

## MCPairs Online 'Tips And Tricks' quick guide

Introduction	2
About the MCPairs Online Database	2
Example medicinal chemistry	4
Compound-To-Measurements	5
What are all the Synonyms for a compound?	5
Searching for Assays (Endpoints) and Goal nicknames	6
Searching for Assays (Endpoints)	6
Goal searching	7
Pair Searching	9
Viewing Structures, Changing Units, understanding 'Fold Change' and Delta AlogP98	9
Changing Units and Fold Changes	9
Viewing structures	10
Find the most efficient changes	11
What do all the ? marks mean?	11
Finding matched pairs from sub-structures starting with ChEMBL	12
Finding matched pairs from publications	13
Workflow from patents - Searching sureChembl by substructure	15
Compounds-From-Rules	17
Setting the Substructure Lock	17
Charge Filter (Advanced filtering)	20
I do not seem to have any results?	21
What are Ugliers, phys-props filters and the such like?	21
SpotDesign™	26
Workflows to explore your ideas	26
What if I don't get any results?	27

## Introduction

MCPairs Online was designed to be simple and involve as few inputs and clicks as possible to obtain results and export them. For many cases this works well and chemists can explore the results via the medium of Excel or another visualization tool. Our basic on-line training will have covered the basics. This 'Tips and Tricks' guide covers more advanced features and methods to get more out of MCPairs. The aim is to be rich in screen shots and few in words.

## About the MCPairs Online Database

The compounds, measurements and knowledge are derived from public and proprietary sources. There are currently nearly 2 million structures and identifiers, representing 1.73 million unique compounds. For the most part these compounds have some measurement in a biological, ADMET or physical property assay. MCPairs does not aim to be comprehensive across known chemical space, but aims to be comprehensive over 'measured' chemical space. If you cannot find a compound in our database, let us know and we will see if we can find it with some additional quality knowledge.

Goal / Endpoint	Number of Rules
LogD TM	41557
PPB rat log(proportion Free)	9626
PPB dog log(proportion Free)	2294
PPB hum log(proportion Free)	25324
PPB mouse log(proportion Free)	3624
PPB cyno log(proportion Free)	306
MDCK dog perm log(ER) dog	5300
Hep CI rat mL.min-1.10-6cells	20122
Hep CI hum mL.min-1.10-6cells	7708
Hep CI mouse mL.min-1.10-6cells	3252
Hep CI dog mL.min-1.10-6cells	3448
Hep CI cyno mL.min-1.10-6cells	3508
Mic CI hum uL.min-1.mg-1	30925
Mic CI mouse uL.min-1.mg-1	16602
Mic CI rat uL.min-1.mg-1	21246
Mic CI dog uL.min-1.mg-1	18538
Mic CI cyno uL.min-1.mg-1	19074
CYP inhibs 1A2 pIC50 hum	7380
CYP inhibs 2C9 pIC50 hum	12922
CYP inhibs 2C19 pIC50 hum	6638
CYP inhibs 2D6 pIC50 hum	10202
CYP inhibs 3A4 pIC50 hum	14997
hERG hum inhib pIC50	14022
Nav1.5 hum inhib pIC50	3248
SOL generic log(M)	33185

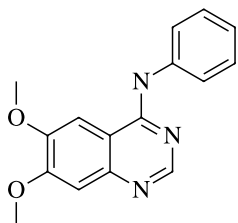
Current number of ADMET rules per assay type

Protein Class or assay type	Number of Targets
Anti-target_DDI	131
Bacterial activity	24
Cell lines	215
Dehydrogenase	43
Epigenetic_reg	143
GPCR	1195
Ion channels	416
Kinase	1341
Phosphatase	629
Protease	464
Transporter	118
Transcription factor nuclear Receptor	195
Virus activity	132
Other	1313

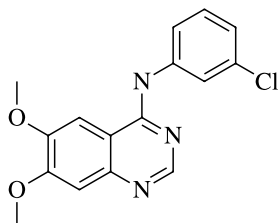
Current numbers of protein targets per class

## Example medicinal chemistry

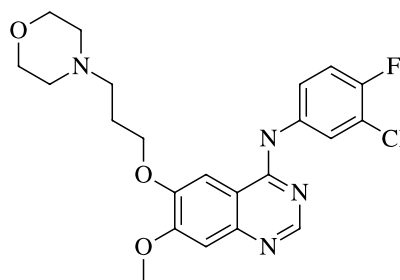
Our on-line training, and this document, use the same example from EGFR kinase inhibitor research. The compound CHEMBL94191 was optimized, on potency, to CHEMBL7917, which was a good inhibitor of EGFR but had poor solubility and sub-optimal stability. The compound Gefitinib (Iressa) was the resultant compound that went into the clinic and was approved.



CHEMBL94191

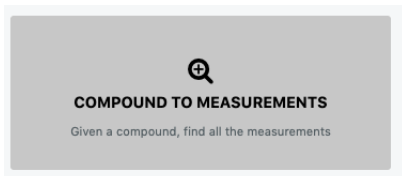


CHEMBL7917



Iressa Gefitinib CHEMBL939 BDBM5447 ZD-1839

## Compound-To-Measurements



### What are all the Synonyms for a compound?

Use Compound-To-Measurements to find all the Synonyms for a compound name. Click Compound-To-Measurement and enter your compound ID. Click the endpoints list, select an Endpoint, and all of the synonyms appear.

**Endpoints:**

Tyrosine-protein\_kinase\_HCK\_Homo\_sapiens\_

**iressa**

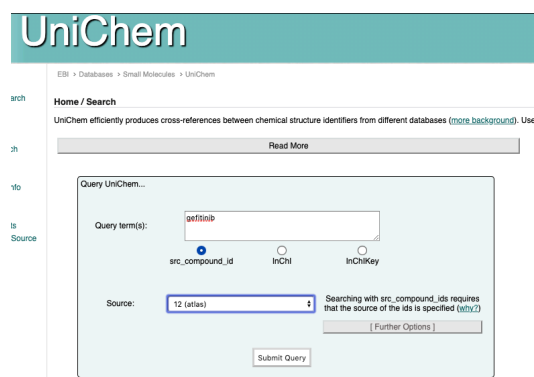
**Adducts:**

Compound Name		Compound Structure
CHEMBL939	<input type="button" value="Q"/>	<chem>COc1cc2c(cc1OCCCN3CCOCC3)c(ncn2)Nc4ccc(c(c4)C)F</chem>
BDBM5447	<input type="button" value="Q"/>	<chem>COc1cc2c(cc1OCCCN3CCOCC3)c(ncn2)Nc4ccc(c(c4)C)F</chem>
ZD-1839	<input type="button" value="Q"/>	<chem>COc1cc2c(cc1OCCCN3CCOCC3)c(ncn2)Nc4ccc(c(c4)C)F</chem>
Iressa	<input type="button" value="Q"/>	<chem>COc1cc2c(cc1OCCCN3CCOCC3)c(ncn2)Nc4ccc(c(c4)C)F</chem>
Gefitinib	<input type="button" value="Q"/>	<chem>COc1cc2c(cc1OCCCN3CCOCC3)c(ncn2)Nc4ccc(c(c4)C)F</chem>

Notice the ZD-1839 - hyphens can be an issue when searching for a compound.

If you think we should have a compound identifier, try looking in ChEMBL's UniChem search engine on-line and let us know what you find. It might be possible that the compound you want has a ChEMBL or BDBM identifier?

UniChem link : <https://www.ebi.ac.uk/unichem/>



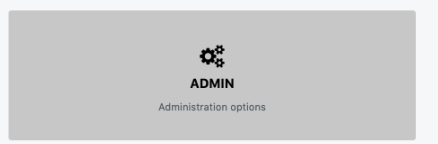
There is a drop down for different database types to search - a hit looks like the screen below. We can now use a ChEMBL ID or BDBM number in MCPairs

## Query Results...

Search terms and Sorted-by columns are highlighted. [Default sort lists EBI sources first]  
 For clarity, Standard InChIs are omitted, but may be [toggled on/off](#).  
 Use the drop downs on the table footer to filter by individual columns.

src_id	Source Name	src_compound_id	Currently Assigned	LR *	UCI **	Standard InChI
1	chembl	CHEMBL939	Yes		228021	XGALLCVXEZP†
3	pdbe	IRE	Yes		228021	XGALLCVXEZP†
7	chebi	49668	Yes		228021	XGALLCVXEZP†
12	atlas	gefitinib	Yes		228021	XGALLCVXEZP†
15	surechembl	SCHEMBL7866	Yes		228021	XGALLCVXEZP†
2	drugbank	DB00317	Yes		228021	XGALLCVXEZP†
2	drugbank	DB07998	No	28-MAY-2011	228021	XGALLCVXEZP†
4	qtodb	4941	Yes		228021	XGALLCVXEZP†

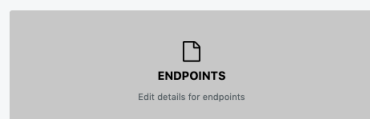
## Searching for Assays (Endpoints) and Goal nicknames



## Searching for Assays (Endpoints)

Start by clicking the Admin tab - most users of MCPairs will not be able to edit any information under this tab, but we can search it. Usually this will be to find an endpoint for a specific protein target, the exactly assay conditions and importantly cross reference the full name to the 'nickname' in the Goal menu.

Click ENDPOINTS to search



Enter 'bruton' (case insensitive) in the search box and press return.

Endpoints

Combined

Goals

bruton

Tyrosine-protein\_kinase\_BTK\_Homo\_sapiens\_pIC50\_[303.3]

Description: This BTK competition assay measures compound potency (IC50) for the inactivated state of Brutons Tyrosine Kinase using FRET (Förster/Fluorescence Resonance Energy Transfer) technology. The BTK-Eu complex was incubated on ice one hour prior to use at a starting concentration of 50 nM BTK-Bioscience 10 nM Eu-streptavidin (Perkin-Elmer Catalog#AD0062). The assay buffer consisted of 20 mM HEPES (pH 7.15) 0.1 mM DTT 10 mM MgCl2 0.5 mg/ml BSA with 3% Kinase Stabilizer (Fremont Biosolutions Catalog #STB-K02). After 1 h the reaction mixture from above was diluted 10 fold in assay buffer to make 5 nM BTK: 1 nM Eu-Streptavidin complex (donor fluorophore). 18 µl of a mixture of 0.11 nM BTK-Eu and 0.11 nM Kinase Tracer 178 (Invitrogen Catalog #PV5593) with BTK-Eu alone as no negative control was then dispensed into 384-well flat bottom plates (Greiner 784078). Compounds to be tested in assay were prepared as 10x concentrations and serial dilution in half-log increments was performed in DMSO so as to generate 10 point curves. To initiate the FRET reaction compounds prepared as 10x stock in DMSO was added to the plates and the plates were incubated 18-24 h at 140C. After the incubation the plates were read on a BMG Pherastar Fluorescent plate reader (or equivalent) and used to measure the emission energy from the europium donor fluorophore (620 nm emission) and the FRET (665 nm emission).

Code: 303.3

Unit: pIC50

Species: Homo sapiens

Route: in\_vitro

Minimum pairs for rule:

Creator: Roche

Owner: contact@medchemica.com

Delete

Modify

Add To Combined

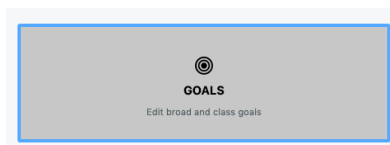
Clone Endpoint

We get all of the Bruton Tyrosine kinase assays registered into MCPairs. Again, most blue buttons are for editing or adding more endpoints, so will not work for MCPairs Online users.

**TopTip** - The search automatically searches any text in the Endpoint information - name, description, species and so on.

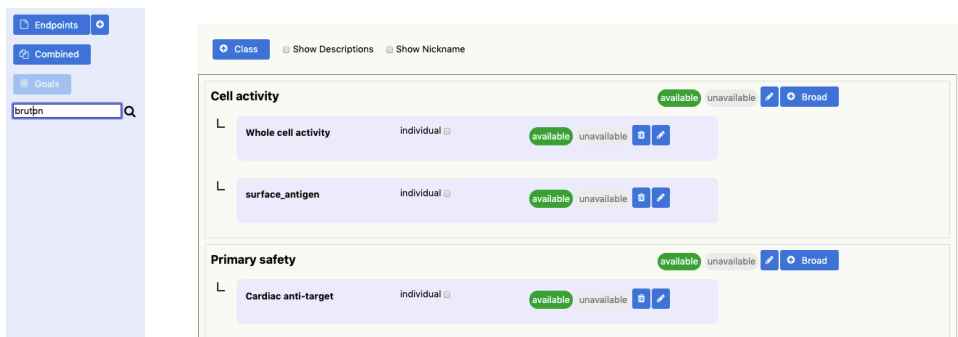
## Goal searching

Click Home -> Admin -> Goal



It may take a second or two to load all the Goals - we have a lot!

Enter 'bruton' in the search box and press return



What are we looking at?

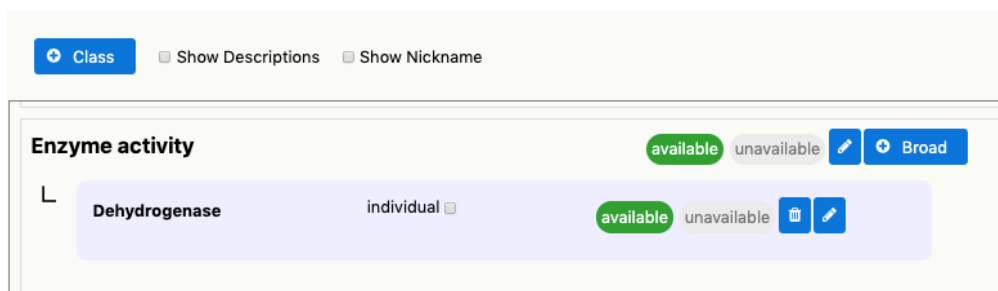
Goals are grouped via a hierarchy

Class -> Broad -> Individual

The first view is Class -> Broad only - there are too many Individuals to show.

An Individual Goal maps 1:1 with an individual assay.

Scroll down to 'Enzyme activity'.

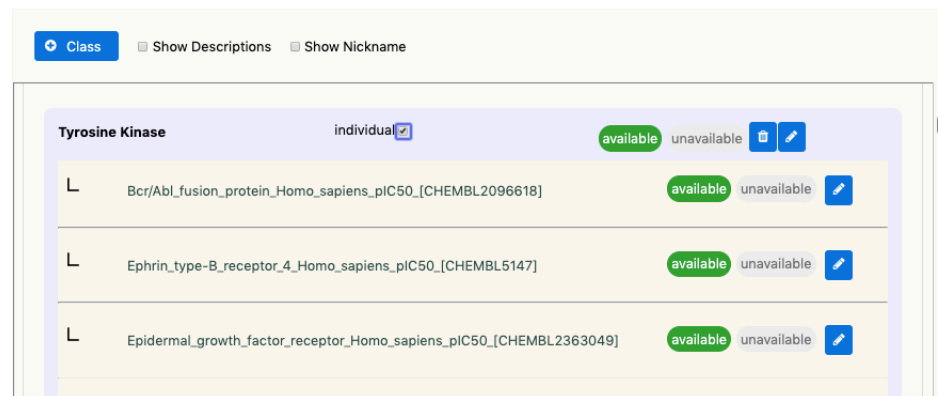


...and find 'Tyrosine Kinase'



Our search looked for 'bruton' in any part of the Goal information and has narrowed the list down. It was quite a large set.

Notice the individual tick box - click that - now all the Individual / Goals assays are displayed.

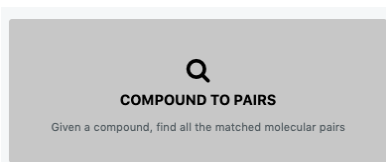


**TopTip** - Clicking the Show Descriptors and Show Nicknames gives a snippet of the Goal description and the Nickname. The Nicknames usually have the protein target acronyms e.g. EGFR. Several acronyms can exist for the same protein. The above is a workflow to find out which one MCPairs uses. Nicknames are used in the Goal menus in SpotDesign™ and Compounds-From-Rule.

For further details, click the Pencil Icon - you will not be able to make changes but you can see the detail.



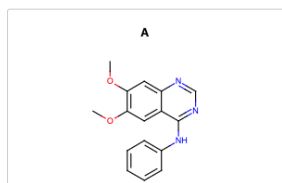
## Pair Searching



Basic searching for matched pairs from a compound identifier is covered in our on-line training. Here are some extra tips and workflows

**TopTip:** The identifier search is not case sensitive (e.g. chembl, ChEMBL, ChEmBl, bdbm all work)

Compound A

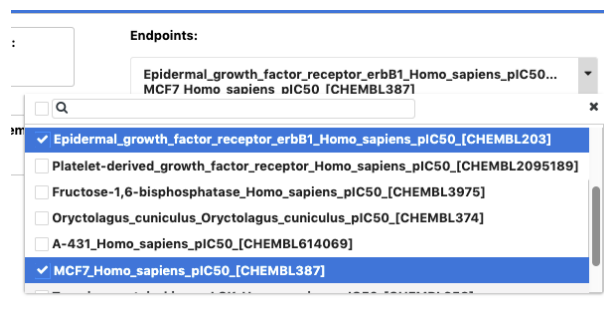
## Viewing Structures, Changing Units, understanding 'Fold Change' and Delta AlogP98

We have revamped Compound-To-Pairs and added some new features from v1.5 onwards

### Changing Units and Fold Changes

Enter a search compound (CHEMBL7917), check the structure and click the Green button. There are 203 matched pairs; a second box indicated there are 140 pairs where there is a measurement for both compounds in an assay. So 63 matched pairs where there is one or both measurements missing - there could be unknown SAR?

Select EGFR (CHEMBL203) and MCF7 cell line (CHEMBL387) from the Endpoints list

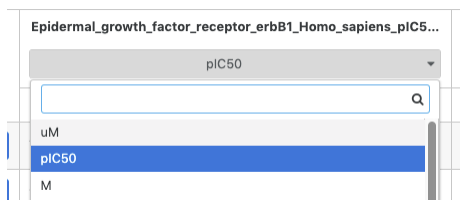


Columns of data appear, and by default the data is on a linear scale [pIC50, pKi, log10(solubility Molar)].

We do not show the data for A initially, but this can be toggled On/Off by clicking the blue button:



Change units with the drop down from each column



One of the columns changes:

- With linear units B-A is shown ( $\text{pIC50}_B - \text{pIC50}_A$ ) - so say 0.5 log unit change (3 fold)
- With nonlinear units the column changes to 'Fold Change' ( $\text{pIC50}_A / \text{pIC50}_B$ ) is shown and a helper arrow

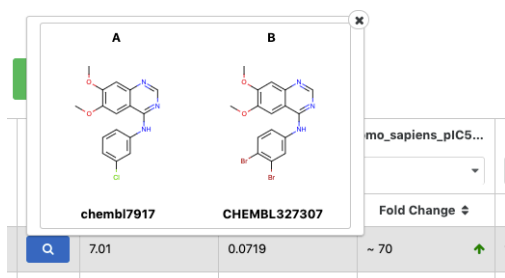
uM		
B	Fold Change	
0.000	~ 70	↑
0.000	~ 35	↑

**TopTip:** Notice the  $\mu\text{M}$  potency of B reads 0.000 as numbers are rounded to 3 decimal places. Change the units to nM, click the A data back on, and look at the data again.

nM		
A	B	Fold Change
7.01	0.0719	~ 70 ↑
7.01	0.250	~ 35 ↑

## Viewing structures

Click on any blue magnifying glass icon to see a matched pair.



Better still, try

Show Structures

With all the structures in the table, the data can be sorted by any column, so enabling rapid SAR analysis.

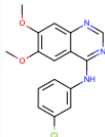
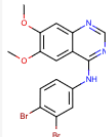

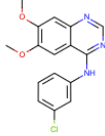
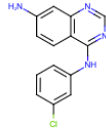

Find the most efficient changes

**TopTip:** Let's just switch the EGFR data back to a linear scale (pIC50) and click the Show AlogP98 button

Show  
AlogP98

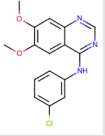
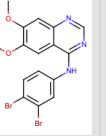

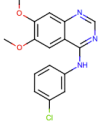
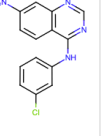

The AlogP98 column is added to the right of the pairs, next to EGFR data.  
The column B-A or Delta AlogP98 are close together.

If we click, and sort on EGFR B data, the most potent is at the top, then we can compare the increase in potency against structural change AND the change in AlogP98.

A ?	B ↕		AlogP98			Epidermal_growth_factor_receptor_erbB1_Homo...		
			B ↕	A ↕	B-A ↕	A ↕	B ↕	B-A ↕
CHEMBL7917 	CHEMBL327307 		4.916	4.044	0.872	8.154	10.143	1.989
CHEMBL7917 	CHEMBL328955 		3.609	4.044	-0.435	8.154	9.602	1.448

Consider the pairs above: 2 bromines compared to a chlorine give a big increase in potency, 2 log units 100 fold, (assuming we believe the assay - let's do that) for only 0.87 delta AlogP98. The second pairs we have 1.5 log units (30 fold) increase but a decrease in lipophilicity of -0.44 - interesting stuff.

What do all the ? marks mean?

A ?	B ↕		Epidermal_growth_factor_receptor_erbB1_Homo_sapiens_pIC50...			MCF7_Homo_sapiens_pIC50_[CHEMBL387]		
			A ↕	B ↕	B-A ↕	A ↕	B ↕	B-A ↕
CHEMBL7917 	CHEMBL327307 		8.154	10.143	1.989	?	?	?
CHEMBL7917 	CHEMBL328955 		8.154	9.602	1.448	?	?	?

**TopTip:** This where MCPairs has found matched pairs but the data is 'missing' for both of the compounds in an endpoint selected. Given these data are from the literature we have plenty of '?' marks - not everything is published, and a matched pair from one study at one institute will not have been tested in another. Those with Enterprise MCPairs may have a lot more data and the opportunity to 'fill in the gaps' with more testing. In any case these compounds have been synthesized and could be obtained from the original source (try the ChEMBL ids in a search from ChEMBL website).

## Exporting the matched pairs

**TopTip:** Exporting the Excel sheet has all of the pairs and all of the datasets, and includes all of the gaps. Sorting this sheet can yield a list of compounds to test and pull in the missing data. A second table has the data with structures, which is easier to work with for plotting data. A text file export maybe easier to import into another tool (DataWarrior, SpotFire, JMP).

## Finding matched pairs from sub-structures starting with ChEMBL

What if we are not sure about the compound identifiers and want several sets of matched pairs from a single core? Try a sub-structure search on the ChEMBL website to yield appropriate identifiers.

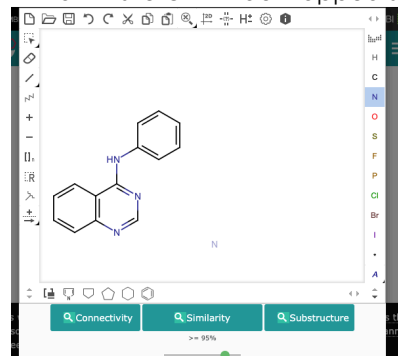
The ChEMBL website and functionality has had an upgrade....

<https://www.ebi.ac.uk/chembl/>

For a substructure search go to the top right-hand corner and click 'Draw a Structure'



A Marvin sketch window appears and we can draw the quinazoline core



Click Substructure and let's see what comes back:

The screenshot shows the ChEMBL search interface. On the left, there are filters for 'Type' (Small molecule: 5406), 'Max Phase' (0: 5387, 1: 3, 2: 7, 3: 4, 4: 5), and '#ROS Violations' (0: 2477, 1: 1759, 2: 1101, 3: 16). The main area displays 'Showing 1-6 out of 5,406 records' with 'Records per page: 6'. Three compound cards are shown: CHEMBL7917 (Name: TYRPHOSTIN AG...), CHEMBL14627 (Name: No Data), and CHEMBL14699 (Name: No Data). Each card includes a chemical structure and a 'Select All' checkbox.

The new view has 'cards' for each compound, and now we can scroll through these, take appropriate ChEMBL ids, and use these in a Pair Search e.g. CHEMBL7917

The screenshot shows the ChEMBL Pair Search interface. On the left, there is a search box for 'Compound A' with 'CHEMBL7917' entered. Below it is a chemical structure of CHEMBL7917. To the right, there are buttons for 'Show AlogP98' and 'Show Structures'. The main area displays 'Pairs Found: 286' and 'Endpoints: Tyrosine-protein kinase SRC'. Below this, there is a list of 'Pairs with measure' with 15 items, including 'Tyrosine-protein\_kinase\_ABL\_Homo', 'NIH3T3\_Mus\_musculus\_piC50\_[CHE', 'Protein\_kinase\_C\_alpha\_Homo\_sapie', 'Tyrosine-protein\_kinase\_SRC\_Homo' (selected), 'Epidermal\_growth\_factor\_receptor\_e', and 'Platelet-derived\_growth\_factor\_rece'. At the bottom, there is a table with columns 'A', 'B', and 'pIC50'. The first row shows 'CHEMBL7917' and 'CHEMBL306315' with a 'pIC50' of 4.000.

Now we can explore all 286 matched pairs of this compound but with organized data.

## Finding matched pairs from publications

**TopTip** - In a similar manner we can search out the matched pairs triggered from a publication and potential get other matched pairs from other sources.

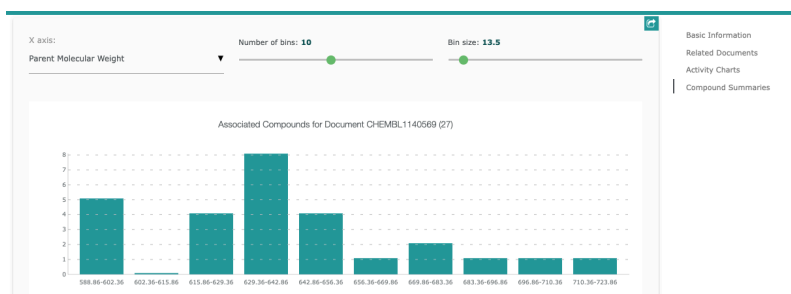
In ChEMBL lets search the papers of 'Dossetter'

The screenshot shows the ChEMBL search results page for 'Dossetter'. The search box contains 'Dossetter' and the results are filtered by 'Documents'. The results list shows four entries, each with a 'Go to Document' link and a ChEMBL ID: CHEMBL2163205, CHEMBL2069241, CHEMBL3217707, and CHEMBL1165992.

Click Documents to browse them all:

The screenshot shows the ChEMBL search results page for 'Dossetter'. The search box contains 'Dossetter' and the results are filtered by 'Documents'. The results list shows four entries, each with a 'Go to Document' link and a ChEMBL ID: CHEMBL2163205, CHEMBL2069241, CHEMBL3217707, and CHEMBL1165992.

Let's scroll down to view a paper on LHRH antagonists - click the id CHEMBL1140569 (this is a document id not a compound id - this can be confusing).



A new tab opens with lots of information on the paper. If we keep scrolling down to 'Compound Summaries' we see a chart as above. Clicking the header 'Associated Compounds for Document', another tab opens, and shows us all the compounds in the paper. Here we can grab an identifier, say ChEMBL249772

Back to MCPairs Online and plug this into the Compound-to-Pairs search.

Compound A:   Pairs Found: **11** Endpoints: **4 items selected**

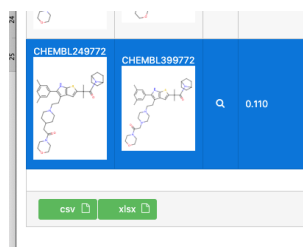
Pairs with measurements: **11**

**Show ALogP98** **Hide Structures**

		Gonadotropin-releasing_hormone_receptor_...		Gonadotropin-releasing_hormone_receptor_...	
		uM		uM	
A	B	B	Fold Change	B	Fold Change
ChEMBL249772	ChEMBL401081	0.030	~ 1.7	0.034	~ 3.6

Exporting the matched pairs as an Excel sheet gives us all of the pairs in the SAR tables of the paper and others from related papers too!

**TopTip:** - don't forget to click the 'Show Structures' to see them in MCPairs, and you can change Units and sort column by clicking at the top of columns. Usefully additional data from other assays not in the publication may be available.



compound name	compound name	Depiction	Depiction	AlignPS	AlignPS	delta	alignPS	alignPS	alignPS
A	B	A	B	A	B	A	B	A	B
CHEMBL249772	CHEMBL399772			6.448	6.383	0.165	0	1	

## Workflow from patents - Searching sureChembl by substructure

**TopTip:** We can get the matched pairs from a patent by searching for compound identifiers with sureChEMBL.

<https://www.surechembl.org/search/>

We can search for sub-structures in patent space and yield compounds and SureChEMBL ids (chembl\_id). These we can translate into ChEMBL ID with UniChem and then see if we have matched pairs in MCPairs pre-processed

Export the spreadsheet to yield compound name and chembl\_id

	A	B	C	D	E
1	patent_id	annotation_reference	schembl_id	smiles	type
2	WO-2011058149-A1	integrins	SCHEMBL1515	COC1=CC2=C	TEXT
3	WO-2011058149-A1	Wortmannin	SCHEMBL4530	COC[C@H]1C	TEXT
4	WO-2011058149-A1	Wortmannin	SCHEMBL4531	COC[C@H]1C	TEXT
5	WO-2011058149-A1	wortmannin	SCHEMBL4530	COC[C@H]1C	TEXT
6	WO-2011058149-A1	wortmannin	SCHEMBL4531	COC[C@H]1C	TEXT
7	WO-2011058149-A1	benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium	SCHEMBL132515	C1CCN(C1)P	TEXT
8	WO-2011058149-A1	benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium	SCHEMBL132516	OP(N1CCCC1	TEXT
9	WO-2011058149-A1	sodium ethoxide	SCHEMBL676	[Na].CCOC1=	TEXT
10	WO-2011058149-A1	bis(cyclooct-1,5-dien)rhodium(I)tetrafluoroborate	SCHEMBL11521	[Rh+3].F[B-]	TEXT
11	WO-2011058149-A1	sulfoxides	SCHEMBL2716	CCCCCCCCS(=	TEXT
12	WO-2011058149-A1	Triphenylphosphite	SCHEMBL13289	O(P(OC1=CC	TEXT
13	WO-2011058149-A1	dibenzoyltartaric acid	SCHEMBL7684	OC(=O)C(O)	TEXT
14	WO-2011058149-A1	dibenzoyltartaric acid	SCHEMBL7685	OC(=O)C(OC	TEXT
15	WO-2011058149-A1	dibenzoyltartaric acid	SCHEMBL7686	OC(C(O)=O)	TEXT
16	WO-2011058149-A1	dinitrobenzoylphenylglycine	SCHEMBL7695	OC(=O)CN(C	TEXT
17	WO-2011058149-A1	dinitrobenzoylphenylglycine	SCHEMBL7696	OC(=O)C(N	TEXT
18	WO-2011058149-A1	dinitrobenzoylphenylglycine	SCHEMBL7697	OC(=O)CN(C	TEXT
19	WO-2011058149-A1	Fingolimod	SCHEMBL7445	CCCCCCCCC1	TEXT
20	WO-2011058149-A1	mycophenolate mofetil	SCHEMBL4195	CCOC1=C(C)C	TEXT
21	WO-2011058149-A1	diflucortolone valerate	SCHEMBL3015	CCCCC(=O)O	TEXT
22	WO-2011058149-A1	diflucortolone valerate	SCHEMBL13157	CCCCC(=O)O	TEXT
23	WO-2011058149-A1	diflucortolone valerate	SCHEMBL13158	CCCCC(=O)O	TEXT

Clicking on individual compounds and tracking through can yield a page with UniChem identifiers - another route to finding identifiers.

Showing 1-50 of 33,374 total structure results

View results as: [Matrix](#) | [Table](#)

Check	Structure image	Chemical information	Mol weight	UniChem Cross References
<input type="checkbox"/>		<p><a href="#">SCHEMBL7866</a></p> <p>Name: N-(3-chloro-4-fluorophenyl)-7-methoxy-6-[3-(morpholin-4-yl)propoxy]quinazolin-4-amine</p> <p>SMILES:  <chem>COC1=C(C(=O)N2C(=O)C(=O)C2)C=C2C(NC3=CC(C)=C(F)C=C3)=NC=NC2=C1</chem></p> <p>InChI Key: XGALLCVXZPNRQ-UHFFFAOYSA-N</p> <p>InChI=1S/C22H24ClFN4O3/c1-29-20-13-19-16(12-21(20)31-8-2-5-28-6-9-30-10-7-28)22(26-14-25-19)27-15-3-4-18(24)17(23)11-15/n3-4,11-14H,2,5-10H2,1H3,(H,25,26,27)</p>	446.902	<p>Fetch UniChem cross references</p>

UniChem Cross References

Related structures

Patent hits

UniChem cross references (Click to expand)

CHEMBL	DrugBank	PDB	Guide to Pharmacology
CHEMBL939	PubChem: Drugs of the Future	ChEBI	ZINC
IBM Patent System	Atlas	FDA SRS	PharmGKB
Selleck	PubChem: Thomson Pharma	PubChem	Module
MolPort	Nikku	BindingDB	EPA CompTox Dashboard
		5447	DrugCentral
		Brenda	ChemicalBook

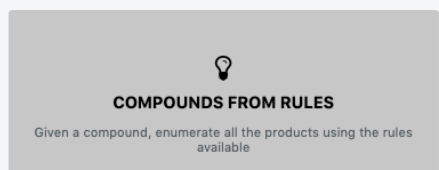
So, we arrive at a ChEMBL939 and Binding DB (BDMB5447) identifiers we can use in MCPairs to retrieve data

It is also possible to search out a specific patent, if the identifier is known, and go through the compounds to find identifiers.

We are interested in MedChemica on how users find this workflow as we would like to automate this process in future versions of MCPairs - please give us feedback.



## Compounds-From-Rules



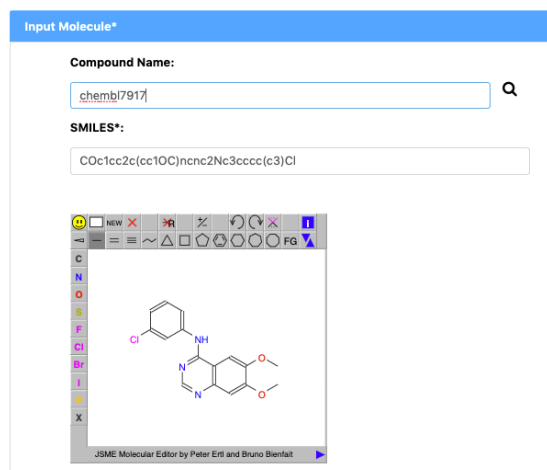
The best approach when looking for great compound suggestions from the rule engine is to start broad and narrow down; no different than a simple sub-structure search and adding groups until it is a manageable set. Let's look at the workflows and TopTips for narrowing down results.

The two best methods to reduce the number of suggestions and make them more specific for your SAR are:


- Sub-structure Lock
- Charge filter

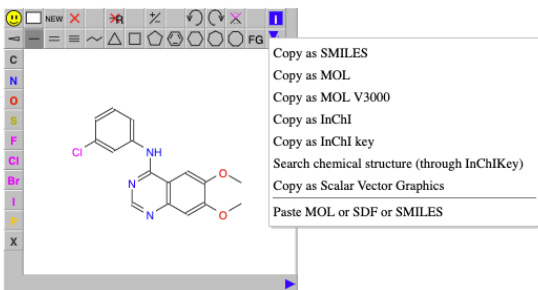
## Setting the Substructure Lock

Let's set our example compound running and tackle the issue it is facing - poor solubility.



Molecules can be entered by identifier, SMILES string or drawing the compound.

**TopTip:** Use an identifier to get the molecule into the drawing tool and modify it to the structure you wish to run. Also, there are other methods to enter a molecule into the drawing tool - click the  icon.



Set the Goal and by default it should be Increase Solubility.

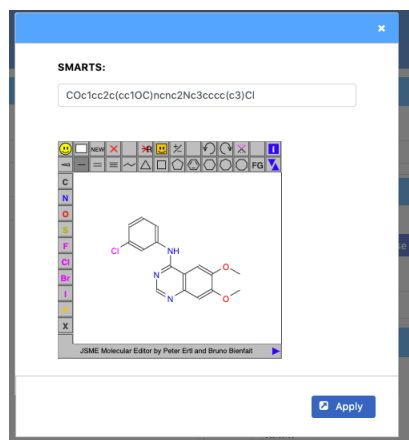
**TopTip:** The Goal menu tree is test searchable

Let's  and we should get 832 or so results - lots of ideas but we can narrow this down.

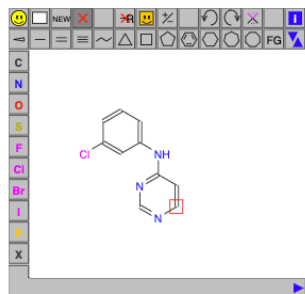
**TopTip:** If your project is a structure based design then consider exporting a full .txt of results, minimizing and docking. The suggestions with bad clashes are easily sorted to the bottom of the pile.

Click the Blue Draw button of the Sub-structure lock

A light box appears with a copy of the submitted structure. Typically, we have some SAR and know where we can and cannot substitute, in addition we can focus the results to part of the molecule. In this case can we substitute on the quinazoline a group that will increase solubility? Let us define what parts of the molecule must be 'Locked' in all of the suggested molecules.



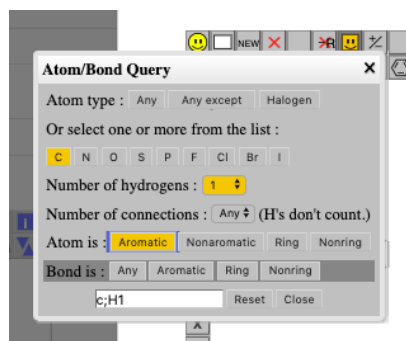
Click the delete key  and 'DELETE' what you want to change.



To block off positions and stop substitution we need to specify and mark explicit hydrogens.

Click the square Smiley face  and a new pop-up appears.

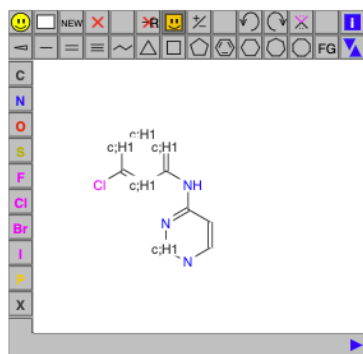
The following settings define an aromatic carbon atom bearing 1 hydrogen.



This feature is quite powerful and most drawing packages do not have this facility to help generate SMARTS patterns from structures.

Click close and we have a new effect defined on the mouse pointer

Click all aromatic carbons we want to block substitution - like so:



..and click 

A SMARTS lock has been generated.

Sub-structure Lock

SMARTS:

Clc2[c;H1]c(Nc1ccn[c;H1]n1)[c;H1][c;H1][c;H1]2

Draw

**TopTip:** Any SMARTs pattern can be entered here. Try dotted SMARTs to lock two different parts of your molecule.

Click 

Submit

 and this returns a much smaller set of compounds: 364.

**TopTip:** A history of your session gathers at the bottom of the screen:

#	Timestamp	Status	Number of Products	
0	Jul 7, 2019, 12:50:21 PM	Complete	139	<div>Save</div>
1	Jul 7, 2019, 12:51:07 PM	Complete	832	<div>Save</div>
2	Jul 7, 2019, 1:05:39 PM	Complete	364	<div>Save</div>

The output of any run can be saved, at any time, but all tracking of these is lost at the end of your session.

**TopTip:** An exported spreadsheet is sorted in increasing Relative Molecular Mass (RMM or Molecular Weight) so the smallest changes, to increase solubility, will be at the top. So big solubilizing groups will be at the bottom.

## Charge Filter (Advanced filtering)

Advanced Filters

☒ Molecular charge:
 

neutral

☐ HBA:
 ☐ HBD:
 ☐ CLogP:
 ☐ RMM:
 ☐ PSA:

Applying a charge filter is easy. Click the box next to Molecular charge and neutral is activated by default. What does this mean? MCPairs will now only return neutral (neither acids or bases) suggested molecules. In our example, where we asked for 'Increase Solubility' this often suggests basic and acidic solubilizing groups, but be undesirable.


Click 

Submit

 - our rapid filtering has turned 832 suggestions into a focused 99.

**TopTip:** On charge filtering it is important to consider the input molecule. If it is already basic then setting neutral could eliminate all results. In the same way, without a filter, could yield many suggestions where a second base or even an acid (thus zwitterion) is added. If your molecule has one basic group, consider setting the slide to one cation. Similar for acids with anion count.

Advanced Filters

☒ **Molecular charge:** cation 

☐ **HBA:**

☐ **HBD:**

☐ **CLogP:**

☐ **RMM:**


☐ **PSA:**

*I do not seem to have any results?*

*What the little blue arrow does?*

There will be a time when your run produces just one product - in fact it's the input molecule... what happened?

On the right-hand side of the history of your session is a blue arrow for each row. Click the arrow and a summary of what happened on the run appears. On this run ClogP and RMM filters were so tight that most of the products were filtered on these phys-props cut-offs. In fact, 2195 results.

5	Jul 7, 2019, 1:20:33 PM	Complete	3	 Save
Job Submission		Run Information		
<b>Job:</b>	9XVCH-KX997_2019-07-07_12_20_28_507529_1562502028237	<b>Number of Errors:</b>	0	
<b>SMILES:</b>	COc1cc2c(cc1OC)ncnc2Nc3cccc(c3)Cl	<b>Number Filtered by Goal:</b>	0	
<b>SMARTS:</b>	Clc2[c;H1]c(Nc1ccn[c;H1]n1)[c;H1][c;H1]2	<b>Number Filtered by Phys Properties:</b>	2195	
<b>Direction:</b>	Increase	<b>Number Filtered by Specificity:</b>	0	
<b>Endpoint:</b>	Solubility	<b>Number Filtered by SMARTS:</b>	706	
<b>Molecular charge:</b>	0	<b>Number Filtered by Uglies:</b>	30	
<b>HBA:</b>	None	<b>Number Adjusted for Protonation:</b>	0	
<b>HBD:</b>	None	<b>Number of Products (after Filters and Deduplication):</b>	3	
<b>CLogP:</b>	[2,3]			
<b>RMM:</b>	[10,300]			
<b>PSA:</b>	None			

**What are Uglies, phys-props filters and the such like?**

Looking at the right hand list of 'Run Information':-

Filtered by Goal:

- Products adjusted by the Goals we selected

Filtered by Phys Properties:

- Charge state, RMM and ClogP in combination can greatly reduce the output.

Filtered by SMARTS:

- is the sub-structure lock we entered
- if you have zero products then combination of lock, charge and phys-props is just too tight.

Filtered by Uglies:

- Similar to PAINS filters plus some additions. (e.g. electrophiles)
- Occasionally a chemical series as a whole can hit these filters.
- If there is an issue, quote the Job reference on the left when reporting an issue.

Filtered by Specificity:

- is an old feature we are going to bring back - thanks to demand
- this only applies the most specific rules to the input molecule.

Adjusted for Protonation state:

- does not filter molecules, but is a chem-infomatic adjustment that sometimes is required.

**TopTip:** *Exploring changes for just one group?*

Consider drawing a simple core structure and attaching to it only the group of interest. This also helps focus the results down. The quinone default structure, on opening Compounds-From-Rules, is useful for this.

Example:

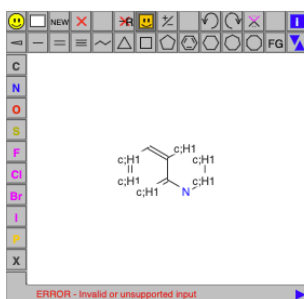
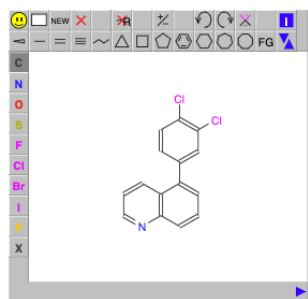
Isosteres for 3,4-dichloro can be hard, just draw onto the quinone the ring, and then edit the quinone in sub-structure lock with all positions blocked, except the ring attachment point.

SMILES\*:

Clc3ccc(c1cccc2ncccc12)cc3Cl

SMARTS:

c2c1c[n[c;H1][c;H1][c;H1]1][c;H1][c;H1][c;H1]2

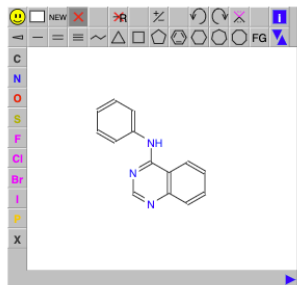


Yields 320 suggestions

**TopTip:** *Changing your input molecule for more results*

Sometimes your input molecule can be overcomplex and MCPairs may not match some of the specific groups. Consider simplifying your molecule, for example by removing all the groups from a benzene or aromatic ring

Example:

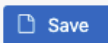
**TopTip:** *Try different Goals to get a more diverse set of output*

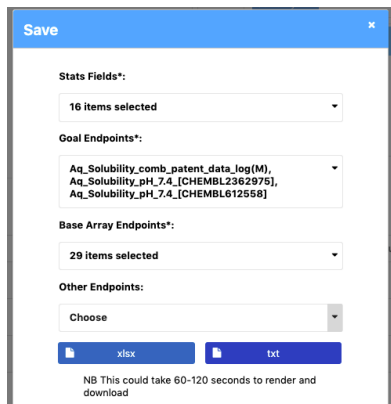
Some rules sets are bigger than others, simply as a result of the frequency of testing. The largest set in MCPairs is LogD. It can be useful to run both increase and decrease LogD to get a much broader set of suggestions. Remember the output in Excel will have solubility, metabolism and other ADMET properties too.

**TopTip:** *Exporting to spreadsheet (refining the output to fewer columns)*

New for v1.5 is a more flexible export of spreadsheet rendering.

It is possible to pick specific Rule sets, statistics fields and thus customize the output. For those that find the default output of MCPairs too dense this feature is very useful.

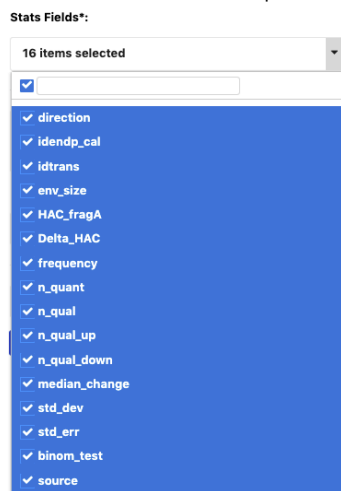
Click  to get the export menu.



Clicking xlsx or txt will yield a spreadsheet with default MCPairs settings.

There are four drop-down menus to customize the spreadsheet.

Stats Fields\* - are the repeat statistic fields for each endpoint.



frequency - proportion of pair improving a property.

n\_quant - number of pairs with quantified measurement (within range)

n\_qual - number of pairs with qualifier measurements (including out-of-range)

n\_qual\_up - number of pairs where delta measurement is +ve

n\_qual\_down - number of pairs where delta measurement is -ve

The rest are self-explanatory

Goal Endpoints\*:

These are the endpoints for the actual Goal you selected - you can choose to narrow this down.

Base Array Endpoints\*:

The Base Array is the basic ADMET endpoints that chemists in compound design would like to see most of the time. They appear to the right of the Goals selected UNLESS the Goals are part of the base array. They

appear in a specific order, grouped together by broad goal setting. It may be worth reducing this list to just what you are interested in, for example, the solubility we just ran and only HLM.

Base Array Endpoints\*:



Other Endpoints:

Nothing is set here by default, but there is a list of every other endpoint that has a rule relevant to the transformation suggestions: One of the most powerful features of MCPairs.

Example: Pick an anti-target you may have concerns about e.g. competing GPCR? The list can be text searched.

**TopTip:** *Looking and the Rule data stats - what I am looking for?*

Once you have your output, how should you examine the statistics - particularly the NED (blue) rules? Remember there are four Rule types 'Increase', 'Decrease', 'Neutral' and 'NED - No Effect Determined'. The Rules are usually defined at 95% confidence; a very high threshold, so NEDs are displayed because there are 6 or more matched pairs but confidence is <95%. So how do we judge these?

The stats data lives in hidden columns between the coloured direction indicators that yield the heat map.

	FR	GH	GX	
MDCK				RLM
dog		HLM	HLM	Clint
perm		Clint	Clint	piC50
logER		[CHEMBL2367379]	[CHEMBL613373]	[CHI
dog				
direction	direction	direction	direction	direction
15				
11	increase	NED	decrease	

We can tell by the column labels at the top.

Grab some columns, right click and select Unhide to reveal them





Let's look at the HLM data

	HLM Clint [CHEMBL2367379]	HLM Clint [CHEMBL2367379]	HLM Clint [CHEMBL2367379]	HLM Clint [CHEMBL2367379]	HLM Clint [CHEMBL2367379]	HLM Clint [CHEMBL2367379]	HLM Clint [CHEMBL2367379]	HLM Clint [CHEMBL2367379]	HLM Clint [CHEMBL2367379]	HLM Clint [CHEMBL2367379]	HLM Clint [CHEMBL2367379]	
	identp cal	idtrans	env size	HAC fragA	Delta HAC	frequency	n quant	n qual	n qual up	n qual down	median change	std dev
3 NED	1647	5	1	3	-2	64	12	14	3	9	-0.0325	0.280719

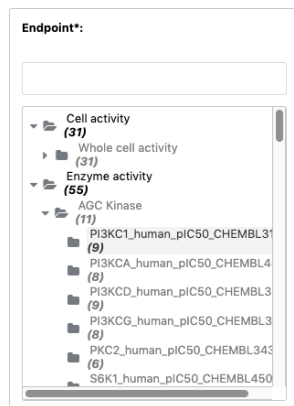
Here there are 14 matched pairs in total (n\_qual) and 12 in range ( $9 + 3 = 12$ ) for up and down. So encouragingly, 9 of the matched pairs decrease metabolism, sounds promising, but the Medium change is close to zero and the std-dev small at 0.28. So maybe there might be an improvement but, if anything, it will not do any harm - might be worth a try if lots of other things look good.

**TopTip:** *Drill backs to original data.*

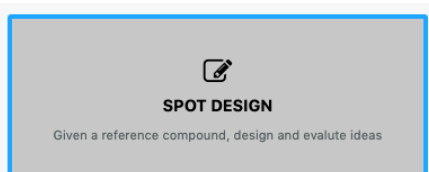
Make sure you look back to the original matched pair data the rules have been derived from. In v1.4 we reduced the hyperlink to just a single link to the right of the molecule.

<chem>ncnc2Nc3ccccc3</chem>	<a href="#">matched pair data</a>	2
	<a href="#">matched pair data</a>	2

The sheet opens without selecting any specific matched pairs for a given assay. We are free to browse using the Tree menu to the right. The number in brackets, against each Endpoint name, is the number of matched pairs available. In this window you are free to explore all of the med chem knowledge for a given chemical transformation and export a spreadsheet of any dataset. This is one of our most powerful features.



## SpotDesign™



## Workflows to explore your ideas

**TopTip:** Consider a simple comparison to get more results. Say we wished to explore a specific substitution on a benzene to improve metabolism and compare this to a leading molecule that is quite lipophilic due to a 3,4-dichloro group.

By all means enter the 3,4-dichloro and your ideas. However, this may not yield too many results because MCPairs will look for your idea against 3,4-dichloro - it is quite specific. Consider drawing just phenyl as the group to give more results? Compare the windows below starting with 3,4-dichlorobenzene and phenyl as reference molecules.

**SMILES\*:**

Clc3ccc(c1ccnc2ccccc12)cc3Cl

[Copy to Compounds From Rules](#)

**Endpoint\*:**

Metabolism [Select](#)

**Idea Molecule\***

**SMILES\*:**

COc3ccc(c1ccnc2ccccc12)cc3OC

	Idea	Delta
ClogP	3.85	-1.29 ↓
HBA	3	2.00 ↑
HBD	0	
PSA	31.35	18.46 ↑
RMM	265.31	-8.83 ↓

[Submit](#)

Other Rules:	
increase	0
decrease	0
remaining	0
total	0

**TopTip:** To change reference click Home at the top and SpotDesign again.

SMILES\*:

Copy to Compounds From Rules

Idea Molecule\*

SMILES\*:

Endpoint\*:

Select

	Idea	Delta
ClogP	3.85	-0.79 ↓
HBA	3	2.00 ↑
HBD	0	
PSA	31.35	18.46 ↑
RMM	265.31	60.06 ↑

Submit

Other Rules:

increase	0
decrease	2
remaining	1
<b>total</b>	<b>3</b>

0 / 10

Hep\_CL\_rat\_mL\_min-1.10-6cells:  
direction: NED  
examples: 9  
examples up: 3  
examples down: 6

## What if I don't get any results?

SpotDesign provides a simple interface to explore your ideas against the rules in the database. Gather your ideas and export a sheet of all of the results. Your ideas are compared against the rules for a given Goal and these may not give any results.

**TopTip:** Try changing the Goal to LogD as this is the biggest Rule set. On finding something click on any coloured wedge and go into the Pair Browser. If you want metabolism info you can now explore whatever pairs are available.

Endpoint\*:

- Cell activity (19010)
- DMPK (26482)
- Enzyme activity (81118)
- integrins Ion-channels and Epigenet
- Phys\_Props (50818)
  - ☐ Fraction-unbound (13790)
  - ☒ LogD (24388)
    - ☒ LogD\_TM (24388)
  - ☐ Solubility (12640)
- Primary safety (3696)

**TopTip:** New for v1.5 - Copy to Compound-From-Rules button

Copy to Compounds  
From Rules

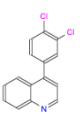
**Reference Molecule\* and Goal\***

**SMILES\*:**

**Endpoint\*:**

**Idea Molecule\***

**SMILES\*:**

**Chemical Structure:**  


**Properties Table:**

ClogP	3.85
HBA	3
HBD	0
PSA	31.35
RMM	265.31

**Other Rules:**

This copies the reference compound and Goal directly into Compounds-From-Rules in a new tab - you can set that running while also still exploring other ideas in the first tab.

Version

Author

Al Dossetter

Date

1<sup>st</sup> July 2019

MCPairs Version

v1.5