The Use of Matched Molecular Pair Analysis (MMPA) to Share Knowledge Between Multiple Companies

Al G. Dossetter, Andrew G. Leach, Ed J. Griffen, Lauren Reid (MedChemica)
Martin Stahl, Jerome Hert, Christain Kramer, Torsten Schindler (Roche)
Jeff Blaney, Hao Zheng, Alberto Gobi (GNE)
Attilla Ting, Steve St-Gallay (AZ)

MedChemica
contact@medchemica.com

Problem
Can we learn more about the link between molecular structure and ADMET properties by sharing large pharma knowledge without sharing confidential structures?

Solution
Use **MMPA** to share fragments of molecules (SMIRKS[1]) and delta property values, rather than full structures and data measurements

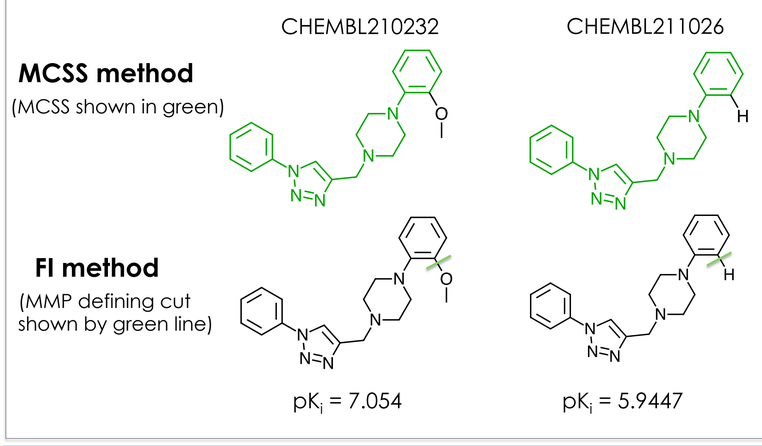
MMPA - a method of determining structure activity relationships (SAR's) within sets of compounds. Matched molecular pairs (MMP's) are identified and differences in their measured ADMET data are used to link properties to structure.[2] The process is 'data hungry' and merging multiple company data sets to create 'big data' offers statistically robust results. **MMPA explained in 3 steps:**

1) MMP's

Molecules that differ only by a particular, well-defined, structural transformation[3]

- CHEMBL211026 and CHEMBL210232 are an example of an MMP found within the ChEMBL18 database[4]
- They both have measured binding in the Dopamine D2 receptor assay
- Two methods exist to identify MMP's; the maximum common substructure (MCSS) method[5] and the fragment and index (FI) method[6]
- MCPairs, developed by MedChemica, allows the user to specify MCSS, FI or a combined method to find MMP's within very large data sets

A MMP found by both methods:



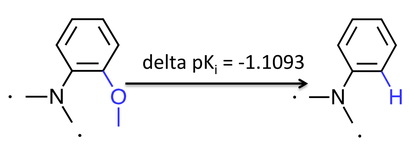
2) Environment Capture

Essential for understanding the context of the transformation[7]

- MCPairs records MMP's as chemical transformations
- Transformations are encoded as SMIRKS and recorded along with their delta property value/s
- The SMIRKS contain the structural change along with the chemical environment spanning up to 4 atoms out
- **There is no way to identify the parent compounds or the individual pK_i values from the SMIRKS and delta pK_i**

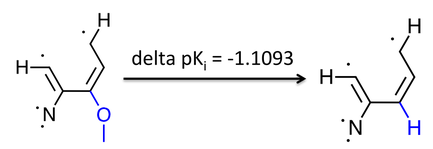
The MMP as a transformation:

4 atom environment:



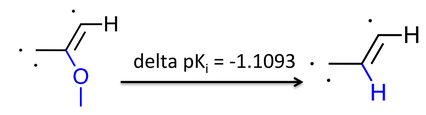
[C:8]([H])([H])[N:7]([C:9]([H])([H]))[c:2]1[c:3]([H])[c:4]([H])[c:5]([H])[c:6]([H])[c:1]1[O][C]([H])([H])([H])>>[C:8]([H])([H])[N:7]([C:9]([H])([H]))[c:2]1[c:1]([H])[c:6]([H])[c:5]([H])[c:4]([H])[c:3]1([H])

3 atom environment:



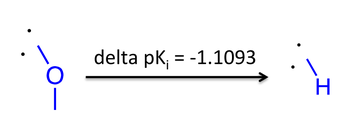
[c:4]([H])[c:5]([H])[c:1]([c:2]([c:3]([H]))[N:6])[O][C]([H])([H])([H])>>[c:4]([H])[c:5]([H])[c:1]([H])[c:2]([c:3]([H]))[N:6]

2 atom environment:



[c:2]([H])[c:3]([c:1])[O][C]([H])([H])([H])>>[c:2]([H])[c:3]([H])[c:1]

1 atom environment:

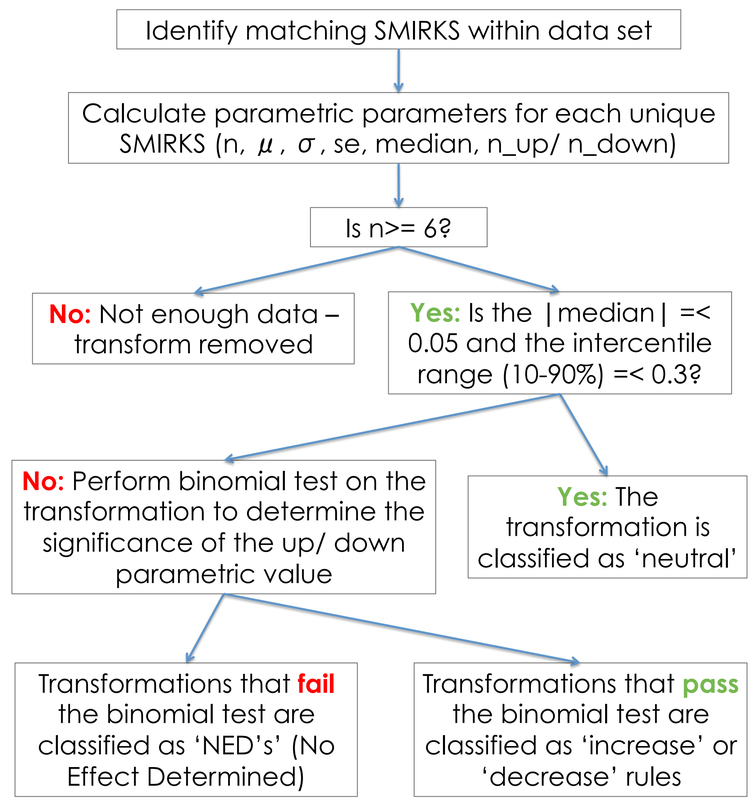


[c:1][O][C]([H])([H])([H])>>[c:1]([H])

3) Knowledge Extraction

Medicinal chemistry rules are extracted from statistical analysis of the transformation data set

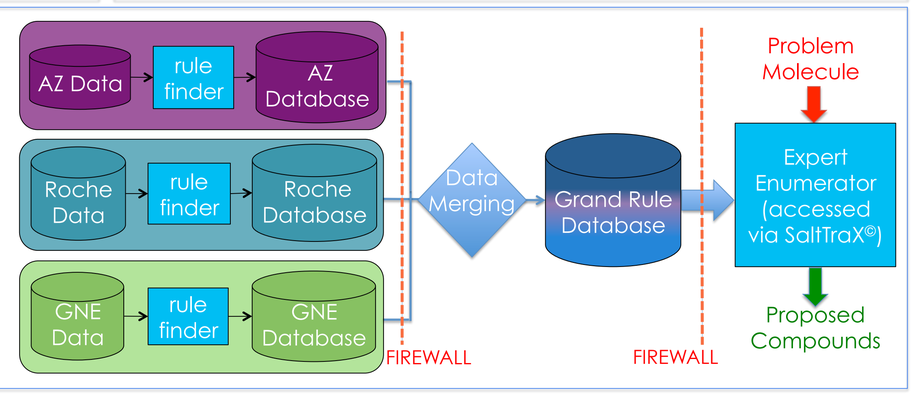
Statistical methods are used on large data sets of SMIRKS to extract chemical transformation rules that increase, decrease or maintain a given property:



The SALT Consortium

Roche · Genentech (A Member of the Roche Group) · MedChemica

- A collaboration between MedChemica, AstraZeneca, Roche and Genentech to perform **Big Data MMPA** on ADMET properties by sharing SMIRKS and delta property values
- The knowledge extracted is combined into the **Grand Rule Database** and a copy of this is supplied back to each company
- It is also made available to smaller companies and universities through an online interface called **SaltTraX©**
- The purpose of the consortium is to enhance the drug design process by speeding up the **design-make-test-analyse** cycle to generate better ideas the first time



Case Study - MMPA on ChEMBL toxicity data

Bowes, J. et al. have analysed the targets that comprise the in vitro pharmacological profiling panels of 4 large pharmaceutical companies. As a result they have published a panel of recommended safety targets for in vitro profiling.[8] We have used MMPA on ChEMBL data to find transformation rules that decrease binding to these targets.

Method

1) Data Selection
- The recommended safety targets were searched in the ChEMBL18 assay database
- A data curation technique, similar to that described by Kramer, C. et al., was used to filter the heterogeneous data[9]
- The assays with > 2000 compound binding measurements were selected

2) MMP Identification
- MCPairs was used to find MMP's within each ChEMBL assay data set
- MMP's were recorded as transformations (SMIRKS) along with their delta binding measurements
- Both MMP finding methods were employed to allow comparison of the results

3) Assay Analysis
- An assay analysis python script, which follows the knowledge extraction method described above, was run on each MCPairs output
- This generated a list of chemical transformation rules for each binding assay

Results

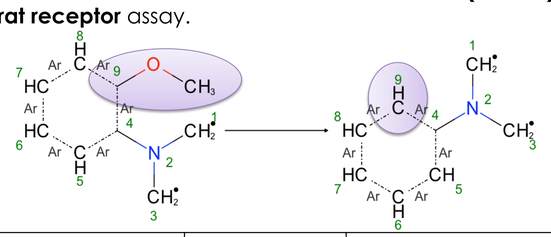
Recommended safety targets with >2000 ChEMBL binding measurements:

Assay	No. of binding measurements (pK_i or PIC50)	No. of decrease rules (MCSS method)	No. of decrease rules (FI method)
Acetylcholinesterase human	2895	58	204
Adenosine A2a receptor human	3984	128	146
Cannabinoid CB1 receptor human	2782	54	86
Cannabinoid CB2 receptor human	2800	75	209
Cyclooxygenase-1 human	2939	40	39
Cyclooxygenase-2 human	3782	96	115
Dopamine D2 receptor rat	2706	74	18
Dopamine D2 receptor human	4532	100	193
Dopamine transporter rat	2587	157	51
GABA-A receptor; anion channel rat	2008	88	100
Mu opioid receptor human	2888	135	63
Serotonin 1a (5-HT1a) receptor rat	3249	144	258
Serotonin 1a (5-HT1a) receptor human	2408	115	164
Serotonin transporter rat	2655	155	104
Serotonin transporter human	2463	177	82
Tyrosine-protein kinase LCK human	2066	11	9

The different methods produced different numbers of decrease rules within the same data sets. Using a combined method would capture all of the possible MMP's and more rules would be found.

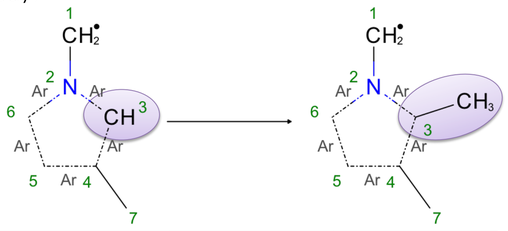
Examples of rules found within the ChEMBL toxicity data:

The MMP example discussed above contributed to the generation of the following **Dopamine D2 human receptor** transformation rule. This transformation was also identified as a rule for the **Seretonin 1a (5-HT1a) rat receptor** assay.



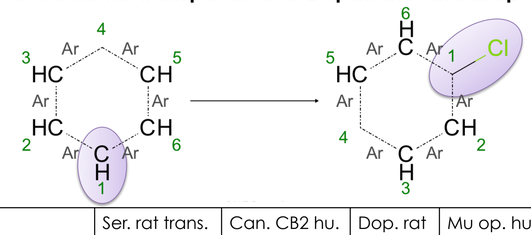
Assay	Dop. D2 hu. rec.		Ser. 1a (5HT1a) rat rec.	
MMP finding method	MCSS	FI	MCSS	FI
Median pK_i change	-0.828	-0.765	-0.75	-0.755
No. of examples	20	16	56	51
Increase examples	1	1	3	3
Decrease examples	19	15	53	48
Binomial p value	0.00004	0.00052	0	0

The following transformation was found to reduce binding in the **Cannabinoid CB1 human receptor** assay and **Cannabinoid CB2 human receptor** assay.



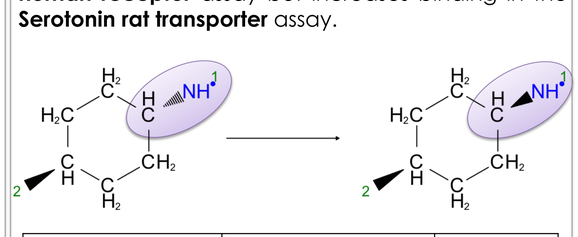
Assay	Can. CB1 hu. rec.	Can CB2 hu. rec.
MMP finding method	MCSS	MCSS
Median pK_i change	-0.377	-0.259
No. of examples	13	34
Increase examples	1	5
Decrease examples	12	29
Binomial p value	0.00342	0.00004

The following transformation was found to have **varying effects** in different assays; it decreases binding in the **Cannabinoid CB2 human receptor**, **Mu opioid human receptor** and **Serotonin 1a (5-HT1a) rat receptor** assays, but it increases binding in the **Serotonin rat transporter** and **Dopamine rat transporter** assays.



Assay	Ser. rat trans.	Can. CB2 hu. rec.	Dop. rat trans.	Mu op. hu. rec.	Ser. 1a (5HT1a) rat rec.
MMP finding method	MCSS	MCSS	MCSS	MCSS	MCSS
Median pK_i/pIC50 change	0.455	-0.252	0.388	-0.47	-0.388
No. of examples	69	33	45	13	26
Increase examples	44	8	32	1	7
Decrease examples	25	25	13	12	19
Binomial p value	0.02949	0.00455	0.00661	0.02246	0.02896

This is an example of a transformation between two MMP's with **opposite chirality**. The transformation decreases binding in the **Serotonin 1a (5-HT1a) human receptor** assay but increases binding in the **Serotonin rat transporter** assay.



Assay	Ser. 1a (5-HT1a) hu. rec.	Ser. rat trans
MMP finding method	FI	FI
Median pK_i change	-0.79	0.86
No. of examples	22	22
Increase examples	1	21
Decrease examples	21	1
Binomial p value	0.00001	0.00001

References

[1]Daylight Chemical Information Systems, Inc. *Daylight Toolkit.* [Online] 2008 [Accessed 7th April 2015]. Available from: http://www.daylight.com/products/toolkit.html
[2]Griffen, E. et al. *Matched Molecular Pairs as a Medicinal Chemistry Tool.* J. Med. Chem. 2011, **54**(22), pp.7739-7750.
[3]Leach, A.G. et. al. *Matched Molecular Pairs as a Guide in the Optimization of Pharmaceutical Properties; a Study of Aqueous Solubility, Plasma Protein Binding and Oral Exposure.* J. Med. Chem. 2006, **49**(23), pp.6672-6682.
[4]Bento, A. et. al. (2014) 'The ChEMBL bioactivity database: an update.' Nucleic Acids Res., 42 1083-1090.
[5]Warner, D.J. et. al. *WizePairZ: A Novel Algorithm to Identify, Encode, and Exploit Matched Molecular Pairs with Unspecified Cores in Medicinal Chemistry*. J. Chem. Inf. Model. 2010, **50**(8), pp.1350-1357.
[6]Hussain, J. and Rea, C. *Computationally Efficient Algorithm to Identify Matched Molecular Pairs (MMPs) in Large Data Sets.* J. Chem. Inf. Model. 2010, **50**(3), pp.339-348.
[7]Papadatos, G. et al. *Lead Optimization Using Matched Molecular Pairs: Inclusion of Contextual Information for Enhanced of hERG Inhibition, Solubility, and Lipophilicity.* J. Chem. Inf. Model. 2010, **50**(10), pp. 1872-1886.
[8]Bowes, J. et al. *Reducing safety-related drug attrition: the use of in vitro pharmacological profiling.* Nature Reviews, Drug Discovery. 2012, **11**(12), pp.909-922.
[9]Kramer, C. et al. *The Experimental Uncertainty of Heterogeneous Public K_i Data.* J. Med. Chem. 2012, **55**(11), pp.5165-5173.

MedChemica