

# MedChemica

CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

MCPairs Online AI tools to accelerate projects  
Generic training Course

**Twitter @MedChemica**  
**Twitter @covid\_moonshot**

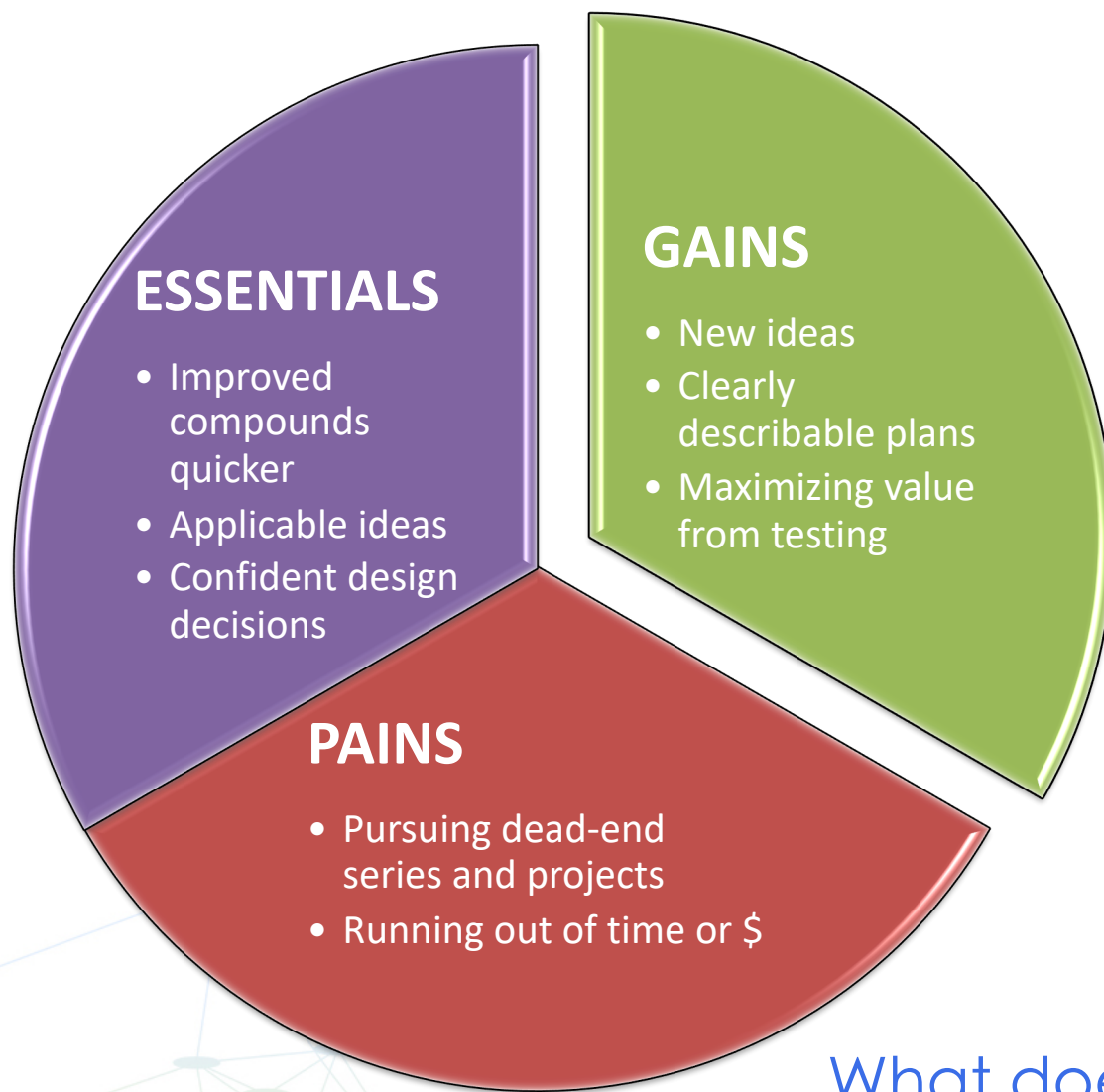
**Twitter #BucketListPapers**  
[www.medchemica.com/bucket-list/](http://www.medchemica.com/bucket-list/)

Last update – 21<sup>st</sup> June 2020

June 2020

## For the training today:

- Going to present Powerpoint slides, then live demos with MCPairs Online in Google Chrome
- What MCPairs gives to the experienced medicinal chemist?
- From SAR to Matched Molecular Pair Analysis
- How MCPairs works?
- SAR analysis with Compound-To-Pairs
- Testing your ideas with SpotDesign™
- **Ideas and new directions with RuleDesign**



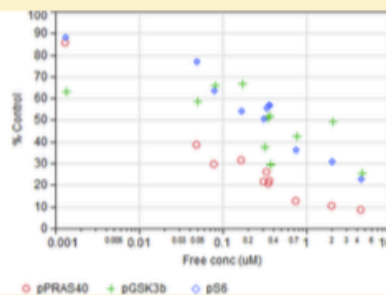
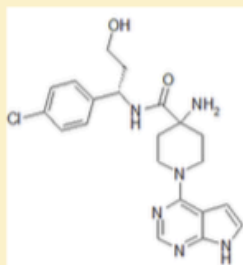
What does a project need?

## Training will use a literature example

**Discovery of 4-Amino-*N*-[(1*S*)-1-(4-chlorophenyl)-3-hydroxypropyl]-1-(7*H*-pyrrolo[2,3-*d*]pyrimidin-4-yl)piperidine-4-carboxamide (AZD5363), an Orally Bioavailable, Potent Inhibitor of Akt Kinases**

Matt Addie, Peter Ballard, David Buttar, Claire Crafter, Gordon Currie, Barry R. Davies, Judit Debreczeni, Hannah Dry, Philippa Dudley, Ryan Greenwood, Paul D. Johnson, Jason G. Kettle,\* Clare Lane, Gillian Lamont, Andrew Leach, Richard W. A. Luke, Jeff Morris, Donald Ogilvie,<sup>†</sup> Ken Page, Martin Pass, Stuart Pearson, and Linette Ruston

Oncology iMed, AstraZeneca, Alderley Park, Macclesfield SK10 4TG, United Kingdom

**S** Supporting Information

**ABSTRACT:** Wide-ranging exploration of analogues of an ATP-competitive pyrrolopyrimidine inhibitor of Akt led to the discovery of clinical candidate AZD5363, which showed increased potency, reduced hERG affinity, and higher selectivity against the closely related AGC kinase ROCK. This compound demonstrated good preclinical drug metabolism and pharmacokinetics (DMPK) properties and, after oral dosing, showed pharmacodynamic knockdown of phosphorylation of Akt and downstream biomarkers in vivo, and inhibition of tumor growth in a breast cancer xenograft model.

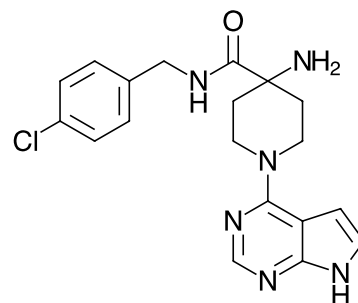
<http://dx.doi.org/10.1021/jm301762v>

# Capivasertib (AZD5363) – AKT inhibitors

AKT pIC<sub>50</sub> 7.89 (13nM)  
 LogD 2.9  
 Sol (pSol) -5.3 (5 – 150μM)  
 hERG pIC<sub>50</sub> 5.2 (5.2 μM)

**Potent enough**

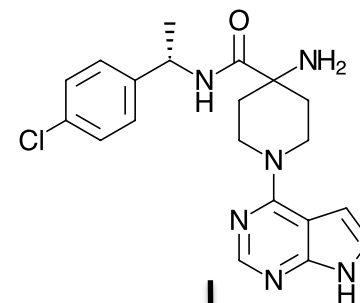
**hERG and improved solubility**



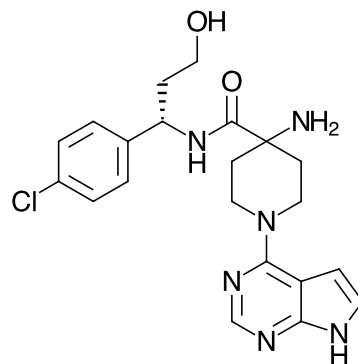
CHEMBL598194

ΔpIC<sub>50</sub> + 0.2  
 ΔLogD - 0.2 (?)  
 ΔpSol + 0.5  
 ΔhERG ~ 0.2

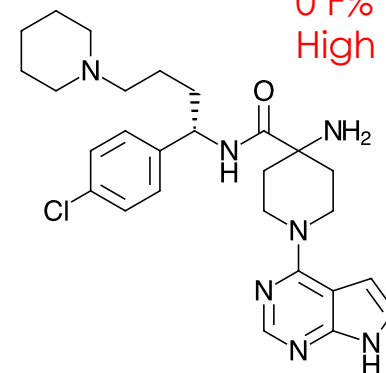
CHEMBL2325742



ΔpIC<sub>50</sub> ~ 0.0  
 ΔLogD - 0.3  
 ΔpSol + 1.4  
 ΔhERG - 0.5



CHEMBL2325741



CHEMBL2325729

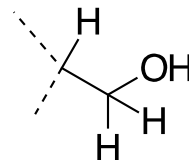
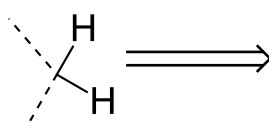
0 F%  
 High Cl

AKT pIC<sub>50</sub> 8.4 (3nM)  
 LogD 2.5  
 Sol (pSol) -3.1 (780μM)  
 hERG pIC<sub>50</sub> <4.0 (>100 μM)

**Potent**

**No hERG and improved solubility**

## From SAR to MMPA.....

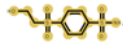
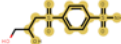
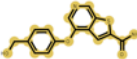
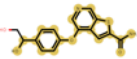
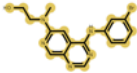
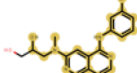
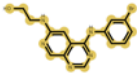
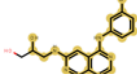
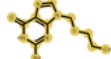
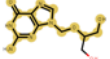
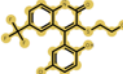
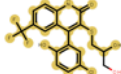
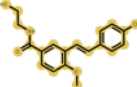
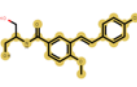


**MedChemica**  
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

A	B	pSol A ( $\mu\text{M}$ )	pSol B ( $\mu\text{M}$ )	$\Delta\text{pSol}$
 CHEMBL2325742	 CHEMBL2325741	- 4.1 (77 $\mu\text{M}$ )	- 3.1 (870 $\mu\text{M}$ )	1.0
 CHEMBL3356658	 CHEMBL218767	- 6.0 (1.0 $\mu\text{M}$ )	- 3.7 (178 $\mu\text{M}$ )	2.3
 CHEMBL456802	 CHEMBL456322	-5.7 (2.0 $\mu\text{M}$ )	- 4.1 (82 $\mu\text{M}$ )	1.6
				3 pairs +ve Sol Median 1.6

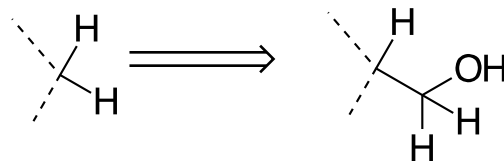
MCPairs Rule finder required 6 matched pairs for 95% confidence

## From SAR to MMPA.....

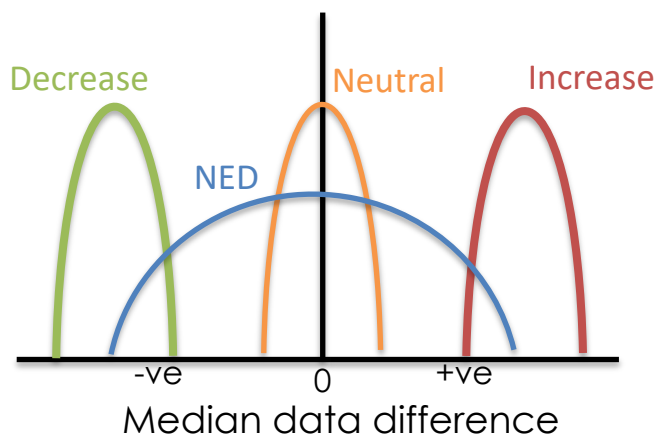
A	B	E	F	H	I	J	K	L	M
compound name A	compound name B	Depiction A	Depiction B	Aqueous Solubility at pH 7.4 solubility [CHEMBL2362975] unit	Aqueous Solubility at pH 7.4 solubility [CHEMBL2362975] qualifier A	Aqueous Solubility at pH 7.4 solubility [CHEMBL2362975] measurement A	Aqueous Solubility at pH 7.4 solubility [CHEMBL2362975] qualifier B	Aqueous Solubility at pH 7.4 solubility [CHEMBL2362975] measurement B	Aqueous Solubility at pH 7.4 solubility [CHEMBL2362975] measurement delta
CHEMBL104459	CHEMBL316800			log10(M)	=	-1.6042	=	-1.2971	0.3071
CHEMBL118022	CHEMBL115462			log10(M)	=	-4.63875	=	-4.3343	0.30445
CHEMBL161956	CHEMBL165547			log10(M)	=	-9.9586	=	-9.7447	0.2139
CHEMBL165864	CHEMBL166093			log10(M)	=	-10.699	=	-9.6778	1.0212
CHEMBL184	CHEMBL182			log10(M)	=	-5.2353	=	-4.7115	0.5238
CHEMBL184521	CHEMBL439660			log10(M)	=	-5.0168	=	-3.3704	1.6464
CHEMBL1949786	CHEMBL1949790			log10(M)	=	-4.31675	=	-3.1588	1.15795

Actual Rule from MCPairs  
Endpoint:  
Aqueous Solubility at pH 7.4  
[CHEMBL2362975]

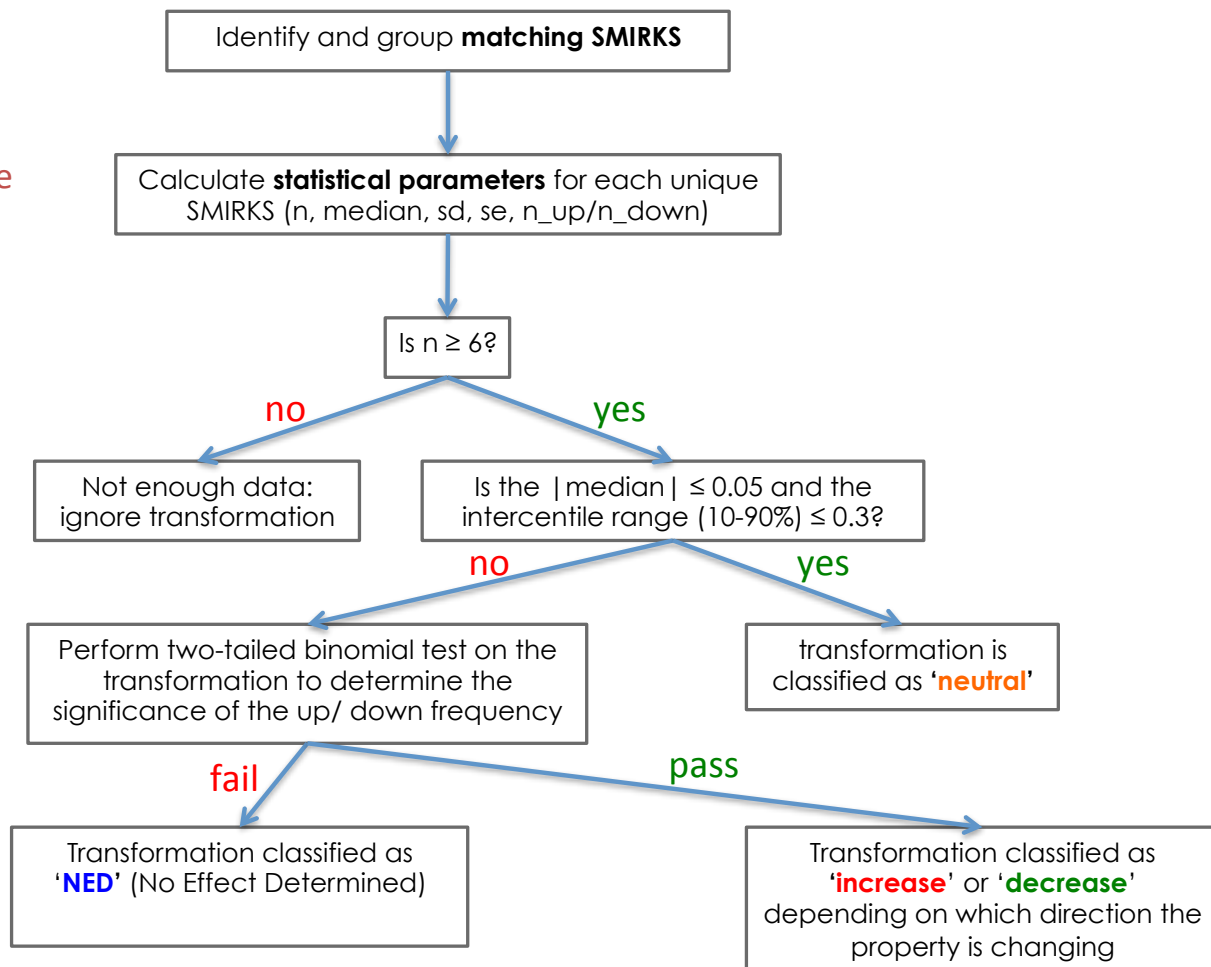
n-qual 69  
n-qual-up 47  
n-qual-down 21  
median  $\Delta pSol$  0.26  
std dev +/- 0.636



# Rule selection



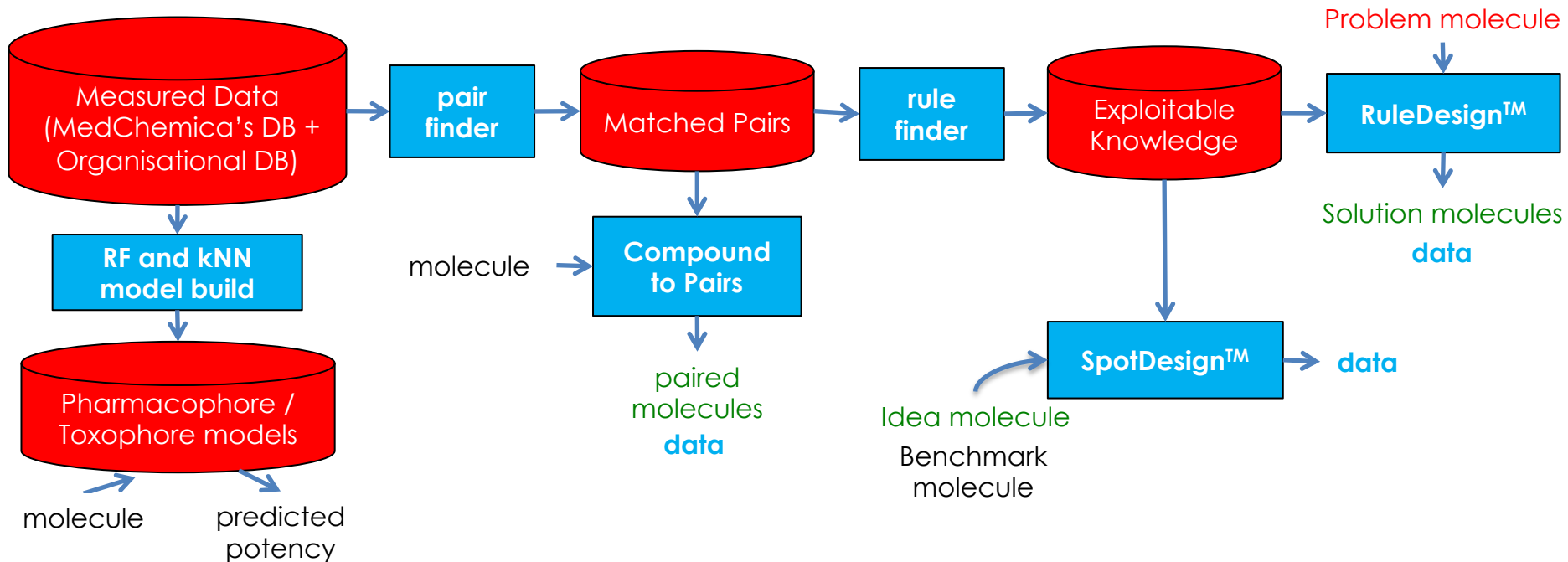
- No assumption of normal distribution
- Manages 'censored' = qualified / out-of-range data



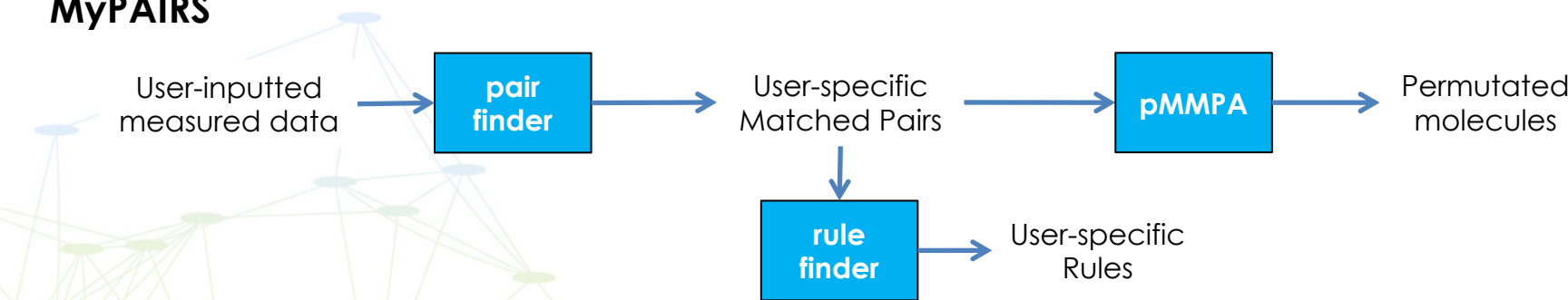
Leach et al. *J. Chem. Inf. Model.* **2017**, **57**, 2424 - 2436

# Exploiting data derived ADMET Knowledge

## DATABASE TOOLS

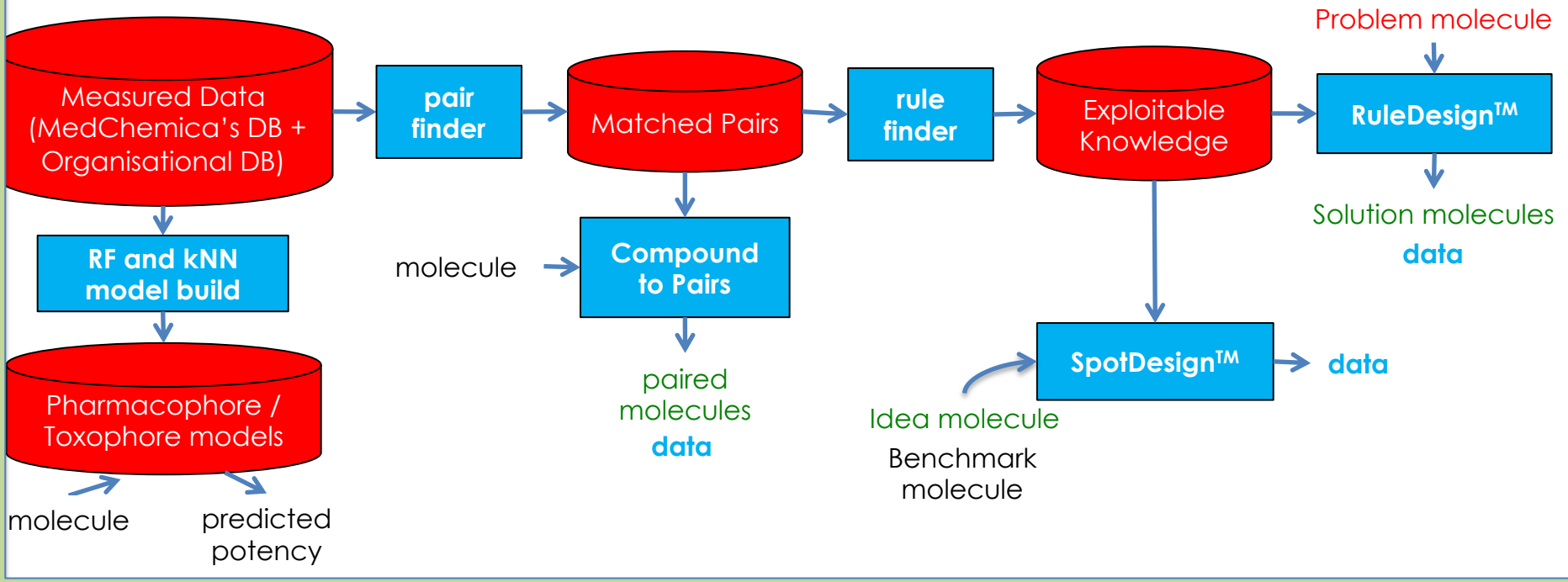


## MyPAIRS

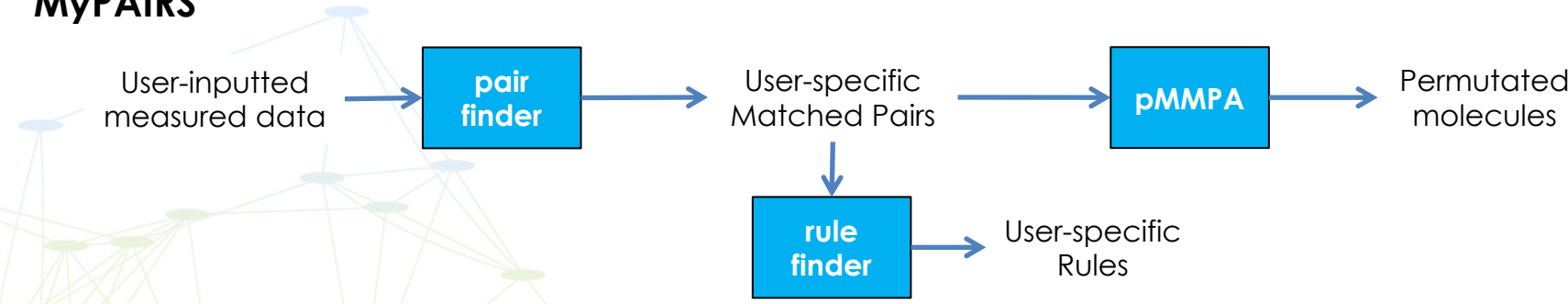


# Exploiting data derived ADMET Knowledge MedChemica

## DATABASE TOOLS

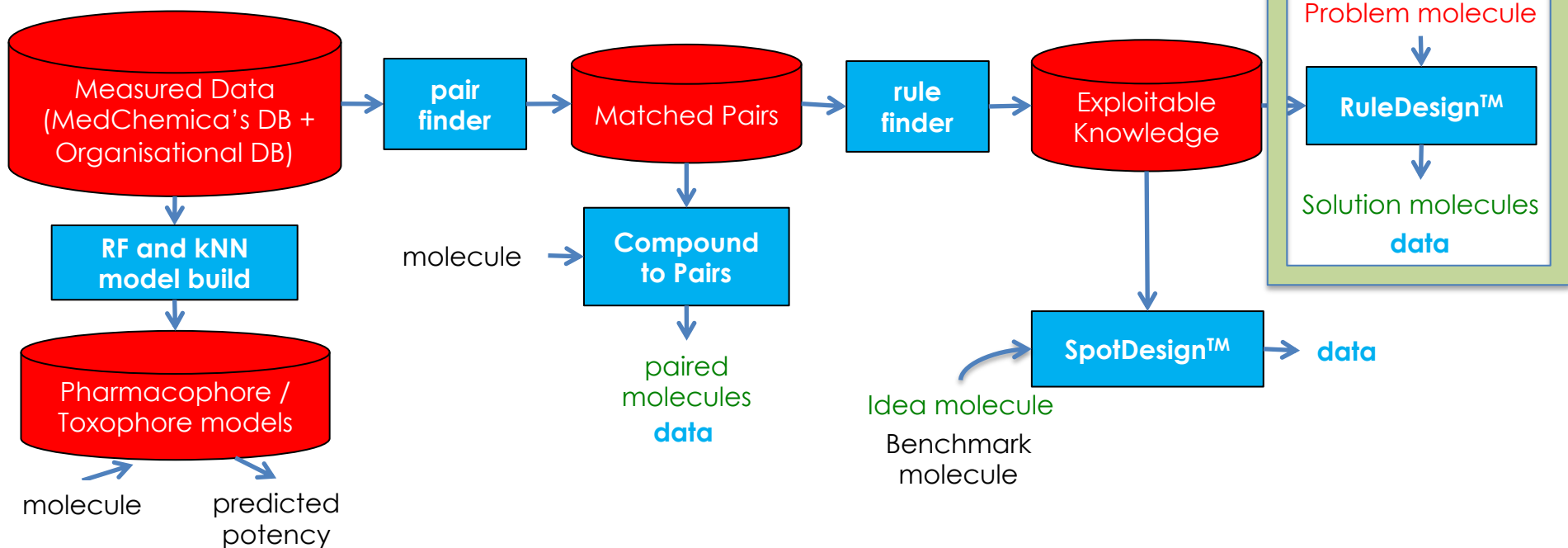


## MyPAIRS

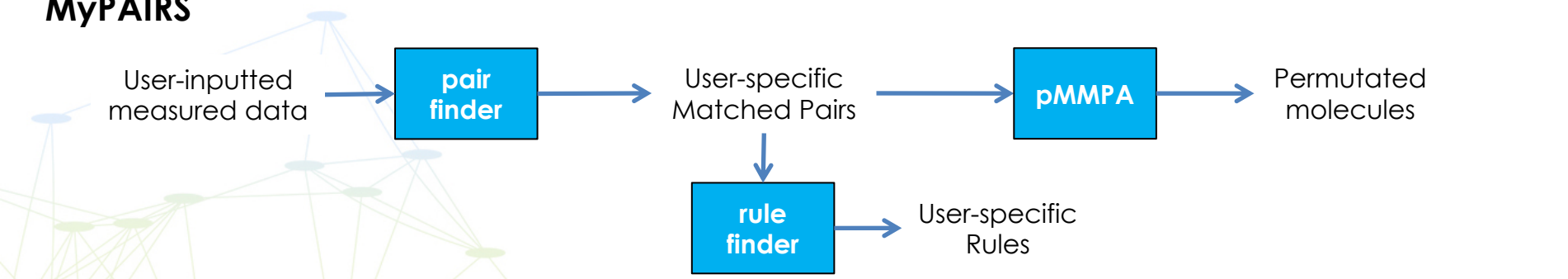


# Exploiting data derived ADMET Knowledge

## DATABASE TOOLS



## MyPAIRS



## Exercise – New Idea based on rules (Show me the Gems'!)

- Input chemical name for lead molecule CHEMBL2325997
  - Choosing the property to improve (setting the Goal)
    - Find Phys\_Prop -> Solubility (Broad Goal)
  - Filtering the results down by substructure and phys-props
  - Export the results – controlling the output
  - Exploring results in Excel
    - what do the numbers mean?
    - How do I use these to make a knowledge based decision?
  - Drilling back to the original source (looking at the matched pairs)



## Matched Pairs

Icon indicates that these tools use the central database of knowledge

Access to MCPairs Database Matched Molecular Pairs Search



### RULEDESIGN™

*Formerly Known as Compounds From Rules*

Submit compound(s), enumerate products using Rules from the central database



### SPOTDESIGN™

Submit a reference compound and explore your ideas seeking support from the central database



### COMPOUND TO PAIRS

Search the central database of the matched pairs of a compound



### COMPOUND TO MEASUREMENTS

Search the central database for the current aggregated data of a compound



Input Molecule\*

Single

Multi

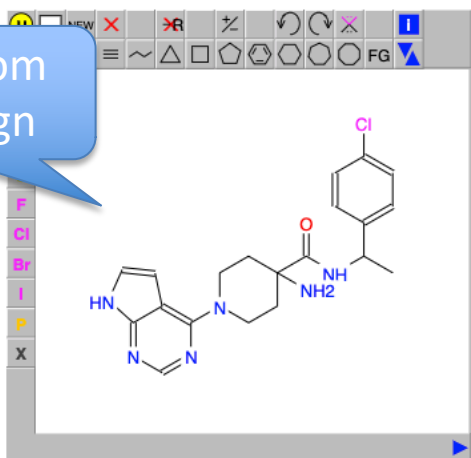
**Compound Name:**

CHEMBL2325997

**SMILES\*:**

CC(c1ccc(cc1)Cl)NC(=O)C2(CCN(CC2)c3c4cc[nH]c4ncn3)N

Copied from  
SpotDesign



### Goal\*

**Direction\*:**

Increase

**Endpoint\*:**

solubility ✕

## Check the direction

Select

**Specificity:**

## Goal from SpotDesign

## Advanced Filters

 Molecular charge:

☐ **HBA:**

**HBD:**

**CLogP:**

**RMM:**

PSA:

**Substructure Lock:**

### Input Compound Phys. Properties

## Anions

0

HBA

5

LogP

2.79

RMM

398.90

### Cations

1

HBD

3

PSA

99.93

GO!

Submit

# Rule Design

Input Molecule\*

Single Multi

Compound Name:

CHEMBL2325997

SMILES\*:

CC(c1ccc(cc1)Cl)NC(=O)C2(CCN(CC2)c3c4cc[nH]c4ncn3)N

C

N

O

S

F

Cl

Br

I

P

X

Goal\*

Direction\*:

Increase

Endpoint\*:

solubility x

Select

☐ Specificity:

Advanced Filters

☐ Molecular charge:
 

☐ HBA:
 ☐ HBD:
 ☐ CLogP:
 ☐ RMM:
 ☐ PSA:
 ☐ Substructure Lock:

Input Compound Phys. Properties

Anions	0	Cations	1
HBA	5	HBD	3
LogP	2.79	PSA	99.93
RMM	398.90		

Focus your design

Save Your results

Submit

Save

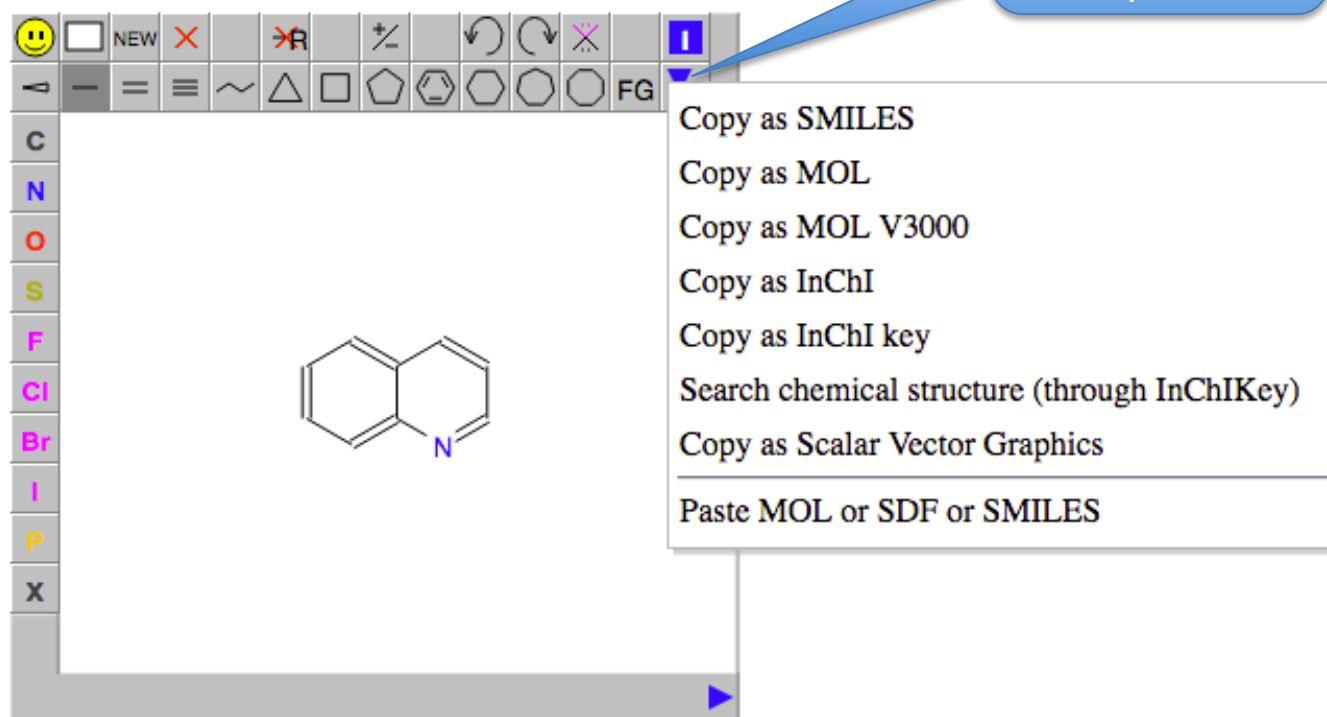
#	Timestamp	Status	Number of Products
0	Feb 1, 2021, 11:46:13 AM	Complete	612

Exploiting medicinal chemistry knowledge to accelerate projects

June 2020

## TopTip : Other methods to get a structure in

- **Pasting structures in and out**



The screenshot shows a chemical structure editor interface. In the center is a chemical structure of 1,2,3,4-tetrahydronaphthalene (a bicyclic system with a benzene ring fused to a cyclohexane ring). To the left is a vertical toolbar with buttons for elements: C, N, O, S, F, Cl, Br, I, P, and X. Above the structure is a horizontal toolbar with various editing tools, including a blue square button with a white 'I' icon. A blue callout bubble points to this button with the text "Click blue triangle square". A context menu is open, listing the following options: Copy as SMILES, Copy as MOL, Copy as MOL V3000, Copy as InChI, Copy as InChI key, Search chemical structure (through InChIKey), Copy as Scalar Vector Graphics, and Paste MOL or SDF or SMILES.

# TopTip : Other methods to get a structure in

- **Upload a file of multiple compounds**
  - File can contain compounds + measurements or just compounds

The screenshot shows a web interface for uploading multiple compounds. The main heading is "Input Molecule\*". Below it, there are two tabs: "Single" and "Multi". The "Multi" tab is selected. A blue button labeled "+ Add Compounds" is visible. A dashed box highlights the text "Available file types: Compound Data file or SMILES file" and "Maximum 50 compounds". A callout bubble points to the "Multi" tab with the text "Click here to input a multi-compound file". Another callout bubble points to the "Available file types" text with the text "Hover over for info on the different file formats accepted". A third callout bubble points to the "SMILES file" link with the text "Click here to input a multi-compound file". Below the dashed box, a "Compound Data file" section is shown, detailing accepted extensions (.txt, .tsv, .csv) and file format requirements. To the right, there is a "Direction\*" section with a dropdown menu set to "Increase" and a "Select" button. Below that, there is a "Specificity" checkbox. At the bottom, there is an "Advanced Filters" section with checkboxes for "Molecular charge:", "HBA:", "HBD:", "CLogP:", "RMM:", "PSA:", and "Substructure Lock:".

Input Molecule\*

Single Multi

+ Add Compounds

Available file types: Compound Data file or SMILES file  
Maximum 50 compounds

**Compound Data file**

**Accepted Extensions:** .txt .tsv .csv

Files are tab (\t) or comma (,) separated, and can be submitted without headers but the order of the columns must be fixed as: compound\_name, compound\_structure, measurement, qualifier. Measurement is on the log scale (e.g. pIC50), and the qualifier is either ">", "<" or "=".

Check the documentation for more details.

Direction\*:  
Increase

solubility x Select

☐ Specificity:

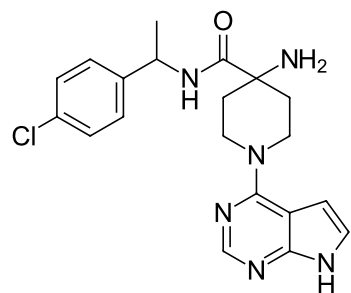
**Advanced Filters**

- ☐ Molecular charge:
- ☐ HBA:
- ☐ HBD:
- ☐ CLogP:
- ☐ RMM:
- ☐ PSA:
- ☐ Substructure Lock:

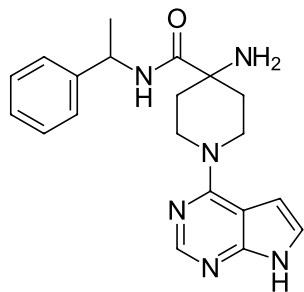
# What molecule should I put in?

The more specific the 'seed' molecule, the fewer rules will apply to it

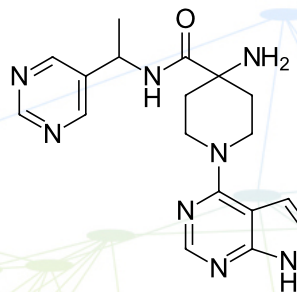
Suggestions to increase solubility



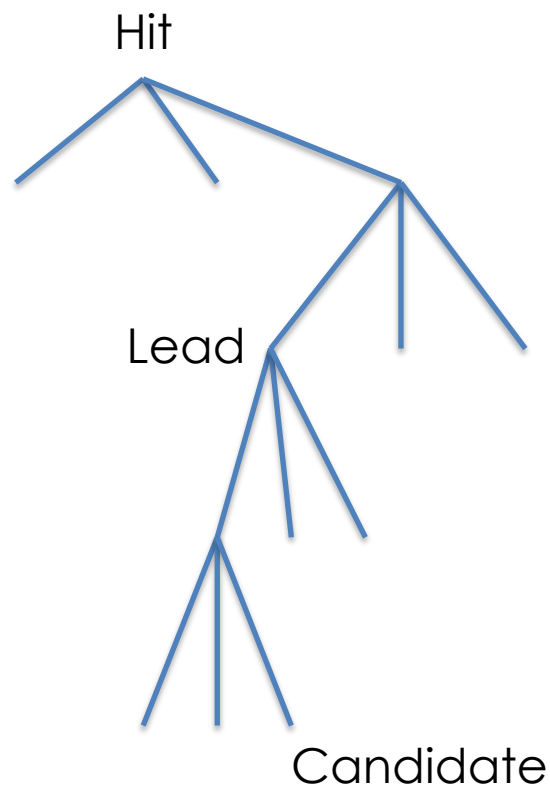
612



664



216



## Filtering down the output

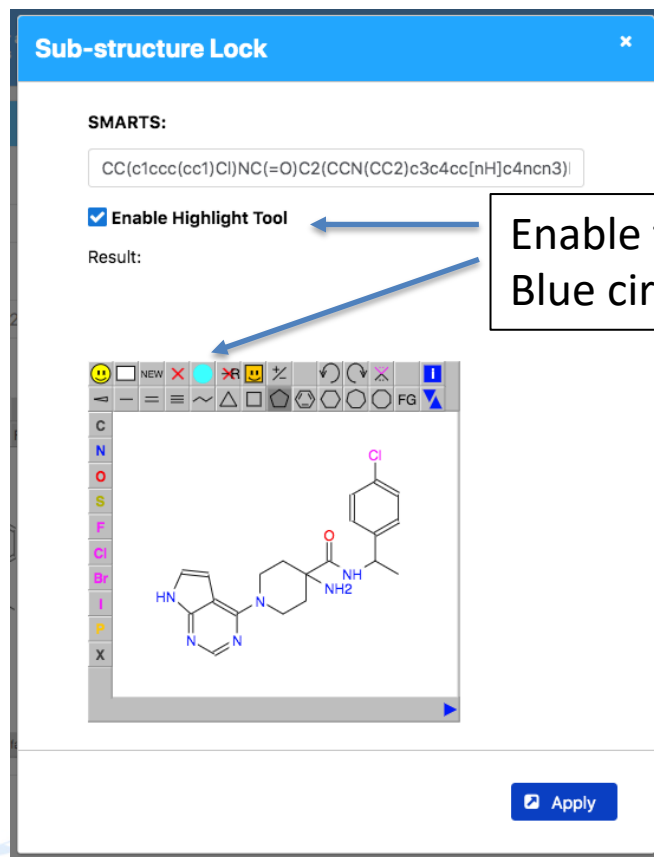
- Adding a locked structure (eg pharmacophore)
  - Highlighting the atoms to changes
  - Blocking H substitution
  - Variable atoms
  - 'Complete SMARTS patterns'
  - Also see Tips'n'Tricks document for a detailed work flow
- Charge control  
Filtering by the charge of the final compounds (neutral, acidic, basic)
- Simple Lipinski property filters



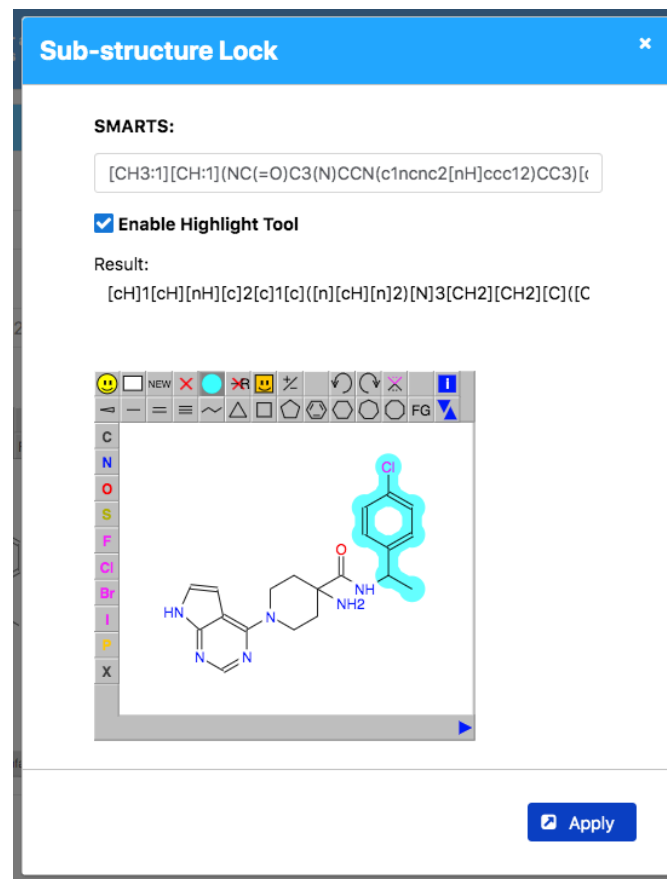
Draw

# Highlight the part you want to change

- Add a locked structure with Substructure-Lock (eg pharmacophore)



Enable tool, click  
Blue circle

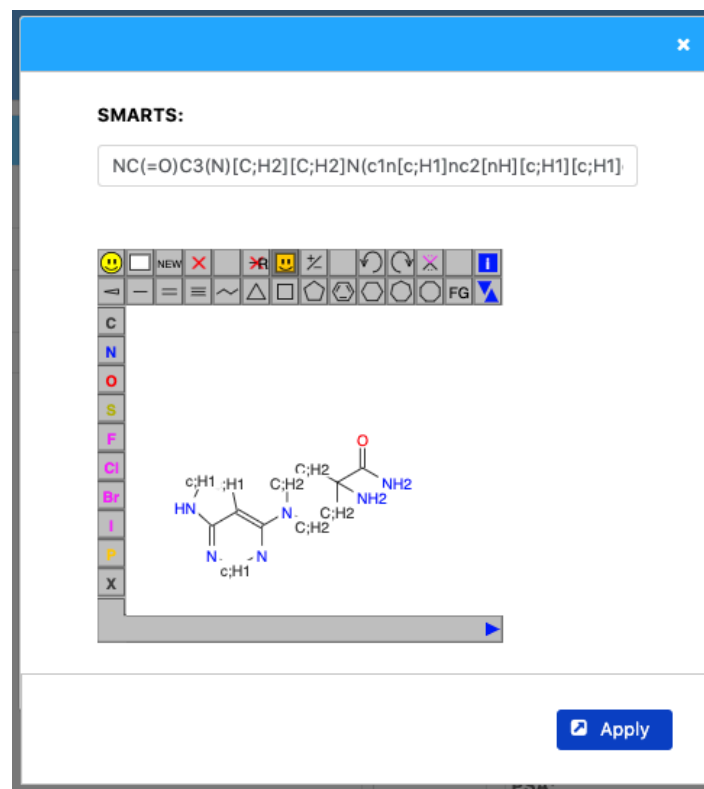
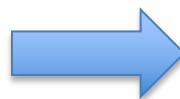
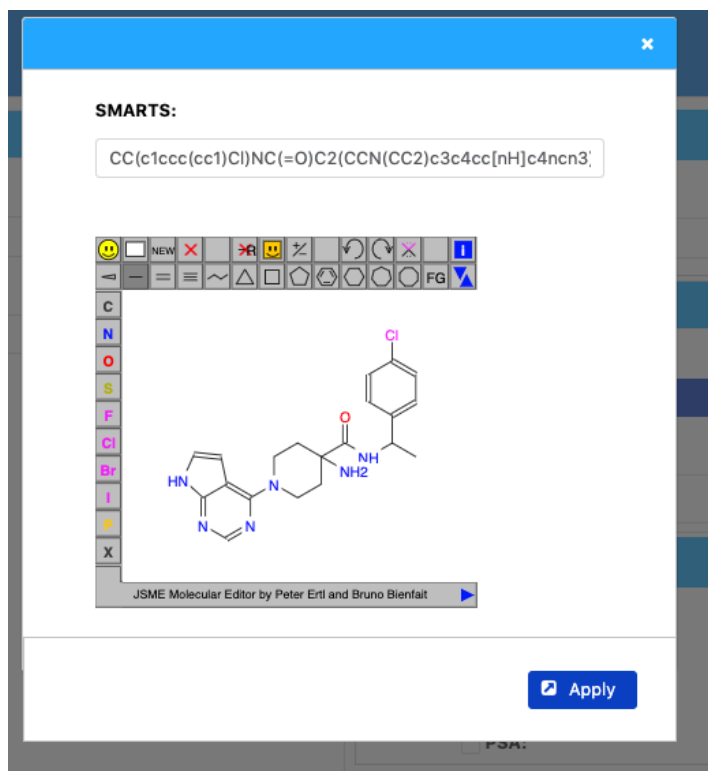


Click atoms with highlighter tool

# Alternatively lock out specific hydrogens

 Draw

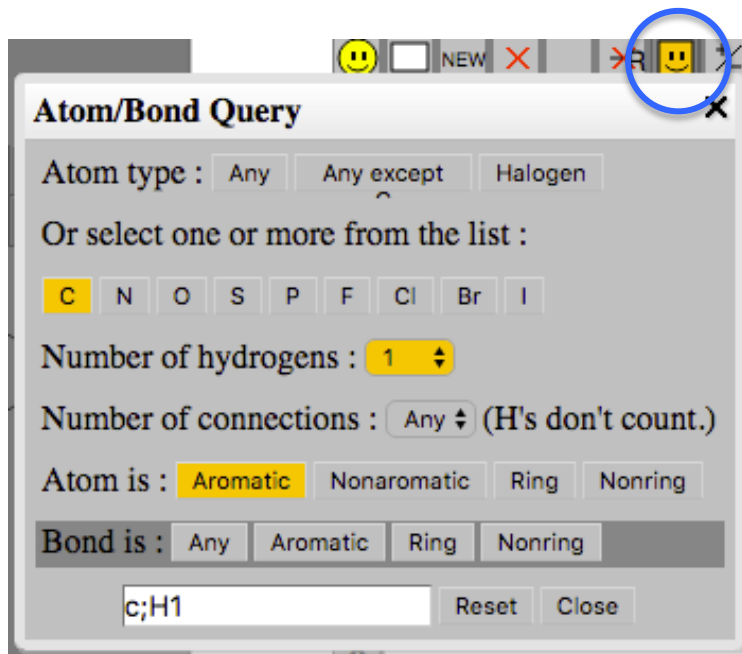
- Add a locked structure with Substructure-Lock (eg pharmacophore)



Delete to groups you want to change  
Lock out hydrogens to stop substitution

# Defining H-atoms to block out

Blocking H



Atom/Bond Query

Atom type : Any Any except Halogen

Or select one or more from the list :

C N O S P F Cl Br I

Number of hydrogens : 1

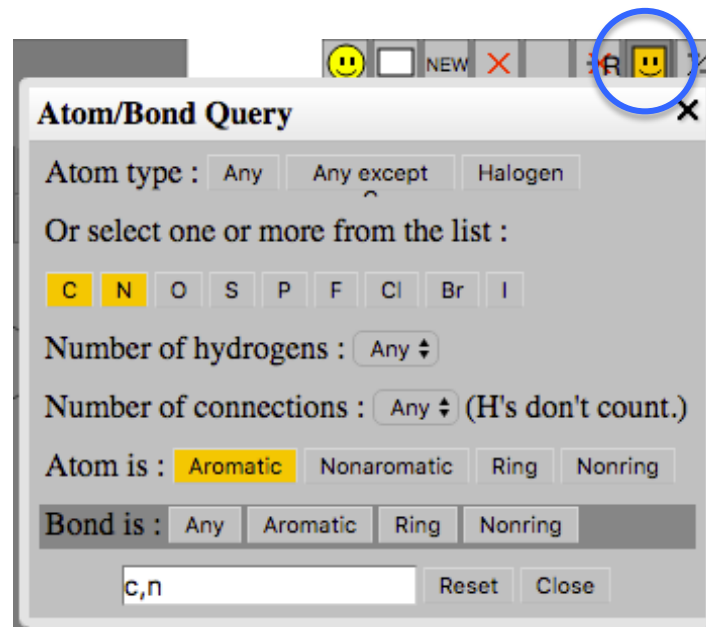
Number of connections : Any (H's don't count.)

Atom is : Aromatic Nonaromatic Ring Nonring

Bond is : Any Aromatic Ring Nonring

c;H1 Reset Close

Variable atoms



Atom/Bond Query

Atom type : Any Any except Halogen

Or select one or more from the list :

C N O S P F Cl Br I

Number of hydrogens : Any

Number of connections : Any (H's don't count.)

Atom is : Aromatic Nonaromatic Ring Nonring

Bond is : Any Aromatic Ring Nonring

c,n Reset Close

Close the box, then click on atoms in the picture to change their filter types



Sub-structure Lock

SMARTS:

NC(=O)C3(N)[C;H2][C;H2]N(c1n[c;H1]nc2[nH][c;H1][c;H1]c12)[C;H2][C;

Draw

# Combined Filters

Goal Selection

Charge and Calc  
Props

Sub-Structure Lock

## History Box


1st run Solubility ideas	612
2 <sup>nd</sup> run with Sub-Structure Lock	338
3 <sup>rd</sup> Run no neutral molecules	1
4 <sup>th</sup> with single cation	123

# MedChemica

CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY


**Goal\***

Direction\*:

Endpoint\*:   Select

☐ Specificity:

**Advanced Filters**

☒ Molecular charge:  

☐ HBA:


☐ HBD:

☐ CLogP:

☐ RMM:





☐ PSA:

☒ Substructure Lock:

 Draw

**Input Compound Phys. Properties**

Anions	0	Cations	1
HBA	5	HBD	3
LogP	2.79	PSA	99.93
RMM	398.90		

	Number of Products	
	612	 Save
	338	 Save
	1	 Save
	123	 Save

# Common Goal Patterns

Example	Goal	Directions
1	Solubility	increase
2	Metabolism	decrease
3	HLM_Clint_generic_uL.min-1.mg-1	decrease
4	Absorption	increase
5	Caco-2_A_to_B_perm_CHEMBL614058	increase
6	Transporter	decrease
7	MDCK-dog-perm-log(ER)-dog	decrease
8	Fraction unbound	increase
9	PPB_hum_log(proportion_Free)	increase
10	LogD_TM	decrease (usually)
11	Anti-target_cardiac	decrease
12	hERG_hum_inhib_pIC50	decrease
13	Anti-target_DDI	
14	CYP_inhibs_3A4_pIC50_hum	
15	CCR5_human_pIC50_CHEMBL274	decrease (reduce off target)

# How many pairs? – deeper Goal setting

## Specific Goal settings

### 'All Rules'

- all of the Increase and Decrease Rules for all datasets
- warning output can be large
- not suitable for Excel spreadsheet

### 'Hit to Lead'

- most frequent transformations chemists perform

## Non-rules transformations from pair counts

### 'Min 3 pair Trans'

- all transformations with 3 OR MORE matched pairs

### 'Min 6 pair Trans'

- all transformations with 6 OR MORE matched pairs
- Actually Increase, Decrease, Neutral and NED

**Goal\***

**Direction\*:**

Increase

**All Rules**

All Rules

Hit to Lead

**Direction**

Increase

Decrease

**Select**

**Goal\***

**Direction\*:**

3 Pair Trans Only

**Min Pair Rules**

Min 3 Pair Trans

Min 4 Pair Trans

Min 5 Pair Trans

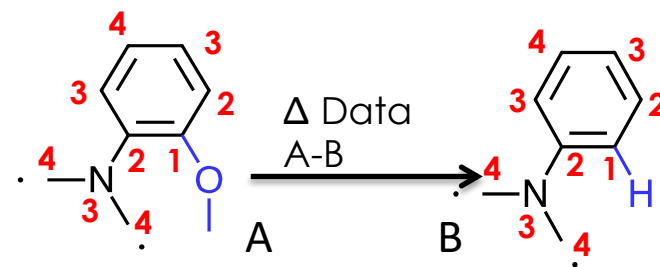
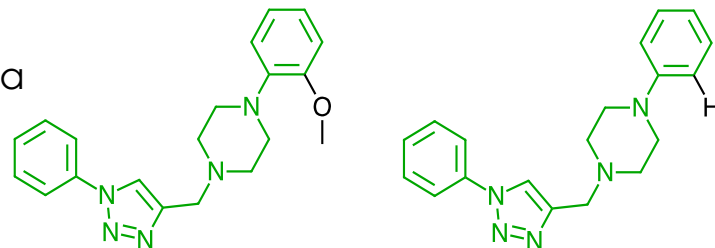
Min 6 Pair Trans

**Specific Pair Rules**

**Select**

# How much environment is captured?

- **Matched Molecular Pairs** – Molecules that differ only by a particular, well-defined structural transformation
- **Capture the change and environment** – MMPs can be recorded as transformations from A  $\rightarrow$  B
- **Statistical analysis to define “medicinal chemistry rules”**  
Defined transformations with high probability of improving properties of molecules
- High specificity are environment size 3 and 4, low is 1 and 2.



Griffen, E. *et al.* *J. Med. Chem.* **2011**, *54*(22), pp.7739-7750.

# Specificity of Transformation

Specificity setting

## 'Specificity slider'

- Range 1 to 4
- 4 being the most specific
- setting 3-4 (or 3,4) only apply the most specific rules.
- The number refers to the amount of environment captured with the change in the group

## Useful combinations

- Min 3 pairs Trans with high specificity
  - with Kinase – best rules for kinase potency
- All Rules with high specificity
  - The most specific ADMET solutions

**Goal\***

**Direction\*:**  
Min 3 Pair Trans

**Endpoint\*:**  
solubility

☒ **Specificity:** [3 - 4]

Click the Blue circle to see run results in Browser

Complete

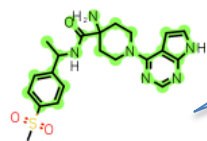
108

Hover mouse over to see larger structures

Products No Sort

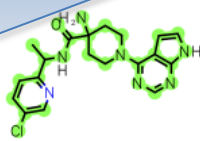
Save Selection

CHEMBL2325990



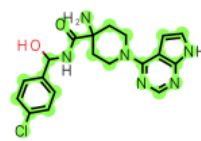
Goals

1627307787759\_2



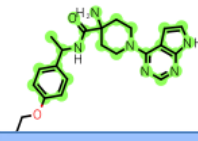
Goals

1627307787759\_3



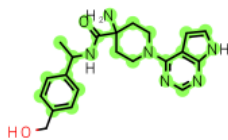
Goals

1627307787759\_4



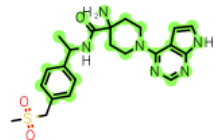
Goals

1627307787759\_5



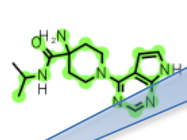
Goals

1627307787759\_6



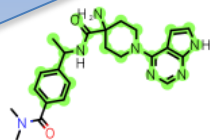
Goals

1627307787759\_7



Goals

1627307787759\_8



Goals

Results are paginated

1 2 3 4 5

Sort the results – select option here

Submit

#	Timestamp	Status	Number of Products	Actions
0	Jul 26, 2021, 2:56:50 PM	Complete	108	Save

Products No Sort

- LogP
- PSA
- RMM
- Aq\_Solubility\_pH\_7.4\_[CHEMBL2362975]
- Aq\_Solubility\_generic\_log(M)
- Aq\_Solubility\_pH\_7.4\_[CHEMBL612558]

Select one of the dataset chosen by the Goal

CHEM

7787759\_2

Goals

1627307787759\_3

Goals

1627307787759\_4

Goals

1627307787759\_5

Goals

1627307787759\_6

Goals

1627307787759\_7

Goals

1627307787759\_8

Goals

1 2 3 4 5

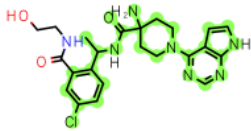
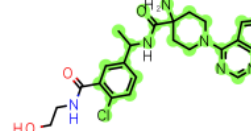
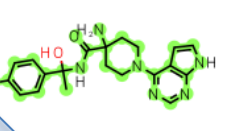
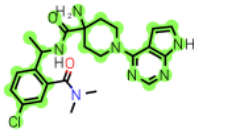
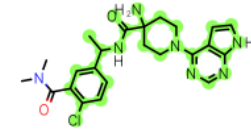
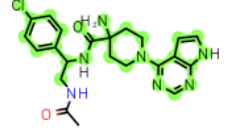
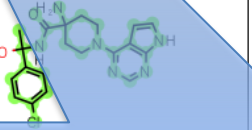
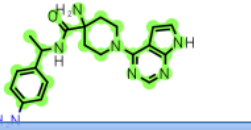
Selected Dataset / Goal

Select sort order – Median Change is the size of the change expected

#	Timestamp	Status	Number of Products	Actions
0	Jul 26, 2021, 2:56:50 PM	Complete	108	<a href="#">Save</a>

Products: Aq\_Solubility\_pH\_7.4\_[CHEMBL2362975]    Median Change    [Save Selection](#)

<div>1627307787759_74</div>  <div>1.15 Med. Change</div> <div>Goals</div>	<div>1627307787759_73</div>  <div>1.15 Med. Change</div> <div>Goals</div>	<div>1627307787759_70</div>  <div>1.09 Med. Change</div> <div>Goals</div>	<div>1627307787759_31</div>  <div>1.00 Med. Change</div> <div>Goals</div>
<div>1627307787759_30</div>  <div>1.00 Med. Change</div> <div>Goals</div>	<div>CHEMBL2325732</div>  <div>0.82 Med. Change</div> <div>Goals</div>	<div>1627307787759_19</div>  <div>0.78 Med. Change</div> <div>Goals</div>	<div>1627307787759_34</div>  <div>0.63 Med. Change</div> <div>Goals</div>

1 2 3

Select preferred example to export

Click the Goal view to see more detailed data

	#	Timestamp	Status	Number of Products	Actions
	0	Jul 26, 2021, 2:56:50 PM	Complete	108	Save

Products: Aq\_Solubility\_pH\_7.4\_[CHEMBL2362975] | Median Change |

Save Selection

1627307787759\_74

1.15  
Med. Change

Goals

1627307787759\_73

1.15  
Med. Change

Goals

1627307787759\_70

1.09  
Med. Change

Goals

1627307787759\_31

1.00  
Med. Change

Goals

1627307787759\_30

1.00  
Med. Change

Goals

CHEMBL2325732

0.82  
Med. Change

Goals

1627307787759\_19

Med. Change

1627307787759\_34

Med. Change

Save Selection is available once compounds have been selected

1 2 3

Input molecule and suggested molecule

Results Viewer

Reference: 1627307787759

Product: 1627307787759\_74

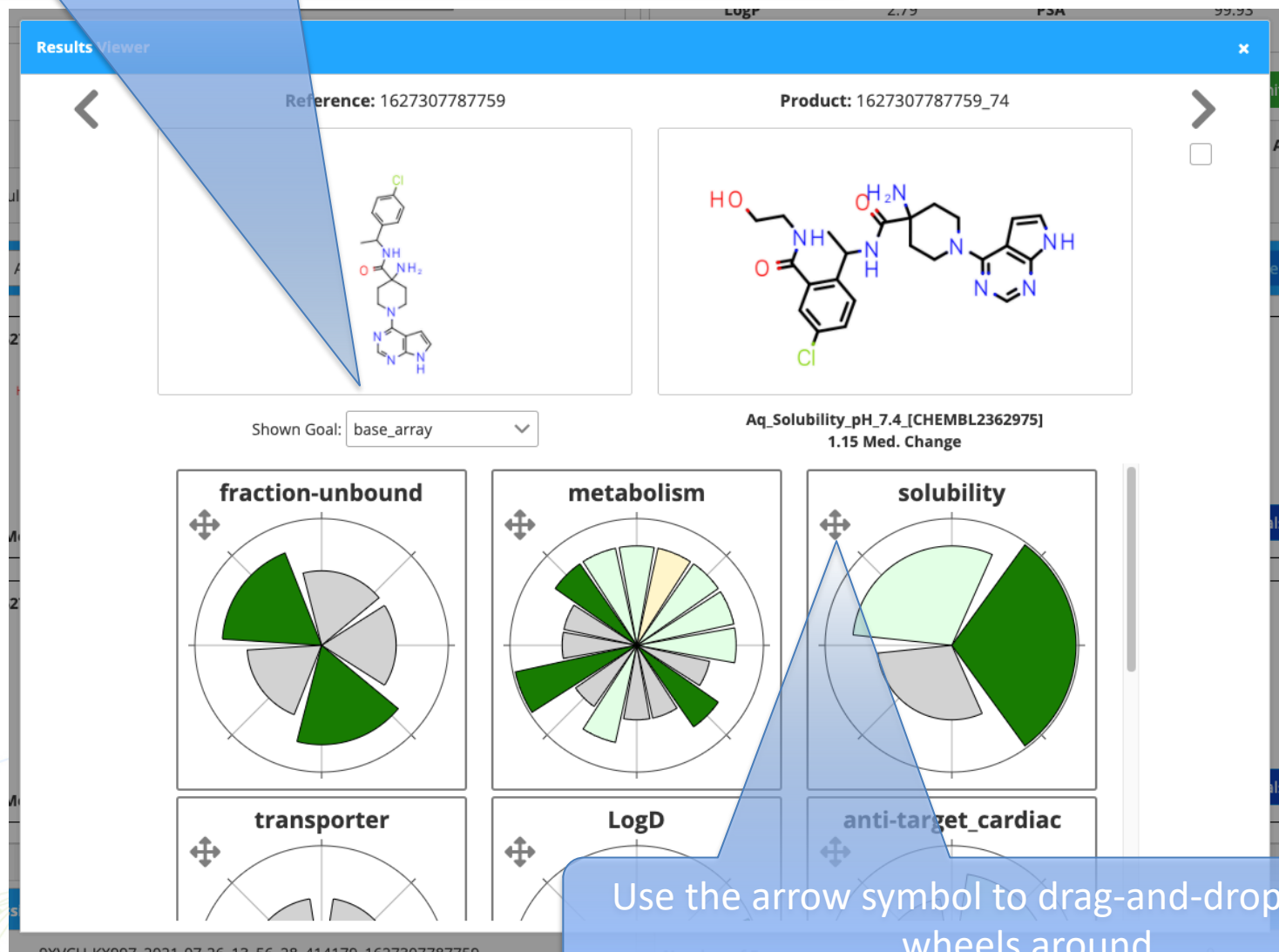
Shown Goal:

Aq\_Solubility\_pH\_7.4 [CHEMBL2362975]  
1.15 Med. Change

Select compound. Use Right and Left arrows to browse through.

Hover over to see more detailed matched pair data

Select base\_array to get the basic ADMET properties



# Export results

The screenshot shows a 'Save' dialog box with the following sections:

- Stats Fields\***: 16 items selected
- Goal Endpoints\***: Aq\_Solubility\_generic\_log(M), Aq\_Solubility\_pH\_7.4\_[CHEMBL23...
- Base Array Endpoints\***: 30 items selected
- Other Endpoints**: No Endpoints Selected
- Selected Compounds**: 4 compounds(s) selected
- File Type\***: Spreadsheet (.xlsx) (selected), Flat File (.txt), CSV (.csv), Spotfire Text Data (.stdf)

Select preferred stats fields

Customise which datasets are exported. Default is your Goal and the Base Array

Select output file style

## Using the results

- Download the Excel file
- Inspect the structures
  - Filtering and sorting is possible as structures locked to cells
- Look at the breadth of data
- Unhide columns for more data
- Or export csv and use Spotfire, Vortex, Datawarrior....

# Looking at the results

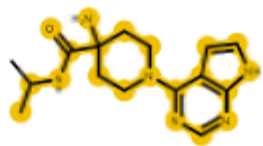
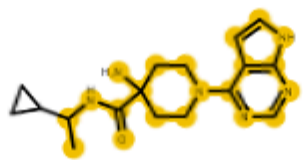
## MedChemica

CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

Results sorted in increasing RMM (Mol Weight)

One column per assay  
– colour and direction  
- LogD decrease, Sol increase

Yellow highlight is the overlap with the input compound

G		H	I	J	K	L		AD	AT
								Aq Solubility pH 7.4 [CHEMBL236297 5]	Aq Solubility pH 7.4 [CHEMBL612558 ]]
pair data		CLogP	HBA	HBD	PSA	RMM	LogD TM direction	direction	direction
									
<a href="#">matched pair data</a>		0.35	7	3	99.93	302.4	decrease	NED	.
									
<a href="#">matched pair data</a>		0.72	7	3	99.93	328.4	NED	.	.

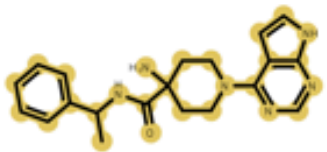
# Looking at the Stats

G	H	I	J	K	L	M	BJ	BQ	BR	BS	BT	BU	BV	BW
Depiction	pair data	ClO gP	HB A	HB D	PSA	RMM	Aq Solubility comb patent data log(M) direction	Aq Solubility comb patent data log(M) n quant	Aq Solubility comb patent data log(M) n qual	Aq Solubility comb patent data log(M) n qual up	Aq Solubility comb patent data log(M) n qual down	Aq Solubility comb patent data log(M) median change	Aq Solubility comb patent data log(M) std dev	Aq Solubility comb patent data log(M) std dev
	<a href="#">matched pair</a>	0.4	7	3	100	302.4	increase							
<chem>C1(CCN(CC1)c2c3cc[nH]c3ncn2)N</chem>	<a href="#">matched pair</a>	0.7	7	3	100	328.4	increase	9	9	9	0	1.701	0.586	

Click the hyperlink  
to see the source  
data

n-qual – number of matched pairs used in analysis  
n-qual-up – number of +ve delta measurements  
n-qual-down – number of -ve delta measurements  
n-median result – middle delta measurement  
Std-dev – standard deviation – give an idea of the spread of data

# Looking at the Stats

G	H	I	J	K	L	M	QD	QK	QL	QM	QN	QQ	QP	QR	QS
Depiction	pair data	Cl gP	HB A	HB D	PSA	RMM	hERG hum inhib pIC50 direction	hERG hum inhib pIC50 n quant	hERG hum inhib pIC50 n qual	hERG hum inhib pIC50 n qual up	hERG hum inhib pIC50 n qual down	hERG hum inhib pIC50 median change	hERG hum inhib pIC50 std dev	hERG hum inhib pIC50 std err	hERG hum inhib pIC50 binom test
 <chem>NC(=O)C2(CCN(CC2)c3c4cc[nH]c4ncn3)N</chem>	<a href="#">matched pair</a>	1.6	7	3	100	364.4	decrease	133	133	19	114	-0.411	0.357	0	0

Looking at this example – Cl to H to decrease hERG binding

- 133 matched pairs
- 114 pairs are –ve so pIC50 is lower for the phenyl
- 19 pairs are +ve so pIC50 is higher for the phenyl
- 0.411 – so half of the pairs are -0.411 or greater improvement
- 0.357 – std dev – small spread – very good chance of 3 fold improvement

## Summary

- Choose a 'reasonable' start point
- Focus your results by iterative runs
- Use Substructure locking

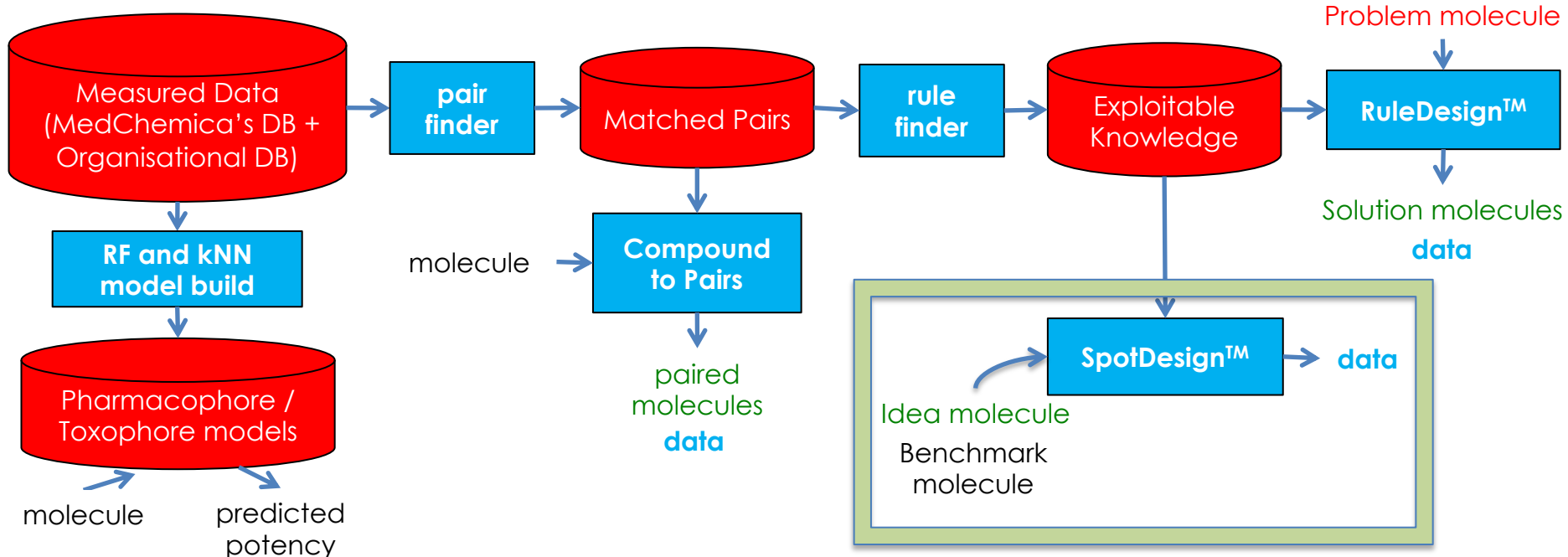
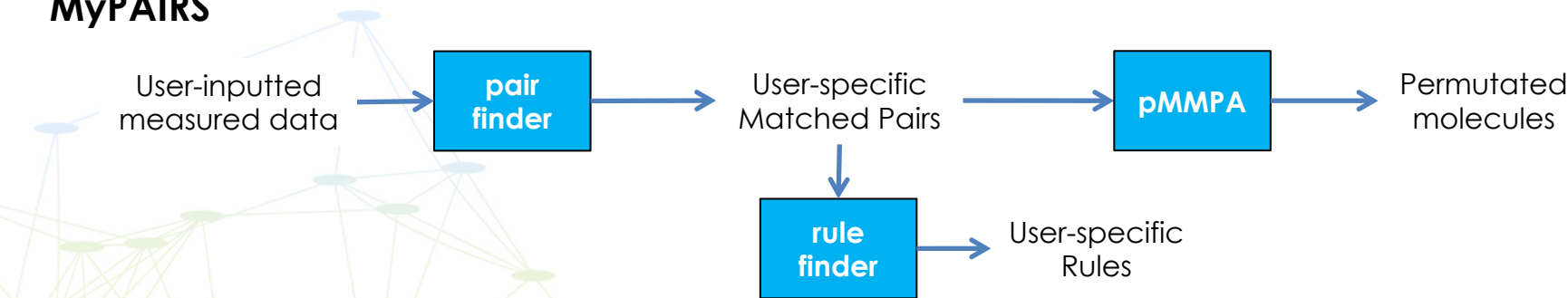
**Remember:**

**Always download your results – server stores nothing!**

## Questions on RuleDesign?

Please use Webinar tools if you are on-line

## Exploiting data derived ADMET Knowledge

**DATABASE TOOLS****MyPAIRS**

## Exercise – Explore ideas to improve solubility

- Input chemical name for lead molecule CHEMBL2325997
  - Choosing the property to improve (setting the Goal)
    - Find Phys\_Prop -> Solubility (Broad Goal)
  - Returning results, exploring results
    - what do the numbers mean?
  - Drilling back to the original source (looking at the matched pairs)
  - Running multiple ideas
- Search – Inspect – Export
  - Save with structures or as .csv for other applications
    - (Structures saved as SMILES)

## Matched Pairs

Access to MCPairs Database Matched Molecular Pairs Service



### RULEDESIGN™

*Formerly Known as Compounds From Rules*

Submit compound(s), enumerate products using Rules from the central database



### SPOTDESIGN™

Submit a reference compound and explore your ideas seeking support from the central database



### COMPOUND TO PAIRS

Search the central database of the matched pairs of a compound



### COMPOUND TO MEASUREMENTS

Search the central database for the current aggregated data of a compound



# SpotDesign : How good is my idea?

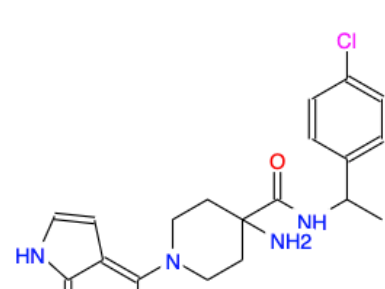
## Entering a reference compound

Endpoint\*:

**Reference Molecule\***

**Compound Name:**

**SMILES\*:**



**Copy as SMILES**  
Copy as MOL  
Copy as MOL V3000  
Copy as InChI  
Copy as InChI key  
Search chemical structure (through InChIKey)  
Copy as Scalar Vector Graphics  
Paste MOL or SDF or SMILES

**Next >>**

Compound name?

SMILES structure

Other methods

# SpotDesign : How good is my idea?

## Selecting a Goal

Endpoint\*:

**Goal\***

Endpoint\*:

solu

Phys\_Props (50316)

- ☒ Solubility (12138)
  - ☒ Aq\_Solubility\_generic\_log(M) (10832)
  - ☒ Aq\_Solubility\_pH\_7.4\_[CHEMBL2362975] (1)
  - ☒ Aq\_Solubility\_pH\_7.4\_[CHEMBL612558] (20)

Apply

Text search

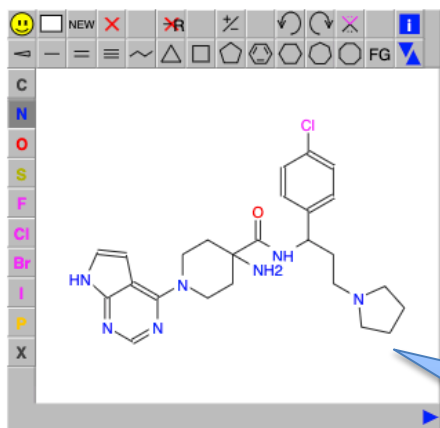
Select Goal

## SpotDesign : How good is my idea?

## Idea Molecule\*

Single Multi

SMILES\*:

NC5(C(=O)NC(CCN1CCCC1)c2ccc(Cl)cc2)CCN(c3ncnc4[nH]ccc34)CC5


Draw your  
idea molecule

	Idea	Delta	
ClogP	2.77	0.51	↑
HBA	8.00	1.00	↑
HBD	3.00	0.00	●
PSA	103.17	3.24	↑
RMM	482.02	83.13	↑

Submit

## Rules

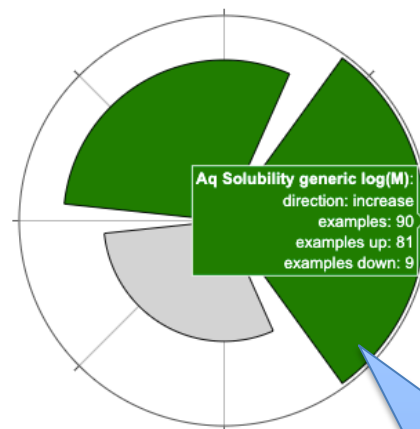
Phys\_Props &gt; Solubility

Endpoints with Rules:

2 / 3

Other Rules:

increase	2
decrease	2
remaining	37
total	41



Tip! Clicking an endpoint

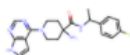
Hover to  
explore results  
Click to drill  
deeper

GO!

## Quick tip...

## Reference Molecule\* and Goal\*

SMILES\*:

CC(c1ccc(cc1)Cl)NC(=O)C2(CCN(CC2)c3c4cc[nH]c4ncn3)N


Endpoint\*:

Solubility ✕

Select

Copy to Rule Design

## Idea Molecule\*

## Rules

Single    Multi

SMILES\*:

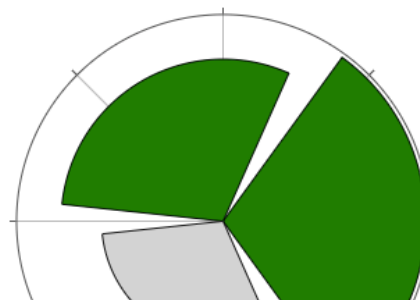
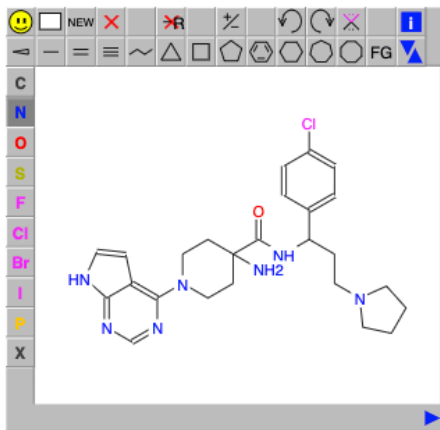
NC5(C(=O)NC(CCN1CCCC1)c2ccc(Cl)cc2)CCN(c3ncnc4[nH]ccc34)CC5

Shortcut to RuleDesign

2 / 3

Other Rules:

increase	2
decrease	2
remaining	37
total	41

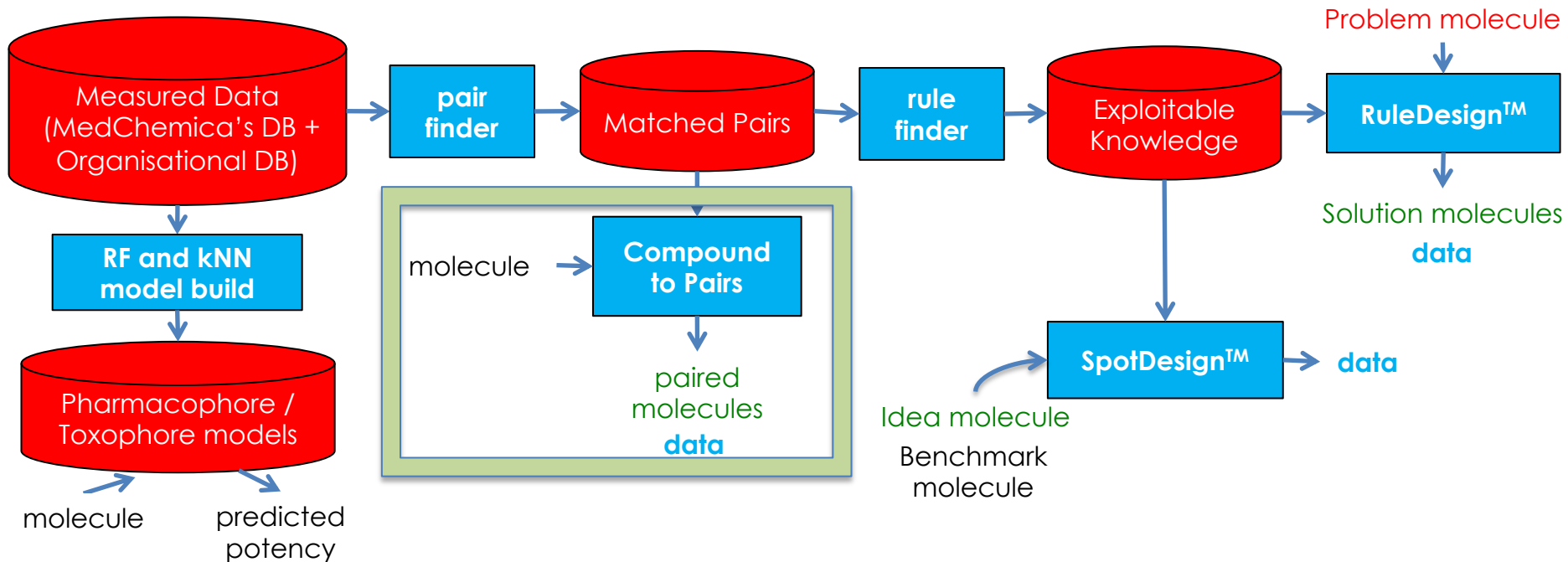


# Questions on SpotDesign™?

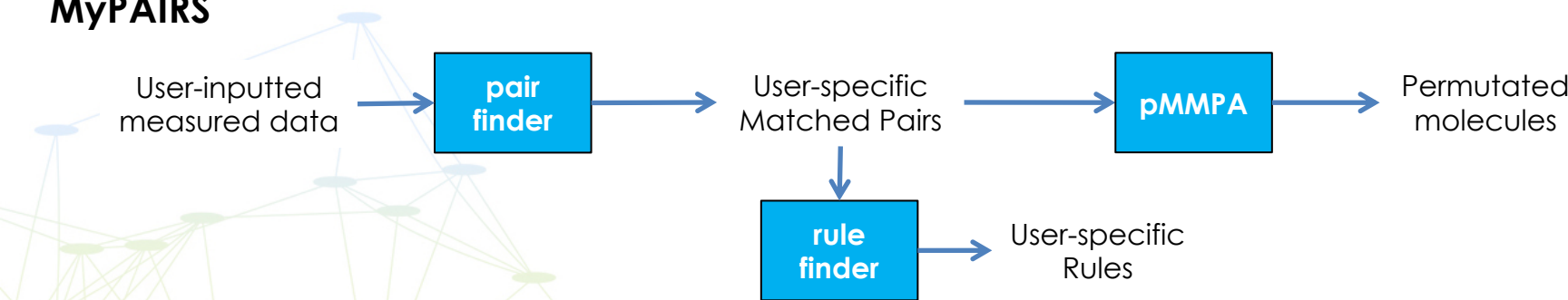
Please use Webinar tools if you are on-line

# Exploiting data derived ADMET Knowledge

## DATABASE TOOLS



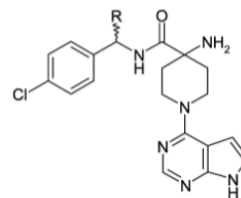
## MyPAIRS



## Fast SAR understanding via Matched pairs

Table 2. Akt Enzyme and Cell Potency, Selectivity, and hERG Activity for  $\alpha$ -Alkyl-Substituted Benzylamide Analogues

Table 3. Akt Enzyme and Cell Potency, Selectivity, and hERG Activity for Selected Aryl-Substituted Benzylamide Analogues

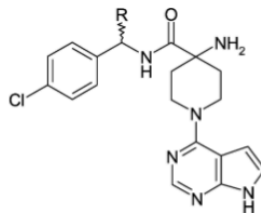
Table 4. Akt Enzyme and Cell Potency, Selectivity, and hERG Activity for  $\alpha$ -Substituted Benzylamide Analogues Carrying a Basic Side Chain

R	IC <sub>50</sub> (nM)							log D	solubility <sup>e</sup> (μM)
	Akt1 <sup>a</sup>	Akt2 <sup>a</sup>	Akt3 <sup>a</sup>	Cell <sup>b</sup>	ROCK2 <sup>c</sup>	hERG <sup>d</sup>			
45	2	14	7	96	31 [16]	>100000			
46	3	15	4	126	71 [25]	>100000	1.4		> 1700
47	4	96	36	209	104 [29]	>100000	1.3		> 1500
48	5	42	15	156	134 [25]	74807	2.7		1400
49	5	30	10	81	112 [24]	>33300			
50	3			110	34 [12]	>33300			2320
51	4	31	18	43	39 [10]	21030			>2470
52	4	13	6	78	34 [10]	29367	2.4		>2090

<sup>a</sup>All IC<sub>50</sub> of phosphatase  
IonWorks

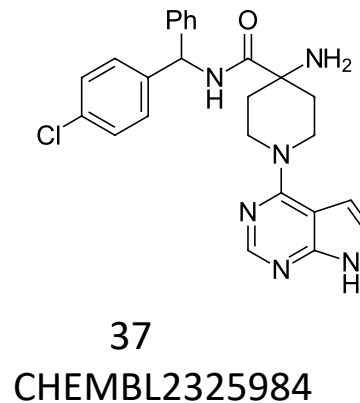
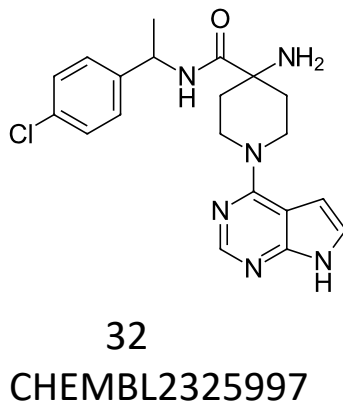
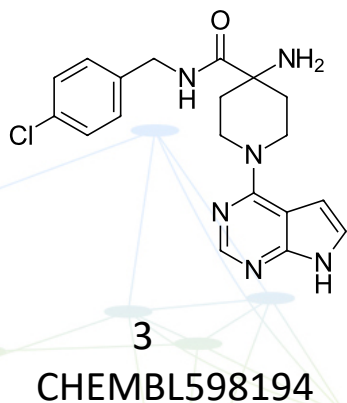
<sup>a</sup>All IC<sub>50</sub> data are the mean of at least 3 experiments mediated by Akt in MDAMB468 cells. <sup>b</sup>Cell potency. <sup>c</sup>ROCK2 potency. <sup>d</sup>hERG activity. <sup>e</sup>Solubility in 0.1 M phosphate buffer.

## SAR worked example: AKT inhibitors

**Table 2. Akt Enzyme and Cell Potency, Selectivity, and hERG Activity for  $\alpha$ -Alkyl-Substituted Benzylamide Analogues**


	R	IC <sub>50</sub> (nM)					log D	solubility <sup>e</sup> ( $\mu$ M)
		Akt1 <sup>a</sup>	Akt2 <sup>a</sup>	Akt3 <sup>a</sup>	cell <sup>b</sup>	ROCK2 <sup>c</sup>		
3	H	13	66	57	328	66 [5]	5235	2.9
32	Me	8	40	30	197	101 [13]	7200	2.7
33	R-Me	276	836	523	4594	1396 [5]	9092	3.3
34	S-Me	4	20	16	134	55 [15]	6747	2.7
35	Et	7	23	15	144	126 [19]	6495	3.5
36	c-Pr	5	30	24	208	261 [52]	2600	3.4
37	Ph	41	210	270	1620	576 [14]	1600	4.1
38	Bn	31	190	150	1650	586 [19]	3500	4.1

<sup>a</sup>All IC<sub>50</sub> data are the mean of at least  $n = 2$  independent measurements. Each has a standard error of measurement (SEM)  $\pm 0.2$  log unit. <sup>b</sup>Inhibition of phosphorylation of GSK3 $\beta$  mediated by Akt in MDAMB468 cells. <sup>c</sup>Value in brackets indicates enzyme selectivity ratio to Akt1. <sup>d</sup>CHO cells, IonWorks assay. <sup>e</sup>Thermodynamic solubility in 0.1 M phosphate buffer at pH 7.4 (25 °C).



## Exercise – Rapid exploration of SAR for literature compounds

- Input chemical name to return matched pairs (ChEMBL598194)
  - Add assay data and explore the SAR tables
  - Add structures, Add ALogp98
  - Change units and sort the data
  - Understand the “gaps” in data
- Search – Inspect – Export
  - Save with structures or as .csv for other applications
    - (Structures saved as SMILES)
- Extras:
  - Finding compounds in ChEMBL and patent literature

## Matched Pairs

Access to MCPairs Database Matched Molecular Pairs Service



### **RULEDESIGN™**

*Formerly Known as Compounds From Rules*

Submit compound(s), enumerate products using Rules from the central database



### **SPOTDESIGN™**

Submit a reference compound and explore your ideas seeking support from the central database



### **COMPOUND TO PAIRS**

Search the central database of the matched pairs of a compound



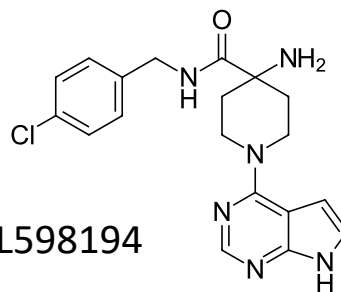
### **COMPOUND TO MEASUREMENTS**

Search the central database for the current aggregated data of a compound



# SAR worked example: AKT inhibitors

3  
CHEMBL598194



MedChemica  
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

MedChemica  
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

### Compound to Pairs

Modules My Account

**Input**

**Compound**

CHEMBL598194

**Select Endpoint**

Choose Endpoint(s)

**Pairs Found:**

**77**

**Pairs with measurements:**

**50**

**Filter by Core**

No Cores

Show ALogP98

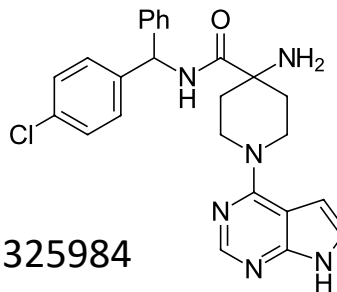
Show Structures

Click to show structures  
in results

	A ?	B ↑↓	
▶	CHEMBL598194	CHEMBL2325988	🔍

# SAR worked example: AKT inhibitors

37  
CHEMBL2325984



MedChemica  
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

MedChemica  
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

Compound to Pairs

Modules My Account

**Input**

**Compound**

CHEMBL2325984

**Select Endpoint**

Choose Endpoint(s)

**Chemical Structure**

**Pairs Found:**

**55**

**Pairs with measurements:**

**55**

**Filter by Core**

No Cores

Show ALogP98 Show Structures

	A ?	B ↑↓	
▶	CHEMBL2325984	CHEMBL2325728	🔍

# SAR worked example

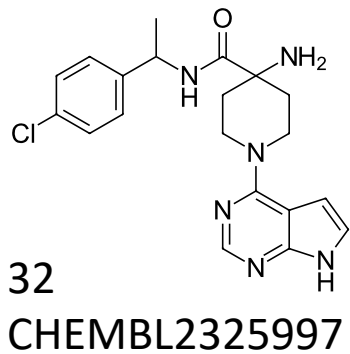


Table 2. Akt Enzyme and Cell Potency, Selectivity, and hERG Activity for  $\alpha$ -Alkyl-Substituted Benzylamide Analogues

	R	IC <sub>50</sub> (nM)					log D	solubility <sup>c</sup> ( $\mu$ M)
		Akt1 <sup>a</sup>	Akt2 <sup>a</sup>	Akt3 <sup>a</sup>	cell <sup>b</sup>	ROCK2 <sup>c</sup>		
3	H	13	66	57	328	66 [5]	2.9	150
32	Me	8	40	30	197	101 [13]	2.7	180

MedChemica  
CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

Compound to Pairs

Modules My Account

Input

Compound

CHEMBL2325997

Select Endpoint

Choose Endpoint(s)

- ☐ Serine/threonine-protein\_kinase\_AKT3\_Homo\_sapien
- ☐ Serine/threonine-protein\_kinase\_AKT2\_Homo\_sapien
- ☐ Rho-associated\_protein\_kinase\_2\_Homo\_sapiens\_pIC
- ☐ Serine/threonine-protein\_kinase\_AKT\_Homo\_sapiens
- ☐ HERG\_Homo\_sapiens\_pIC50\_[CHEMBL240]

Pairs Found: 68

Pairs with measurements: 68

Filter by Core

No Cores

Click to see data

# AKT inhibitors

Input

Compound

CHEMBL2325997

Q

Select Endpoint

3 Endpoint(s) Selected

▼

Pairs Found:

68

Pairs with measurements:

32

Filter by Core

CN(C)C(=O)N1CCCN1c2nc3ccccc3n2

Click to show all structures

Or click to show just one pair

Show ALogP98

Show Structures

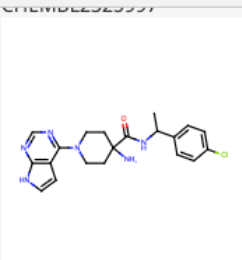
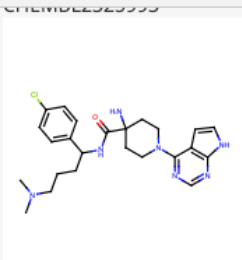

	A ?	B ↑↓		Serine/threonine-protein kinase_AKT...			HERG_Homo_sapiens_pIC50_[CHEMBL...		
				uM			uM		
				B ↑↓	Fold Change ↑↓		B ↑↓	Fold Change ↑↓	
▶	CHEMBL2325997	CHEMBL2325988	Q	4.260	~ 140	↓	33.297	~ 4.6	↓
▶	CHEMBL2325997	CHEMBL2325728	Q	0.018	~ 1.7	↑	21.028	~ 2.9	↓

# Instant SAR table

Click to add lipophilicity data

Click to change units

Click to sort pair with large change to the top

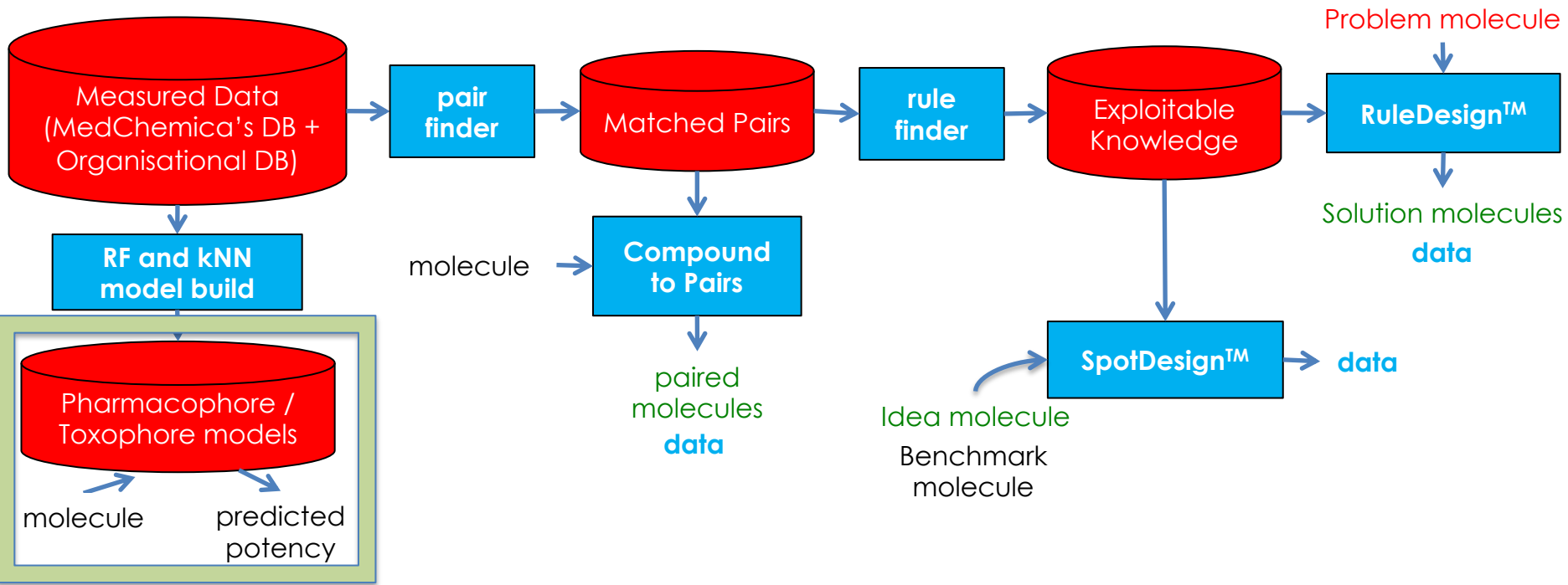
Show ALogP98		Hide Structures					
A ?		B ?		Serine/threonine-protein_kinase ...		HERG_Homo_sapiens_pIC50_[CHEMBL...	
				uM		uM	
				B ?	Fold Change ?	B ?	Fold Change ?
				0.004	~ 7.5 ↑	100.000	~ 14
				0.006	~ 5 ↑	29.370	~ 4.1
				0.007	~ 4.3 ↑	100.000	~ 14

## Questions on Compound-To-Pairs?

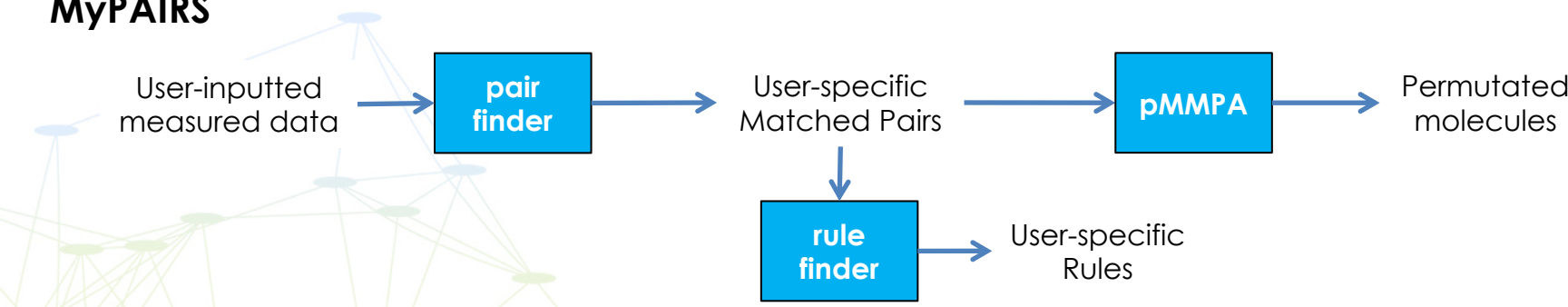
Please use Webinar tools if you are on-line

# Exploiting data derived ADMET Knowledge

## DATABASE TOOLS



## MyPAIRS



## Toxophore and Pharmacophore prediction models

- Models that utilise knowledge from the central database to predict potency
  - Potency at a toxicologically-linked Endpoint = Toxophore
  - Potency at a target Endpoint = Pharmacophore
- 2 types of models are used:
  - Regression Forest (RF)
    - Decision trees trained on central database data
    - Descriptors = pairs of features connected by shortest bond path
  - k Nearest Neighbour (kNN)
    - Compares query molecule with central database molecules
    - Descriptor = Morgan fingerprint
    - Distance metric = Tanimoto distance

## Exercise – Assess potential toxophore activity of CHEMBL2325997

- Input compound name: CHEMBL2325997
- Select Toxicologically-linked Endpoints
- Run Toxophore prediction
- View results
  - Take a minute to understand the predictions using the help information provided (click ?)
- Export as ready-to-use PowerPoint presentation



## Predict

Use pre-built database Machine Learning Models to make predictions



### TOXOPHORES

Predict potential issues, using models from database data, for a compound



### PHARMACOPHORES

Predict properties (potency), using models built from database data, for a compound

## Chemistry Toolbox

A selection of chemistry tools



### MyPAIRS

Take your data, find matched pairs, rules and permute new compounds



### CLEAN, CALC and FILTER

Take your data, clean up the SMILES, calculate properties, and uses SMARTS to search or filter



### DEPICT

A straight forward SMILES / SMARTS checker

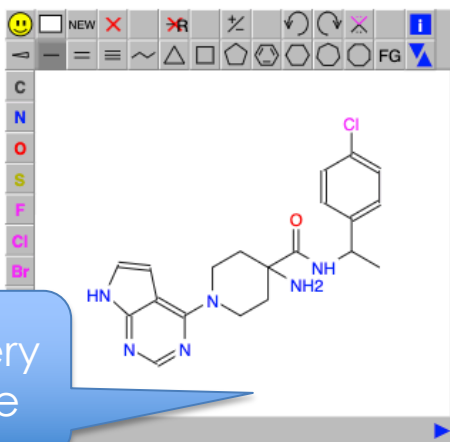
# Toxicity predictions for CHEMBL2325997

**Input Molecule\*****Single** Multi

Compound Name:

CHEMBL2325997

SMILES\*:

CC(c1ccc(cc1)Cl)NC(=O)C2(CCN(CC2)c3c4cc[nH]c4ncn3)N

Input query molecule

Go!

Submit

**Assay Selection****Toxophore Assays**

	Assay	Labels
<input checked="" type="checkbox"/>	Glutamate NMDA receptor; GRIN1/GRIN2B Homo sapiens pIC50 [CHEMBL1907603]	CNS
<input checked="" type="checkbox"/>	Voltage-gated L-type calcium channel alpha-1C subunit Oryctolagus cuniculus pIC50 [CHEMBL2830]	CVS
<input checked="" type="checkbox"/>	Neuronal acetylcholine receptor; alpha4/beta2 Homo sapiens pIC50 [CHEMBL1907589]	CNS
<input checked="" type="checkbox"/>	Serotonin 1b (5-HT1b) receptor Homo sapiens pIC50 [CHEMBL1898]	CNS
<input checked="" type="checkbox"/>	Phosphodiesterase 3A Homo sapiens pIC50 [CHEMBL241]	CVS
<input checked="" type="checkbox"/>	Serotonin 1a (5-HT1a) receptor Homo sapiens pIC50 [CHEMBL214]	CNS
<input checked="" type="checkbox"/>	Dopamine D1 receptor Homo sapiens pIC50 [CHEMBL2056]	CNS
<input checked="" type="checkbox"/>	Muscarinic acetylcholine receptor M2 Homo sapiens pIC50 [CHEMBL211]	CNS
<input checked="" type="checkbox"/>	Sodium channel protein type V alpha	

Select toxicity Endpoint(s)

# Toxicity predictions for CHEMBL2325997

Red = predicted potency at toxic Endpoint

Scroll through results

Green = predicted inactivity at toxic Endpoint

the highest ordered biosystem label assigned.

CNS

Acetylcholinesterase Torpedo electricus pIC50 [CHEMBL4078]	Cannabinoid CB1 receptor Homo sapiens pIC50 [CHEMBL2325997]
Delta opioid receptor Homo sapiens pIC50 [CHEMBL236]	Dopamine D1 receptor Homo sapiens pIC50 [CHEMBL2056]
Dopamine D2 receptor Homo sapiens pIC50 [CHEMBL217]	Dopamine transporter Homo sapiens pIC50 [CHEMBL238]
GABA-A receptor; anion channel Rattus norvegicus pIC50 [CHEMBL1907607]	Glutamate NMDA receptor; GRIN1/GRIN2B Homo sapiens pIC50 [CHEMBL1907603]
Kappa opioid receptor Homo sapiens pIC50 [CHEMBL237]	Monoamine oxidase A Homo sapiens pIC50 [CHEMBL1951]
Mu opioid receptor Homo sapiens pIC50 [CHEMBL233]	Muscarinic acetylcholine receptor M1 Homo sapiens pIC50 [CHEMBL216]
Muscarinic acetylcholine receptor M2 Homo sapiens pIC50 [CHEMBL211]	Neuronal acetylcholine receptor; alpha4/beta2 Homo sapiens pIC50 [CHEMBL1907589]
Norepinephrine transporter Homo sapiens pIC50 [CHEMBL2325997]	Phosphodiesterase 4 Homo sapiens pIC50 [CHEMBL2093863]

Click on Endpoint for breakdown

Tip: click ? for more info

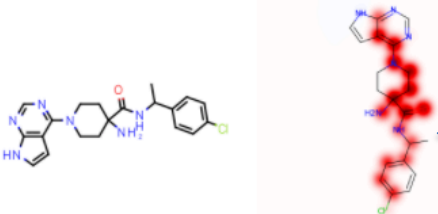
Tip: click to design less toxic compounds

Kappa opioid receptor Homo sapiens pIC50 [CHEMBL237]

CC(c1ccc(cc1)Cl)NC(=O)C2(CCN(CC2)c3c4cc[nH]c4ncn3)N



Unit  
To Rule Design



Overall heat map of atom contributions

RF prediction (based on Cohen's d and R<sup>2</sup>)

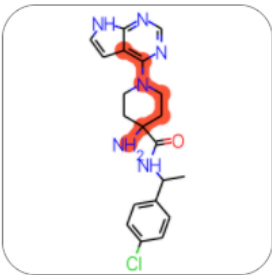
2 Pharmacophore(s)

Active

In domain

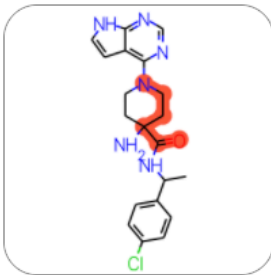
pIC50: 6.229

Range: 5.346 - 7.112



AR\_5atom\_BG

median with [without]: 6.602[5.045]  
n with / total: 459 / 1185



HA\_4atom\_RR

median with [without]: 6.495[5.039]  
n with / total: 515 / 1185

Most important descriptors from RF

kNN prediction (based on activity of NN and Tanimoto distance to NN)

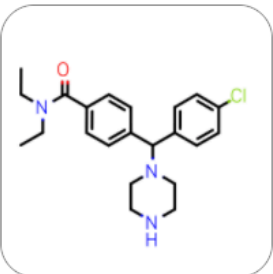
Nearest Neighbour(s)

Active

Out of domain

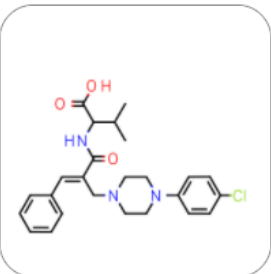
pIC50: 6.114

Range: 5.945 - 6.283



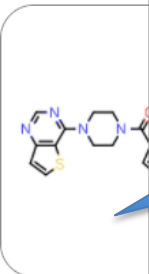
CHEMBL128006

pIC50: 5.000  
Tan. distance: 0.730



CHEMBL4067441

pIC50: 6.032  
Tan. distance: 0.732



CHEMBL3

pIC50:  
Tan. distance:

Nearest neighbours in database

## Predict

Use pre-built database Machine Learning Models to make predictions



### TOXOPHORES

Predict potential issues, using models from database data, for a compound



### PHARMACOPHORES

Predict properties (potency), using models built from database data, for a compound

## Chemistry Toolbox

A selection of chemistry tools



### MyPAIRS

Take your data, find matched pairs, rules and permute new compounds



### CLEAN, CALC and FILTER

Take your data, clean up the SMILES, calculate properties, and uses SMARTS to search or filter



### DEPICT

A straight forward SMILES / SMARTS checker

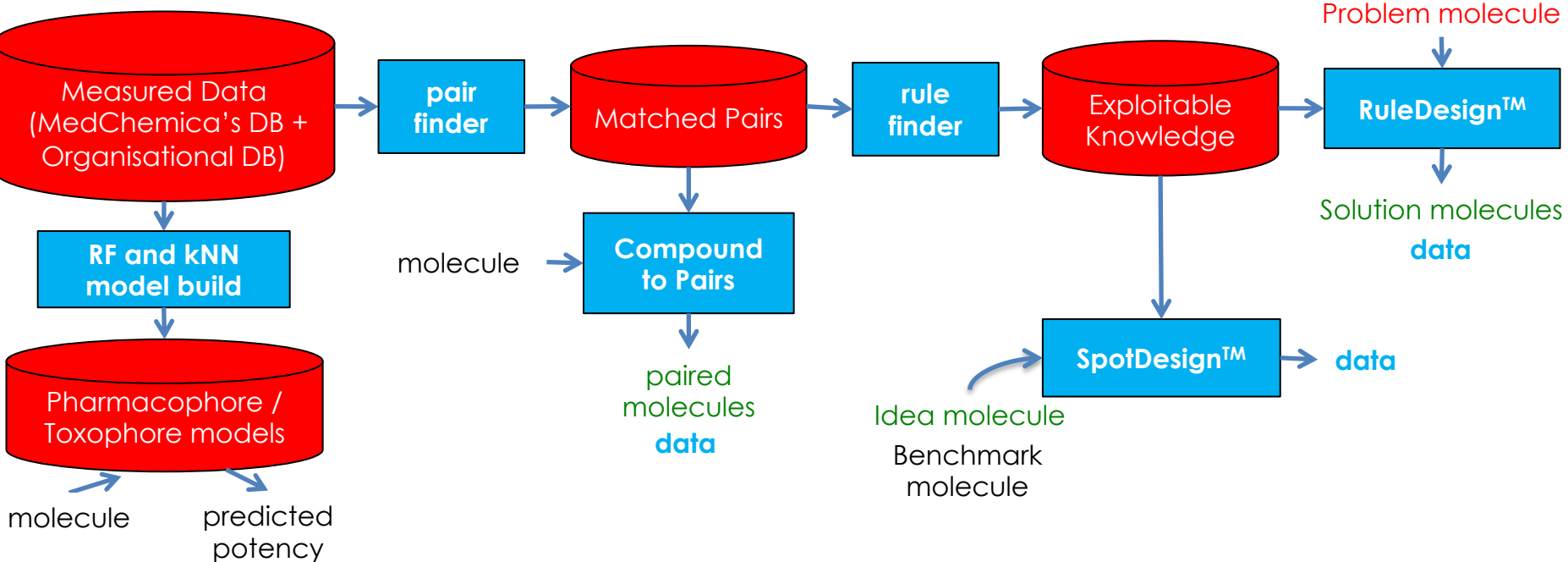
Same methods as Toxophores but  
green = active  
red = inactive

# Questions on Toxophores / Pharmacophores?

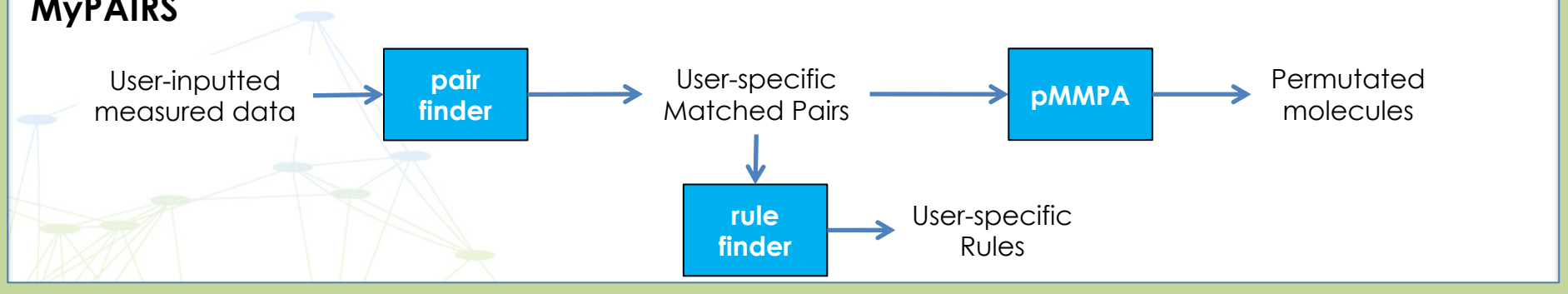
Please use Webinar tools if you are on-line

# Exploiting data derived ADMET Knowledge

## DATABASE TOOLS



## MyPAIRS



These tools do not interact with the central database of knowledge

## Chemistry Toolbox

A selection of chemistry tools



### MyPAIRS

Take your data, find matched pairs, rules and permute new compounds



### CLEAN, CALC and FILTER

Take your data, clean up the SMILES, calculate properties, and uses SMARTS to search or filter



### DEPICT

A straight forward SMILES / SMARTS checker

# MyPairs

- Applies MCPairs methods to user-inputted data
  - Does not interact with central database
- Available methods:
  - Pair Finding
  - Rule Finding
  - Permutative MMPA (molecule suggestions based on application of transformations)
- Example use on a live project:
  - COVID moonshot
    - Open-source antiviral discovery project
    - Targeting SARS-COV-2 main protease
  - Input data = protease binding data dated 14-12-2020
  - Does pMMPA used on historical project data suggest our current lead compound (3-02-2021)?

## 1 Matched Molecular Pairs Finder - MyPairs

Select a file:

Input data file

Current Selection: **FI\_agg\_data\_14\_12\_2020\_23\_26\_47.txt**
**Available file types:** [Compound Data file](#) or [SMILES file \(Direct Pairs only\)](#)

Pair Finding Method ⓘ

FI and MCSS ▾

Choose pair finding method

Environment Size Cut-Off ⓘ

4

- ☐ Ignore Chirality
- ☐ Show Hydrogens
- ☐ Disengage RDKit Calculations

FI Settings

+

MCSS Settings

+

## 2 Calculate Rules

☐ Add Rules Calculated from Matched Pairs

## 3 Permutative Matched Molecular Pair Analysis (pMMPA)

☐ Apply pMMPA

Click here to edit pair finding settings

#

Job ID

Timestamp

Progress

Save

No Jobs Available

Hover over to read  
information on  
each parameter

### FI Settings

Max Batch Size ⓘ

Fragmentation SMARTS Pattern ⓘ

Max Heavy Atoms in Fragment ⓘ

Ratio, by heavy atoms, for Fragmentation ⓘ

☒ Apply fragment size cutoff using ratio to size of whole molecule

☒ Apply largest frag algorithm - selects the best fragmentation pattern (called Filargest on the command line)

Fragments are generated by disconnecting single rotatable bonds (see SMARTS pattern). We require one fragment to be smaller than the other. The smaller fragment is a 'modification' and the larger a 'common core'. Define the maximum size of the modification fragment (8 to 16 Ha is typical).

### MCSS Settings

MCSS Heavy Atom Overlap ⓘ

☐ Rapid method?

Click here to  
enable rule  
finding

### 1 Matched Molecular Pairs Finder - MyPairs

Select a file:

[+ Choose](#)

Current Selection: **FI\_agg\_data\_14\_12\_2020\_23\_26\_47.txt**

Available file types: [Compound Data file](#) or [SMILES file \(Direct Pairs only\)](#)

Pair Finding Method ⓘ

FI and MCSS

Environment Size Cut-Off ⓘ

4

- ☐ Ignore Chirality  
☐ Show Hydrogens  
☐ Disengage RDKit Calculations

FI Settings

+

MCSS Settings

+

### 2 Calculate Rules

☐ Add Rules Calculated from Matched Pairs

### 3 Permutative Matched Molecular Pair Analysis (pMMPA)

☐ Apply pMMPA

Click here to  
enable  
permutation of  
molecules

Submit

#

Job ID

Timestamp

Progress

Save

No Jobs Available

### 1 Matched Molecular Pairs Finder - MyPairs

Select a file:  
[+ Choose](#)

Current Selection: **FI\_agg\_data\_14\_12\_2020\_23\_26\_47.txt**

**Available file types:** [Compound Data file](#) or [SMILES file \(Direct Pairs only\)](#)

Pair Finding Method **i**  
FI and MCSS

Environment Size Cut-Off **i**

☐ Ignore Chirality  
☐ Show Hydrogens  
☐ Disengage RDKit Calculations

**FI Settings**  
**MCSS Settings**

[+ Add Seed Compounds](#)

**Go!**

### 2 Calculate Rules

☒ Add Rules Calculated from Matched Pairs  
Minimum Pairs required for a Rule

### 3 Permutative Matched Molecular Pair Analysis (pMMPA)

☒ Apply pMMPA

Number of Generations

Prediction Threshold **i**

[+ Add Seed Compounds](#)

**Available file types:** [Compound Data file](#)

[+ Add Additional Transformations](#)

**Available file types:** [MCPairs Rules file](#) or [pMMPA Rules input file](#)

[+ Add Comparison Compounds](#)

**Available file types:** [Compound Data file](#) or [SMILES file](#)

**Physical Property Filters**  
*Filter out compounds outside of the selected bounds*

☐ Lipophilicity  
☐ PSA  
☐ HBA  
☐ HBD  
☐ Molecular Weight  
☐ Charge

**Submit**

Hover over for description of each parameter

Default is to apply 1 generation of permutations

Option to add seed molecules (subset of input data you wish transformations to be applied to)

Option to add additional transformations for permutation

Option to filter results by physical property

Input measurement data is on a logarithmic scale e.g. pIC50. A cutoff can be applied to only yield new molecules with a likely prediction above this value.

- ☐ Show Hydrogens  
☐ Disengage RDKit Calculations

FI Settings



MCSS Settings



3.3

☐ minimise

+ Add Seed Compounds

Available file types: Compound Data file

+ Add Additional Transformations

Available file types: MCPairs Rules file or pMMPA Rules input file

+ Add Comparison Compounds

Available file types: Compound Data file or SMILES file**Physical Property Filters**

Filter out compounds outside of the selected bounds

- ☐ Lipophilicity  
☐ PSA  
☐ HBA  
☐ HBD  
☐ Molecular Weight  
☐ Charge

Submit

#	Job ID	Timestamp	Progress	Save
0	BQZIA-U0FBK_2021-02-03_10_39_54_609375	Feb 3, 2021, 10:55:42 AM	Finished	<a href="#">Save</a>

Export zip file  
of results

- task\_message.txt = breakdown of results
- pmmpa\_results\_gen\_1.csv = suggested compounds

Generation 0 =  
seed compounds

Scroll down for  
generation 1  
predictions

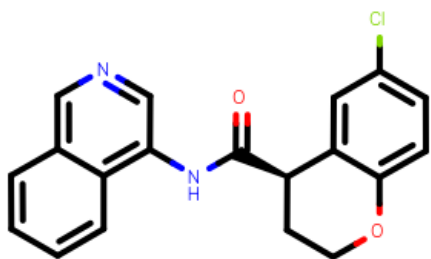
Scroll across for prediction statistics

	A	B	C	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1	id	structure	generation	LogP	PSA	RMM	HBA	HBD	measurement	qualifier	combined_predicted_and_real_measurement	prediction_median	prediction_max	prediction_min	number_of_pairs	median_measurement_delta	environment_size
1393	prod_gen_1_236	[H][C@]1[O]C2c(cc(cc2)~	1	3.74	51.22	352.81	4	1		~	4.85	4.85	6.9	2.8	8	0.15	4
1394	prod_gen_1_237	c1cc2c(cc1f)cnc2NC(=O)~	1	3.39	51.22	356.78	4	1		~	6.9	6.9	6.9	6.6	4	0	4
1395	prod_gen_1_238	c1ccc2c(c1)cnc2NC(=O)~	1	2.51	54.02	303.36	4	2		~	6.6	6.6	6.6	6.6	1	0	4
1396	prod_gen_1_239	[H][C@]1[C]C[O]c2(cc(cc2)~	1	3.79	51.22	352.81	4	1		~	6.15	6.15	8.4	3.3	16	0.35	4
1397	prod_gen_1_240	[H][C@]1[O]c2c(cc(cc2)~	1	3.33	60.45	368.81	5	1		~	4.75	4.75	5.9	3.6	4	-1.45	4
1398	prod_gen_1_241	COc1ccc2cnc(c2c1)N3C~	1	2.46	80.76	409.82	7	1		~	5.9	5.9	5.9	5.8000001	3	0	4

# MyPairs – COVID results

- Use excel to apply project-relevant filters e.g:
  - Sort suggestions by predicted median binding measurement
  - Filter for CLogP < 4
  - And we find our current lead compound as a suggestion:

	A	B	C	I	J	K	L	M	N	O	P	Q	R	T
1	id	structure	generation	precursor	precursor_mcs	pair_dat	CLogP	PSA	RMM	HB <sub>A</sub>	HB <sub>B</sub>	measurem	qualifier	prediction_median
1405	prod_gen_1_3930	Cc1cnccc1NC(=O)CN2CCN(CC2)C(=O)CCl	1	Cc1cnccc1NC(=O)CN2CCN(CC2)C(=O)C	Cc1cnccc1NC(=O)CN2CCN(CC2)		-0.43	65.54	310.78	6	1		~	7.5
1409	prod_gen_1_284	[H][C@H]1[C@H](c2c(ccc(c2C)Cl)OC1)C(=O)Nc3c4c(cnc3)cccc4)[H	1	C[C@H]1COc2ccc(cc2[C@H]1C(=O)Nc3cncc4c3cccc4)Cl	[H][C@H]1COc2ccc(cc2[C@H]1C(=O)Nc3cncc4c3cccc4)Cl		3.55	51.22	352.81	4	1		~	7.3
1432	prod_gen_1_3667	Cc1cccc1NC(=O)CC2C(=O)N(C(=O)S2)c3cccc(c3)C(F)(F)F	1	Cc1cccc1NC(=O)CC2C(=O)N(C(=O)S2)c3cccc(c3)C(F)(F)F	Cc1cccc1NC(=O)CC2C(=O)N(C(=O)S2)c3cccc(c3)C(F)(F)F		3.89	66.48	408.39	5	1		~	7.2
1435	prod_gen_1_1755	c1ccc2c(c1)cncc2NC(=O)[C@H]3CCNc4c3cc(cc4)Cl	1	c1ccc2c(c1)cncc2NC(=O)[C@H]3CCOc4c3cc(cc4)Cl	CC[C@H](c1cccc(c1)Cl)C(=O)Nc2cnc		3.12	54.02	337.8	4	2		~	7.1
1436	prod_gen_1_1791	c1ccc2c(c1)cncc2NC(=O)C3CCNc4c3cc(cc4)Br	1	c1ccc2c(c1)cncc2NC(=O)C3CCOc4c3cc(cc4)Br	CCC(c1cccc(c1)Br)C(=O)Nc2cnc		3.31	54.02	382.25	4	2		~	7.1
1438	prod_gen_1_2384	c1cc(c2c(c1)c(c[nH]2)CCNC(=O)CC)CC3CCN(CC3)C(=O)CCl	1	CC(=O)NCCc1c[nH]c2c1cccc2CN3CCN(CC3)C(=O)CCl	CC(=O)NCCc1c[nH]c2c1cccc2CN3CCN(CC3)C(=O)CCl		1.34	68.44	411.33	6	2		~	7.1
1439	prod_gen_1_2717	c1ccc2c(c1)cncc2NC(=O)[C@H]3CCS(=O)(=O)c4c3cccc4	1	c1ccc2c(c1)cncc2NC(=O)[C@H]3CCOc4c3cc(cc4)Cl	CC[C@H](c1cccc(c1)Cl)C(=O)Nc2cnc		1.89	76.13	352.41	5	1		~	7.1

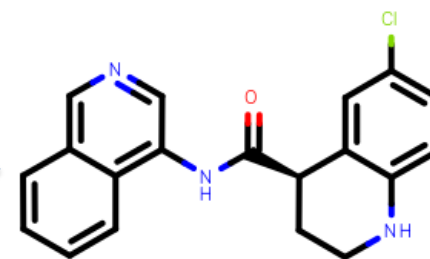


Lead Compound 14-12-2020

pIC<sub>50</sub> = 6.7

Moderate antiviral cell activity

High microsomal clearance



Lead Compound 3-2-2021

pIC<sub>50</sub> = 6.6

Good antiviral cell activity

Moderate microsomal clearance

- Project benefit:

- Alternative lead compound found with better antiviral and ADMET properties

# MyPairs – COVID results

- pmmpa\_evidence\_gen\_1.csv
  - Provides matched pair evidence for suggestions

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	product_id	product_structure	precursor_id	precursor_structure	idsmirks	smirks	environment	compound_name_A	compound_structure_A	measurement_A	qualifier_A	compound_name_B	compound_structure_B	measurement_B	qualifier_B
2	prod_gen_1_236	[H][C@]1(Oc2c(cc(c EDJ-MED-e4bC C[C@H]1C[C@H](c2c MAT-POS [H][C:1]1([						4 MAT-POS-8a69d52e-3 C[C@H]1C[C@H](c2cc(		4.3 =		CHEMBL94657	C[C@]1(CCOc2c1cc(cc	4.7 =	
3	prod_gen_1_236	[H][C@]1(Oc2c(cc(c EDJ-MED-e4bC C[C@H]1C[C@H](c2c MAT-POS [H][C:1]1([						4 MAT-POS-8a69d52e-3 C[C@H]1C[C@H](c2cc(		4.3 =		CHEMBL2182047	C[C@]1(CCOc2c1cc(cc2	6.5 =	
4	prod_gen_1_236	[H][C@]1(Oc2c(cc(c EDJ-MED-e4bC C[C@H]1C[C@H](c2c MAT-POS [H][C:1]1([						4 MAT-POS-8a69d52e-2 C[C@H]1C[C@H](c2c		6.2 =		CHEMBL94657	C[C@]1(CCOc2c1cc(cc	4.7 =	
5	prod_gen_1_236	[H][C@]1(Oc2c(cc(c EDJ-MED-e4bC C[C@H]1C[C@H](c2c MAT-POS [H][C:1]1([						4 MAT-POS-8a69d52e-2 C[C@H]1C[C@H](c2c		6.2 =		CHEMBL2182047	C[C@]1(CCOc2c1cc(cc2	6.5 =	
6	prod_gen_1_236	[H][C@]1(Oc2c(cc(c EDJ-MED-e4bC C[C@H]1C[C@H](c2c MAT-POS [H][C:1]1([						4 EDJ-MED-e4b030d8-3 C[C@H]1C[C@H](c2cc(cc		4.7 =		CHEMBL94657	C[C@]1(CCOc2c1cc(cc	4.7 =	
7	prod_gen_1_236	[H][C@]1(Oc2c(cc(c EDJ-MED-e4bC C[C@H]1C[C@H](c2c MAT-POS [H][C:1]1([						4 EDJ-MED-e4b030d8-3 C[C@H]1C[C@H](c2cc(cc		4.7 =		CHEMBL2182047	C[C@]1(CCOc2c1cc(cc2	6.5 =	
8	prod_gen_1_236	[H][C@]1(Oc2c(cc(c EDJ-MED-e4bC C[C@H]1C[C@H](c2c MAT-POS [H][C:1]1([						4 EDJ-MED-e4b030d8-2 C[C@H]1C[C@H](c2cc(cc		6.6 =		CHEMBL94657	C[C@]1(CCOc2c1cc(cc	4.7 =	
9	prod_gen_1_236	[H][C@]1(Oc2c(cc(c EDJ-MED-e4bC C[C@H]1C[C@H](c2c MAT-POS [H][C:1]1([						4 EDJ-MED-e4b030d8-2 C[C@H]1C[C@H](c2cc(cc		6.6 =		CHEMBL2182047	C[C@]1(CCOc2c1cc(cc2	6.5 =	
10	prod_gen_1_237	c1cc2c(cc1F)cncc2N MAT-POS-b3e c1ccc2c(c1)cncc2NC( MAT-POS [c:1][c:2]1						4 MAT-POS-0c8fa4a7-3 c1ccc2c(c1)c(nn2C(=O)C		4 <		CHEMBL259181	c1cc(cc(c1)C)CC(=O)n2c	4 <	
11	prod_gen_1_237	c1cc2c(cc1F)cncc2N MAT-POS-b3e c1ccc2c(c1)cncc2NC( MAT-POS [c:1][c:2]1						4 EDJ-MED-c8e7a002-1 c1ccc2c(c1)c(n[nH]2)NC(		4 <		CHEMBL1667924	c1cc(cc(c1)C)CC(=O)Nc	4 <	
12	prod_gen_1_237	c1cc2c(cc1F)cncc2N MAT-POS-b3e c1ccc2c(c1)cncc2NC( MAT-POS [c:1][c:2]1						4 VLA-UCB-1dbca3b4-1 c1ccc2c(c1)cncc2NC(=O)N		6.4 =		CHEMBL3407863	c1cc2cncc(c2cc1F)NC(=O	6.1 =	
13	prod_gen_1_237	c1cc2c(cc1F)cncc2N MAT-POS-b3e c1ccc2c(c1)cncc2NC( MAT-POS [c:1][c:2]1						4 BEN-DND-7e92b6ca-7 c1ccc(cc1)S(=O)(=O)N2C		4 <		CHEMBL48412	c1ccc(cc1)S(=O)(=O)N2c	4 <	
14	prod_gen_1_238	c1ccc2c(c1)cncc2NC ALP-POS-477d c1ccc2c(c1)cncc2NC( JAG-UCB- [C:1]([H])(f						4 JAG-UCB-52b62a6f-11 Cc1ccn2cnnc2c1C(=O)C		4 <		CHEMBL1667898	Cc1ccn2cnnc2c1NC(=O)N	4 <	
15	prod_gen_1_239	[H][C@]1(COc2c(cc MAT-POS-8a6 C[C@H]1COC2cccc( MAT-POS [H][C:1]1([						4 MAT-POS-8a69d52e-5 C[C@H]1COC2cccc( MAT-POS [H][C:1]1([		4 <		CHEMBL133016	C[C@H]1C[C@H](c2cc	6.6 =	

Matched pair structures and data listed for each suggestion

## Questions on MyPairs?

Please use Webinar tools if you are on-line

# And Finally...

- Problems, suggestions, ideas for interface / calculation engine
  - [contact@medchemica.com](mailto:contact@medchemica.com)