

# RDKit Analytics

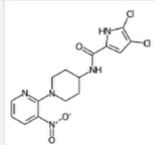
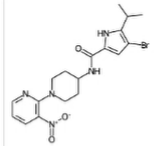
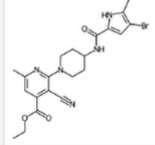
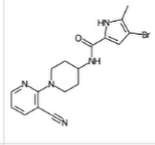
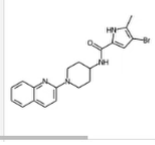
Matt Stahl

Glysade, LLC

RDKit UGM 2022

# Reaction Matched Pairs

Lead Discovery ChemCharts

	Structure <MOLFILE>	Title	First Sample	SAU Gyr IC50	Sau516 M
1		AZ1001	6/10/2002	>400	>64
2		AZ1002	8/16/2002	>50	>64
3		AZ1003	8/16/2002	7.14	>64
4		AZ1004	8/16/2002	9.93	>64
5		AZ1005	8/16/2002	3.31	>64

Lead Discovery ChemCharts

Search...

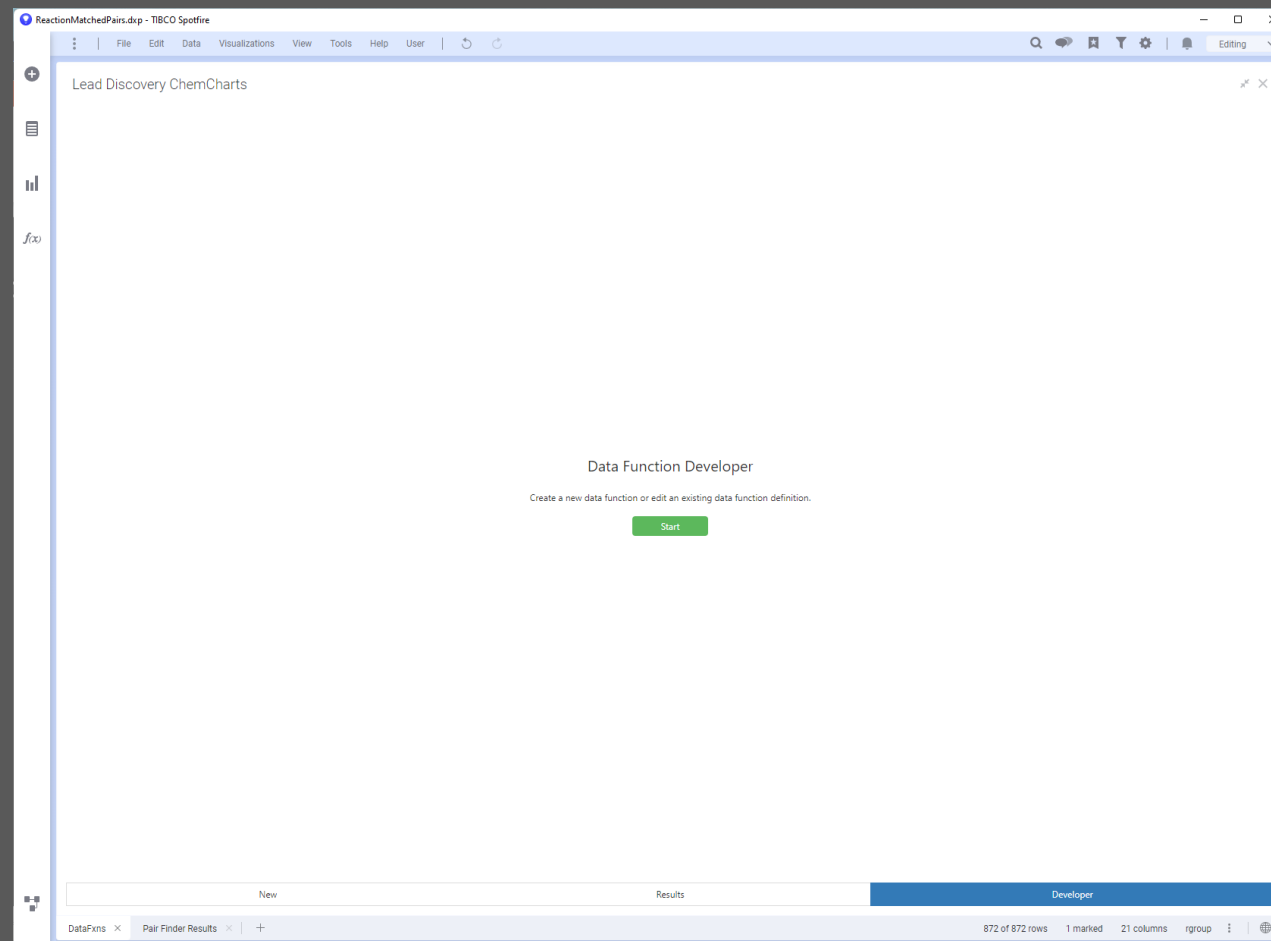
Search tags...

- Biopolymer
  - Ab Component Analysis
  - Antibody Numbering
  - Blast Local Column Search
  - Blast Local Text Search
  - Blast Table Search
  - Blast Web Search
  - Extract Genbank Regions
  - Identify Enzyme Restriction Sites
  - IG Blastn Local Text Search
  - IG Blastp Local Text Search
  - Multiple Sequence Alignment
  - Pairwise Sequence Alignment
  - Sequence Pattern Match
  - Sequence Regular Expression Match
  - Translate DNA to Protein Script
  - Translate Sequences (bidirectional)
  - Translate Sequences From All Reading Frames
- Chemistry
  - ADMET Properties
  - Apply Depiction Template (ChemDraw)
  - Apply Depiction Template (RDKit)
  - Core Decomposition Analysis
  - Deprotect Structures
  - Exact Mass (.NET)
  - Exact Mass (Python)
  - HBond SMARTS
  - Maximum Common Substructure (MCS)
  - MMPDB Database Search
  - MMPDB Project Search

New Results Developer

872 of 872 rows 1 marked 21 columns rgroup

# DataFxn Development



# DataFxn Development

ReactionMatchedPairs.dxp - TBCO Spotfire

File Edit Data Visualizations View Tools Help User

Lead Discovery ChemCharts

Step 2

Edit the data function definition details. The preview will update to reflect changes where possible.

Name: Exact Mass (Python)

Description: Calculates exact mass for structures using Python code in data function definition

Category: Chemistry

Tags: MAB, PipelinePilot, analyze

Enter new tag text Add tag

Version: 1.0.0

Executor ID: Glyside.CPythonDataFxn

Service name: Script

Allowed clients: ☒ Analyst ☒ WebPlayer

Limit by: ☒ None ☐ Filtered rows ☐ Marked rows ☐ Filtered and marked rows

Service URI: glyside.python

Maximum number of output columns: 1

Maximum number of output tables: 0

Update behavior: Previous Next

Input fields Output fields Data fn script Visual configuration script

Data Function

Exact Mass (Python) Chemistry Caltopro

Calculates exact mass for structures using Python code in data function definition

Data table: rgroup

Select structure column: Structure <MOLFILE>

OK Cancel

New Results Developer

DataFxn x Pair Finder Results x +

872 of 872 rows 1 marked 21 columns rgroup

# DataFxn Development

ReactionMatchedPairs.dxp - TBCO Spotfire

File Edit Data Visualizations View Tools Help User

Lead Discovery ChemCharts

Step 3

Edit the data function input fields, output fields, script and visual configuration script by selecting the corresponding tab in the preview panel on the right-hand-side. The preview will update to reflect changes where possible.

**Input fields**

Control	ID	Label
Column select box	structureColumn	Select structure column

ID: structureColumn

Control: Column select box

Data type: string

Label: Select structure column

Tooltips:

Validation rules:

- ☒ required Must select column of 2D structures
- ☐ oneRequired Exactly one value required
- ☐ oneOrMoreRequired One or more values required

Column filters:

Data type	Content type (optional)
string	chemical/x-mdl-molfile chemical/x-mdl-molfile-v3000 chemical/x-smiles chemical/x-daylight-smiles
binary	chemical/x-mdl-molfile chemical/x-mdl-molfile-v3000

Previous Finish

**Data Function**

**Exact Mass (Python)**

Calculates exact mass for structures using Python code in data function definition

**Data table**

rgroup

**Select structure column**

Structure <MOLFILE>

OK Cancel

New Results Developer

DataFxn x Pair Finder Results x +

872 of 872 rows 1 marked 21 columns rgroup

# DataFxn Development

ReactionMatchedPairs.dxp - TBCO Spotfire

File Edit Data Visualizations View Tools Help User

Lead Discovery ChemCharts

Step 3

Edit the data function input fields, output fields, script and visual configuration script by selecting the corresponding tab in the preview panel on the right-hand-side. The preview will update to reflect changes where possible.

**Script**

Use code editor to modify script

Previous Finish

Input fields Output fields Data fn script Visual configuration script

```
1 from rdkit.Chem.Descriptors import ExactMolWt
2 from df.chem_helper import column_to_molecules
3 from df.data_transfer import DataFunctionRequest, DataFunctionResponse, DataType, ColumnData, \
4   string_input_field
5
6
7 def execute(request: DataFunctionRequest) -> DataFunctionResponse:
8     column_id = string_input_field(request, 'structureColumn')
9     input_column = request.inputColumns[column_id]
10    mols = column_to_molecules(input_column)
11    weights = [None if m is None else ExactMolWt(m) for m in mols]
12    output_column = ColumnData(name=f'{input_column.name} Exact Mass', dataType=DataType.DOUBLE, values=weights)
13    response = DataFunctionResponse(outputColumns=[output_column])
14    return response
15
```

New Results Developer

DataFxn x Pair Finder Results x +

872 of 872 rows 1 marked 21 columns rgroup

# Acknowledgements

- Gareth Jones
- Paul Watson