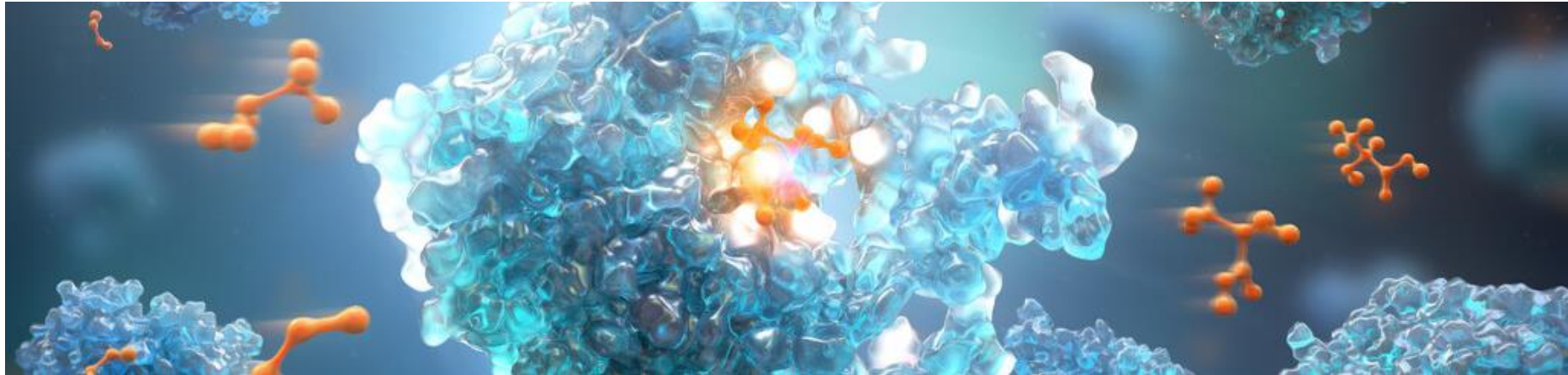


A modern tool for integrating cheminformatics data science in Spotfire

Lars Brive, Justin Morley, Maxime Guitet

PerkinElmer EMEA Nexus 2019



Acknowledgements



Wolfgang Klute
Johan Ulander
Nick Tomkinson



Eric Le Roux
Claire Wallon
Benjamin Pannetier

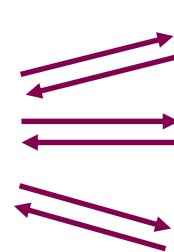
Michael Bodnarchuk
Peter Brandt
Sameer Kawatkar
Susanne Winiwarter
Kun Song
Magnus Polla
Ekaterina Ratkova
SpotOn team



Spotfire



Biovia Pipeline Pilot



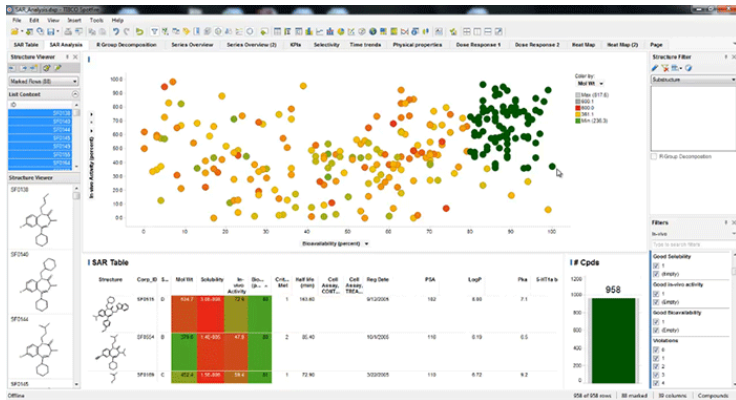
Databases

Calculation services

AI tools

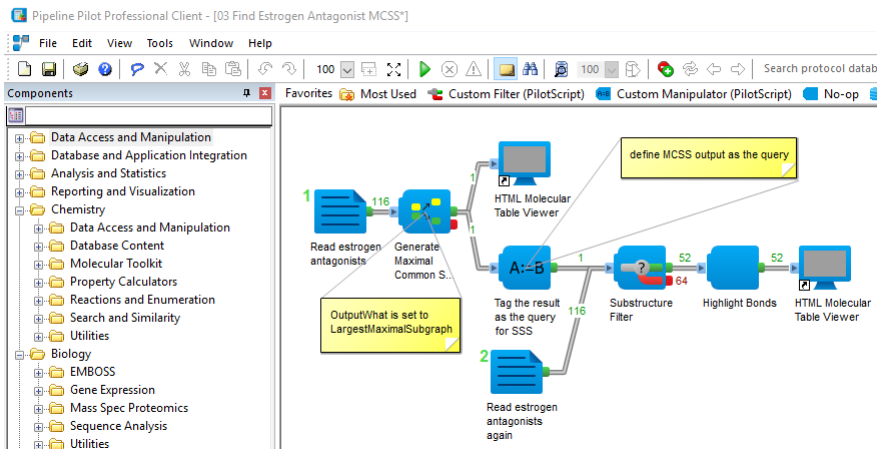
....

Spotfire:
Analysis, visualisations, PerkinElmer Lead Discovery



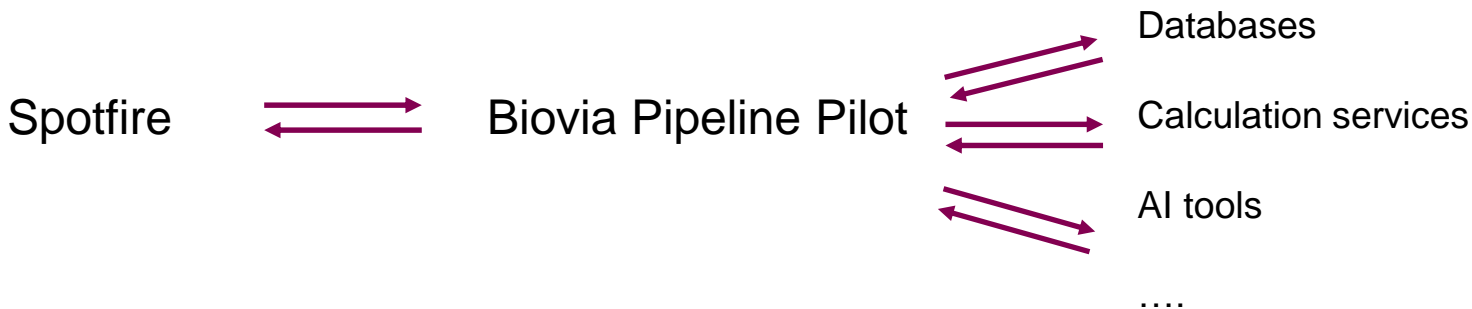
PerkinElmer.com

Biovia Pipeline Pilot: Data processing



Pipeline Pilot example protocol





Requirements

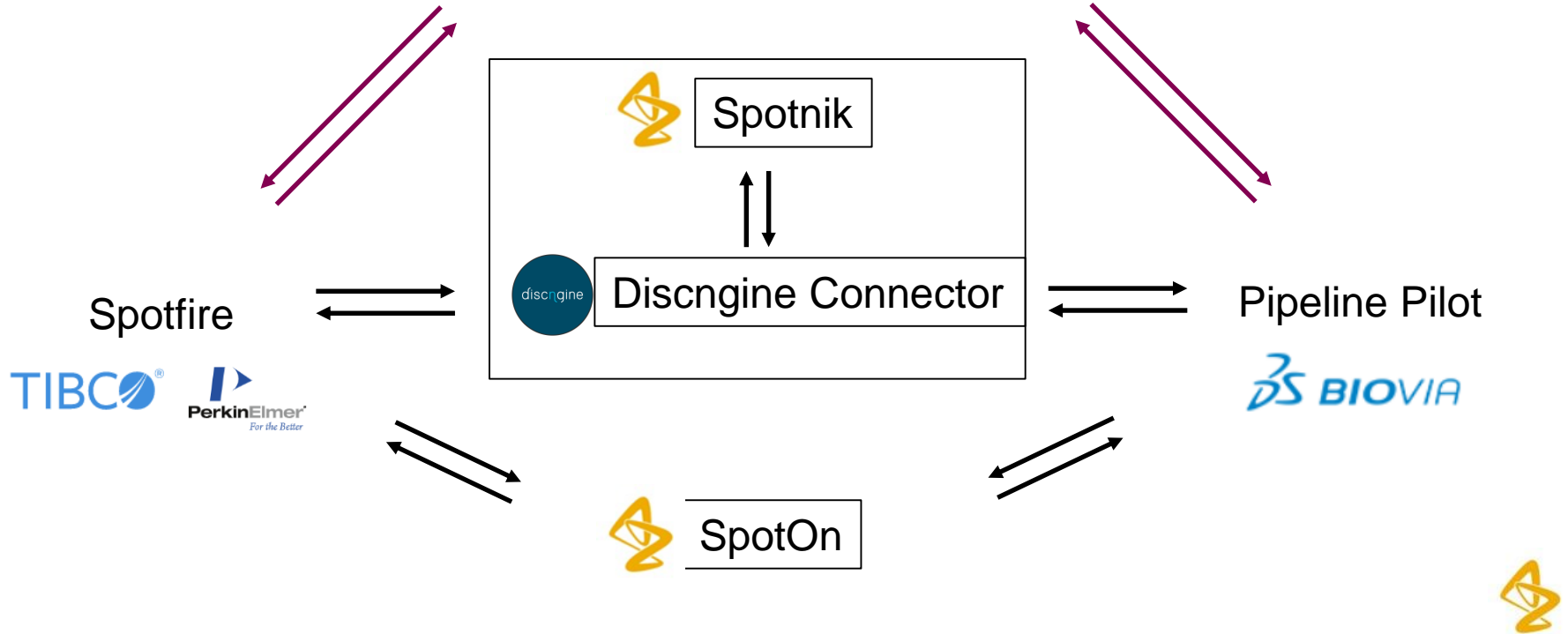
- Interactive and flexible analysis of data
- Complex queries made easy.
- Default visualisations.
- Range of user expectations: One green button or full control over parameters.
- Allow advanced users to implement their own protocols.
- Modular design
- Rapid implementation. Global usage.

”Are these compounds available in our compound collection”

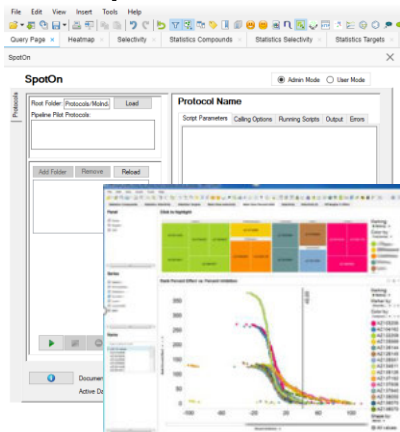
”Given a list of compound names, extract the bioactivity data in test, comparison with internal and external data. Group by target type.”



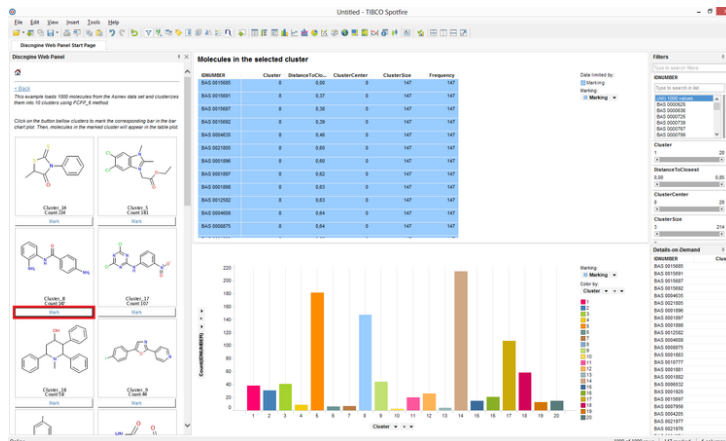
The Future



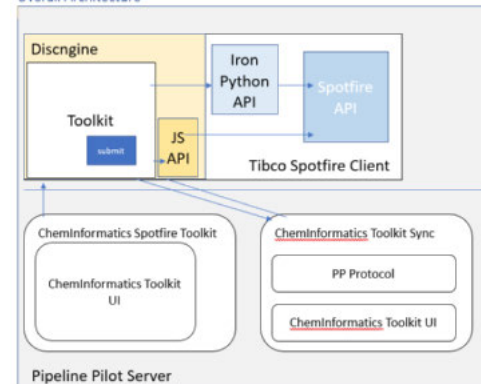
AZ SpotOn



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






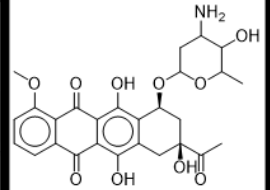
3D

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

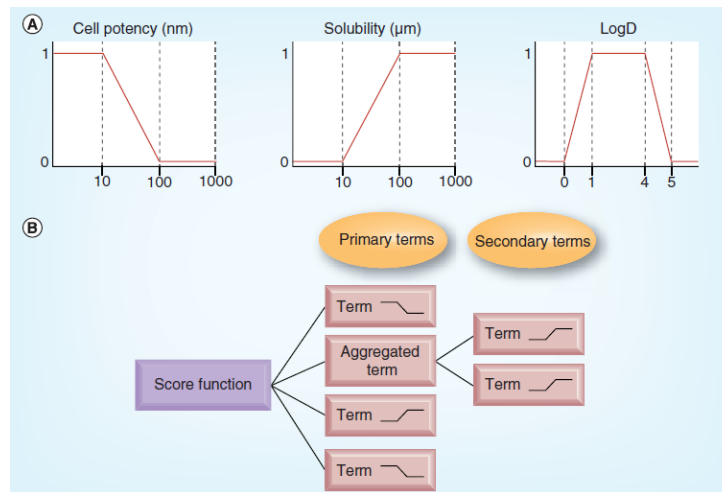
Query ID

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

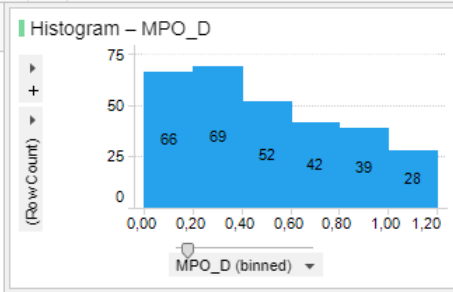
Run

Toolkit Version: 0.1.8

[View help for the plugin](#)

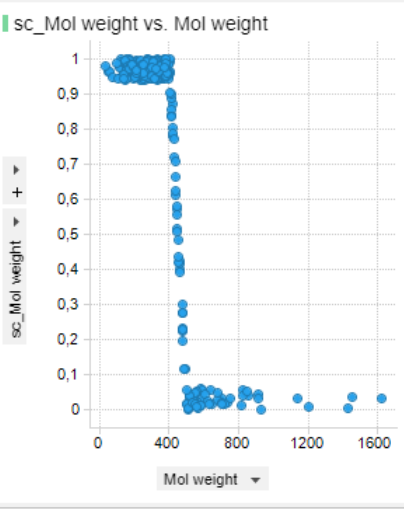
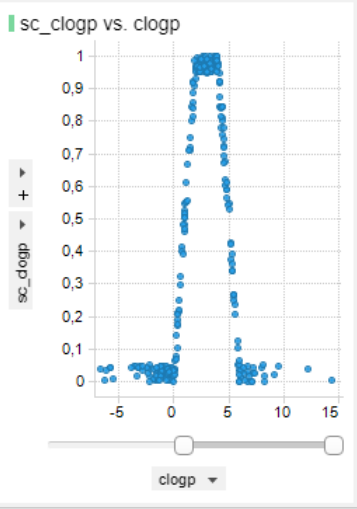
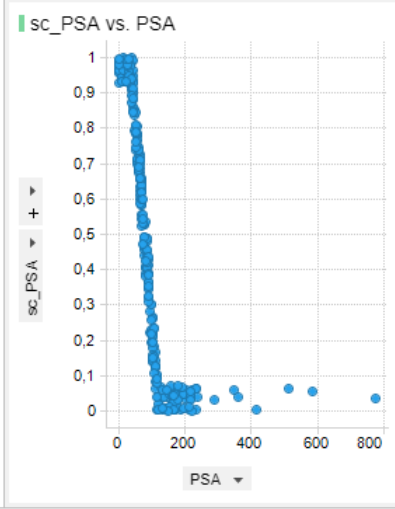
Find more help [here](#) ⓘ

[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

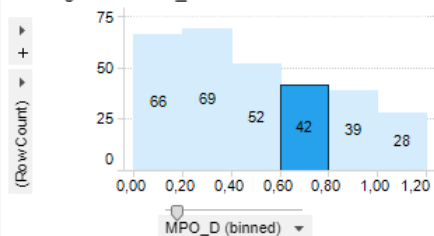
Toolkit Version: 0.1.8

View help for the plugin

Find more help [here](#) ⓘ

Refresh the current view

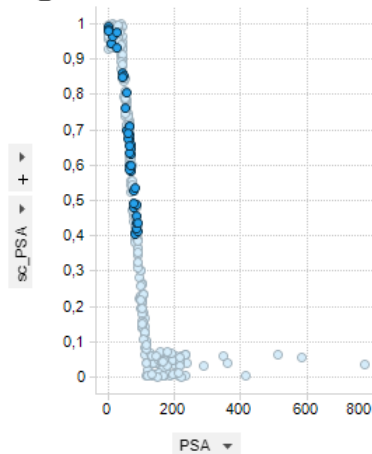
Histogram – MPO_D



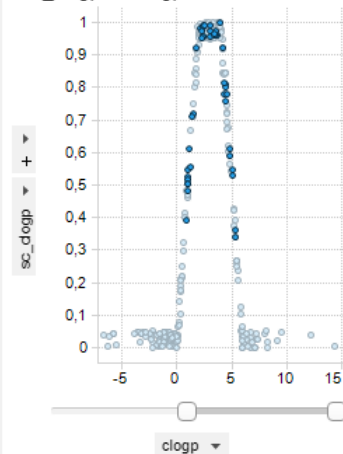
ParameterTable

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		

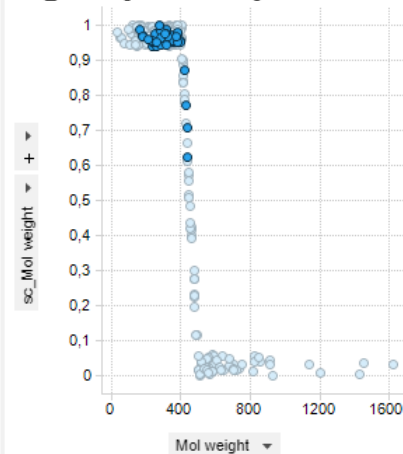
sc_PSA vs. PSA



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

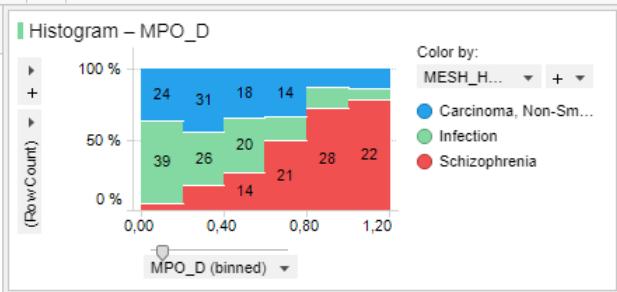
Run

Toolkit Version: 0.1.8

View help for the plugin

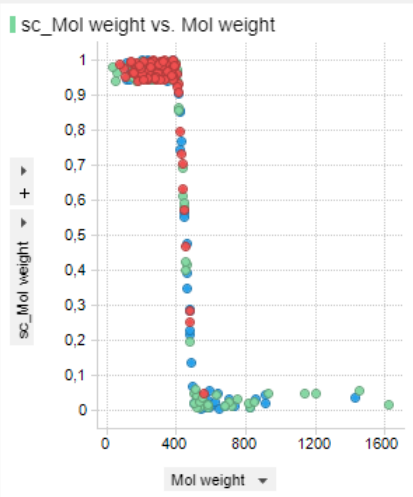
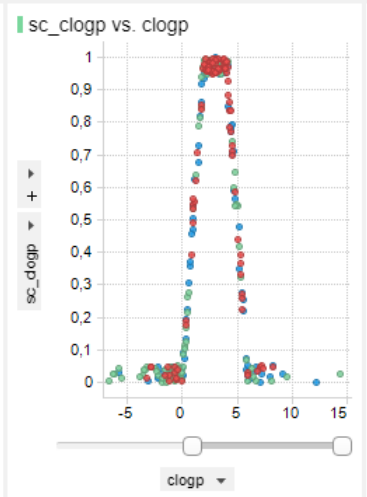
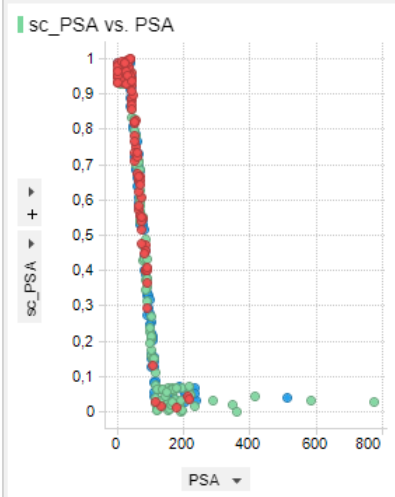
Find more help [here](#) ⓘ

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

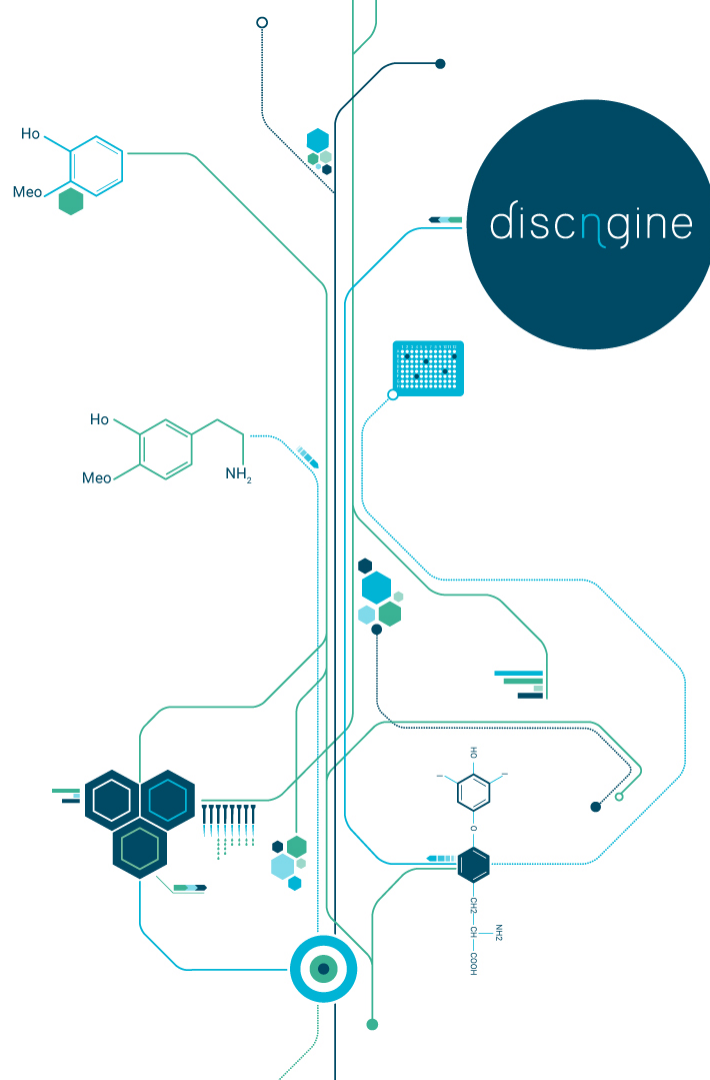


Confidentiality Notice

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



Spotfire Web Application for Pipeline Pilot



Connector

<https://connector.discngine.com>

Client Automation



spotfireDocumentEditor

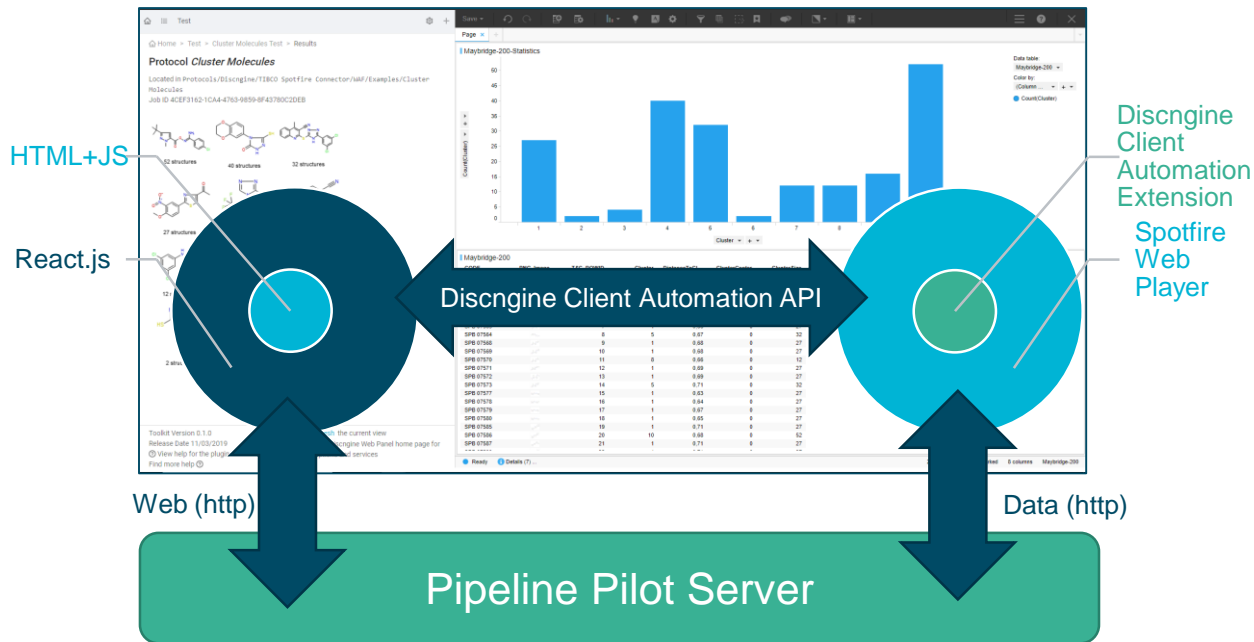
```
.addDataTable(sdbfFileURL)  
.addPage("Comparison Of Mean Ranges")  
.addBoxPlot(jsonConfig);
```



Web Panel for Spotfire Analyst

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 image displays the S.W.A.P.P. software interface, divided into three main sections:

- Left Panel (Property Computation):** Shows configuration options for data and calculators. Under "Input parameters", the "Data table name" is set to "Maybridge-300". Under "Calculators", the "ADMET" section has "EXT CYP2D6", "EXT Hepatotoxic", and "EXT PP" checked. The "Published models" section has "Absorption Level", "BBB Level", "BBB", "Solubility Level", and "Solubility" checked.
- Center Panel (Register new protocol):** A dialog box with a progress bar and three steps: "Select protocol", "Define inputs", and "Create Menu Link". The "Select protocol" step is active, showing a tree view of protocols. "Cluster Molecules" is highlighted in the "SWAPP Examples" folder.
- Right Panel (Data Visualization):** Displays two charts. The top chart is a bar plot titled "300-Clusters" with the x-axis labeled "Cluster" (1-16) and the y-axis representing frequency. The bottom chart is a scatter plot titled "tpe_x" with the x-axis labeled "tpe_x" and the y-axis labeled "tpe_x". Both charts include legends for "Color by: Sum(Cluster)" and "Marker by: (Row Number)".

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 sections: "Select protocol", "Define inputs", and "Create Menu Link".

Select protocol: The "Data Table Name" field is set to "TIBCO Spotfire Data Table".

Define inputs:

- AvgNumberPerCluster:** Type is "Field". Description: "The average number of molecules per cluster. (Overrides $\langle i \rangle \text{NumberOfClusters} \langle /i \rangle$ if both are specified.)"
- Molecular Description:** Type is "Multi Checkboxes". Description: "A predefined set of properties to use for clustering".

Overview / Default value: The value is set to 20.

Advanced parameter options:

- 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

Left Panel (Property Computation):

- Input parameters:** Data table name: Maybridge-300.
- Calculators:** ADMET: Extensible model (EXT CYP2D6, EXT Hepatotoxic, EXT PP2A) - Select all.
- Published models:** Absorption Level, BBB Level, BBB Solubility Level, Solubility - Select all.
- TOPKAT:** Classification models: Aerobic Biodegradability, Ames Mutagenicity, Developmental Toxicity Potential, Mouse Female FDA None vs Carcinogen, Mouse Female FDA Single vs Multiple, Mouse Female NTP, Mouse Male FDA None vs Carcinogen, Mouse Male FDA Single vs Multiple, Mouse Male NTP, Ocular Irritancy Mild vs Moderate Severe, Ocular Irritancy Moderate vs Severe, Ocular Irritancy None vs Irritant, Rat Female FDA None vs Carcinogen, Rat Female FDA Single vs Multiple, Rat Female NTP, Rat Male FDA None vs Carcinogen, Rat Male FDA Single vs Multiple, Rat Male NTP, Skin Irritancy Mild vs Moderate Severe, Skin Irritancy Moderate vs Severe, Skin Irritancy None vs Irritant, Skin Sensitization None vs Sensitizer.

Right Panel (Visualizations):

- 300-Clusters:** A bar chart showing the distribution of clusters. The x-axis is "Cluster" (1-16) and the y-axis is "Count". The bars are colored by "Size/Cluster".
- Scatter Plot:** A scatter plot showing the relationship between "tpe_x" and "tpe_y". The x-axis ranges from 1 to 16, and the y-axis ranges from -6 to 6. The points are colored by "Marker: (Row Number)" and "Color by: Size/Cluster".

Example protocol: intput

The screenshot displays a web interface for Cheminformatics. The left sidebar contains a 'Summary' section for 'Cluster Molecules' with the following input parameters:

- Data table name: Maybridge-200
- Marked rows only:
- 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

A 'Submit' button is located below the parameters. Under 'Shortcut link', there are options for 'Auto Submit' and 'Copy Shortcut Link'.

The main content area shows a table titled 'Maybridge-200' with the following columns: 'Structure <M...>', 'CODE', and 'PIG_image'. The table contains 11 rows of chemical structures, each with a corresponding code (SPB 07552 to SPB 07571). The status of each row is indicated as '[-]'.

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 a login message: 'Logged in to Pipeline Pilot as maxime.guillet'. The bottom right corner shows 'Page', '- 200 of 200 rows', '0 marked', and '3 columns'.

Example protocol: output

Cheminformatics

Home > Cheminformatics > Cluster Molecules > Results

Protocol Cluster Molecules

Located in Cheminformatics
Job ID 4854AB75-7E67-4E7D-9892-CD567FCF87FA

53 structures

45 structures

32 structures

27 structures

16 structures

12 structures

12 structures

4 structures

2 structures

2 structures

Structure <M...>	CODE	PIIG_Image	TSC_ROWID	Cluster	DistanceToCl...	Cluste...
[...]	SPB 07552		e2e115ca-c3d2-414b-9b4f-597f62c3903c	7	0.73	
[...]	SPB 07553		d6fa25cb-71ae-421b-a6ca-2ec3eda52ea9	7	0.70	
[...]	SPB 07554		4bb3df10-8cbf-4817-ae6f-5d3e3b6596800	2	0.67	
[...]	SPB 07560		926f0081-043a-4897c-a11c-6f83e310b29e	1	0.70	
[...]	SPB 07561		0321c8e5-f8e5-4720-ba5d-ae5b167c40fe	1	0.71	
[...]	SPB 07562		32f21d2c-1955-435a-8805-989c7f04b68c	1	0.70	
[...]	SPB 07563		a795241a-a90e-4e3b-a84b-480ea1f0324f	1	0.68	
[...]	SPB 07564		e3d90960-e2e4-4055-9f60-4f5a750c482	5	0.67	
[...]	SPB 07568		295c2b30-074-4acc-836c-f5068a0e3ebf	1	0.68	
[...]	SPB 07569		c19eccd9-4381-4a35-821f-2754a0694385	1	0.68	
[...]	SPB 07570		41db5642-737d-4895-a84c-43171a83460f	8	0.66	
[...]	SPB 07571		68f70b-e-83dd-403b-964b-93c1a67c3f3c	1	0.69	

Maybridge-200-Clusters

Cluster	Count(Cluster)
1	27
2	2
3	4
4	65
5	32
6	2
7	12
8	12
9	16
10	65

Page | 200 of 200 rows | 0 marked | 8 columns

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



Future work

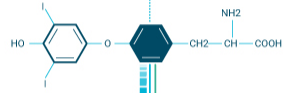
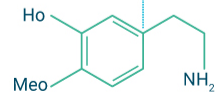
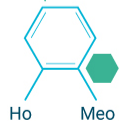
Extend to Knime, Talend

Implement more features

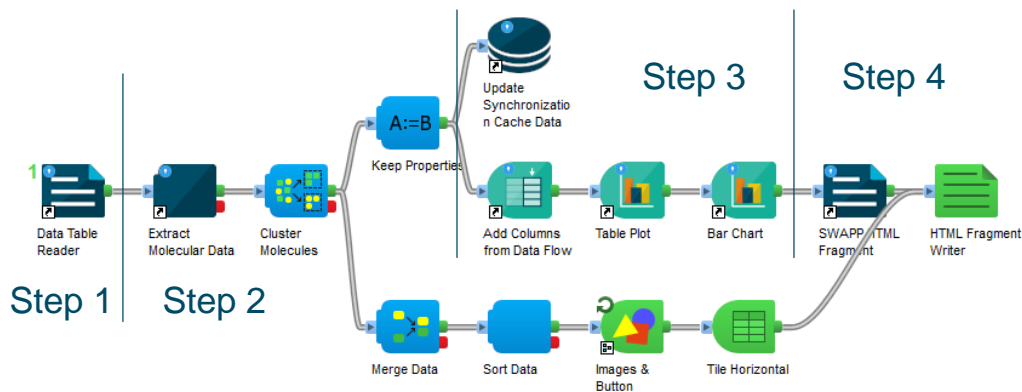
- Chain protocols
- Extend user permissions to control visibility of protocols
- Allow conditional linking of parameters



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