5. In the **Description** field, type a description of the filter.
6. Click **Save**. The Save As dialog displays.
7. Click to select the folder where you want to save the filter element.
8. Type a descriptive filter name (for example, **SubstructureFilter**).
9. Click **Save**. The filter element is saved in the library.

Cartridge	Search Expression
CambridgeSoft CSCartridge	CSCartridge.MoleculeContains(%1, ?structure, 'FULL=NO') = 1
Symyx® Direct Cartridge (MDL)	sss(%1, ?structure)=1
Accelrys Accord Chemistry Cartridge	ChemistryMatches(%1, sssConst.MatchSub, ?structure) = 1
ChemAxxon JChem Oracle Cartridge	jchem.jc_compare(%1, ?structure, 't:s stereoSearch:n')=1
IDBS ChemXtra Cartridge	SSS(%1, ?structure, IDBS_MOLECULE.IGNORE_STEREO) = 1

GGA Bingo	bingo.sub(%1, ?structure,'') = 1

4.4.2 Creating Similarity Search Filters

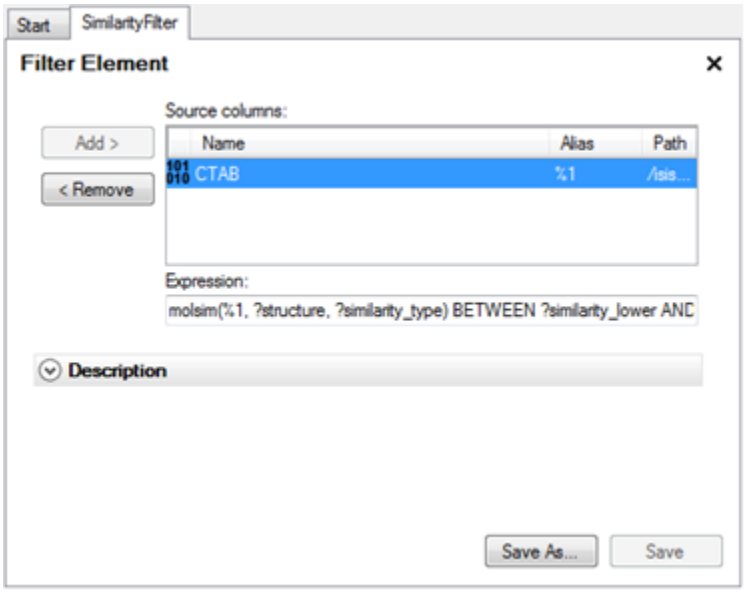
To create a similarity search filter:

1. In the Information Designer, click **New** and select **Filter**. The Filter Element tab opens.
2. In the **Data Sources** tree, select the column containing chemical structures indexed by the cartridge to use in the filter condition.
3. Click **Add >**. The selected column is added to the Source columns list on the Filter Element tab.
4. Define the following filter condition in the Expression field:

   molsim(%1,?structure,?similarity_type) BETWEEN ?similarity_lower AND ?similarity_upper

   Where **molsim** is the Symyx® Direct (MDL) similarity search operator extracting similarity values for chemical structures, **structure** is the parameter for the query structure, **similarity_type** is the parameter for the similarity search type, **similarity_lower** is the parameter for the lower limit of the similarity rate range, and **similarity_upper** is the parameter for the upper limit of the similarity rate range.

Comment: See the table below for a description of the search expressions to use on other cartridges.



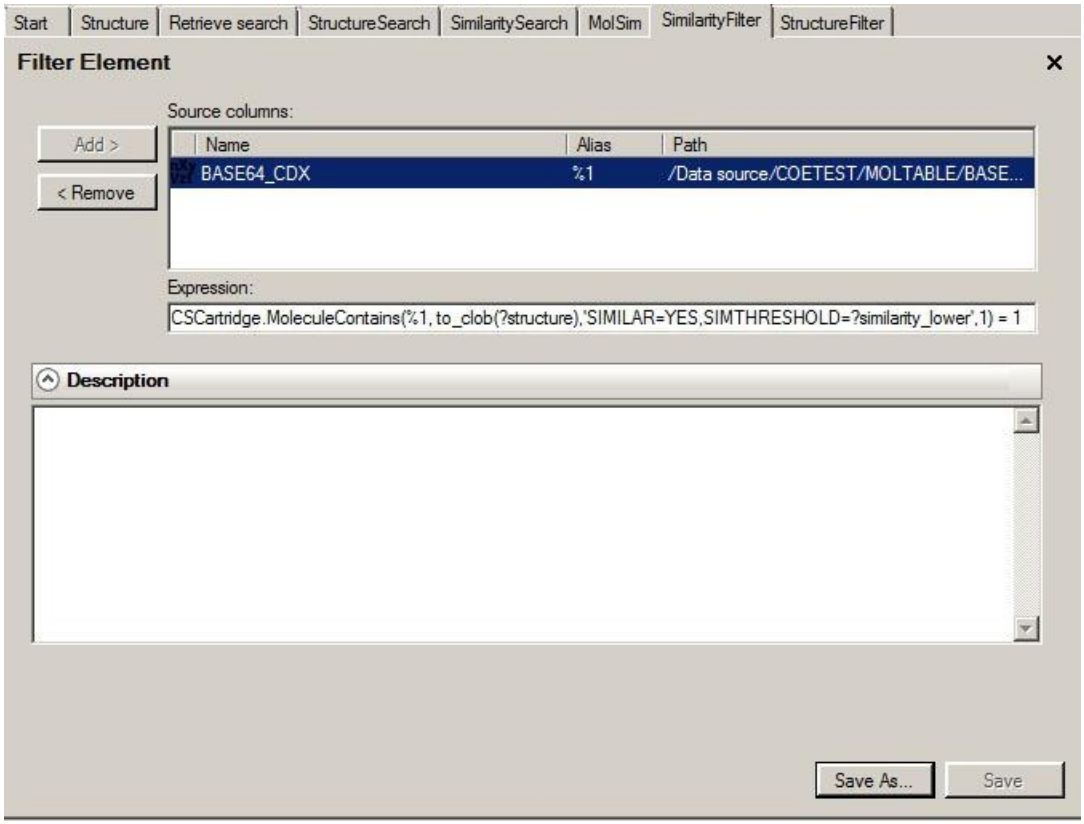
5. In the **Description** filed, type a description of the filter.
6. Click **Save**. The Save As dialog displays.
7. Click to select the folder where you want to save the filter element.
8. Type the filter name, (for example, **SimilarityFilter**).
9. Click **Save**. The filter element is saved in the library.

Cartridge	Search Expression
CambridgeSoft CSCartridge	CSCartridge.MoleculeContains(%1,

	to_clob(?structure),'SIMILAR=YES,SIMTHRESHOLD=?similarity_lower',1) = 1
Symyx® Direct Cartridge (MDL)	molsim(%1,?structure,?similarity_type) BETWEEN ?similarity_lower AND ?similarity_upper
Accelrys Accord Chemistry Cartridge	ChemistryMatches(%1, sssConst.MatchTanimoto, ?structure)*100 BETWEEN ?similarity_lower AND ?similarity_upper
ChemAxxon JChem Oracle Cartridge	jchem.jc_dissimilarity(%1, jchem.jcf.molconvert(?structure, 'smiles:au')) < 1-(?similarity_lower/100)
IDBS ChemXtra Cartridge	C$IDBSCHX8.SIM(%1, ?structure , C$IDBSCHX8.IDBS_MOLECULE.TANIMOTO, 0) > (?similarity_lower/100)
GGA Bingo	Bingo.Sim(%1, ?structure, 'tanimoto') > ?similarity_lower/100

If you are using CSCartridge, the following expression must be entered in the **Expression** field:

CSCartridge.MoleculeContains(%1, to_clob(?structure),'SIMILAR=YES,SIMTHRESHOLD=?similarity_lower',1) = 1



4.5 Creating Information Links

Three information links are required to use the structure search functionality:

- For retrieving structures based on the list of identifiers
- For substructure search

- For similarity search

4.5.1 Retrieving Information Links

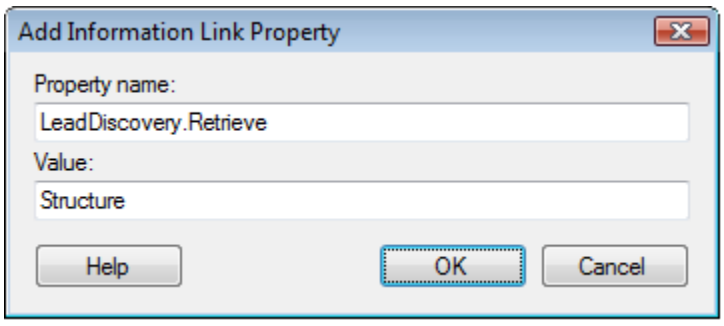
To create information links for retrieving structures:

Note: You must always create this information link. If you want to apply both similarity searches and substructure searches, this information link is used as the retrieve link for both search types, provided that all three information links use the same property prefix.

1. In the Information Designer, click **New** and select **Information Link**. The Information Link tab opens.
2. In the **Elements** tree, select the column element that you created for the column containing identifiers of chemical structures.
3. Click **Add >**. The selected element is added to the Elements list on the Information Link tab.
4. In the **Elements** tree, select the column element that you created for the column containing chemical structures.
5. Click **Add >**. The selected element is added to the Elements list on the Information Link tab.

   Tip: In this information link, you can add other columns from the database. These columns display when you create a new data table from a list of search results.

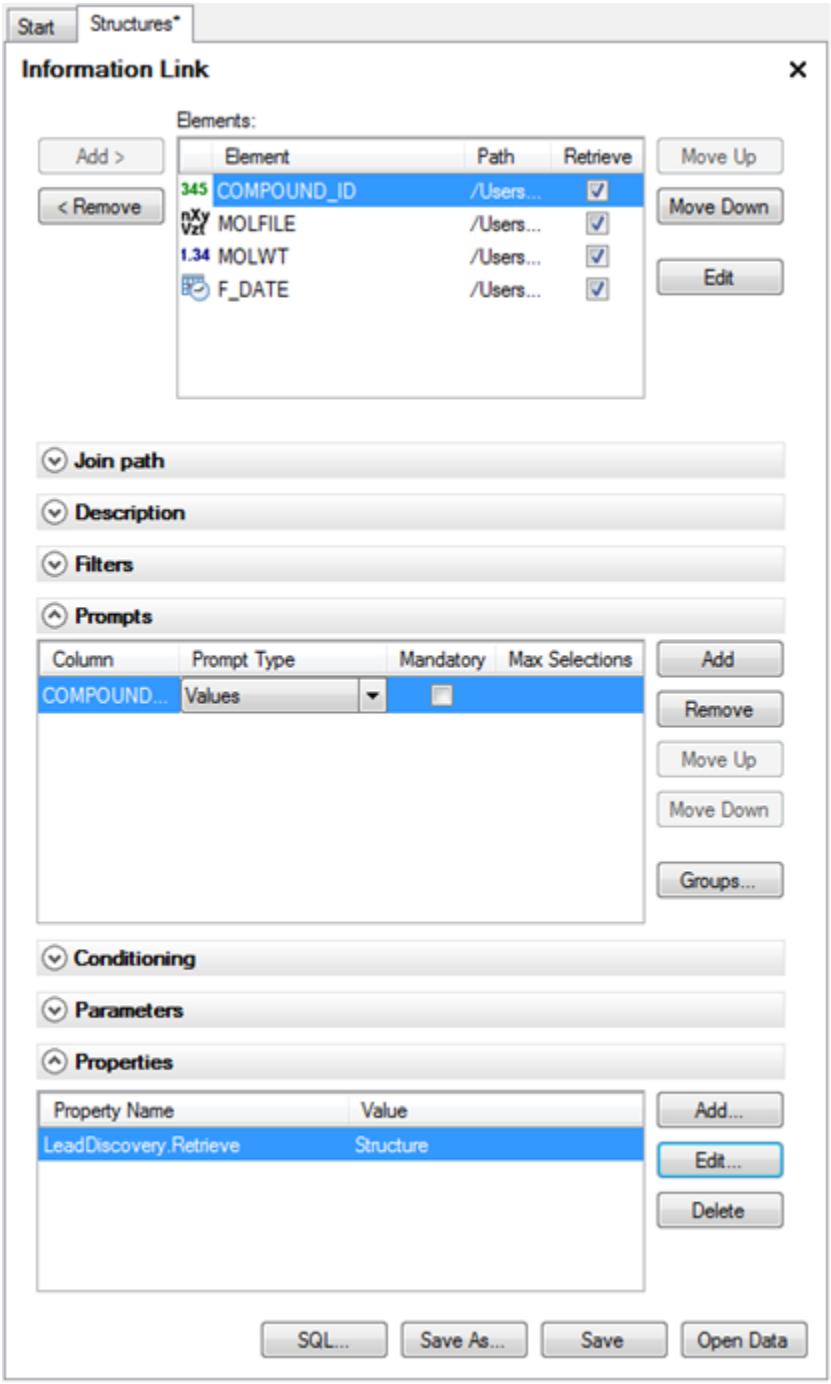
6. Ensure the check box in the **Retrieve** column of the Elements list is selected for both elements.
7. In the **Description** field, type a description of the information link (optional).
8. In the **Prompts** section, click to show the prompting controls.
9. Click **Add**.
   Comment: If you have included more columns than the required ones, the **Add Column** dialog displays. In that case, select the identifier column and click **OK**. The identifier column is added to the Prompts list on the Information Link tab.
10. Set **Prompt Type** to Values.
11. Ensure the **Mandatory** box is cleared.
12. In the **Properties** section, click to show the information link properties controls.
13. Click **Add** to add the information link property.
14. Type **<prefix>.Retrieve** for the **Property name** and **Structure** for the **Value** in the **Add Information Link Property** dialog. This property specifies that the information link will be shown among the available links in the Configure Structure Connection dialog.



Note: While the second part of the property name is pre-defined and you must assign it without changes (for the structure retrieval link, this is **Retrieve**), you can use an arbitrary prefix for the element properties (for example, **LeadDiscovery**).

One set of elements (as shown in the Configuration Overview chapter) should always use one and the same property prefix. If you want to configure alternative information links, you can create more than one set of elements (e.g., to set up information links running against different data tables). However, each set of elements should always use its own property prefix.

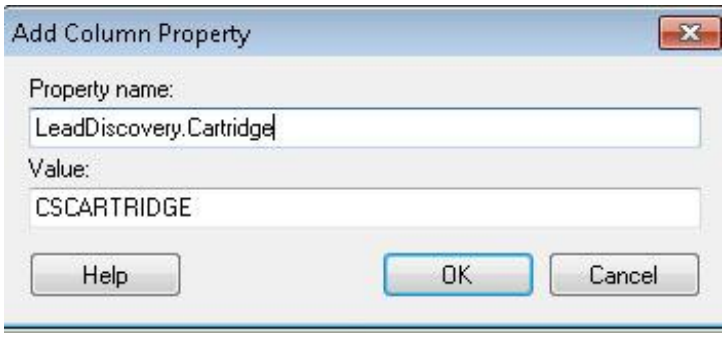
15. Click **OK**. The link property displays in the properties list.



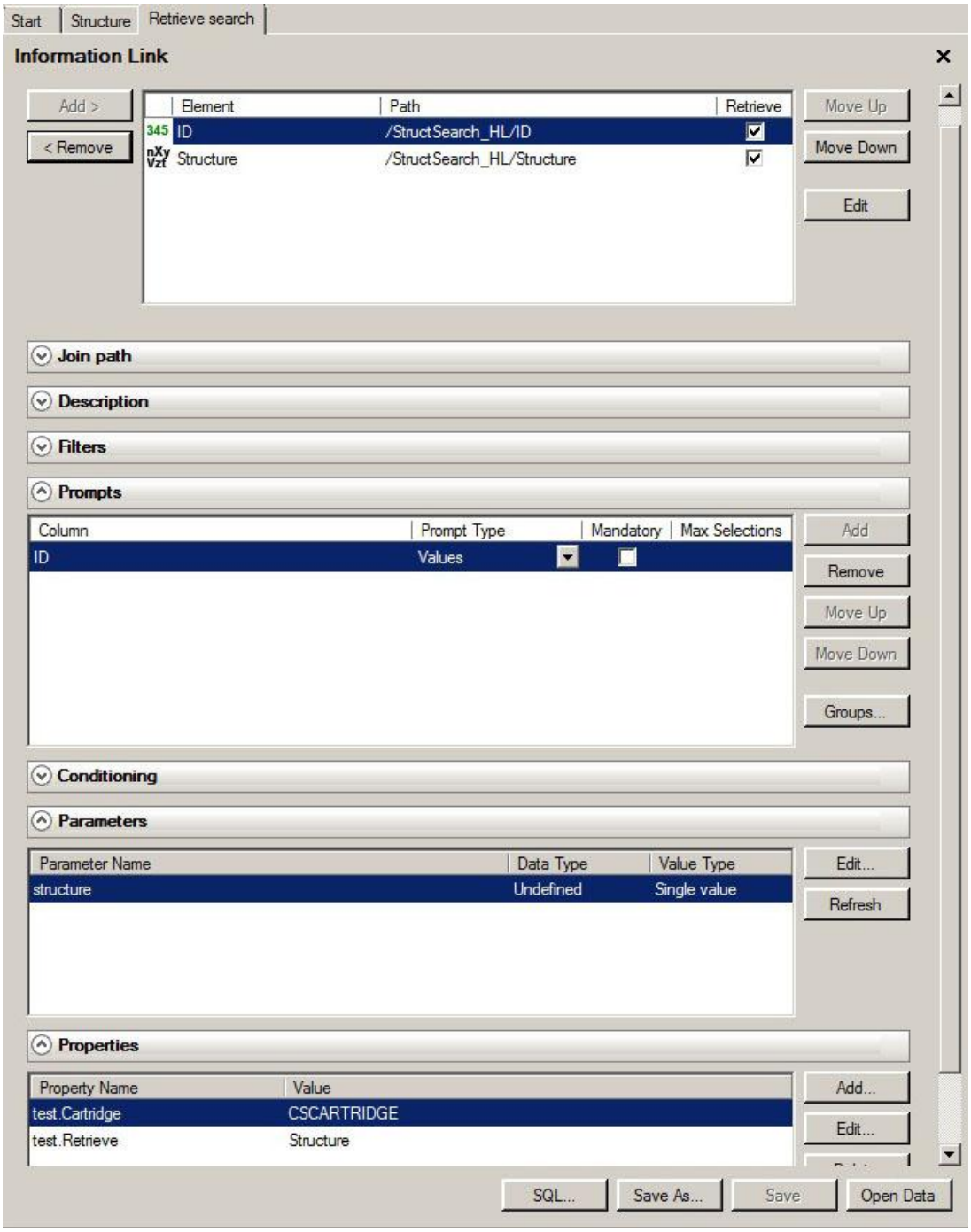
16. Click **Save**. The Save As dialog appears.
17. Select where you want the invitation link saved.
18. In the **Name** field, type a name for the information link (for example, **Structures**).
19. In the **Description** field, type text describing the purpose of the information link (optional).
20. Click **Save**. The new information link is added to the library. Other users can access this link.

If you are using CSCartridge, an additional Information Link property and parameter must be added.

1. In the **Properties** section, click to show the information link properties controls.
2. Click **Add** to add the information link property.
3. Type **<prefix>.Cartridge** for the **Property name** and **CSCartridge** for the **Value** in the **Add Information Link Property** dialog. This property specifies that the information link will be shown among the available links in the Configure Structure Connection dialog.



4. Click **OK**. The link property displays in the properties list.
5. In the Parameters section, click to show the parameters controls.
6. Click **Refresh**. The structure parameter appears in the parameters list.
7. Select the structure parameter and click **Edit**. The Edit Parameter dialog displays.
8. Set the **Data Type** to **Undefined** and the **Value Type** to **Single value**. Click **OK**.



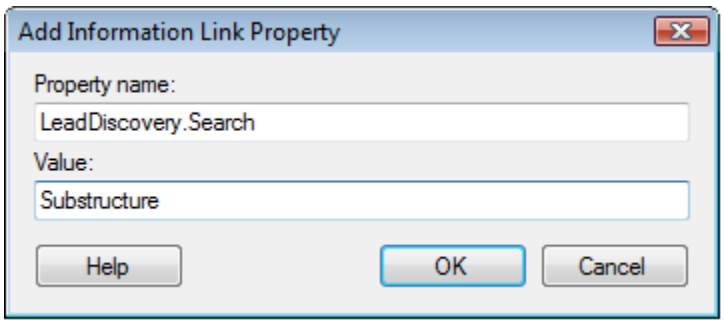
Note: This Retrieve Search Information Link is for use with the Structure Viewer ONLY. It cannot be used to directly add a data table from an Information Link' (**File > Add Data Tables > Add > Information Link**).

4.5.2 Creating Substructures Search Information Links

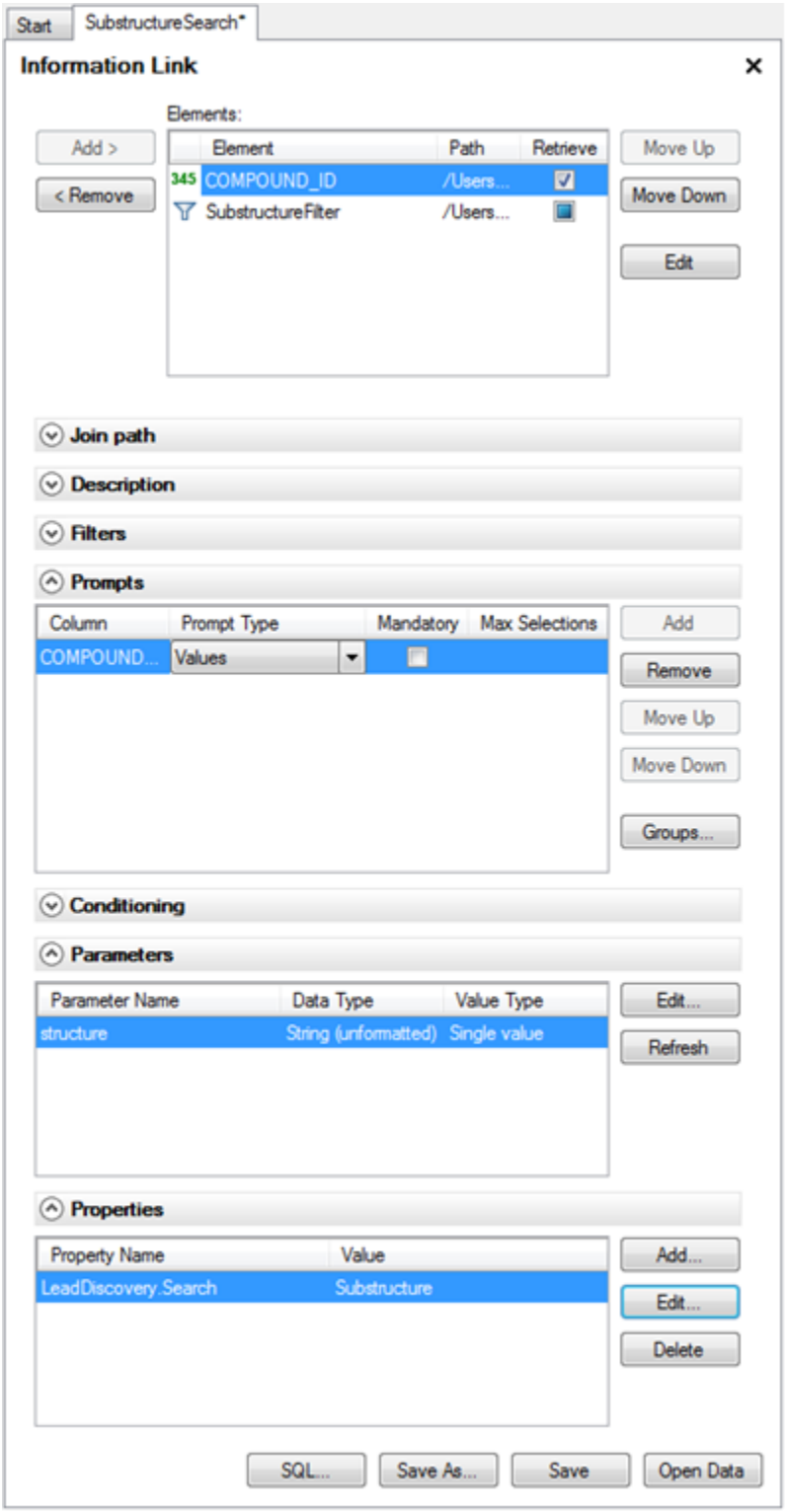
To create an information link for the substructures search:

1. In the Information Designer, click **New** and select **Information Link**. The Information Link tab is opened.
2. In the **Elements** tree, select the column element that you created for the column containing identifiers of chemical structures.
3. Click **Add >**. The selected element is added to the Elements list on the Information Link tab.
4. In the Elements tree, select the filter element that you created for the substructure search.
5. Click **Add >**. The selected element is added to the Elements list on the Information Link tab.
6. Ensure that the check box in the **Retrieve** column of the Elements list is selected for the structure identifier column element.
7. In the Description field, type a description of the information link (optional).
8. In the Prompts section, click to show the prompting controls.

   Comment: You must add a prompt step on the identifier column. This is required for limiting the search to a list, instead of searching against the entire database when performing a structure search.
9. Click **Add**. The identifier column is added to the Prompts list on the Information Link tab.
10. Set **Prompt Type** to Values.
11. Ensure the **Mandatory** box is cleared.
12. In the Parameters section, click to show the parameters controls.
13. Click **Refresh**. The structure parameter appears in the parameters list.
14. Select the structure parameter and click **Edit**. The Edit Parameter dialog displays.
15. Set the **Data type** to **Undefined** and click **OK**.
16. In the **Properties** section, click to show the information link properties controls.
17. Click **Add** to add the information link property.
18. Type **<prefix>.Search** for the **Property Name** and **Substructure** for the **Value** in the **Add Information Link Property** dialog. This property specifies the information link is shown among the available Substructure search links in the Structure Search dialog.



19. Click **OK**. The link property displays in the properties list.

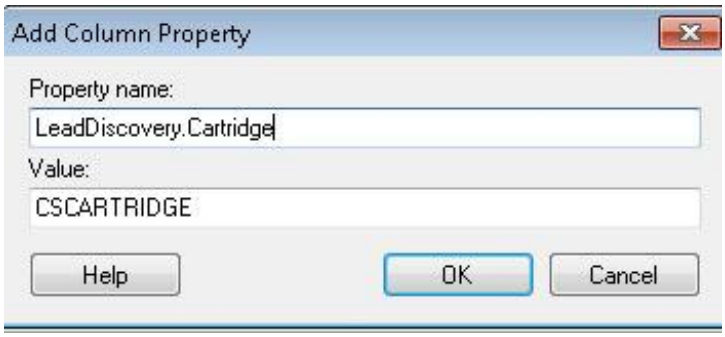


20. Click **Save**. The Save As dialog appears.
21. Select where you want to save the information link.
22. In the **Name** field, type a name for the information link (for example, **SubstructureSearch**).

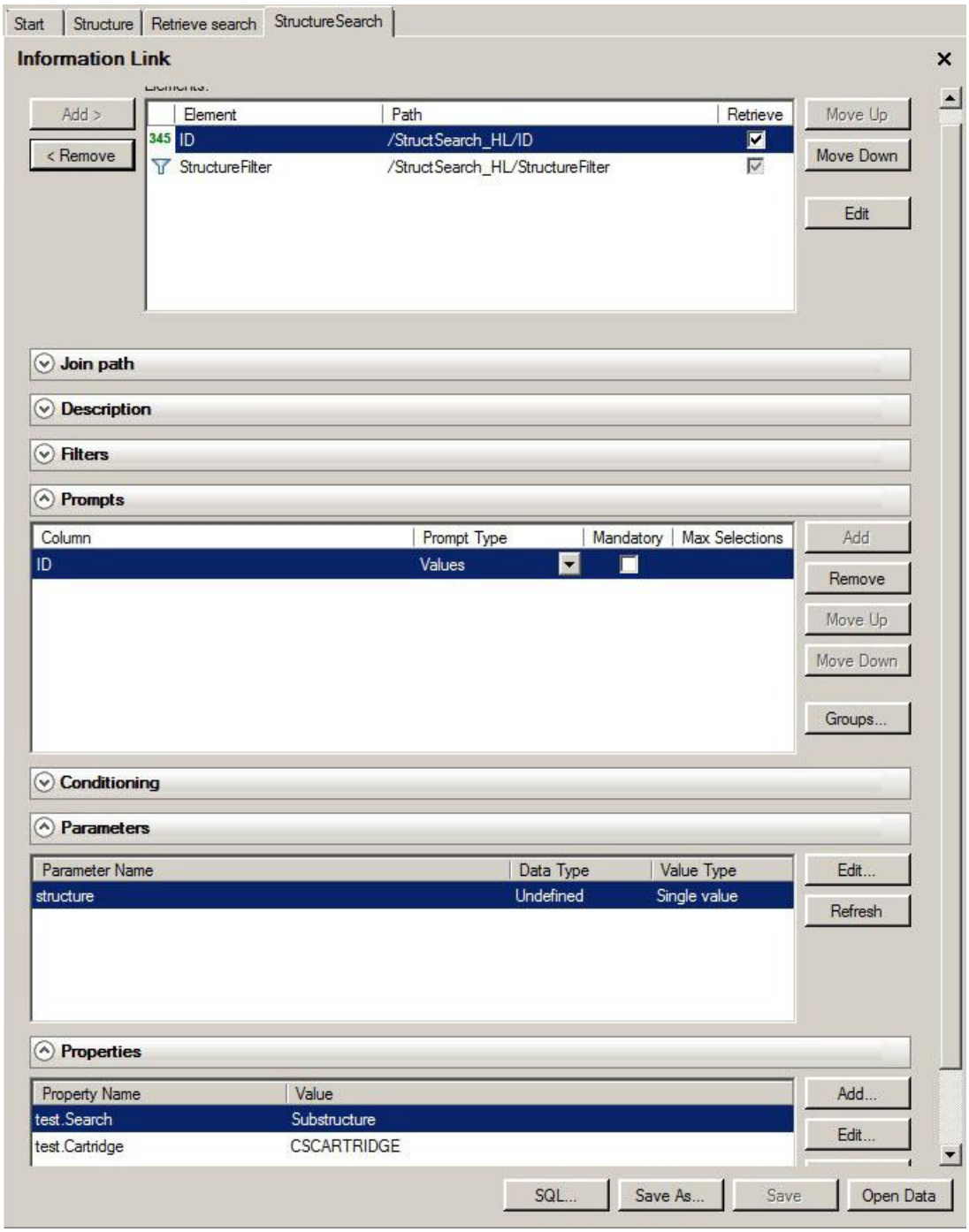
23. In the **Description** field, type text describing the purpose of the information link (optional).
    Comment: It is useful to note that this information link is used in the Structure Search tool only, not to retrieve any data directly into TIBCO Spotfire, which might help to avoid confusion among end-users.
24. Click **Save**. The new information link is added to the library. Other users can access this new information link.

If you are using CSCartridge, an additional Information Link property must be added.

1. In the **Properties** section, click to show the information link properties controls.
2. Click **Add** to add the information link property.
3. Type **<prefix>.Cartridge** for the **Property name** and **CSCartridge** for the **Value** in the **Add Information Link Property** dialog. This property specifies that the information link will be shown among the available links in the Configure Structure Connection dialog.



4. Click **OK**. The link property displays in the properties list.



Note: When setting an Information Link for Substructure Search, Lead Discovery will convert the query molecule to MOLfile of SMILES before using it to replace the ?structure marking. However, if you set <prefix>.Cartridge to CSCartridge, then it will know that the cartridge can support ?structure in CDX format, and it will not convert it.

4.5.3 Creating Similarity Search Information Links

To create an information link for the similarity search:

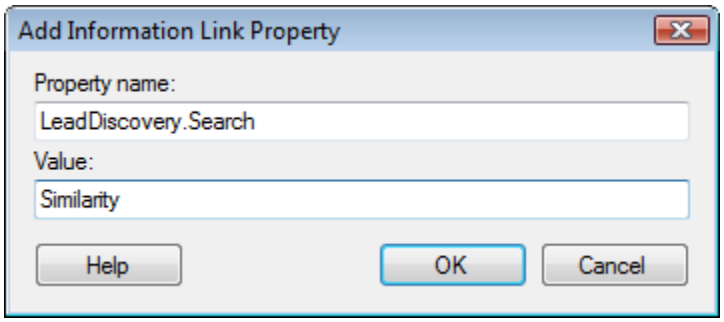
1. In the Information Designer, click **New** and select **Information Link**. The Information Link tab is opened.
2. In the **Elements** tree, select the column element that you created for the column containing identifiers of chemical structures.
3. Click **Add >**. The selected element is added to the Elements list on the Information Link tab.
4. In the **Elements** tree, select the column element containing similarity values for chemical structures.
5. Click **Add >**. The selected element is added to the Elements list on the Information Link tab.
6. In the Elements tree, select the filter element that you created for the similarity search.
7. Click **Add >**. The selected element is added to the Elements list on the Information Link tab.
8. Ensure that the check box in the **Retrieve** column of the Elements list is selected for the structure identifiers and similarity values column elements.
9. In the **Description** field, type a description of the information link (optional).
10. In the **Prompts** section, click to show the prompting controls.
    Comment: You must add a prompt step on the identifier column. This is to limit the search to a list, instead of searching against the entire database, when performing a structure search.
11. Click **Add**. The identifier column is added to the Prompts list on the Information Link tab.
12. Set **Prompt Type** to Values.
13. In the **Parameters** section, click to show the parameters controls.
14. Click **Refresh**. Four filter parameters appear in the parameters list.

Parameter Name	Data Type	Value Type
similarity_lower	String	Single value
similarity_type	String	Single value
similarity_upper	String	Single value
structure	String	Single value

    Comment: In this example, a Symyx® (MDL) cartridge has been used. For other cartridges some of the parameters above may be missing.

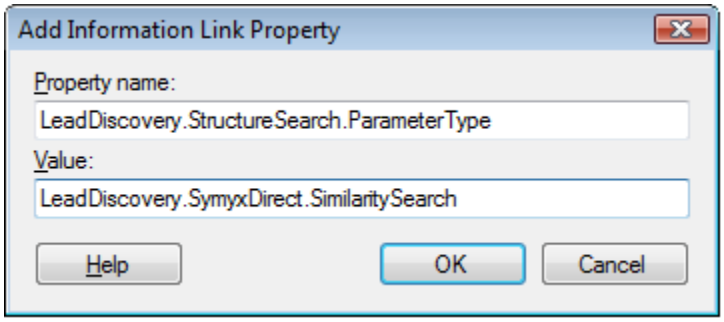
15. Select the **similarity_lower** parameter and click **Edit**. The Edit Parameter dialog displays.
16. Set the **Data type** to **Integer** and click **OK**.
17. Select the **similarity_upper** parameter and click **Edit**. The Edit Parameter dialog displays.
18. Set the **Data type** to **Integer** and click **OK**.
19. Select the **structure** parameter and click **Edit**. The Edit Parameter dialog displays.
20. Set the **Data type** to **Undefined** and click **OK**.

21. In the **Properties** section, click to show the information link properties controls.
22. Click **Add** to add the information link property.
23. Type **<prefix>.Search** for the **Property name** and **Similarity** for the **Value** in the **Add Information Link Property** dialog. This property specifies that the information link will be shown among the available Similarity search links in the Structure Search dialog.



Note: While the second part of the property name is pre-defined and you must assign it without changes (for the similarity search link, this is Search), you can use an arbitrary prefix for the element properties (for example, **LeadDiscovery)**.
One set of elements (as shown in the Configuration Overview chapter) should always use one and the same property prefix. If you want to configure alternative information links you can create more than one set of elements (e.g., to set up information links running against different data tables). However, each set of elements should always use its own property prefix.

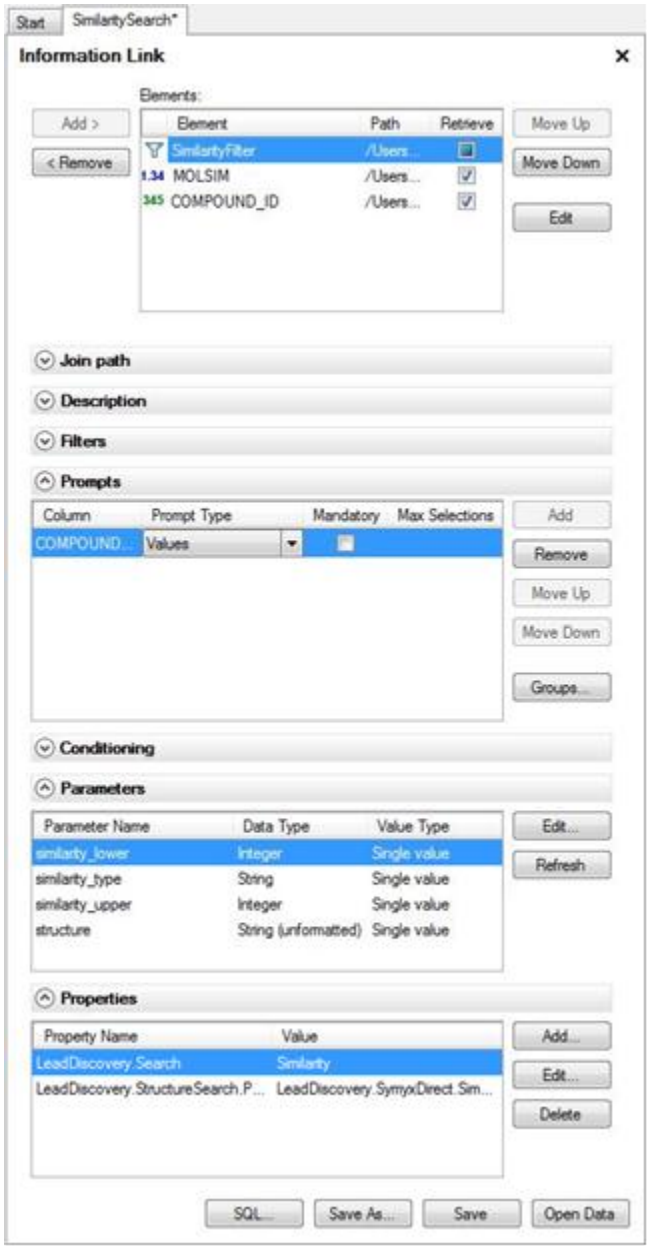
24. Click **OK**. The link property displays in the property list.
25. Again, click **Add** to add another information link property.
26. Type **LeadDiscovery.StructureSearch.ParameterType** for the **Property name** and the expression corresponding to the chosen cartridge found in the table below, for the **Value**. This property ensures that the correct **Parameters** dialog in the **Structure Search** is available for the chosen cartridge.



Cartridge	Value expression
CambridgeSoft CSCartridge	LeadDiscovery.CSCartridge.SimilaritySearch
Symyx® Direct Cartridge (MDL)	LeadDiscovery.SymyxDirect.SimilaritySearch
Accelrys Accord Chemistry Cartridge	LeadDiscovery.Accord.SimilaritySearch
ChemAxxon JChem Oracle Cartridge	LeadDiscovery.JChem.SimilaritySearch
IDBS ChemXtra Cartridge	LeadDiscovery.ChemXtra.SimilaritySearch
GGA Bingo	LeadDiscovery.Bingo.SimilaritySearch

Note: If you do not perform steps 26 and 27 in this list, a default **Parameters** dialog is used.

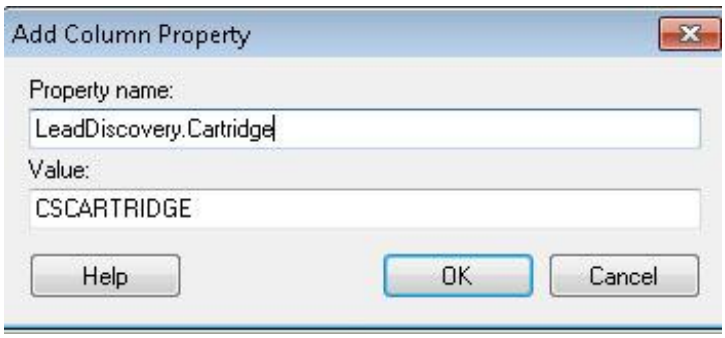
27. Click **OK**. The link property displays in the properties list.



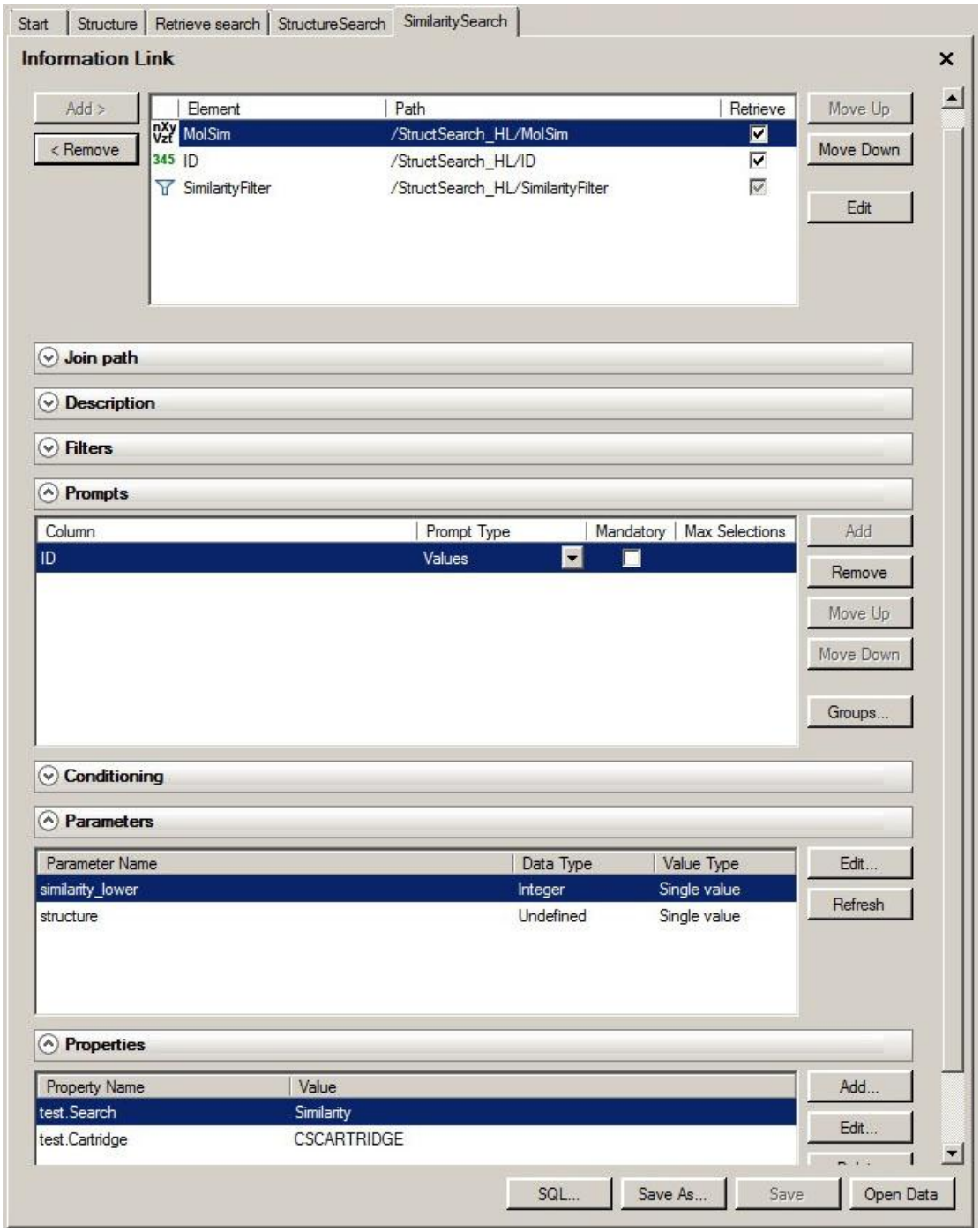
28. Click **Save**. The Save As dialog appears.
29. Select where you want to save the information.
30. In the **Name** field, type a name for the information link (for example, **SimilaritySearch**).
31. In the **Description** field, describe the purpose of the information link (optional).
    Comment: It is useful to note that this information link is used in the Structure Search tool only, not to retrieve any data directly into TIBCO Spotfire, which might help to avoid confusion among end-users.
32. Click **Save**. The new information link is added to the library. Other users can access this new information link.

If you are using CSCartridge, an additional Information Link property must be added.

1. In the **Properties** section, click to show the information link properties controls.
2. Click **Add** to add the information link property.
3. Type **<prefix>.Cartridge** for the **Property name** and **CSCartridge** for the **Value** in the **Add Information Link Property** dialog. This property specifies that the information link will be shown among the available links in the Configure Structure Connection dialog.



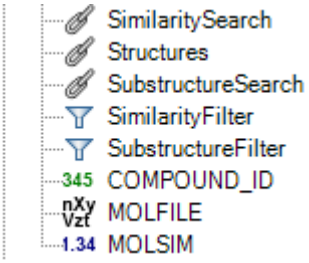
4. Click **OK**. The link property displays in the properties list.



Note: When setting an Information Link for Similarity Search, Lead Discovery will convert the query molecule to MOLfile of SMILES before using it to replace the ?structure marking. However, if you set <prefix>.Cartridge to CSCartridge, then it will know that the cartridge can support ?structure in CDX format, and it will not convert it.

4.6 Verifying the Configuration

In the Elements tree, the elements (columns, filters, and information links) you created display in the tree. For example:



5 Testing the Installation

Perform the following procedures to verify the installation.

5.1 Verifying the Installation

The following five procedures ensure proper installation.

5.1.1 Opening an SDFile:

To open an SDFile:

1. Click **Open** on the toolbar, or select **File** > **Open**.
2. Browse to the SDFile and click **Open**. You will see the table containing the fields of the SDFile, including structures.

5.1.2 Exporting Data from Visualizations

To export data from visualization:

1. Right-click a table with structures and select **Export to SDFile**.
2. Select whether to export all rows, marked rows, or filtered rows only:
   - All rows - export every row in the data table regardless of filtering
   - Marked rows - export the marked rows only
   - Filtered rows - export the rows remaining after filtering only
3. Select the column containing structures from the drop-down list.
   Comment: The SDFile format supports only one structure per record. You can select any column in the table that contains structures. The first column containing structures is displayed as the default choice.
4. Use **Add**, **Remove**, and **Remove All** to select the columns to export with the structure.
5. Use **Move Up** and **Move Down** to specify the order of columns in the SDFile.
6. Click **OK**.

7. In the Save As dialog, specify a filename and where you want to save the file.
8. Click **Save**.

5.1.3 Reaching the Structure Viewer

To reach the Structure Viewer:

1. Select a table visualization that contains structures.
   Comment: If you have not previously opened the visualization with structures, the Structure Viewer button and the corresponding menu item may be inactive.
2. Click **Structure Viewer** in the toolbar  . The Configuration Structure Connection dialog displays.
   Comment: You can also select **View** > **Structure Viewer** from the menu.
3. Select the columns in the analysis containing molecular structures and identifiers. (You can retrieve the structures from an information link.)
   Note: Lead Discovery expects the ID column to uniquely identify a structure, so that no two structures have the same value of ID. If several structures have the same value in the ID column, only one of them will display in the Structure Viewer. To indicate this, the shown structure is marked as “ambiguous”.
4. Click **OK**. The Structure Viewer displays.
5. Mark or highlight compounds in a visualization. When there are no marked rows, the structure from the currently highlighted row in the data table displays in the Structure Viewer pane of the Structure Viewer, and its identifier displays in the List Content pane. When there are marked rows, the structures from the marked rows display in the Structure Viewer pane and their identifiers display in the List Content pane. When you highlight a marked row, its corresponding identifier and structure is highlighted in the Structure Viewer and List Content panes.

5.1.4 Searching for Compounds in the Structure Database

To search for compounds in the structure database:

1. In the Structure Viewer or in a table visualization with structures, click to activate the compound on which you want to base the search.
2. Right-click the structure and select **Structure Search** from the pop-up menu. The Structure Search dialog displays.
   Comment: You can also select **Tools** > **Structure Search** from the main menu.
3. Select one of the alternatives under **Search for**.
4. If the **Parameters** button becomes available, click on it to edit the parameters available for the selected search method.
   Comment: If you select **Similarity search**, you must enter a threshold value in the **Similarity Search Parameters** dialog. The higher the threshold, the shorter is the hit list.
5. If you want to import a structure from a file, click **Import Structure**. The Open dialog appears. Select the file you want to import and click **Open**. The structure from the file displays in the Structure Search dialog.
6. If required, click **Edit**. Your selected structure editor launches. (If you have access to more than one editor, you can select which one to use.) Edit the structure and click **OK** (for ChemDraw) or **Transfer** (for Symyx® Draw (MDL)). Please refer to the documentation for your current structure editor for more details.
7. Select the information link to use for the search in the **Search in** drop-down list.
8. If required, limit the search to the current list of identifiers.

Comment: This option is only available when performing the search from the Structure Viewer and when any list except for the Marked Rows list is selected in the Lists drop-down of the Structure Viewer. When you start a structure search, from table visualization, without activating the Structure Viewer, or, if the Marked Rows list is selected, the Data Source option is the only option presented in the **Limit search to** drop-down.

9. Specify the name for the resulting list.
10. Click **OK**. Structures in the database that match the search criteria display in the Structure Viewer (Structure Viewer pane). A new ID list is added and displayed in the List Content Pane.

5.1.5 Selecting the Renderer

To change the renderer used to display structures in the Structure Viewer:

1. In the Structure Viewer pane of the Structure Viewer panel, or popover, right-click the structure.
2. Select **Renderer** from the menu.
3. Select the renderer to use.
   Comment: You can switch renderers if you have more than one of the supported renderers installed on your machine.

6 Appendix

Below are examples of what the resulting SQL from the specified information links might look like for different cartridges.

6.1 CambridgeSoft CSCartridge

Similarity search:

```
SELECT
   CsCartridge.Similarity(1) AS "MOLSIM",
   M1."MOL_ID" AS "ID"
FROM
   "COETEST"."MOLTABLE" M1
WHERE
   (CSCartridge.MoleculeContains(M1."BASE64_CDX",
   to_clob(?structure),
   'SIMILAR=YES,SIMTHRESHOLD=?similarity_lower',
   1) = 1)
   AND <conditions>
```

Substructure search:

```
SELECT
   M1."MOL_ID" AS "ID"
FROM
   "COETEST"."MOLTABLE" M1
WHERE
   (CSCartridge.MoleculeContains(M1."BASE64_CDX",
   ?structure,
   'FULL=NO') = 1)
   AND <conditions>
```

Retrieve links with no highlighting:

```
SELECT
   T1."ID" AS "ID",
   T1."M" AS "STRUCTURE"
FROM
   "CSCUSER"."TESTSTRUCTURES" T1
WHERE
   <conditions>
```

Retrieve links with highlighting :

```
SELECT
   M1."MOL_ID" AS "ID",
   NVL(CsCartridge.highlight(1),
   M1."BASE64_CDX") AS "STRUCTURE"
FROM
   "COETEST"."MOLTABLE" M1
WHERE
   (CsCartridge.MoleculeContains(M1."BASE64_CDX",
   ?structure,
   'highlight=yes',
```

```
   1) !=2)
   AND <conditions>
```

6.2 GGA Bingo

Similarity search:

```
SELECT
   S1."MOL_ID" AS "MOLID",
   Bingo.Sim(S1."SMILES",
   ?structure,
   'tanimoto') AS "SIMILARITYSCORE"
FROM
   "COETEST"."SFCOMPOUNDS" S1
WHERE
   (Bingo.Sim(S1."SMILES",
   ?structure,
   'tanimoto') > ?similarity_lower/100)
   AND <conditions>
```

Retrieve links:

```
SELECT
   S1."MOL_ID" AS "MOLID",
   S1."SMILES" AS "SMILES",
   Bingo.Molfile(S1."SMILES") AS "STRUCTURE"
FROM
   "COETEST"."SFCOMPOUNDS" S1
WHERE
   <conditions>
```

Substructure search:

```
SELECT
   S1."MOL_ID" AS "MOLID",
   Bingo.Molfile(S1."SMILES") AS "STRUCTURE"
FROM
   "COETEST"."SFCOMPOUNDS" S1
WHERE
   (bingo.sub(S1."SMILES",
   ?structure,
   '') = 1)
   AND <conditions>
```

6.3 Accelrys DataDirect Cartridge (Previously MDL)

Similarity search:

```
SELECT
   I1."CDBREGNO" AS "CDBREGNO",
   molsim(I1."CTAB",
   ?structure,
   ?similarity_type) AS "SIMSCORE"
FROM
   "ISIS"."ISISRC2D_MOL" I1
WHERE
   (molsim(I1."CTAB", ?structure, ?similarity_type) BETWEEN ?similarity_lower
```

```
   AND ?similarity_upper)
   AND <conditions>
```

Retrieve links:

```
SELECT
   I1."CDBREGNO" AS "CDBREGNO",
   molfile(I1."CTAB") AS "MOLFILE"
FROM
   "ISIS"."ISISRC2D_MOL" I1
WHERE
   <conditions>
```

Substructure search:

```
SELECT
   I1."CDBREGNO" AS "CDBREGNO"
FROM
   "ISIS"."ISISRC2D_MOL" I1
WHERE
   (sss(I1."CTAB",
   ?structure) =1)
   AND <conditions>
```