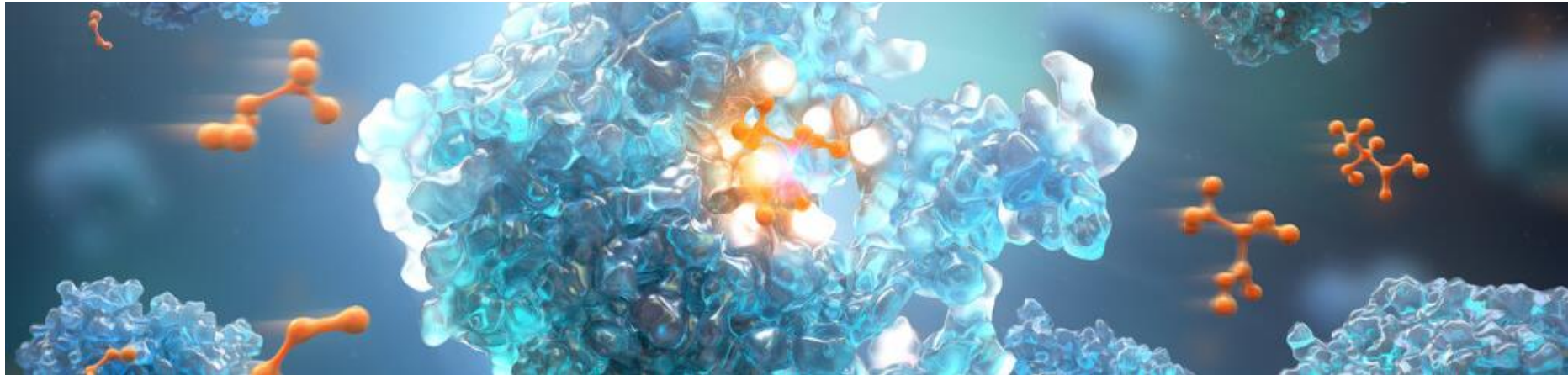


# A modern tool for integrating cheminformatics data science in Spotfire

**Lars Brive, Justin Morley, Maxime Guitet**

PerkinElmer EMEA Nexus 2019



# Acknowledgements



Wolfgang Klute  
Johan Ulander  
Nick Tomkinson



Eric Le Roux  
Claire Wallon  
Benjamin Pannetier

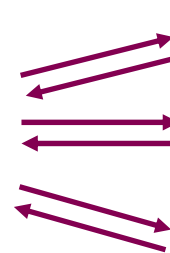
Michael Bodnarchuk  
Peter Brandt  
Sameer Kawatkar  
Susanne Winiwarter  
Kun Song  
Magnus Polla  
Ekaterina Ratkova  
SpotOn team



Spotfire



Biovia Pipeline Pilot



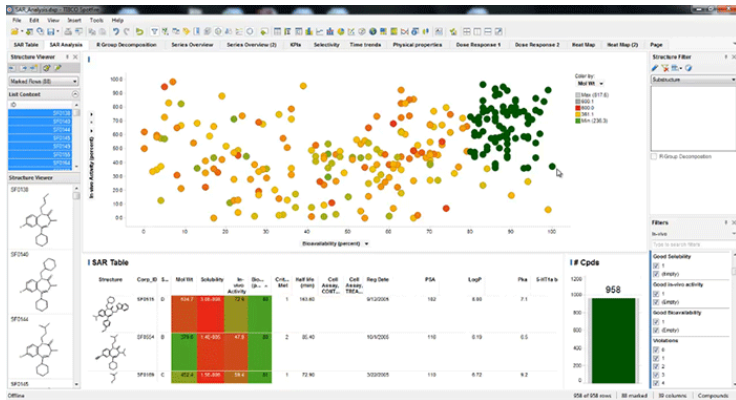
Databases

Calculation services

AI tools

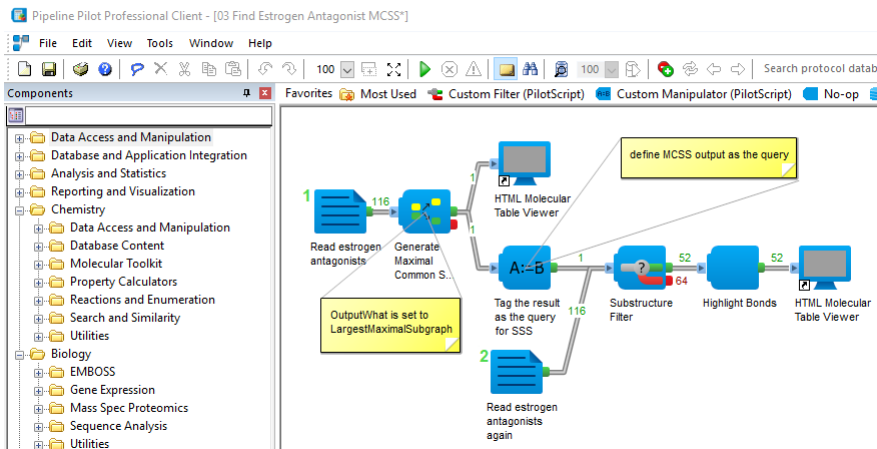
....

Spotfire:  
Analysis, visualisations, PerkinElmer Lead Discovery



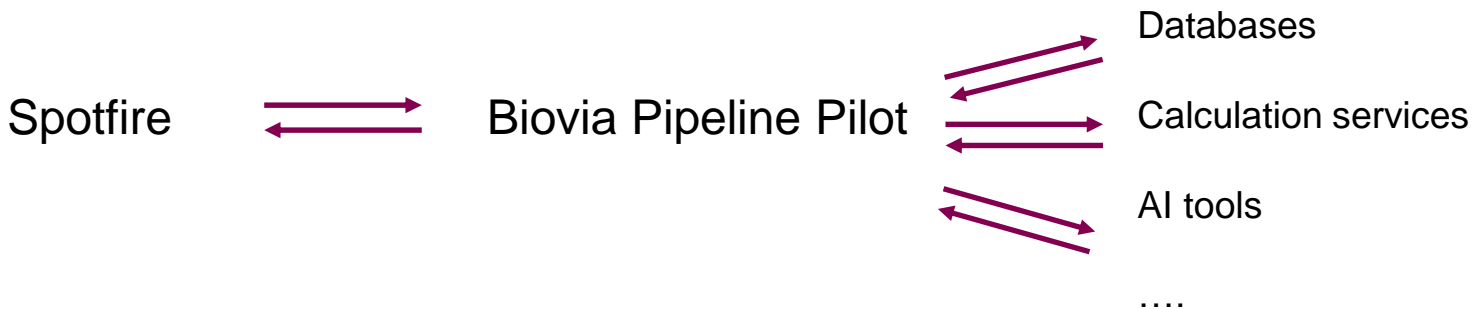
PerkinElmer.com

Biovia Pipeline Pilot: Data processing



Pipeline Pilot example protocol





## Requirements

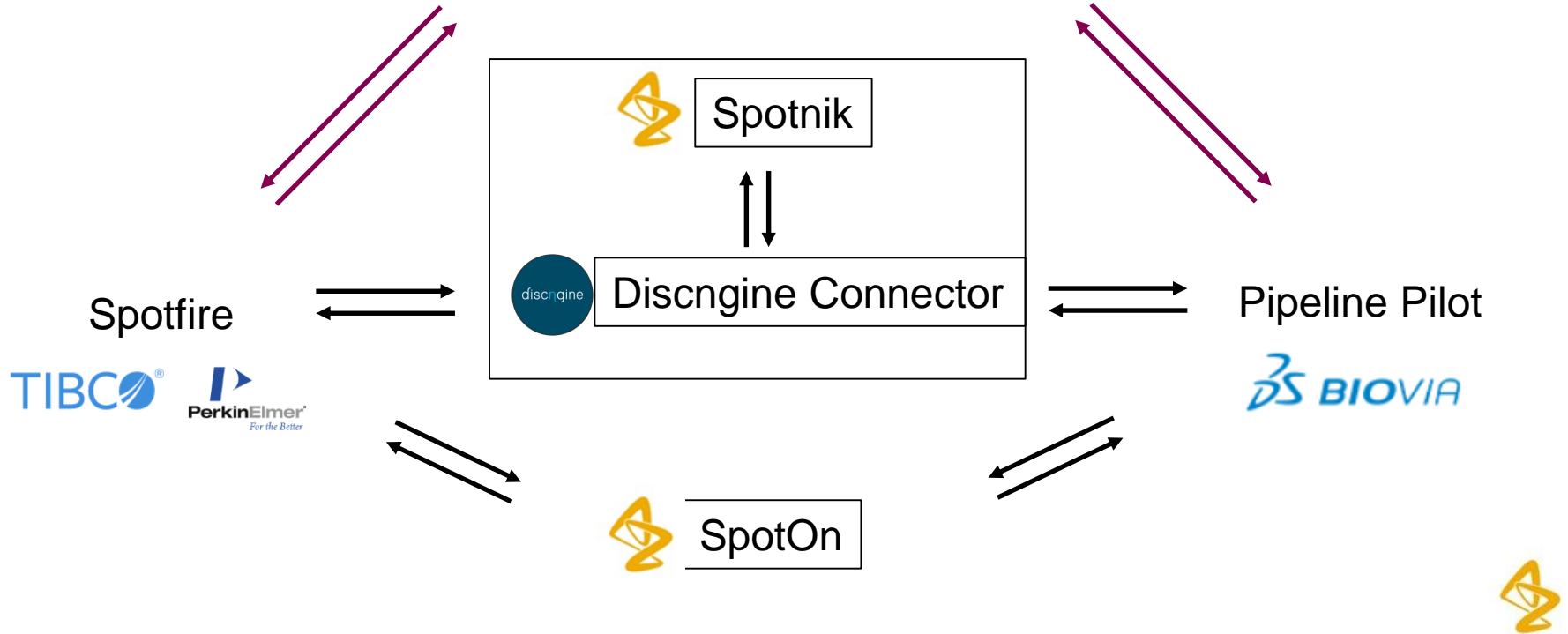
- Interactive and flexible analysis of data
- Complex queries made easy.
- Default visualisations.
- Range of user expectations: One green button or full control over parameters.
- Allow advanced users to implement their own protocols.
- Modular design
- Rapid implementation. Global usage.

”Are these compounds available in our compound collection”

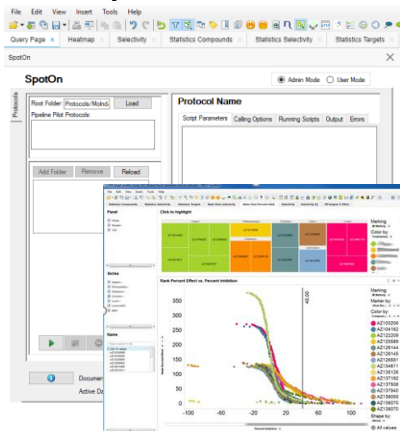
”Given a list of compound names, extract the bioactivity data in test, comparison with internal and external data. Group by target type.”



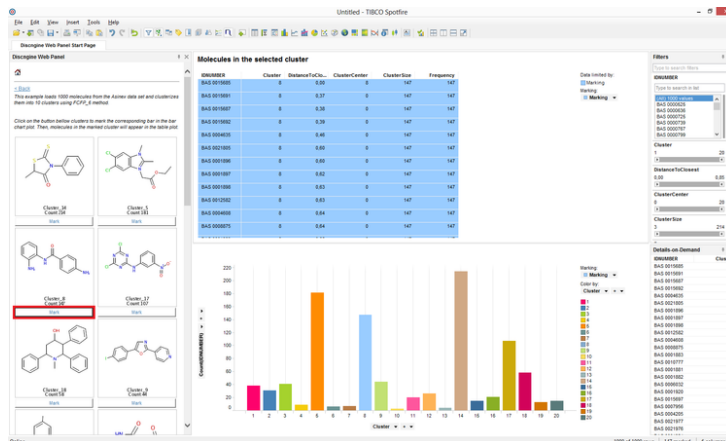
# The Future



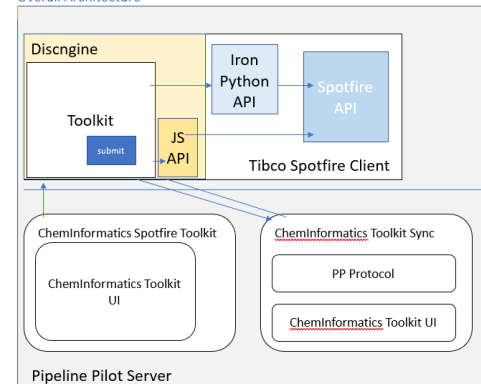
# AZ SpotOn



# Discngine Connector



# AZ Spotnik



Justin Morley



# Discngine: Run Pipeline Pilot protocols from TIBCO Spotfire

## Properties

AZlogD74  
ClogP  
structural alert  
Top 20  
Count bonds and rings  
LogBSF  
PAINS filter

## Analyse

Cluster  
CNS\_MPO  
Diverse selection  
Similarity

## Search

IsCompoundAZ  
IsPublic  
Name to structure  
Structure to name

NT100.dxp - TIBCO Spotfire

File Edit View Insert Tools Help

Page x +

Discngine Web Panel

Properties ▾ Analyse ▾ Search ▾ Dev ▾ Metrics

Please Select a Table: Data Table

Please Select a Structure/ID: compound ▾

C-Lab ClogP ⓘ

Show advanced options ▶

Marked  All Records  Overwrite Columns **Run**

Toolkit Version: 0.1.8

View help for the plugin

Find more help [here](#) ⓘ

Refresh the current view

Visit the Discngine Web Panel home page for more options and services

Online

Data Table

compound
AZ10045959
AZ10046111
AZ10046260
AZ10083919
AZ10093250
AZ10105372
AZ10105928
AZ10133745
AZ10143030
AZ10152488
AZ10155726
AZ10170090
AZ10233142
AZ10236925
AZ10238770
AZ10243096
AZ10276814
AZ10282911
AZ10291655
AZ10291755

Filters

Type to search filters

compound

Type to search in list

(All) 100 values

- AZ10014665
- AZ10014974
- AZ10015284
- AZ10016193
- AZ10016874
- AZ10017306

100 of 100 rows 0 marked 1 columns

# Advanced visualisations

**Near-Neighbour** ⓘ

collection

select all

Chemistry\_Connect

ISAC\_avail

Sigma\_Aldrich

NN

NN Availability

min similarity

fingerprint method

select all

path

tree

circular






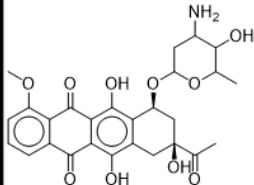
3D

Remove Duplicate NNs

Name\_Result

Marked  All Records  Overwrite Columns

**NN table**

Query ID	NearestNeigh...	Query Smiles	NearestNeigh...	Rank	Similarity	Collection	Method	z_score	TSC_ROWID
	358978748			1	0.99	Sigma_Aldrich	ThreeD	1,29	3
	29716983				0.99	Sigma_Aldrich	ThreeD	1,28	3
	28532017				0.99	Sigma_Aldrich	ThreeD	1,27	3

**NN Scatter Plot**

Query Smiles

Similarity

Marker by: (Row Num... +)

Color by: Method +

- circular
- path
- ThreeD
- tree

Shape by: (None)

All values

Size by: (None)

**NN histogram**

(RowCount)

Marking:  Marking

Color by: Method

- circular
- path
- ThreeD
- tree

Query ID

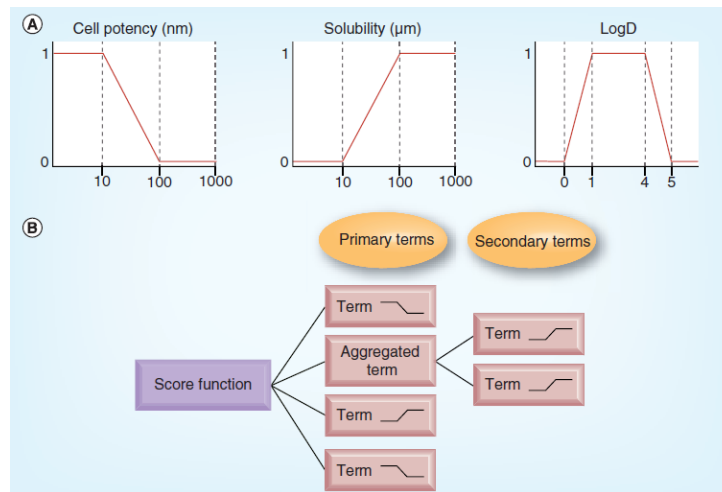


# Support of AI in drug discovery

Generate large numbers of entirely novel compounds

Multi-parameter optimization (MPO) of desirability scores used for selection

Example: MPO with default properties and parameters for a particular project



Nissink and Degorce (2013) Future Med. Chem. 5:753

ColumnName	Weight	function	P1	P2	P3	P4
clogp	1	MiddleValuesGood	0	2	4	6
Mol weight	1	LowValuesGood	400	500		
PSA	1	LowValuesGood	20	120		

CNS MPO; based on Wager et al (2010) ACS Chem. Neurosci. 1:435

Discngine Web Panel

Users ▾ Dev ▾ Analyse ▾ Properties ▾

Custom ▾ Utilities ▾ Search ▾ D360 ▾

Metrics

Please Select a Table: **chemblExampleData** ▾

Please Select a Structure/ID: **CANONICAL\_SMILES** ▾

MPO update score column successful. See Metrics for more information

**MPO auto** ⓘ

Project: **CNS** ▾ ⓘ

Show advanced options ▶

Marked  All Records  Overwrite Columns

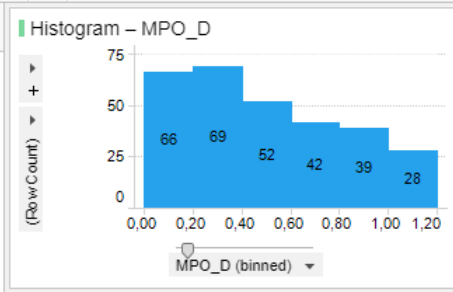
**Run**

Toolkit Version: 0.1.8

[View help for the plugin](#)

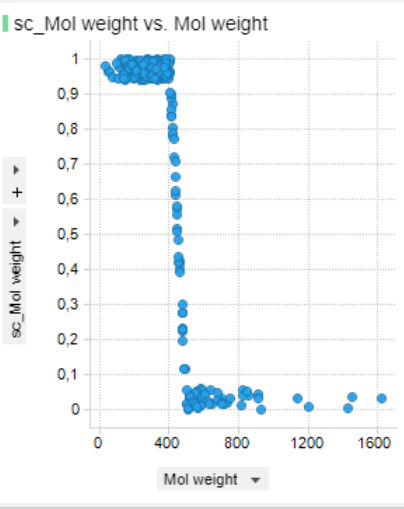
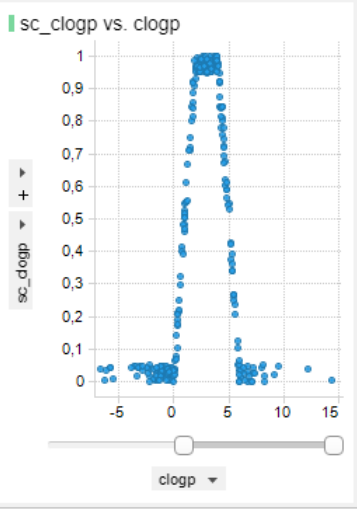
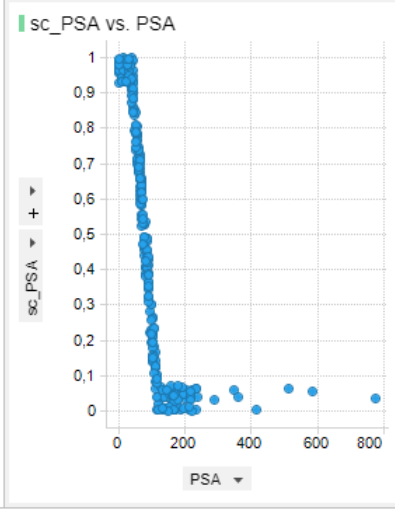
Find more help [here](#) ⓘ

[Refresh the current view](#)



### ParameterTable

ColumnName	MPO_rowNr	Weight	function	P1	P2	P3	P4
clogp	1	1	MiddleValuesGood	0	2	4	6
Mol weight	2	1	LowValuesGood	400	500		
PSA	3	1	LowValuesGood	20	120		





Discngine Web Panel

Users ▾ Dev ▾ Analyse ▾ Properties ▾

Custom ▾ Utilities ▾ Search ▾ D360 ▾

Metrics

Please Select a Table: chemblExampleData ▾

Please Select a Structure/ID: CANONICAL\_SMILES ▾

MPO update score column successful. See Metrics for more information

MPO auto ⓘ

Project

CNS ▾ ⓘ

Show advanced options ▶

 Marked
  All Records
  Overwrite Columns

Run

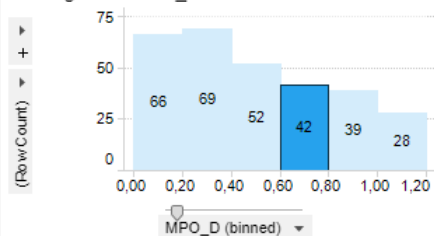
Toolkit Version: 0.1.8

View help for the plugin

Find more help [here](#) ⓘ

Refresh the current view

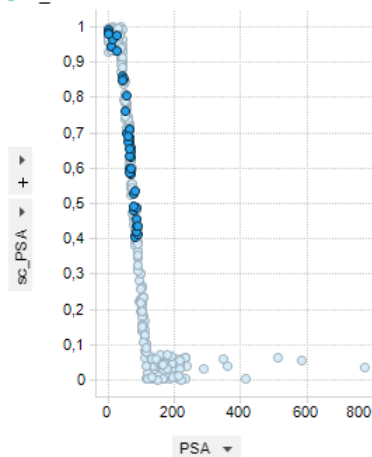
Histogram – MPO\_D



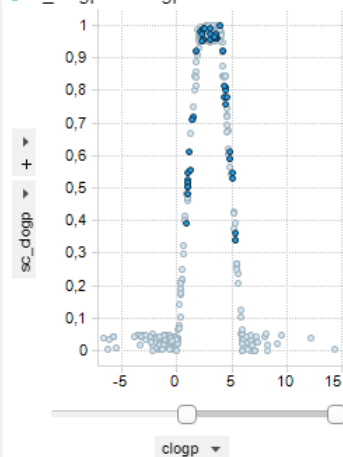
ParameterTable

ColumnName	MPO_rowNr	Weight	function	P1	P2	P3	P4
clogp	1	1	MiddleValuesGood	0	2	4	6
Mol weight	2	1	LowValuesGood	400	500		
PSA	3	1	LowValuesGood	20	120		

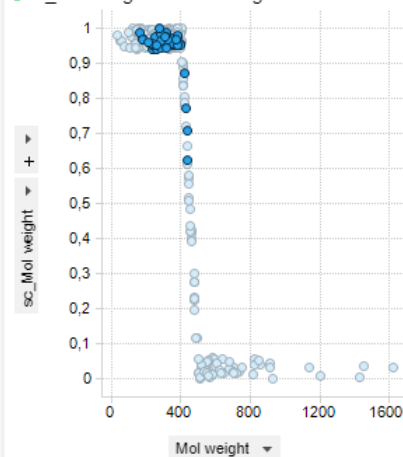
sc\_PSA vs. PSA



sc\_clogp vs. clogp



sc\_Mol weight vs. Mol weight



Discngine Web Panel

Users ▾ Dev ▾ Analyse ▾ Properties ▾

Custom ▾ Utilities ▾ Search ▾ D360 ▾

Metrics

Please Select a Table: **chemblExampleData** ▾

Please Select a Structure/ID: **CANONICAL\_SMILES** ▾

MPO update score column successful. See Metrics for more information

MPO auto ⓘ

Project: **CNS** ▾ ⓘ

Show advanced options ▶

Marked  All Records  Overwrite Columns

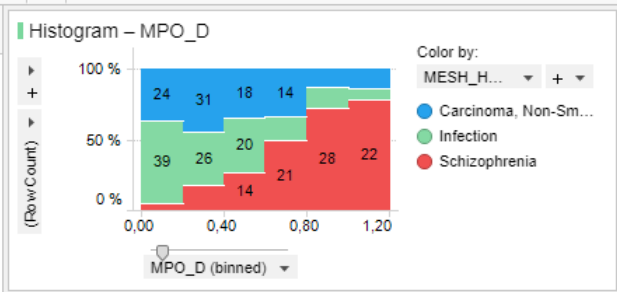
**Run**

Toolkit Version: 0.1.8

View help for the plugin

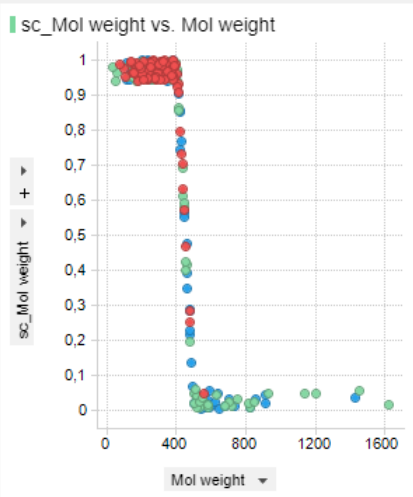
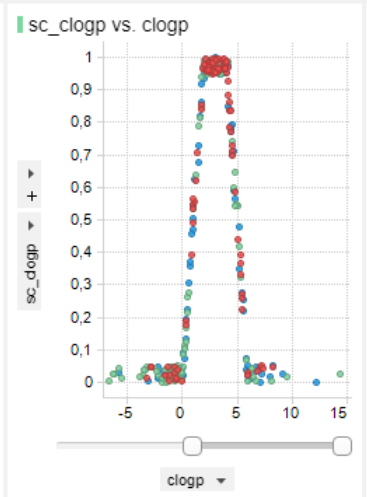
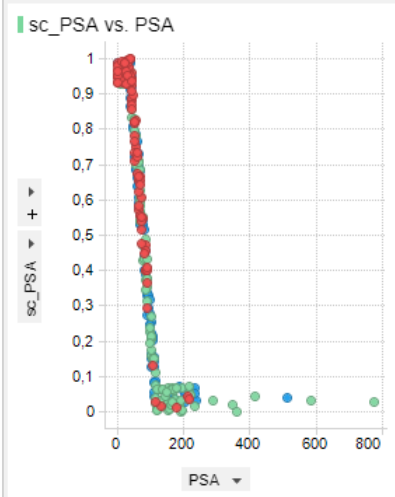
Find more help [here](#) ⓘ

Refresh the current view



### ParameterTable

ColumnName	MPO_rowNr	Weight	function	P1
clogp	1	1	MiddleValuesGood	0
Mol weight	2	1	LowValuesGood	400
PSA	3	1	LowValuesGood	20



# Summary

## Pros:

Advanced chemistry workflows available from Spotfire

Advanced visualisations that gives users a quick start for analysis

Interface that works well for a range of users

Implementation of complex workflows doesn't require programming skills

Good debugging options

Rapid implementation of new tools to scientists

## Cons:

3D structure viewer

Data transfer to Pipeline Pilot limited to one table

Asynchronous jobs not straightforward

Framework needs local support, maintenance and upgrades

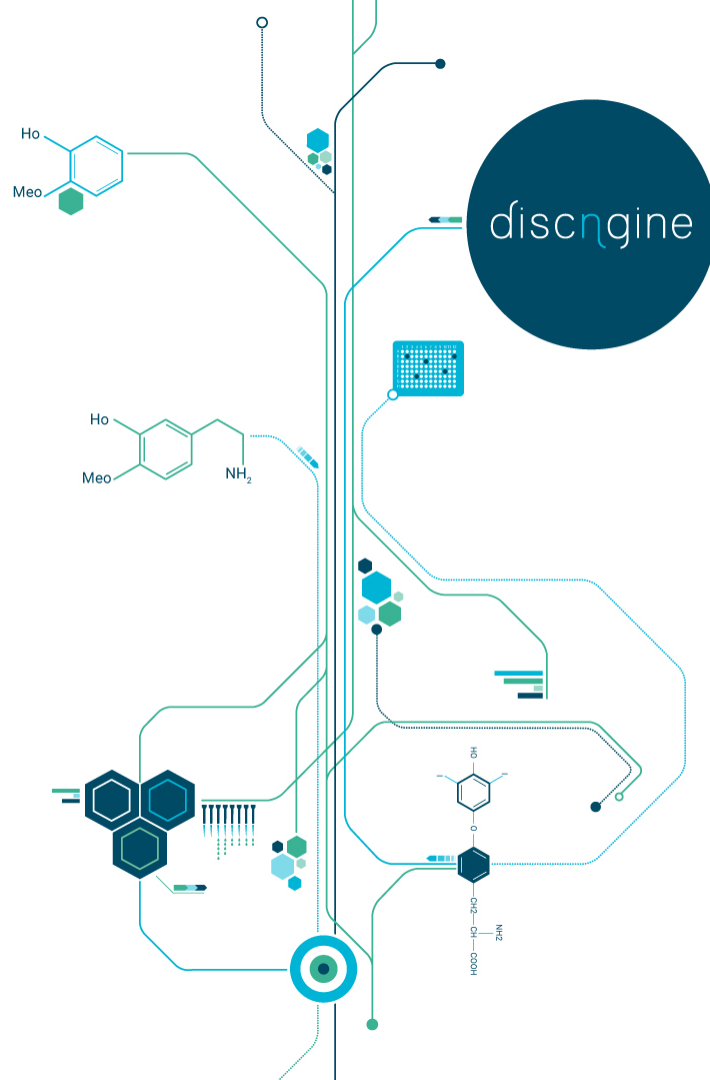


## Confidentiality Notice

This file is private and may contain confidential and proprietary information. If you have received this file in error, please notify us and remove it from your system and note that you must not copy, distribute or take any action in reliance on it. Any unauthorized use or disclosure of the contents of this file is not permitted and may be unlawful. AstraZeneca PLC, 1 Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge, CB2 0AA, UK, T: +44(0)203 749 5000, [www.astrazeneca.com](http://www.astrazeneca.com)



# Spotfire Web Application for Pipeline Pilot



# Connector

<https://connector.discngine.com>

## Client Automation



spotfireDocumentEditor

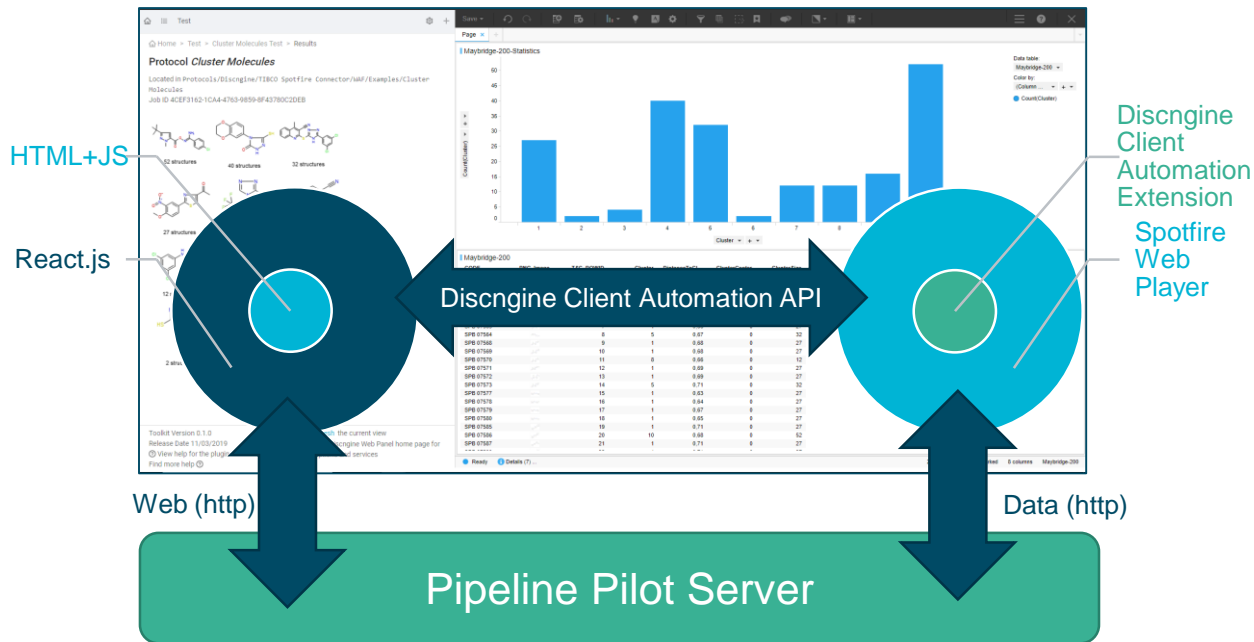
```
.addDataTable(sdbfFileURL)  
.addPage("Comparison Of Mean Ranges")  
.addBoxPlot(jsonConfig);
```



## Mashup for Spotfire Web Player



# S.W.A.P.P: Spotfire Web Application for Pipeline Pilot



in Web Player and Analyst



# SWAPP + Connector Client Automation

A new framework to streamline Pipeline Pilot protocol integration

- **Low code**

1. Declare the Pipeline Pilot protocol
2. Map the parameters to a web form (Wizard)
3. Define actions on the Spotfire document based on protocol output files

## Features

- **Authentication**
- **Job Control (start, poll, stop) including asynchronous run**
- **Push/Pull Data from Web Service endpoint**
- **Allow injection of HTML/JS Snippets**
- > Control Spotfire document
- > Display content

# S.W.A.P.P: Register Protocol 1/2

The screenshot displays the Cheminformatics software interface, divided into three main sections:

- Left Panel (Property Computation):** Shows the 'Summary' and 'Property Computation' sections. Under 'Input parameters', the 'Data table name' is set to 'Maybridge-300'. The 'Calculators' section includes 'ADMET' and 'Published models' with various checkboxes. The 'TOPKAT' section includes 'Classification models' with several checkboxes.
- Center Panel (Register new protocol):** A dialog box titled 'Register new protocol' with a progress bar showing three steps: 'Select protocol', 'Define inputs', and 'Create Menu Link'. The 'Select protocol' step is active, displaying a tree view of protocols. The 'SWAPP Examples' folder is expanded, and 'Cluster Molecules' is selected.
- Right Panel (Data Analysis):** Displays two plots. The top plot is a bar chart titled '300-Clusters' showing the distribution of clusters across 15 categories. The bottom plot is a scatter plot titled 'spe\_x' showing the distribution of 'spe\_x' values across 15 categories.

**300-Clusters Bar Chart Data (Approximate):**

Cluster	Count
1	22
2	8
3	4
4	1
5	41
6	19
7	2
8	45
9	1
10	1
11	81
12	25
13	16
14	1
15	29

**spe\_x Scatter Plot Data (Approximate):**

Cluster	spe_x
1	5.5
2	6.0
3	-1.5
4	-2.5
5	0.5
6	0.5
7	0.5
8	0.5
9	0.5
10	0.5
11	0.5
12	0.5
13	0.5
14	0.5
15	0.5

# S.W.A.P.P: Register Protocol 2/2

The image displays a software interface for registering a new protocol. The main window is titled "Register new protocol: Cluster Molecules" and is divided into three steps: "Select protocol", "Define inputs", and "Create Menu Link".

**Data Table Name \***

- Type: TIBCO Spotfire Data Table
- Description: [Empty text box]

**Overview / Default value**

User will have a select field to pick one of the loaded Data Tables.

**AvgNumberPerCluster**

- Type: Field
- Description: The average number of molecules per cluster. (Overrides <i>NumberOfClusters</i> if both are specified.)
- Advanced parameter

**Overview / Default value**

20

**Molecular Description**

- Type: Multi Checkboxes
- Description: A predefined set of properties to use for clustering
- Advanced parameter

**Overview / Default value**

- FCFP\_2  FCFP\_4  FCFP\_6  ALogP  Molecular\_Weight  Num\_H\_Donors
- Num\_H\_Acceptors  Num\_RotatableBonds  ECFP\_2  ECFP\_4  ECFP\_6  MDLPublicKeys
- Num\_Atoms  Num\_Rings  Num\_AromaticRings  Num\_Fragments
- Molecular\_PolarSurfaceArea  Molecular\_PolarSASA  NPlusO\_Count

Select all

The background shows a dashboard with two charts:

- 300-Clusters**: A bar chart showing the distribution of clusters. The x-axis is labeled "Cluster" (1-15) and the y-axis is labeled "300-Clusters". The bars are colored by "Sum(Cluster)" with a legend ranging from 4.00 to 902.00.
- spe\_x**: A scatter plot showing the distribution of molecules. The x-axis is labeled "spe\_x" and the y-axis is labeled "spe\_x". The markers are colored by "kN" with a legend ranging from 0.00 to 17.00.

At the bottom of the dashboard, it indicates "300 of 300 rows", "0 marked", and "14 columns".

# Example protocol: input

The screenshot displays the Cheminformatics software interface. On the left, the 'Input parameters' section is visible, including a dropdown for 'Data table name' set to 'Maybridge-200', a 'Marked rows only' checkbox, and a 'Molecular description' section with various checkboxes (FCFP\_2, FCFP\_4, FCFP\_6, ALogP, etc.). A 'Submit' button is located at the bottom of this section. The main area on the right shows a table titled 'Maybridge-200' with columns for 'Structure <M...>', 'CODE', and 'PIIG\_image'. The table lists 10 rows of chemical structures with their corresponding codes (SPB 07552 to SPB 07571). At the bottom of the interface, there is a footer with the text 'Powered by Discngine TIBCO Spotfire Connector v5', a 'Reload Page' link, and the user information 'Logged in to Pipeline Pilot as maxime.guitet'.

Cheminformatics

Home > Cheminformatics > Cluster Molecules

Summary

Cluster Molecules

Input parameters

\* Data table name: Maybridge-200

Marked rows only

Hide advanced parameters ^

Avg number per cluster: 20

Molecular description

FCFP\_2  FCFP\_4  FCFP\_6  ALogP

Molecular\_Weight  Num\_H\_Donors

Num\_H\_Acceptors  Num\_RotatableBonds

ECFP\_2  ECFP\_4  ECFP\_6

MDLPublicKeys  Num\_Atoms  Num\_Rings

Num\_AromaticRings  Num\_Fragments

Molecular\_PolarSurfaceArea  Molecular\_PolarSASA

NPlusO\_Count

Select all

Optimize for: Speed and Memory

Submit

Shortcut link

Auto Submit [Copy Shortcut Link](#)

Powered by Discngine TIBCO Spotfire Connector v5

[Reload Page](#)

Logged in to Pipeline Pilot as maxime.guitet

Structure <M...>	CODE	PIIG_image
[...]	SPB 07552	
[...]	SPB 07553	
[...]	SPB 07554	
[...]	SPB 07560	
[...]	SPB 07561	
[...]	SPB 07562	
[...]	SPB 07563	
[...]	SPB 07564	
[...]	SPB 07568	
[...]	SPB 07569	
[...]	SPB 07570	
[...]	SPB 07571	

Page | 200 of 200 rows | 0 marked | 3 columns

# Example protocol: output

Cheminformatics

Home > Cheminformatics > Cluster Molecules > Results

## Protocol Cluster Molecules

Located in ChemoInformatics  
Job ID 4854AB75-7E67-4E70-9892-CD567FCF87FA

53 structures

40 structures

32 structures

27 structures

16 structures

12 structures

12 structures

4 structures

2 structures

2 structures

Powered by Discngine TIBCO Spotfire Connector v5  
[Reload Page](#)  
Logged in to Pipeline Pilot as *maxime.guitet*

### Maybridge-200

Structure <M...	CODE	PHI_Image	TSC_ROWID	Cluster	DistanceToCl...	Cluster
[...]	SPB 07552		e2e115ca- c342-414b- 9bfa-597fd2c390 3c	7	0.73	
[...]	SPB 07553		dc5a65cb- 71ae-4218- a6ca-2ec3eda52 e99	7	0.70	
[...]	SPB 07554		4bb3df10-8cbf- 4817- ae0f-583b38596 906	2	0.67	
[...]	SPB 07560		92fd1081-043a- 497c- a11c-6f83e3f0b2 9e	1	0.70	
[...]	SPB 07561		0321c7a5- dad5-479b-ba5d- ae3b167c40fe	1	0.71	
[...]	SPB 07562		32f21d2c- f955-435a-8805- d88c7f84b68c	1	0.70	
[...]	SPB 07563		a795241a- ad0e-4e38- aa48-480ea1f93 24f	1	0.66	
[...]	SPB 07564		e3490860- e2e4-4035-9fb0- 4f5a75a8c462	5	0.67	
[...]	SPB 07568		386c2b30- f374-4acc-836c- f5068a0e3ebf	1	0.68	
[...]	SPB 07569		c18eccd9-4361- 4a3f- 821f-2754a0694 385	1	0.68	
[...]	SPB 07570		41db5042-737d- 4690- ae4c-43171ae34 60f	8	0.66	
[...]	SPB 07571		093709e5-83dd- 4038-964b- 93c1a97c733c	1	0.69	

Page

### Maybridge-200-Clusters

Cluster	Count(Cluster)
1	27
2	2
3	4
4	40
5	32
6	2
7	12
8	12
9	16
10	53

Count(Cluster)

Cluster

200 of 200 rows 0 marked 8 columns



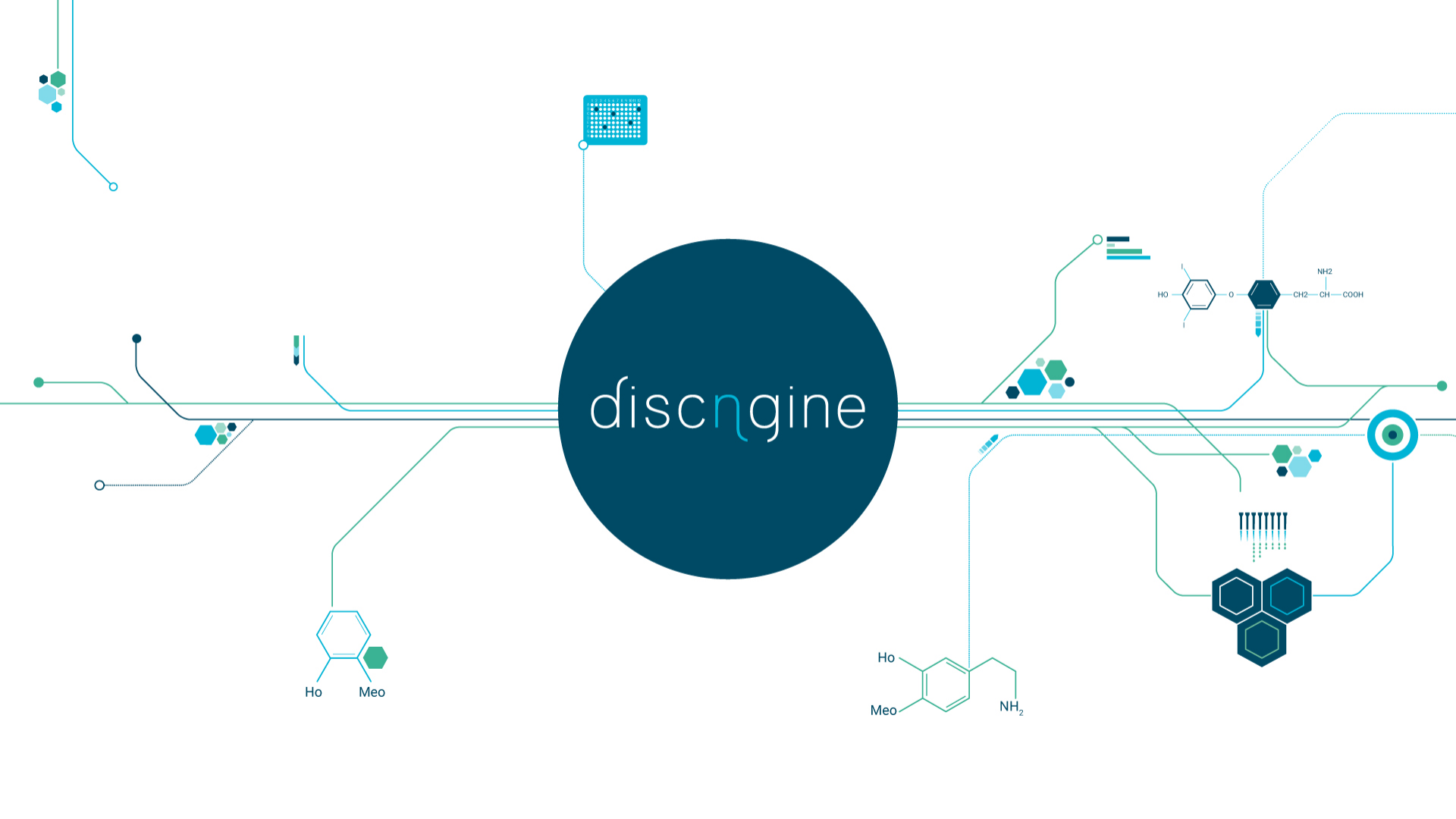
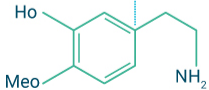
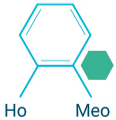
# Future work

Extend to Knime, Talend

Implement more features

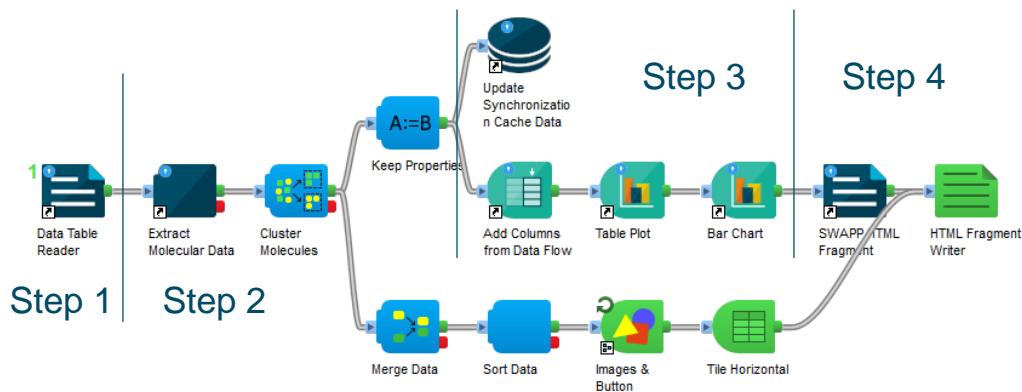
- Chain protocols
- Extend user permissions to control visibility of protocols
- Allow conditional linking of parameters

# discngine





# Existing application: SWAPP - Example



- Step 1: reading data from SBDF
- Step 2: data transformation
- Step 3: preparing API calls
- Step 4: combining API calls and creating HTML/JS file