



Putting RDKit in the hands of the users in a corporate environment: Leveraging the power of Datagrok

Gregori Gerebtzoff
21st of September, 2023
12th RDKit UGM

The Intuence product suite

Datagrok

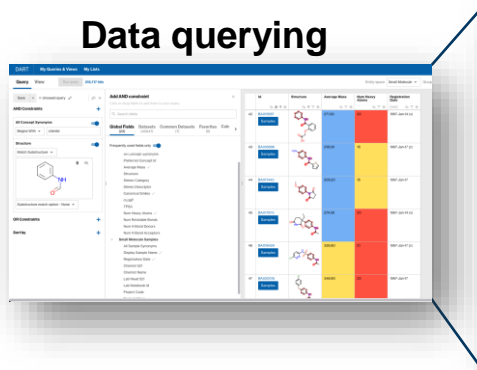
KNIME

Spotfire

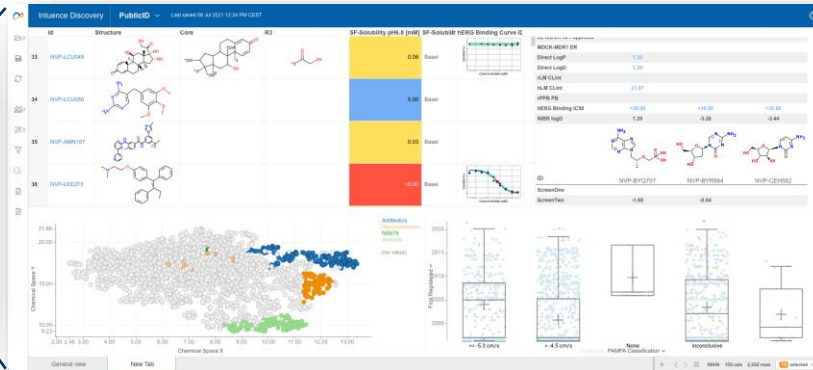
Python



Data querying



Interactive dashboards, decision-making: *Intuence Discovery*



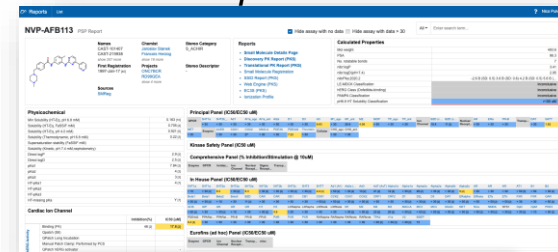
Virtual molecules, new ideas, annotations



ML & Data science



“One-pager” reports: *Intuence Reports*



Data
warehouse



Intuence Discovery: RDKit in back- & front-end

Intuence Discovery PublicID Analysis last saved 6 months ago Data last refreshed 3 hours ago

Primary Series Name 4

- ☒ Antibiotics
- ☒ Benzimidazoles
- ☒ NS676
- ☒ Steroids

Structure

RDKit substructure search

RDKit R-group decomposition

In-line rendering options and export

RDKit fingerprint for UMAP chem. space

RDKit depicter with highlighting

	Id	Structure	Series	Core	R1	R2	R3	Chemical Space X	Chemical Space Y
5	NVP-AFC302		Antibiotics					18.59	9.60
6	NVP-AGC606		Benzimidazoles					19.39	6.56
7	NVP-LFF753		Antibiotics						
8	CGP027827								

Average Mass

ID	NVP-AGC606	NVP-ARL870	NVP-CPR867	NVP-LFG065	NVP-LKS610
ScreenOne	4.06	2.21	4.27	1.81	8.48
ScreenTwo	-3.58	-1.28	5.22	-0.944	-34.0
Screen3	0.890	1.1	0.990	1.04	0.917
HT Solubility pH6.8 [mM]	0.16	0.06	1.70	>1.00	0.28
LE-MDCK v2 PappAtoB			1.60	0.61	
rLM CLint	150.00	240.00	33.31	105.00	220.00
hLM CLint	23.00	27.30	56.20	20.20	35.00
hERG Binding IC50	12.00	>30.00	17.00	>30.00	>30.00
NIBR logP / NIBR logD	3.04 / 2.27	2.15 / -0.376	2.82 / 2.63	3.30 / 3.19	1.82 / -1.62
Series	NS676	Antibiotics	Benzimidazoles	Antibiotics	Antibiotics

General view SimpleCorrelations DataGrid PK Page ModelInfoTable New Tab

MAIN 209 cols Showing 11 of 2342 rows 7 SELECTED

Structure Renderer Options

- ☐ Re-compute 2D layout
- ☒ Align to scaffold
- ☐ Highlight scaffold
- ☐ Abbreviate groups

Formats

SMILES

COc1ccc2nc(S(=O)(=O)Cc3ccc...

Molblock

2510055
RDKit 2D
24 26 0 0 0 0 0

InChI

InChI=1S/C17H19N3O3S/c

PNG SVG

Leveraging RDKit.js in Datagrok

Intence Discovery

PublicID

Analysis last saved 2 days ago

Data last refreshed 2 days ago

4

Primary Series Name

Antibiotics

Benzimidazoles

NS676

Steroids

Structure

N

NIBR logD

Average Mass

	Id	Structure	Series	Core	R1	R2	R3
19	NVP-AFC302		Antibiotics				
20	NVP-AGC606		Benzimidazoles				
21	NVP-LFF753		Antibiotics				
22	CGP027827		NS676				
23	NVP-AFC599		Antibiotics				

ID	CGP027827	NVP-AFC302	NVP-AGC606	NVP-ARL870	NVP-CPR867	NVP-LFG065
ScreenOne	0.20	-0.82	4.06	2.21	4.27	1.81
ScreenTwo	3.58	-1.28	-33.78		5.22	-0.94
ScreenThree	0.89	1.10		0.99	1.00	1.04
HT Solubility pH6.8 [mM]			0.16	0.06	1.70	>1.00
LE-MDCK v2 PappAtoB					1.60	0.61
rLM CLint			150.00	240.00	33.31	105.00
hLM CLint			23.00	27.30	56.20	20.20
hERG Binding IC50	12.00	>30.00	17.00	>30.00	350.00	>30.00
NIBR logP / NIBR logD	3.04 / 2.27	2.15 / -0.38	2.82 / 2.63	3.30 / 3.19	1.82 / -1.62	1.71 / -1.68
Series	NS676	Antibiotics	Benzimidazoles	Benzimidazoles	Antibiotics	Antibiotics

Actions

Add to variables

Copy value

Sketch

Chemistry

CHEMBL similarity

CHEMBL1344: 0.671

CHEMBL1503: 0.643

CHEMBL1201320: 0.643

CHEMBL4117512: 0.643

Interactive js code execution upon structure selection

Leveraging RDKit.js in Datagrok

Adding script as a column (fetching while scrolling)

Intelligence Discovery | PubliCID | Analysis last saved 2 days ago | Data last refreshed 2 days ago

Primary Series Name 4

- ☒ Antibiotics
- ☒ Benzimidazoles
- ☒ NS676
- ☒ Steroids

Structure

NIBR logD

Average Mass

	Id	Structure	Series	Core	R1	R2	R3	Chemistry ChEMBL similarities
19	NVP-AFC302		Antibiotics					<ul style="list-style-type: none">ChEMBL370252:1ChEMBL1672972:1ChEMBL170052:0.678ChEMBL170053:0.678ChEMBL168850:0.678ChEMBL190561:0.647ChEMBL363449:0.647ChEMBL170319:0.643ChEMBL355474:0.643ChEMBL316165:0.635
20	NVP-AGC606		Benzimidazoles					<ul style="list-style-type: none">ChEMBL1344:0.671ChEMBL1503:0.643ChEMBL1201320:0.643ChEMBL4117512:0.643ChEMBL892:0.62ChEMBL4299164:0.584ChEMBL2221248:0.584ChEMBL50030:0.576ChEMBL9890:0.569ChEMBL301137:0.557
21	NVP-LFF753		Antibiotics					<ul style="list-style-type: none">ChEMBL4:1ChEMBL33:1ChEMBL420937:1ChEMBL556371:0.961ChEMBL2237108:0.961ChEMBL966:0.863ChEMBL1459:0.863

ID	CGP027827	NVP-AFC302	NVP-AGC606	NVP-ARL870	NVP-CPR867	NVP-LFG065	NVP-LKS610
ScreenOne	0.20	-0.82	4.06	2.21	4.27	1.81	8.48
ScreenTwo	3.58	-1.28	-33.78		5.22	-0.94	-33.98
ScreenThree	0.89	1.10		0.99	1.00	1.04	0.92
HT Solubility pH6.8 [mM]			0.16	0.06	1.70	>1.00	0.28
LE-MDCK v2 PappAtoB					1.60	0.51	
rLM CLint			150.00	240.00	33.31	105.00	220.00
hLM CLint			23.00	27.30	56.20	20.20	35.00
hERG Binding IC50	12.00	>30.00	17.00	>30.00	350.00	>30.00	
NIBR logP / NIBR logD	3.04 / 2.27	2.15 / -0.38	2.82 / 2.63	3.30 / 3.19	1.82 / -1.62	1.71 / -1.68	2.80 / 2.61
Series	NS676	Antibiotics	Benzimidazoles	Benzimidazoles	Antibiotics	Antibiotics	Benzimidazoles

General view

MAIN 203 cols Showing 39 of 2342 rows 2 SELECTED

Leveraging RDKit.js in Datagrok

https://grok.dev.nibr.novartis.net/script/6c67ecD-3670-11ee-8f8d-23d98c01d3aa

Scripts > Chemistry | ChEMBL similarity

SAVE

Apps

Functions

Scripts

Chemistry | ChEMBL s...

Models

Notebooks

Actions <

Samples <

```
1 //name: Chemistry | ChEMBL similarity
2 //description: This script returns the top n most similar ChEMBL structures
3 //language: javascript
4 //condition: true
5 //tags: panel
6 //input: string structure = "c1ccccc1CCC(NO)" {semType: Molecule} [Molecule to search by]
7 //output: string results
8
9 const rdkitModule = chem.getRdkitModule();
10 const mol = rdkitModule.get_mol(structure);
11 const smiles = mol.get_smiles();
12
13 //console.log("SMILES: ${smiles}");
14 const searchData = await searchArthorForChEMBL(smiles);
15 const similarStructures = getSimilarStructures(searchData);
16 writeResultBlock(similarStructures);
17
18 async function searchArthorForChEMBL(smiles) {
19   const urlTemplate = "https://arthor-svc.nibr.novartis.net/arthor-server//dt/chembl28_may_21/search?type=Similarity&draw=4&start=0&length=40&query="
20
21   const response = await fetch(`${urlTemplate}${smiles}`, {
22     mode: "cors",
23     credentials: "include"
24   });
25   const results = await response.json();
26   return results.data;
27 }
28
29 function getSimilarStructures(results, nStructures = 10) {
30   const topNStructures = results.slice(0, nStructures).map((elem) => {
31     return {
32       structure: elem[1].split('\t')[0],
33       id: elem[1].split('\t')[1],
34       similarity: parseFloat(elem[2])
35     };
36   });
37   return topNStructures;
38 }
39
40 function writeResultBlock(searchResults) {
41   const rootDiv = document.createElement("div");
42
43   let textElem = null;
44
45   searchResults.forEach((item, i) => {
46     const itemDiv = document.createElement("div");
47     itemDiv.style.border = "1px solid #bbb";
48     itemDiv.style.margin = "0 0 10px 0";
49     itemDiv.style.color = "red";
50     if (item.similarity >= 1.0) {
51       itemDiv.style.color = "green";
52     } else if (item.similarity >= 0.7) {
53       itemDiv.style.color = "blue";
54     }
55     const textElem = document.createTextNode(`${item.id}: ${item.similarity}`);
56     itemDiv.appendChild(textElem);
57     const imgDiv = document.createElement("div");
58     const mol = rdkitModule.get_mol(item.structure);
59     const svg = mol.get_svg();
60     imgDiv.innerHTML = svg;
61     itemDiv.appendChild(imgDiv);
62     rootDiv.appendChild(itemDiv);
63   });
64   results = rootDiv;
65 }
66
```

Datagrok

About

Datagrok is a platform for turning data into actionable insights.

Overview

Help

Video lessons

Support

Chemistry | ChEMBL similarity(structure)

This script returns the top n most similar ChEMBL structures

Input:

string structure: Molecule to search by

options: {semType: Molecule, default: "c1ccccc1CCC(NO)"}

Output:

string results

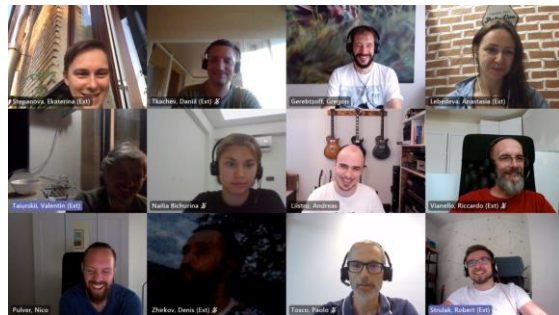
See also:

Scripting

Acknowledgements

Intuence team

- Nico Pulver
- Gregori Gerebtzoff
- Nik Clare
- Andreas Gasser
- Ralph Nijholt
- Recca Gottschall
- Cristina Nieto-Oberhuber
- Jimmy Kromann
- Jianmei Fang
- Daniil Tkachev
- Vimala Selvaraj
- Fang Du
- Riccardo Vianello
- Paolo Tosco
- Kevin Parnow
- EPAM team



- Datagrok team
- Intuence user community
- NIBR Informatics LT
- NIBR Global Discovery Chemistry & CADD
- NIBR Translational Medicine & PKS
- Greg Landrum & RDKit community
- All who have contributed to this tremendous journey...



Thank you