MedChemica

CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

How to give the user what they want?: Challenges in Markush structure visualisation and exploitation

Twitter @MedChemica

www.medchemica.com

lauren.reid@medchemica.com

Sept 2023

Overview

MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

- Brief reminder of the SARkush® algorithm:
  - Clustering algorithm
  - R (and x, X, E) group decomposition algorithm
- Challenges in depicting Markush structures
- Using Markush structures in compound analysis:
  - Project tracking
  - Free Wilson analysis
  - Patent analysis

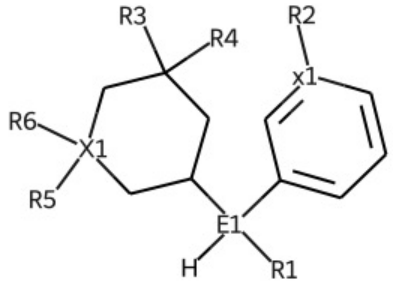
SARkush®

Structure Activity Relationship

Markush

MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

A structural representation designed to communicate SAR



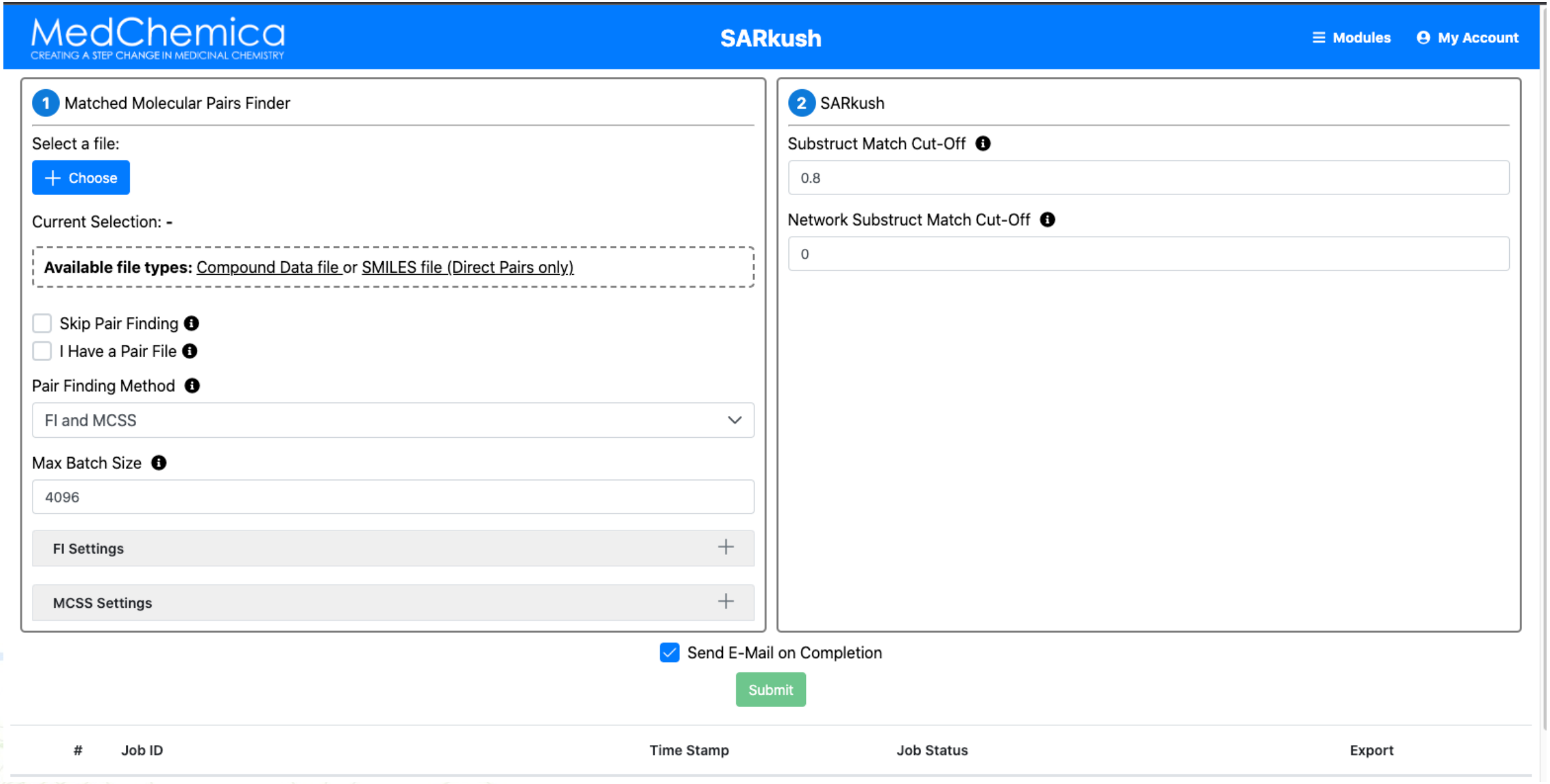
Variable atoms/ groups:
- x = aromatic atom
- X = aliphatic ring atom
- E = linker atom
- R = side chain

compound id	X1	E1	R1	…	data
1	C	N	None	…	measurement1
2	O	C	[H]	…	measurement2
…	…	…	…	…	…

Data taken from J. Med. Chem. 2022, 65, 6513−6540

A look at the GUI…

MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY



Data taken from J. Med. Chem. 2022, 65, 6513−6540

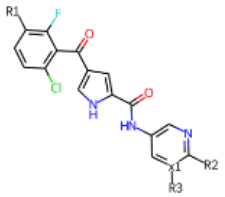
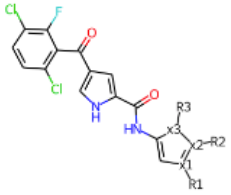
A look at the GUI…

MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

#	Job ID	Time Stamp	Job Status	Export
0	BQZIA-U0FBK_2023-09-12_12_11_43_568121	Sep 12, 2023, 1:11:55 PM	Job Data Retrieved	Download

Overview | Network | SARkush 1 (17) | SARkush 2 (9) | Singletons (0)

Export Overview | View Box Plot | Download Pair File

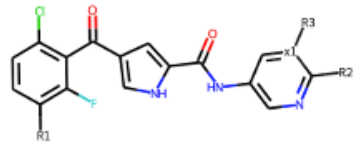
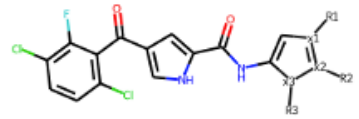
Network	Sarkush	Sarkush Structure	Number of Compounds	Percentage of Compounds	Number of Groups	Measurement Median	CLo
1	1		17	65.38%	4	7.85	
1	2		9	34.62%	6	7.1	

Data taken from J. Med. Chem. 2022, 65, 6513−6540

A look at the output…

MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

- One excel spreadsheet with SARkush structural depictions:

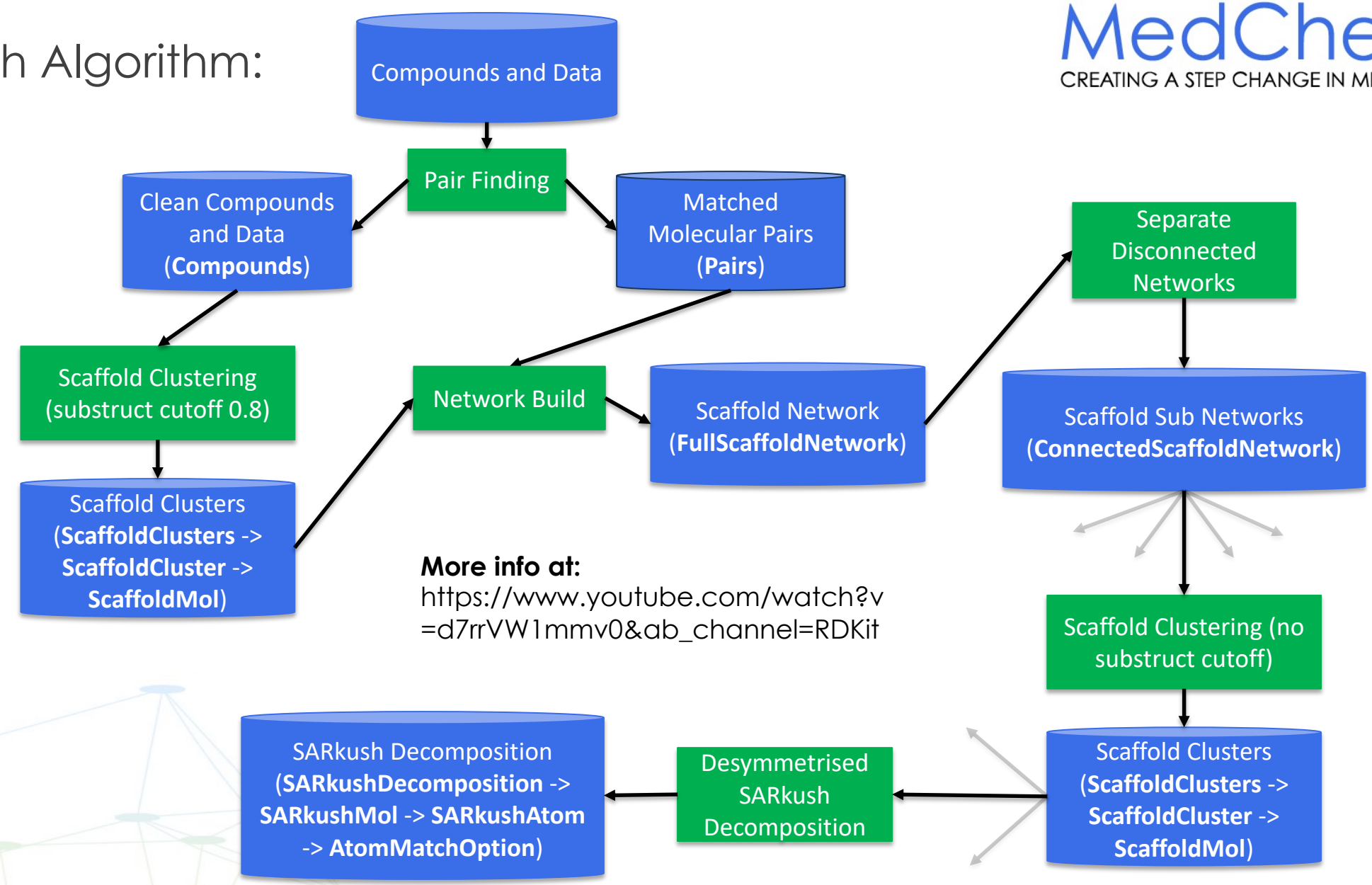
	A	B	C	D	E	F	G	H	I
1	network_id	sarkush_id	sarkush	no_of_compounds	no_of_compounds_cumulative	percentage_of_compounds	measurement_min	measurement_max	measurement_median
2	1	1		17	17	65.38461538	5.455931956	8.301029996	7.853871964
3	1	2		9	26	34.61538462	5.908333042	8.15490196	7.102372909

- A txt file per SARkush structure with compound decomposition:

	A	C	D	E	F	G	H	I	J	K	L	M	N
1	compound_name	SARkush_smiles	x1	R1	R2	R3	qualifier	measurement	CLogP	PSA	RMM	HBA	HBD
2	33j	O=C(Nc1cnc([R2])[x1]([R3])c1)c1cc(C(=O)c2c(Cl)ccc([R1])c2F)c[nH]1	c	[R1]Cl	[R2]C([H])([H])C1([H])C([H])([H])C([H])([H])N(C([H])([H])[H])C([H])([H])C1([H])[H]	[R3][H]	=	8.22184875	4.52	78.09	489.37	6	2
3	33i	O=C(Nc1cnc([R2])[x1]([R3])c1)c1cc(C(=O)c2c(Cl)ccc([R1])c2F)c[nH]1	c	[R1]Cl	[R2]C([H])([H])N1C([H])([H])C([H])([H])N(C([H])([H])[H])C([H])([H])C1([H])[H]	[R3][H]	=	8.301029996	3.09	81.33	490.36	7	2
4	33h	O=C(Nc1cnc([R2])[x1]([R3])c1)c1cc(C(=O)c2c(Cl)ccc([R1])c2F)c[nH]1	c	[R1]Cl	[R2]C([H])([H])N1C([H])([H])C([H])([H])N([H])C([H])([H])C1([H])[H]	[R3][H]	=	8.301029996	2.71	90.12	476.33	7	3
5	33g	O=C(Nc1cnc([R2])[x1]([R3])c1)c1cc(C(=O)c2c(Cl)ccc([R1])c2F)c[nH]1	c	[R1]Cl	[R2]N(C([H])([H])[H])C1([H])C([H])([H])C([H])([H])N(C([H])([H])[H])C([H])([H])C1([H])[H]	[R3][H]	=	7.602059991	4.6	81.33	504.38	7	2
6	33f	O=C(Nc1cnc([R2])[x1]([R3])c1)c1cc(C(=O)c2c(Cl)ccc([R1])c2F)c[nH]1	c	[R1]Cl	[R2]N1C([H])([H])C([H])([H])N([H])C([H])([H])C1([H])[H]	[R3][H]	=	7.886056648	3.09	90.12	462.3	7	3
7	33k	O=C(Nc1cnc([R2])[x1]([R3])c1)c1cc(C(=O)c2c(Cl)ccc([R1])c2F)c[nH]1	c	[R1]Cl	[R2]OC1([H])C([H])([H])C([H])([H])N(C([H])([H])[H])C([H])([H])C1([H])[H]	[R3][H]	=	7.853871964	4.34	87.32	491.34	7	2
8	32m	O=C(Nc1cnc([R2])[x1]([R3])c1)c1cc(C(=O)c2c(Cl)ccc([R1])c2F)c[nH]1	n	[R1]Cl	[R2]C([H])([H])N1C([H])([H])C([H])([H])N(C([H])([H])[H])C([H])([H])C1([H])[H]		=	8.15490196	2.48	94.22	491.35	8	2
9	32l	O=C(Nc1cnc([R2])[x1]([R3])c1)c1cc(C(=O)c2c(Cl)ccc([R1])c2F)c[nH]1	n	[R1]Cl	[R2]C([H])([H])N1C([H])([H])C([H])([H])N([H])C([H])([H])C1([H])[H]		=	8.096910013	2.09	103.01	477.32	8	3
10	32i	O=C(Nc1cnc([R2])[x1]([R3])c1)c1cc(C(=O)c2c(Cl)ccc([R1])c2F)c[nH]1	n	[R1]Cl	[R2]N([H])C1([H])C([H])([H])C([H])([H])N(C([H])([H])[H])C([H])([H])C1([H])[H]		=	7.853871964	3.35	103.01	491.35	8	3
11	32g	O=C(Nc1cnc([R2])[x1]([R3])c1)c1cc(C(=O)c2c(Cl)ccc([R1])c2F)c[nH]1	n	[R1]Cl	[R2]N1C([H])([H])C([H])([H])N(C([H])([H])[H])C([H])([H])C1([H])[H]		=	7.431798276	2.61	94.22	477.32	8	2
12	32k	O=C(Nc1cnc([R2])[x1]([R3])c1)c1cc(C(=O)c2c(Cl)ccc([R1])c2F)c[nH]1	n	[R1]Cl	[R2]N1C([H])([H])C([H])([H])N([H])C([H])([H])C1([H])[H]		=	7.886056648	2.22	103.01	463.29	8	3
13	32b	O=C(Nc1cnc([R2])[x1]([R3])c1)c1cc(C(=O)c2c(Cl)ccc([R1])c2F)c[nH]1	n	[R1][H]	[R2]C([H])([H])[H]		=	5.920818754	2.56	87.74	358.75	6	2
14	32d	O=C(Nc1cnc([R2])[x1]([R3])c1)c1cc(C(=O)c2c(Cl)ccc([R1])c2F)c[nH]1	n	[R1][H]	[R2]N(C([H])([H])[H])C([H])([H])[H]		=	5.455931956	2.56	90.98	387.8	7	2
15	32f	O=C(Nc1cnc([R2])[x1]([R3])c1)c1cc(C(=O)c2c(Cl)ccc([R1])c2F)c[nH]1	n	[R1][H]	[R2]N1C([H])([H])C([H])([H])N(C([H])([H])[H])C([H])([H])C1([H])[H]		=	7.142667504	1.99	94.22	442.87	8	2

SARkush Algorithm:

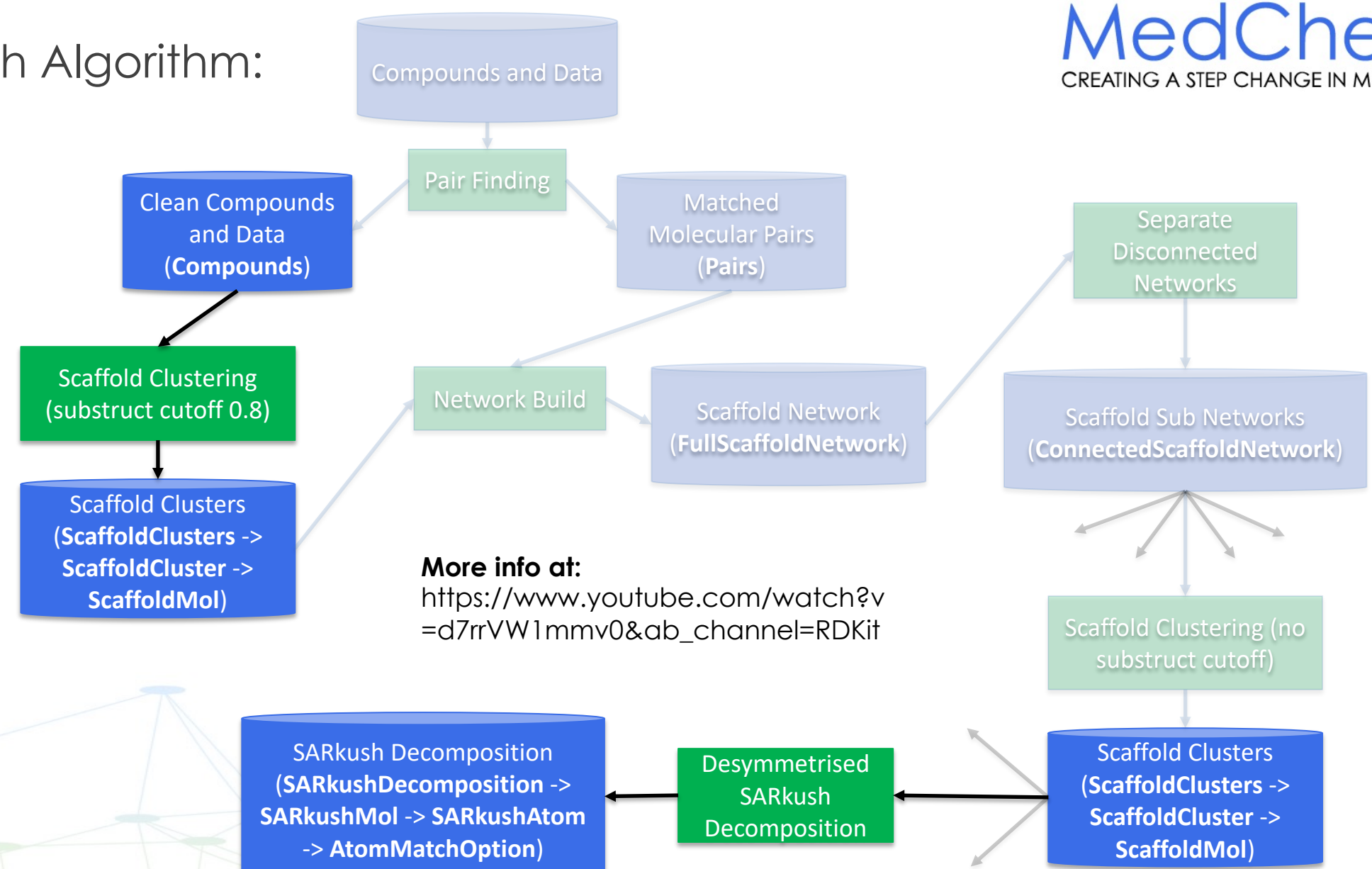
MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY



More info at:
https://www.youtube.com/watch?v=d7rrVW1mmv0&ab_channel=RDKit

SARkush Algorithm:

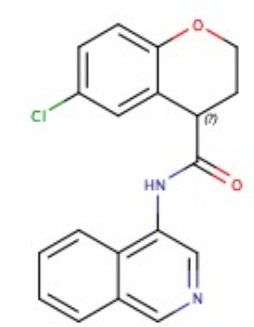
MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY



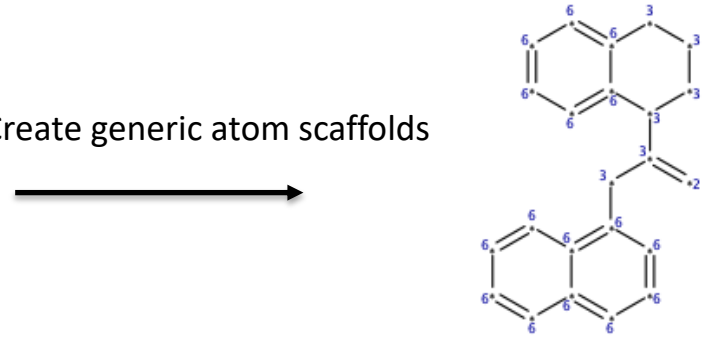
More info at:
https://www.youtube.com/watch?v=d7rrVW1mmv0&ab_channel=RDKit

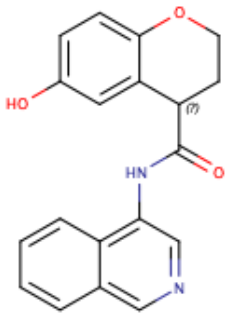
Scaffold Clustering

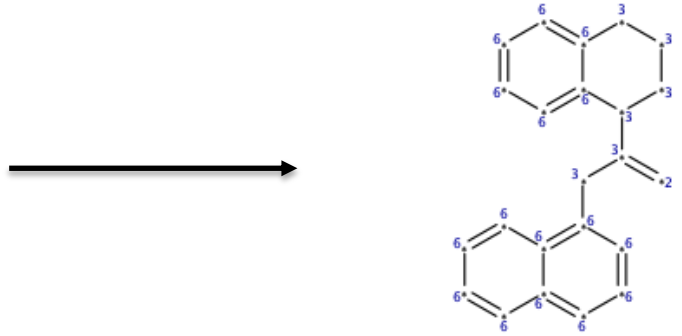
MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

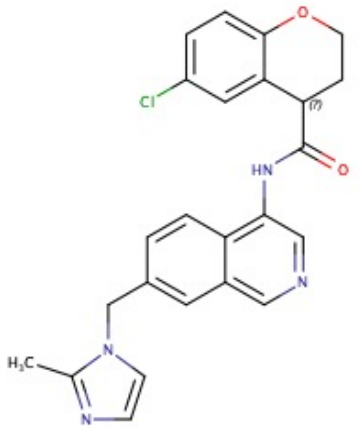


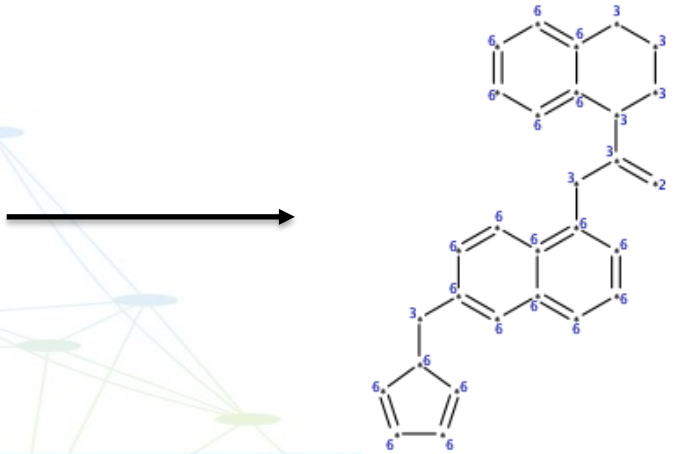
1) Create generic atom scaffolds





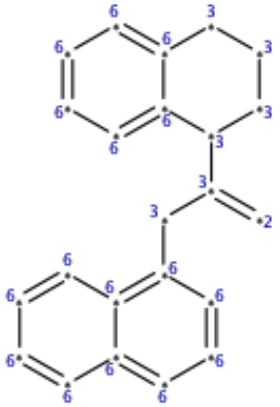






2) Merge compound clusters with matching scaffolds

3) Merge larger scaffolds into smaller if substructure present and atom overlap cutoff met



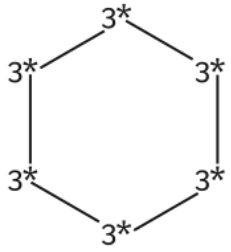
All three compounds are clustered into one scaffold

Atom type	Isotope label
aliphatic chiral	1
carbonyl heteroatom	2
aliphatic non-chiral	3
E bond atom	4
Z bond atom	5
aromatic	6

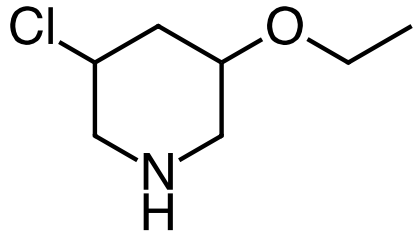
Desymmetrised SARkush Decomposition

MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

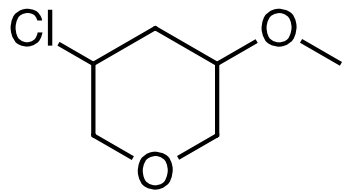
Scaffold



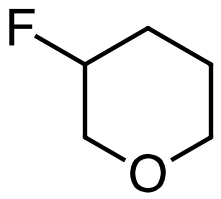
Compounds



1

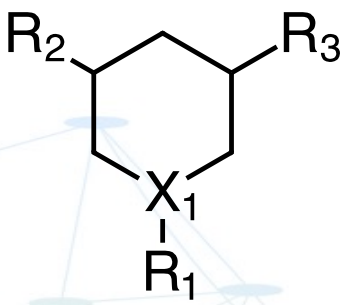


2



3

SARkush



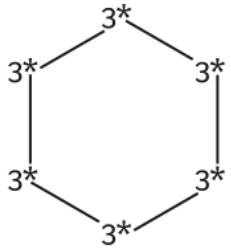
SARkush Decomposition:

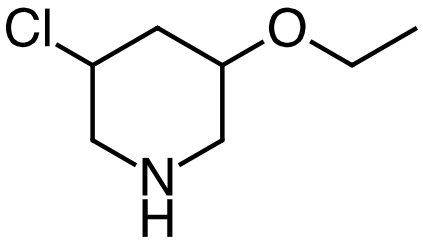
compound	X1	R1	R2	R3
1	N	[R1][H]	[R2]Cl	[R3]OCC
2	O	None	[R2]Cl	[R3]OC
3	O	None	[R2]F	[R3][H]

Desymmetrised SARkush Decomposition

MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

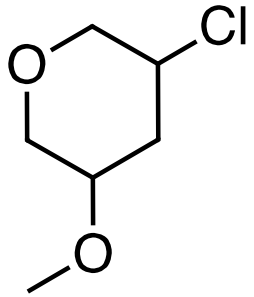
Scaffold



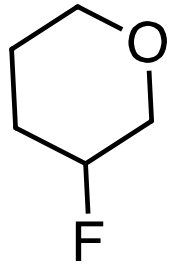


1

Compounds

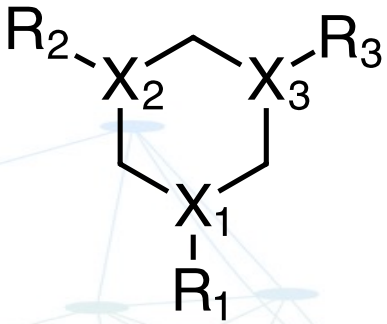


2



3

SARkush

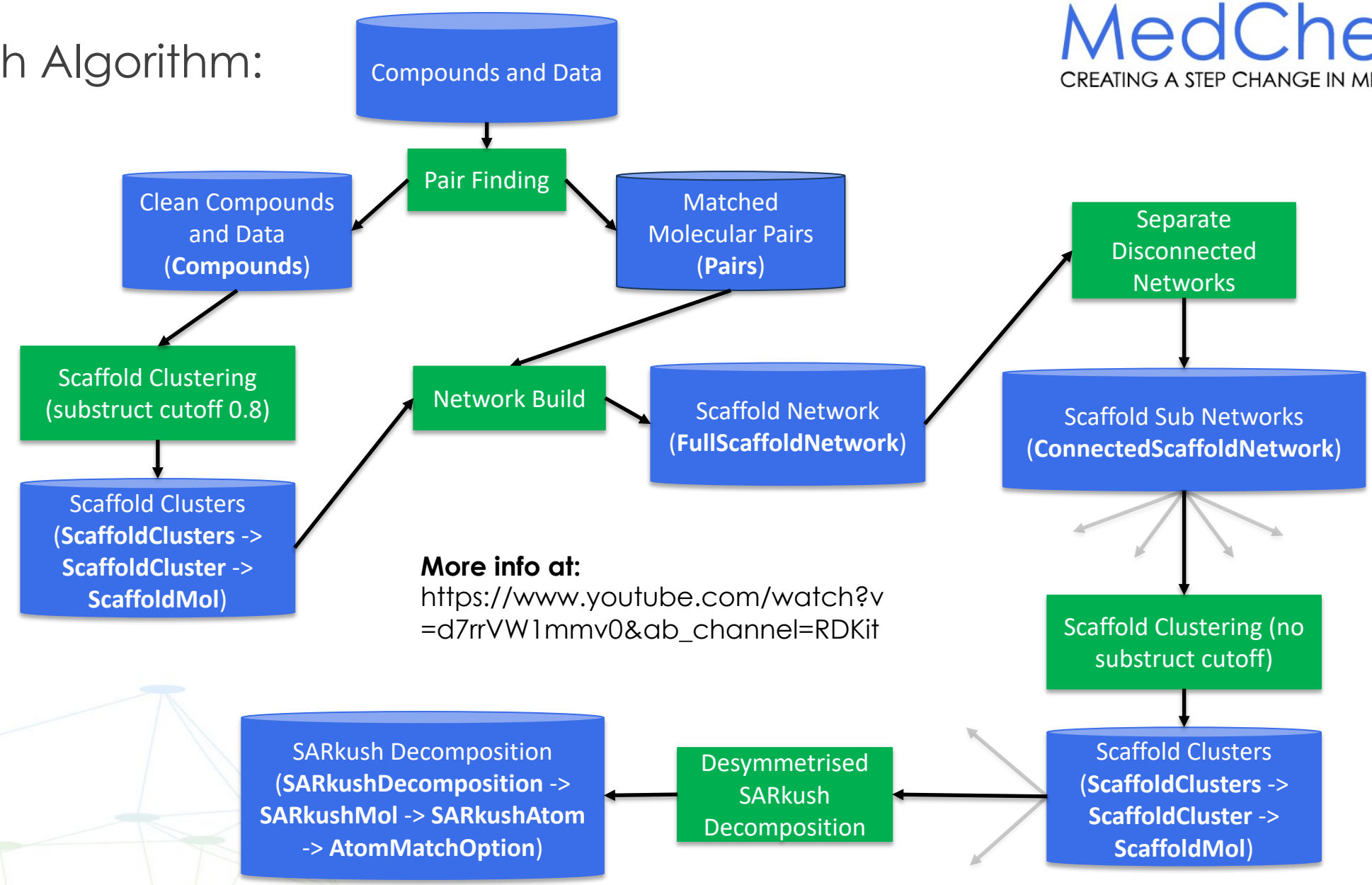


SARkush Decomposition:

compound	X1	X2	X3	R1	R2	R3
1	N	C	C	[R1]H	[R2]Cl	[R3]OCC
2	C	O	C	[R1]OC	None	[R3]Cl
3	C	C	O	[R1]F	[R2][H]	None

SARkush Algorithm:

MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY



More info at:
https://www.youtube.com/watch?v=d7rrVW1mmv0&ab_channel=RDKit

Depicting Markush Structures

MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

- Users need:
  - To understand which atom and bond types form the Markush structure
  - To understand the connectivity of the atoms and bonds in the structure
- Users want:
  - To visualise Markush structures with molecule-like 2D coordinates
  - To visualise the associated member compounds in the same 2D orientation
  - To visualise Markush structures in the users' preferred orientation

Depicting Markush Structures

MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

Output from SARkush algorithm =

- sarkush_mol (RDKit molecule with "sarkush_label" atom properties)
- list of SARkush member molecules
- Atom maps between SARkush scaffold and member compounds

```
for atom in sarkush_mol.GetAtoms():
    if atom.HasProp("sarkush_label"):
        atom.SetProp("atomLabel", atom.GetProp("sarkush_label"))

mol.UpdatePropertyCache()

dm = rdMolDraw2D.PrepareMolForDrawing(mol, kekulize=True)
d2d = rdMolDraw2D.MolDraw2DCairo(width, height)
d2d.DrawMolecule(dm)
d2d.FinishDrawing()
```

Set atomLabel to pre-defined SARkush label

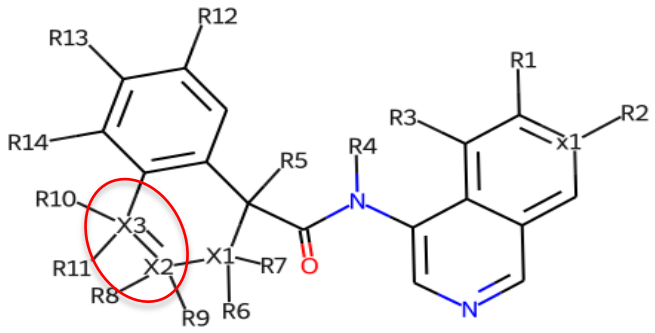
Draw molecule

Depicting Markush Structures

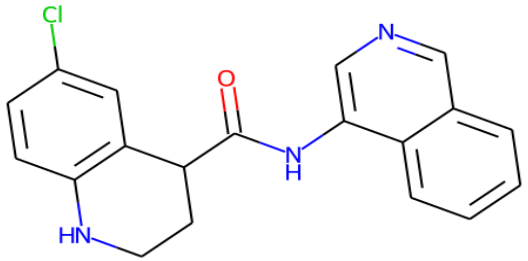
MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

COVID Moonshot example - 90 compounds represented by the SARkush

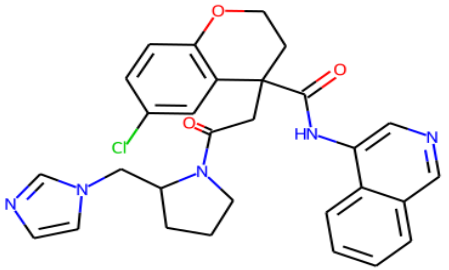
SARkush



Smallest molecule



Largest molecule



- Users need:
  - To understand which atom and bond types form the Markush structure ✗
  - To understand the connectivity of the atoms and bonds in the structure ✓
- Users want:
  - To visualise Markush structures with molecule-like 2D coordinates In this example yes
  - To visualise the associated member compounds in the same 2D orientation ✗

Depicting Markush Structures

MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

```
for atom in sarkush_mol.GetAtoms():
    if atom.HasProp("sarkush_label"):
        atom.SetProp("atomLabel", atom.GetProp("sarkush_label"))

sarkush_mol.UpdatePropertyCache()

dm = rdMolDraw2D.PrepareMolForDrawing(mol, kekulize=True)

for sarkush_mol_bond in sarkush_mol.GetBonds():
    bond_idx = sarkush_mol_bond.GetIdx()
    dm_bond = dm.GetBondWithIdx(bond_idx)
    if sarkush_mol_bond.GetBondType() == Chem.BondType.SINGLE and dm_bond.GetBondType() == Chem.BondType.DOUBLE:
        dm_bond.SetBondType(Chem.BondType.SINGLE)

d2d = rdMolDraw2D.MolDraw2DCairo(width, height)
d2d.DrawMolecule(dm)
d2d.FinishDrawing()
```

If kekulisation error occurs, manually set bond type

https://github.com/rdkit/rdkit/discussions/4716

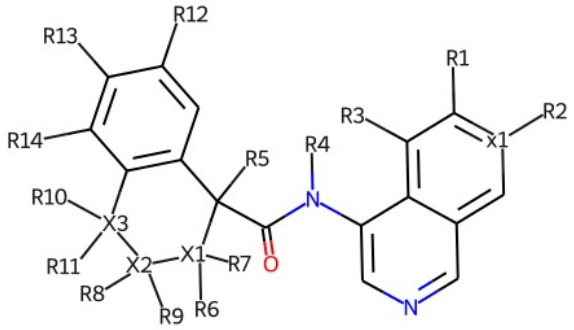
Thanks to Paolo Tosco for the suggestion

Depicting Markush Structures

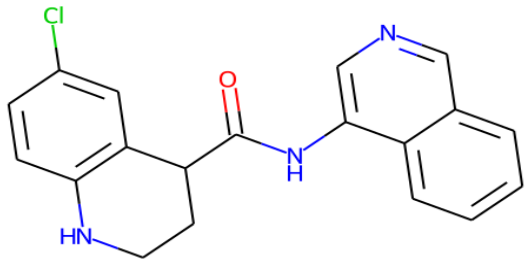
MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

COVID Moonshot example - 90 compounds represented by the SARkush

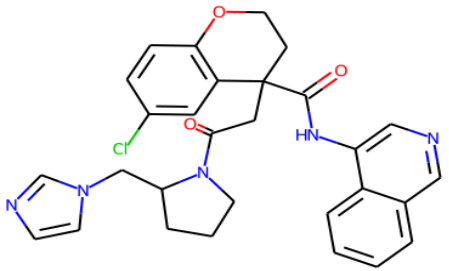
SARkush



Smallest molecule



Largest molecule



- Users need:
  - To understand which atom and bond types form the Markush structure ✓
  - To understand the connectivity of the atoms and bonds in the structure ✓
- Users want:
  - To visualise Markush structures with molecule-like 2D coordinates In this example yes
  - To visualise the associated member compounds in the same 2D orientation ✗

Depicting Markush Structures

MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

```
for atom in sarkush_mol.GetAtoms():
    if atom.HasProp("sarkush_label"):
        atom.SetProp("atomLabel", atom.GetProp("sarkush_label"))

sarkush_mol.UpdatePropertyCache()

for sarkush_mol_bond in sarkush_mol.GetBonds():
    bond_idx = sarkush_mol_bond.GetIdx()
    dm_bond = dm.GetBondWithIdx(bond_idx)
    if sarkush_mol_bond.GetBondType() == Chem.BondType.SINGLE and dm_bond.GetBondType() == Chem.BondType.DOUBLE:
        dm_bond.SetBondType(Chem.BondType.SINGLE)

align_depiction_mol_to_template_mol(dm, align_mol, atom_map)

d2d = rdMolDraw2D.MolDraw2DCairo(width, height)
d2d.DrawMolecule(dm)
d2d.FinishDrawing()
```

Align the SARkush mol to a member mol before drawing

Depicting Markush Structures

MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

```
def align_depiction_mol_to_template_smiles(self, dm, align_mol, atom_map):

    rdDepictor.Compute2DCoords(align_mol)

    coordDict2D = {}
    for dm_atom_id, template_atom_id in atom_map.items():
        pt = align_mol.GetConformer().GetAtomPosition(template_atom_id)
        coordDict2D[dm_atom_id] = Geometry.Point2D(pt.x, pt.y)

    rdDepictor.Compute2DCoords(dm, coordMap=coordDict2D)
```

atom_map between the SARkush and each molecule already exists from the decomposition step

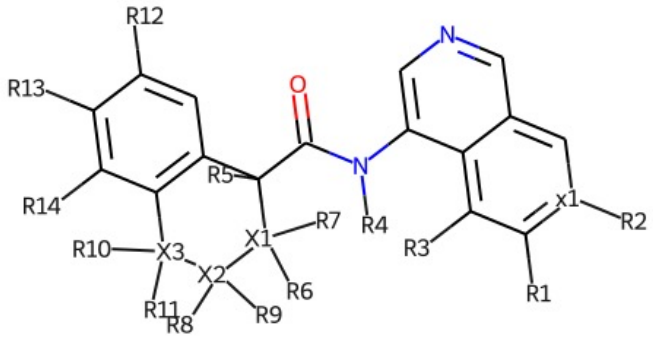
Take coordinates from align_mol and apply to corresponding SARkush atoms

Depicting Markush Structures

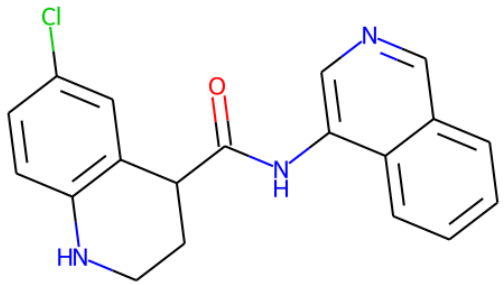
MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

COVID Moonshot example - 90 compounds represented by the SARkush

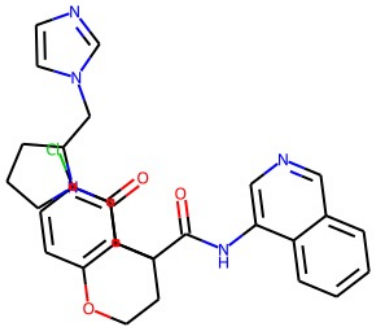
SARkush



Smallest molecule



Largest molecule



- Users need:
  - To understand which atom and bond types form the Markush structure ✗
  - To understand the connectivity of the atoms and bonds in the structure ✗
- Users want:
  - To visualise Markush structures with molecule-like 2D coordinates Kind of
  - To visualise the associated member compounds in the same 2D orientation ✓

Depicting Markush Structures

MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

```
def align_depiction_mol_to_template_smiles(self, dm, align_mol, atom_map):

    rdDepictor.SetPreferCoordGen(True)
    rdDepictor.Compute2DCoords(align_mol)

    coordDict2D = {}
    for dm_atom_id, template_atom_id in atom_map.items():
        pt = align_mol.GetConformer().GetAtomPosition(template_atom_id)
        coordDict2D[dm_atom_id] = Geometry.Point2D(pt.x, pt.y)

    rdDepictor.Compute2DCoords(dm, coordMap=coordDict2D
```

Use CoordGen library

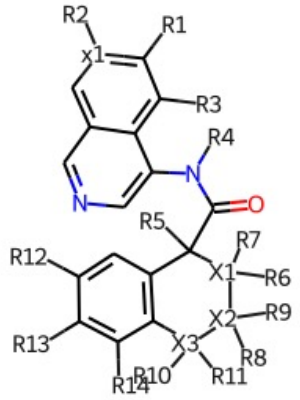
https://github.com/rdkit/rdkit/discussions/6518
Thanks Paolo Tosco for the suggestion (again)

Depicting Markush Structures

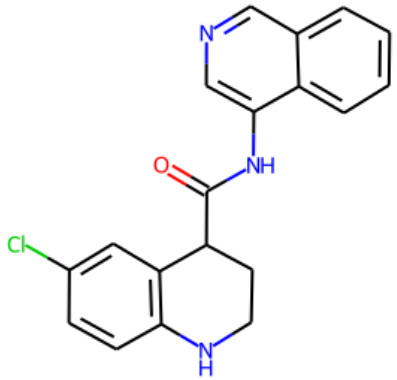
MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

COVID Moonshot example - 90 compounds represented by the SARkush

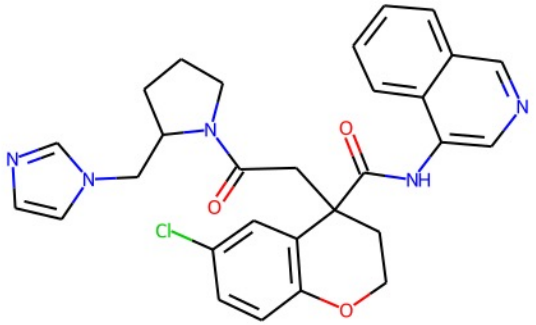
SARkush



Smallest molecule



Largest molecule



- Users need:
  - To understand which atom and bond types form the Markush structure ✓
  - To understand the connectivity of the atoms and bonds in the structure ✓
- Users want:
  - To visualise Markush structures with molecule-like 2D coordinates
  - To visualise the associated member compounds in the same 2D orientation

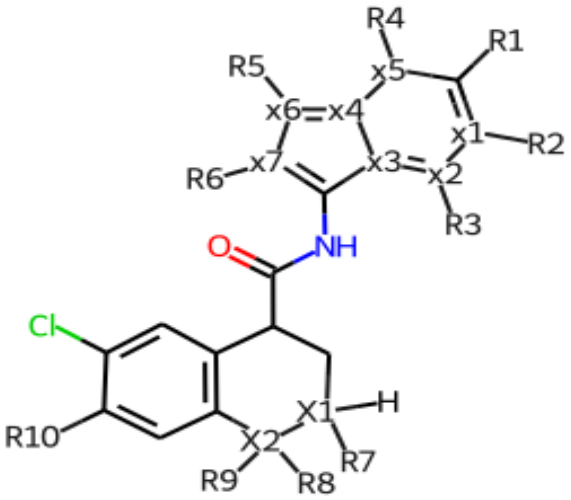
It's trying its best

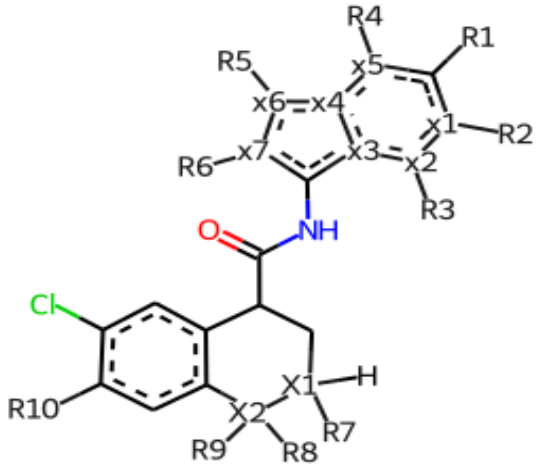
They're trying their best

Depicting Markush Structures

MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

- Sometimes the kekulisation can be misleading or ambiguous
  - The placement of double bonds depends on the values of the x atoms
- We're moving to a delocalized representation in the next release



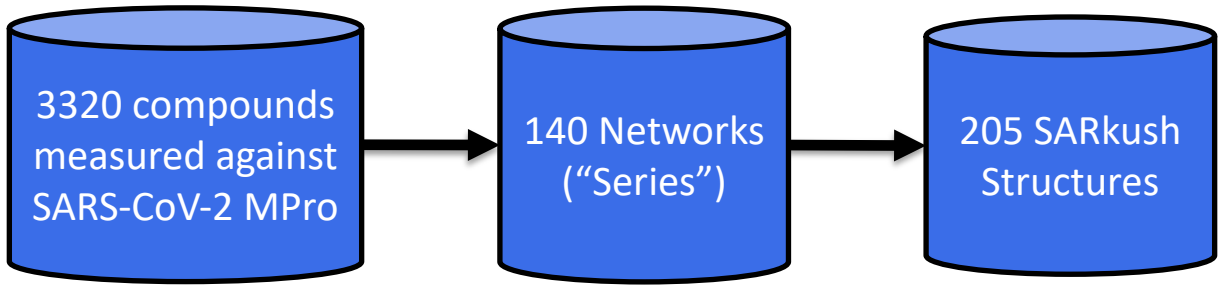



```
dm = rdMolDraw2D.PrepareMolForDrawing(mol, kekulize=False)
```

COVID Moonshot

MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

- Crowd-funded project to develop SARS-CoV-2 MPro inhibitors
- Crowd-sourced ideas resulted in an explosion of data
- SARkush was developed to aid SAR tracking in lead optimisation

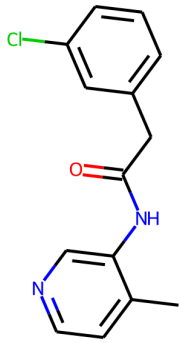


https://dndi.org/research-development/portfolio/covid-moonshot/

Monitoring projects with SARkush structures

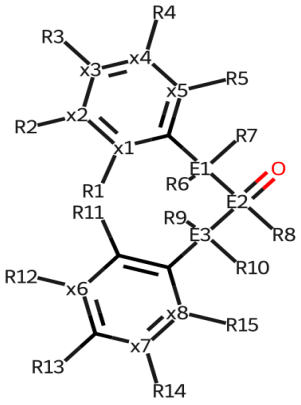
MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

Initial hit



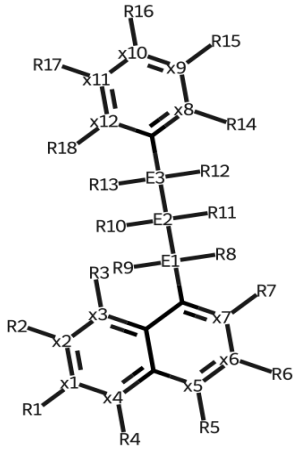
SAR Explored

SARkush 6
143 compounds



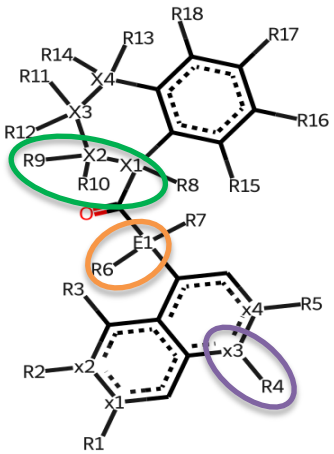
Monocycle to Bicycle

SARkush 1
405 compounds



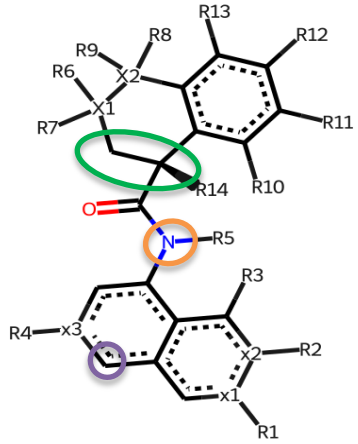
Ring Closure

SARkush 2
365 compounds

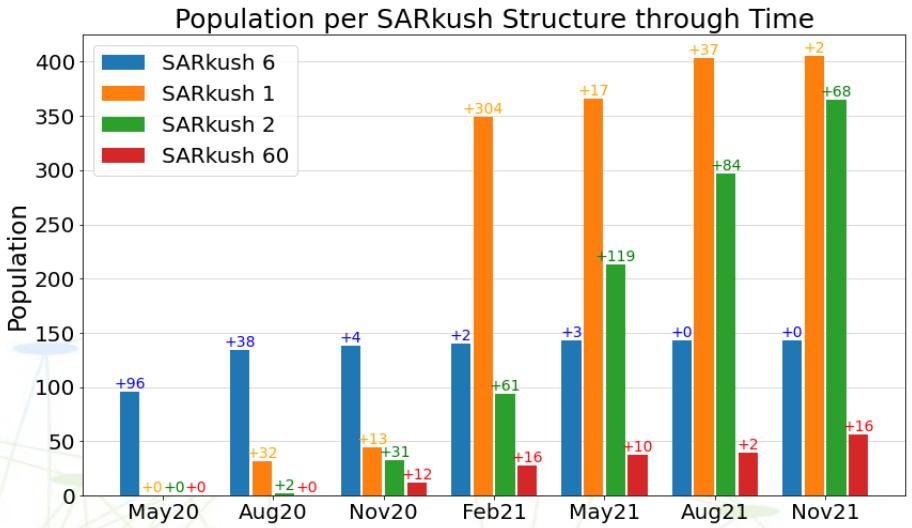


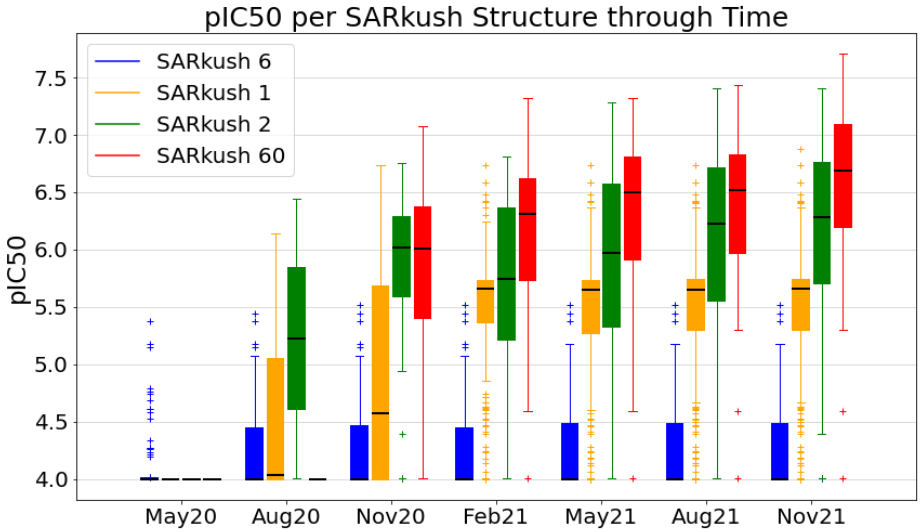
Chiral Separation Identifies Active Isomer

SARkush 60
56 compounds



DNDi-6510



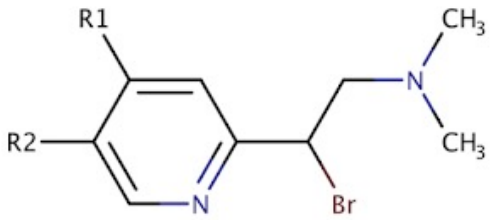


Model Building using Markush Descriptors

MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

- Free Wilson analysis is an established technique for calculating the contribution of R groups to compound activity

- Code is readily available to perform R group decomposition and linear regression
  - See Pat Walters' blog post and GitHub page http://practicalcheminformatics.blogspot.com/2018/05/free-wilson-analysis.html



	R1						R2					
¹▲Name	H	F	Cl	Br	I	CH_3	H	F	Cl	Br	I	CH_3
MOL0001	1	0	0	0	0	0	1	0	0	0	0	0
MOL0002	1	0	0	0	0	0	0	1	0	0	0	0
MOL0003	1	0	0	0	0	0	0	0	1	0	0	0
MOL0004	1	0	0	0	0	0	0	0	0	1	0	0
MOL0005	1	0	0	0	0	0	0	0	0	0	1	0
MOL0006	1	0	0	0	0	0	0	0	0	0	0	1
MOL0007	0	1	0	0	0	0	1	0	0	0	0	0
MOL0008	0	0	1	0	0	0	1	0	0	0	0	0
MOL0009	0	0	0	1	0	0	1	0	0	0	0	0
MOL0010	0	0	0	0	1	0	1	0	0	0	0	0
MOL0011	0	0	0	0	0	1	1	0	0	0	0	0
MOL0012	0	0	1	0	0	0	0	1	0	0	0	0
MOL0013	0	0	0	1	0	0	0	1	0	0	0	0
MOL0014	0	0	0	0	0	1	0	1	0	0	0	0
MOL0015	0	0	1	0	0	0	0	0	1	0	0	0
MOL0016	0	0	0	1	0	0	0	0	1	0	0	0
MOL0017	0	0	0	0	0	1	0	0	1	0	0	0
MOL0018	0	0	1	0	0	0	0	0	0	1	0	0
MOL0019	0	0	0	1	0	0	0	0	0	1	0	0
MOL0020	0	0	0	0	0	1	0	0	0	1	0	0
MOL0021	0	0	0	0	0	1	0	0	0	0	0	1
MOL0022	0	0	0	1	0	0	0	0	0	0	0	1

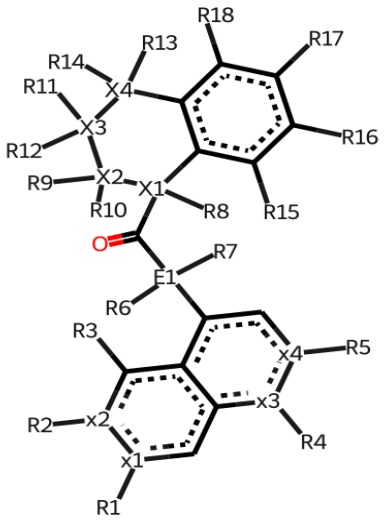
R-Group SMILES	¹▼Coefficient	R-group	Count
[image]	0.386	2	3
[image]	0.302	2	1
[image]	0.228	2	4
[image]	0.193	1	5

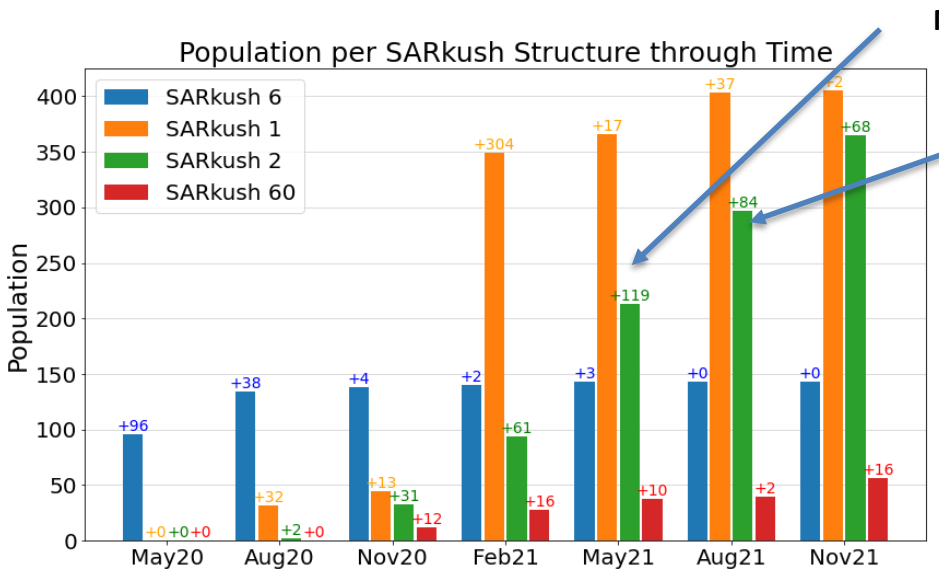
- SARkush can be used to generate Free Wilson descriptors that include x, X and E atoms as well as R groups!

Model Building using Markush Descriptors

MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

Let's look at SARkush 2 from the COVID Moonshot:





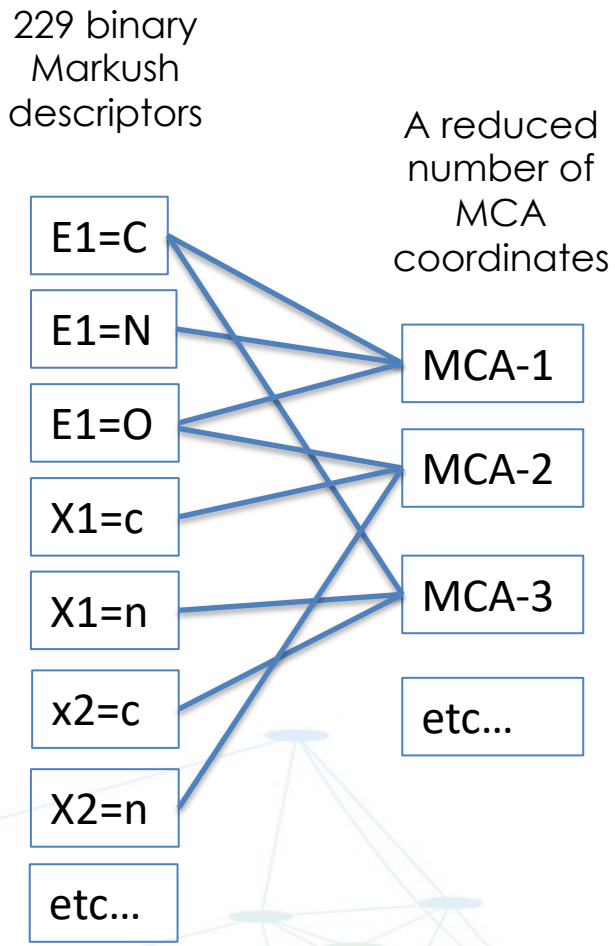
May 2021 = 212 compounds

August 2021 = 296 compounds

- There is a high interdependence of x, X, E and R groups due to their connectivity
- In medicinal chemistry, we make changes in functional groups, not individual atoms
- Multiple correspondence analysis (MCA):
  - Reduces dimensionality
  - Produces non-correlated descriptor variables
  - Is like principal component analysis (PCA) for categorical variables

Model building using Markush descriptors

MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY



Ordinary least squares using MCA coordinate vector

$$\begin{pmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{pmatrix} = \begin{pmatrix} MCA\text{-}1_1, MCA\text{-}2_1, \ldots MCA\text{-}k_1 \\ MCA\text{-}1_2, MCA\text{-}2_2, \ldots MCA\text{-}k_2 \\ . \quad \ldots \quad . \\ . \quad \ldots \quad . \\ . \quad \ldots \quad . \\ MCA\text{-}1_n, MCA\text{-}2_n, \ldots MCA\text{-}k_n \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ \beta_n \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ . \\ . \\ . \\ \epsilon_n \end{pmatrix}$$

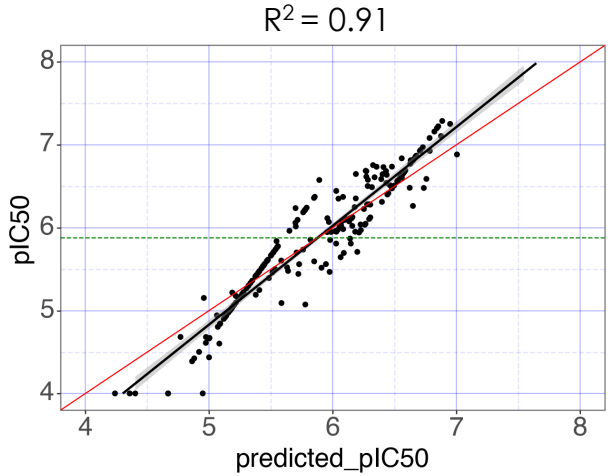
Model building using Markush descriptors

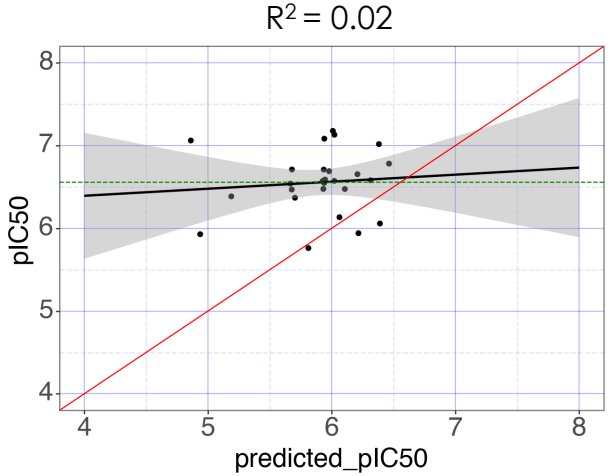
MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

Training data
- Compounds up to May 21

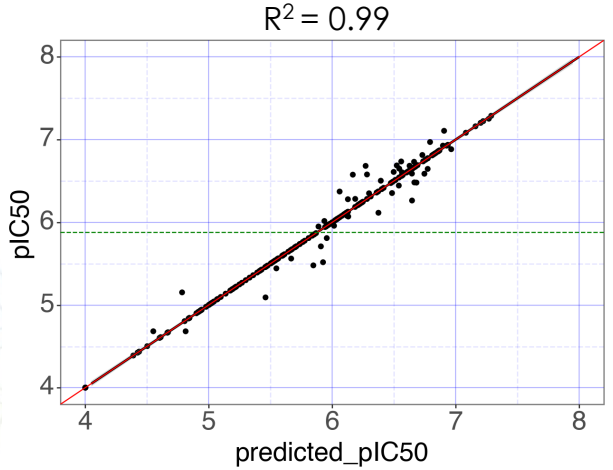
Test data:
- New compounds May-Aug 21
- With existing SARkush groups only

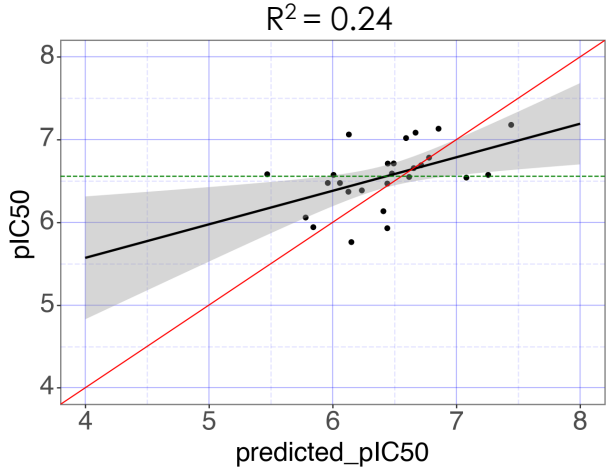
Standard Free Wilson Ridge regression on 229 descriptors:





OLS on 188 MCA coordinates:





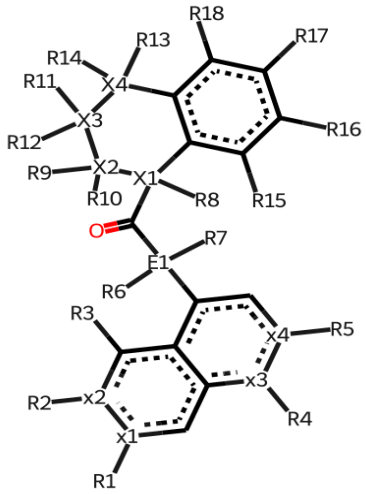
Reduced MCA coordinates provide better descriptors to predict future data

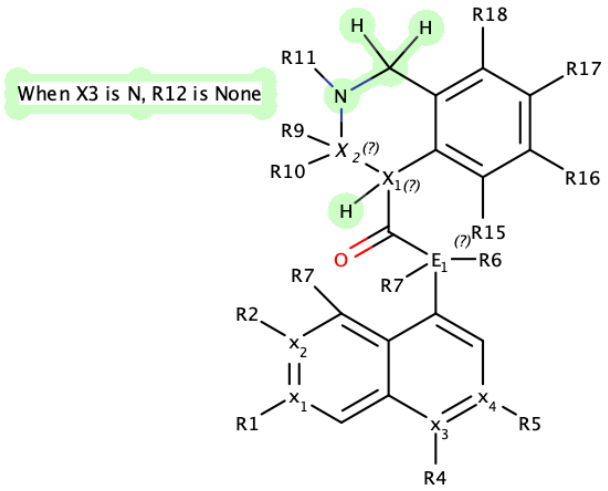
Model building using Markush descriptors

MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

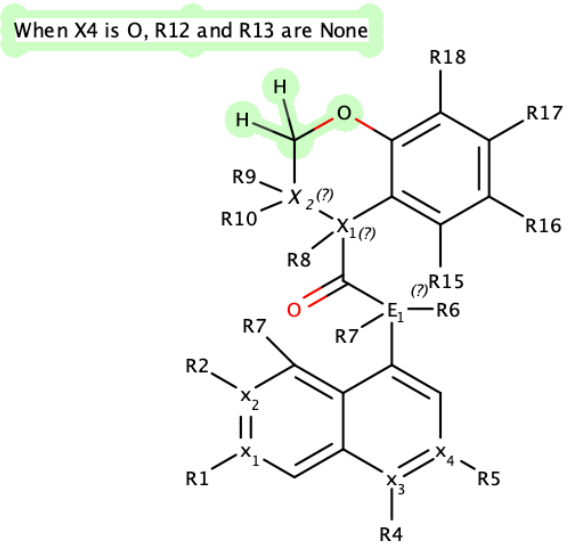
The MCA coordinates and OLS beta coefficients can be used to extract the significant features in the data!

Markush feature	cosine similarity with MCA 1	MCA 1 * OLS coefficient
X4 = C	0.92	1.13
X3 = C	0.90	-0.52
R12 = [H]	0.89	-0.55
R12 = None	0.89	1.07
R14 = [H]	0.86	1.19
X3 = N	0.86	1.15
R11 = [H]	0.83	-0.49
R14 = None	0.81	-0.48
R13 = [H]	0.77	0.98
R13 = None	0.72	-0.63
X4 = O	0.72	-0.63
R8 = [H]	0.40	0.43





Favoured feature

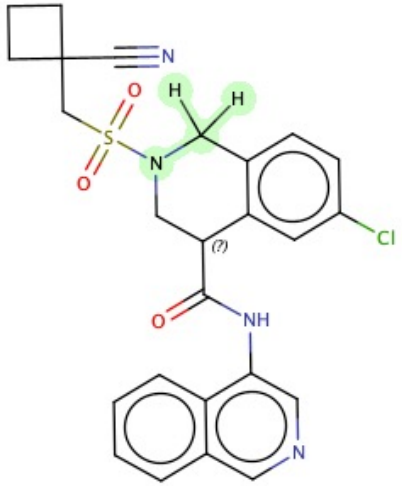


Disfavoured feature

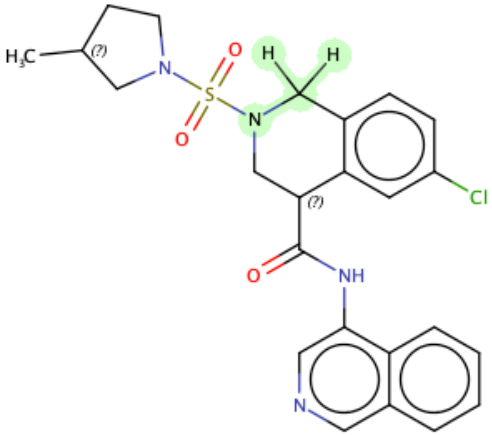
Model building using Markush descriptors

MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

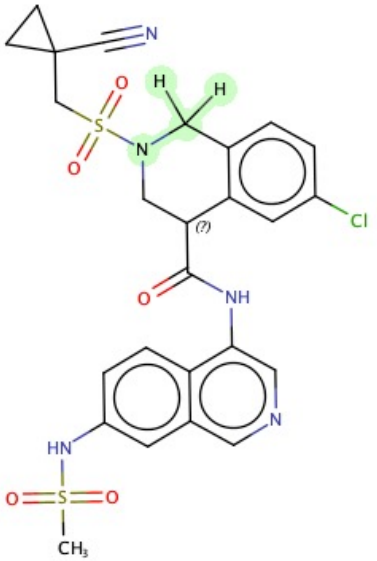
The most active compounds designed between May and August contained this feature…



pIC50 = 7.22



pIC50 = 7.29



pIC50 = 7.40

Patent Corpus Analysis

MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

- MedChemica contributed to a SOS1 inhibitor patent review in 2021:
  - https://www.tandfonline.com/doi/abs/10.1080/13543776.2021.1952984?journalCode=ietp20

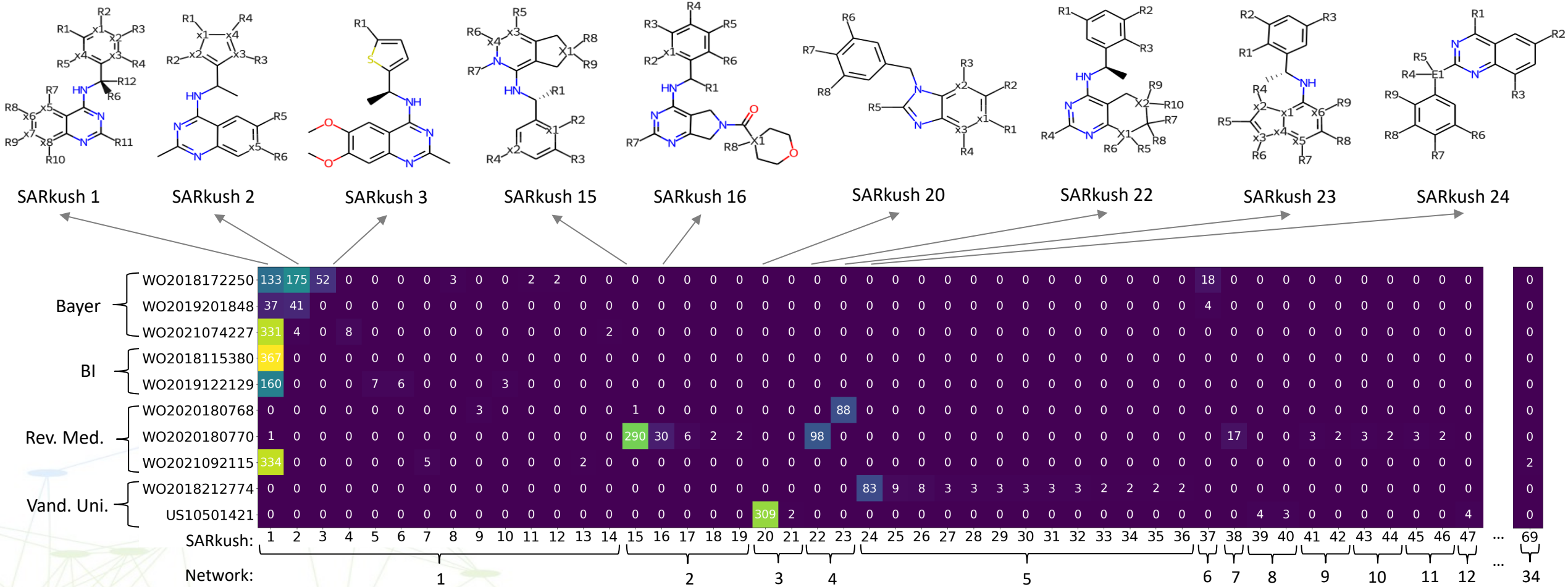
- 10 patents from 4 groups were analysed:
  - The Markush structures were reported
  - Activity data was used to assess the tolerance of varying groups around the structures

- Could we have used SARkush to aid in this analysis?

Analysis of SOS1 Inhibitor Patent Corpus

MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

2832 compounds from 10 patents clustered with SARkush:

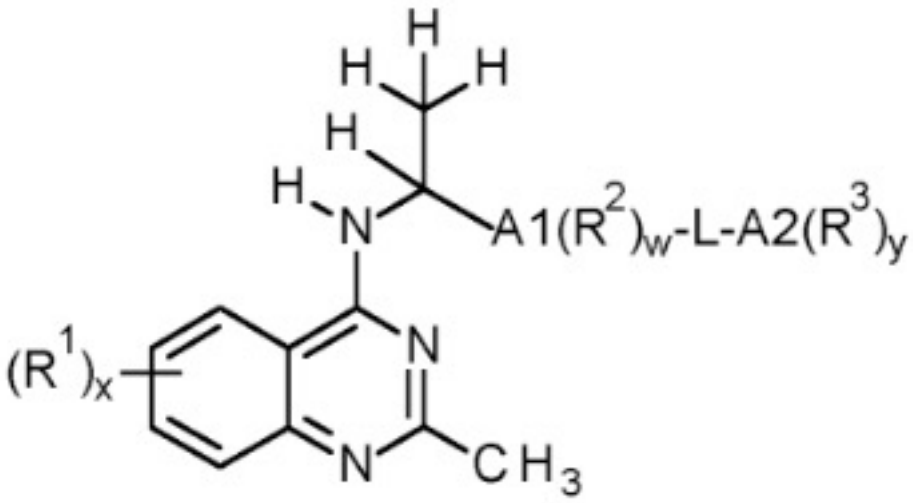


Analysis of WO2018172250

MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

400 compounds from Bayer patent WO2018172250

2-methyl-quinazolines



Analysis of WO2018172250

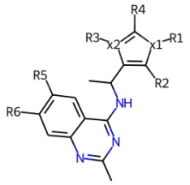
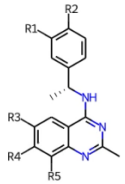
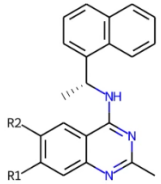
MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

- "Dimethoxyquinazolines appear most frequently and are present in the most potent compound"

0 BQZIA-U0FBK_2023-09-11_08_19_45_775558 Sep 11, 2023, 9:27:56 AM Job Data Retrieved Download

Overview | Network | SARkush 1 (176) | SARkush 2 (133) | SARkush 3 (2) | SARkush 4 (52) | SARkush 5 (18) | SARkush 6 (3) | SARkush 7 (2) | SARkush 8 (2) | SAR

Export Overview | View Box Plot | Download Pair File

Network	Sarkush	Sarkush Structure	Number of Compounds	Percentage of Compounds	Number of Groups	Measurement Median	CLogP Median	P
1	1		176	43.89%	8	6.32	4.97	
2	2		133	33.17%	5	6	4.9	
2	3		2	0.5%	2	5.94	5.02	

Conclusions

MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

- SARkush® automatically clusters compounds into networks and Markush-like structures
- Markush-like structures can be used to:
  - Quickly summarise compound data (for medicinal chemists and patent lawyers)
  - Produce input for further computational analyses (e.g Free-Wilson or QSAR)

- Depicting Markush structures can be challenging – we still have work to do

- Case Studies:
  - Markush structures were used to monitor lead optimisation of the COVID Moonshot series
  - Markush descriptors with MCA were used to build a Free Wilson model on the COVID Moonshot data that identifies the favoured and disfavoured features in the series
  - Markush was used to analyse a SOS1 inhibitor patent corpus

Future Work

MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

- COVID Moonshot graduated into ASAP Discovery (https://asapdiscovery.org/), where antiviral discovery programs are being conducted for:
  - Coronaviruses
  - Flaviviruses
  - Picornaviruses
- SARkush® will be used to aid SAR exploration across each program

- SARkush® GUI to be released later this year!

Acknowledgements

MedChemica
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

The MedChemica team:

- Phillip de Sousa - GUI
- Jessica Stacey
- Al Dossetter
- Dan James
- Ed Griffen
- Jacqui Clarkson
- David Cousins
- Bashy Kahn
- Andrew Leach
- Jason Tierney

- Dave Cosgrove (RDKit community) for algorithm advice
- Allan Jordan (Sygnature) for use-case discussion
- The COVID Moonshot Consortium
- ASAP Discovery