Matched Molecular Sequences: finding the missing pair

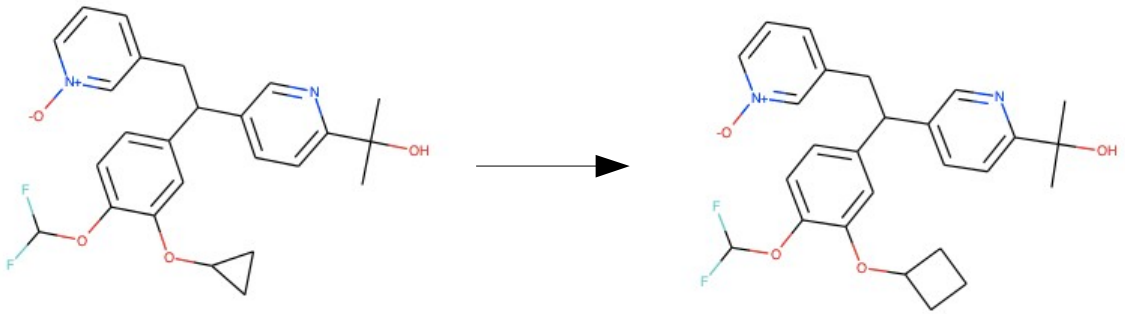
Grégori Gerebtzoff, Ph.D.

3rd RDKit UGM, October 23rd 2014

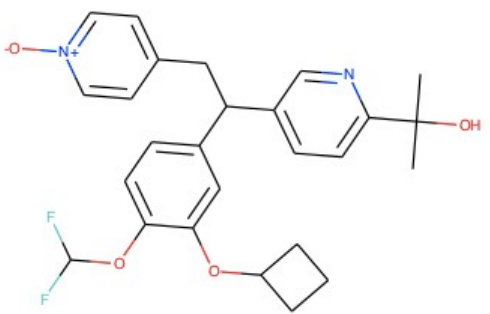
Finding the missing pair

A bug in LUCID or in the chemist's head?

The chemist was searching for a specific transformation in LUCID but couldn't find it...



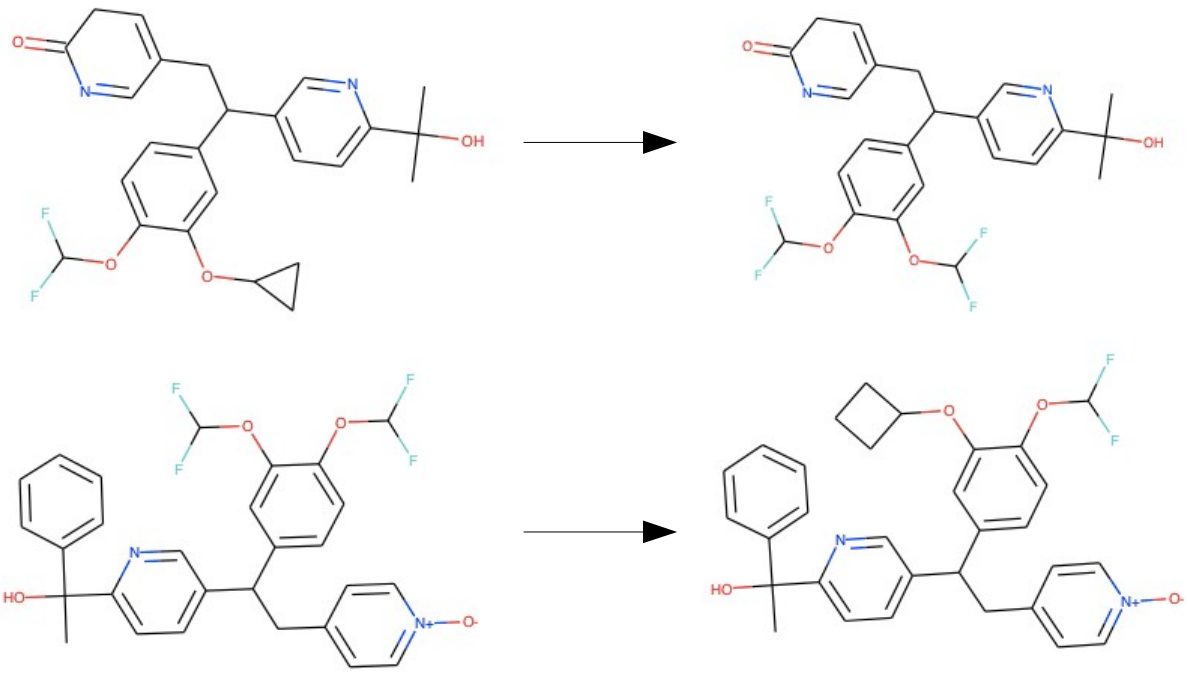
...despite being sure to have synthesized and measure such a pair of compounds.

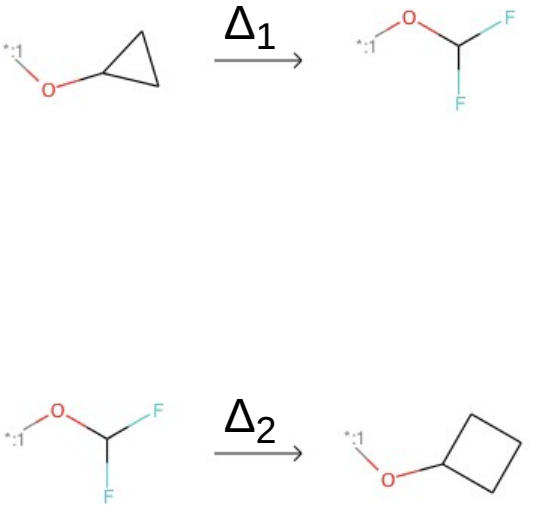


After checking in the corporate database, it was not a bug in LUCID: the pair didn't exist. The chemist mistook it for a slightly different compound.

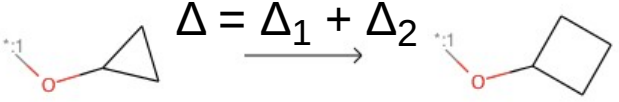
Finding the missing pair

Can we get the missing pair through another path?





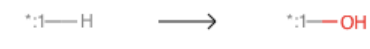
Assuming additivity, we can obtain the missing pair by summing up the contributions of two "networked" MMPs.
Complete paths (A → B_1 → C) are not required.



Matched Molecular Sequences

Validation of the method

Idea: compare the deltas of "normal" vs. "networked" MMPs on common replacements

	# pairs	# variations	Δ pairs	Δ variations
	61	84	-0.152 ± 0.53	-0.211 ± 0.70
	98	71	-0.256 ± 0.40	-0.232 ± 0.41
	131	52	-0.357 ± 0.64	-0.417 ± 0.61
	60	43	-0.409 ± 0.71	-0.381 ± 0.83
	29	33	-0.303 ± 0.52	-0.305 ± 0.92

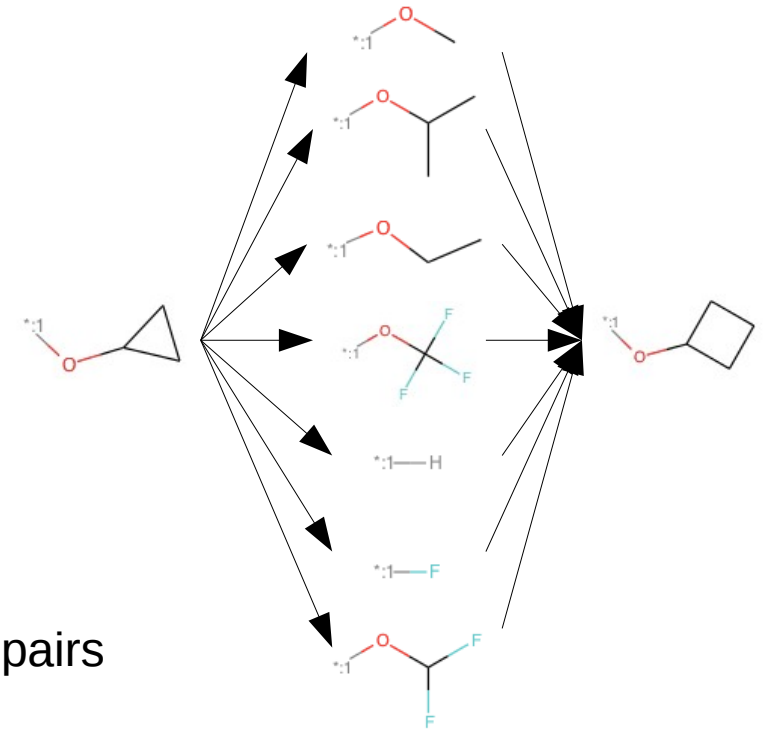
Dataset: chEMBL human hERG pIC_{50}

Matched Molecular Sequences

Usage

- Finding the missing pair

7 variations, Δ = 0.337 ± 0.27



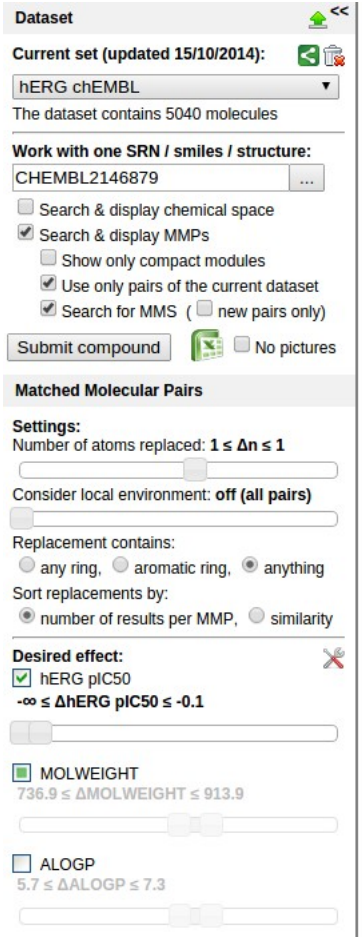
- Getting more confidence on low populated pairs

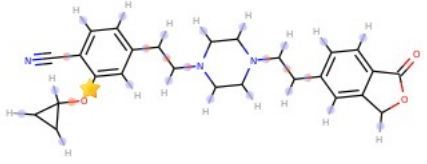
2 pairs, Δ = 0.2 ± 0.43
14 variations, Δ = 0.298 ± 0.36

Dataset: chEMBL human hERG pIC_{50}

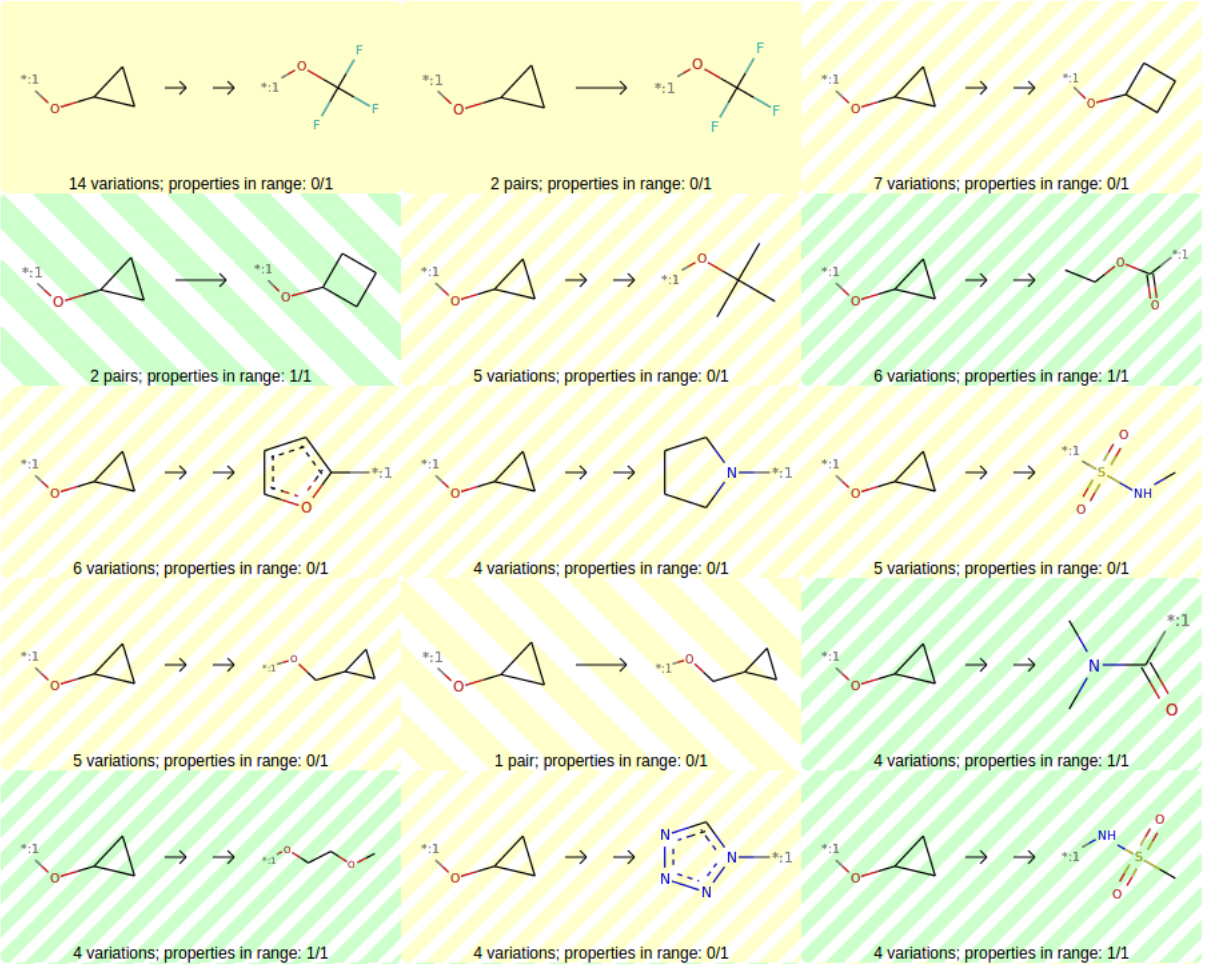
Matched Molecular Sequences

Implementation into LUCID





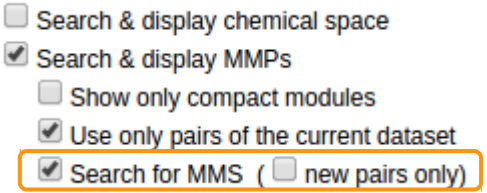
hERG pIC50: 6.33
MOLWEIGHT: 432



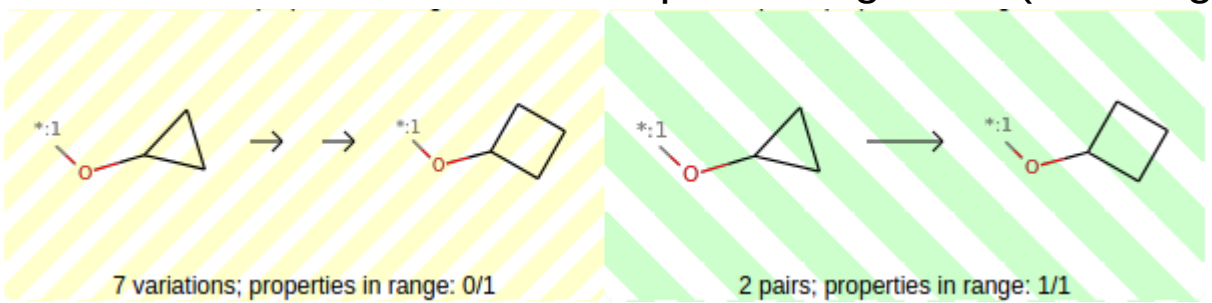
Matched Molecular Sequences

Implementation into LUCID

- A new option allows searching for Matched Molecular Sequences



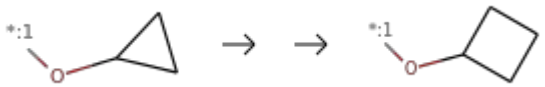
- Matched Molecular Sequences are easily identifiable with the two-arrowed transformation and different striped background (left image)



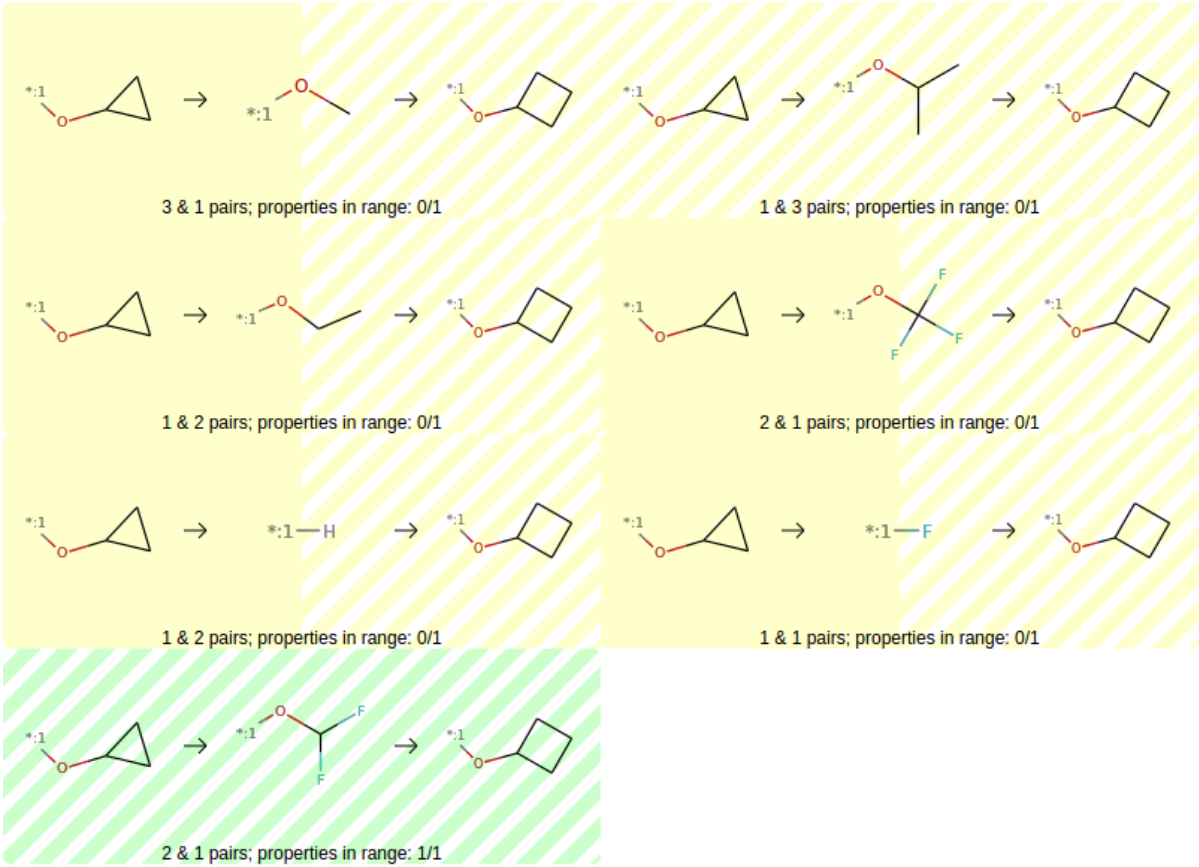
- When searching for both "normal" and "networked" MMPs, if the same transformation is present in both modes they will be displayed next to each other

Matched Molecular Sequences

Implementation into LUCID

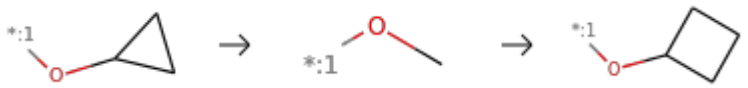


- When selecting one Matched Molecular Sequence, the list of different paths is displayed (see slide #6); as usual statistics for each path are available, and each path is colored according to the whished effect on selected properties.
- When the middle node of a path exists in the dataset, the background of the path is half striped (for instance the first example; see next slide), plain if both middle and end nodes exist.

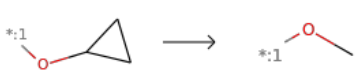


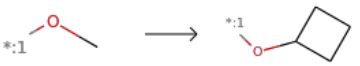
Matched Molecular Sequences

Implementation into LUCID



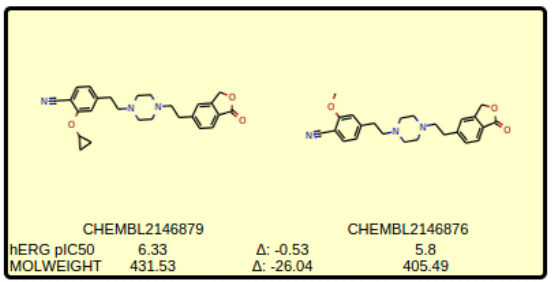
- Selecting one path of a Matched Molecular Sequence will reveal the MMPs for each leg of the path; the left column corresponds to the first leg of the path, the right column the second leg.
- Statistics of each leg are displayed on top of each column.

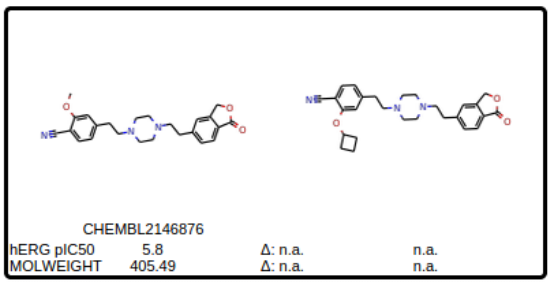


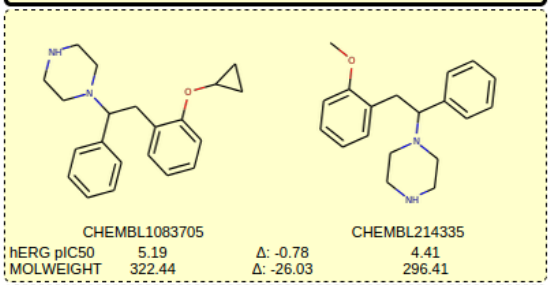


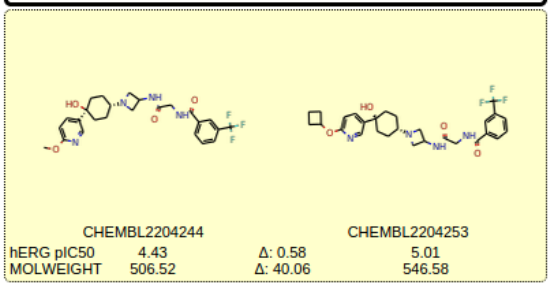
hERG pIC50: 3 pairs
n=3, Δ=-0.377, SD=0.407, ↑:33.33%, ↓:66.67%, ↔:0%
MOLWEIGHT: 3 pairs
n=3, Δ=-26, SD=0.00471, ↑:0%, ↓:100%, ↔:0%

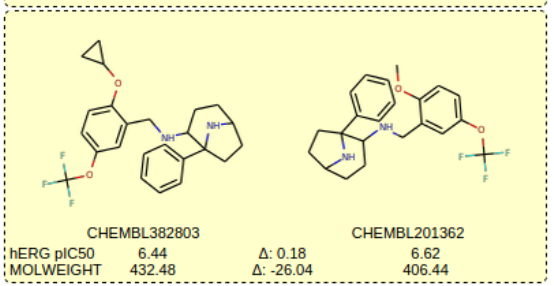
hERG pIC50: 1 pair
n=1, Δ=0.58, ↑:100%, ↓:0%, ↔:0%
MOLWEIGHT: 1 pair
n=1, Δ=40.1, ↑:100%, ↓:0%, ↔:0%











Matched Molecular Sequences

Confirming an example from the literature on chEMBL

Springer and Sokolnicki *Chemistry Central Journal* 2013, **7**:167
http://journal.chemistrycentral.com/content/7/1/167

Chemistry Central Journal

RESEARCH ARTICLE Open Access

A fingerprint pair analysis of hERG inhibition data

Clayton Springer* and Katherine L Sokolnicki

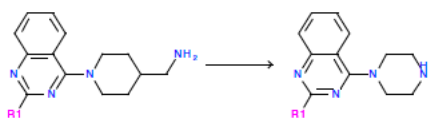
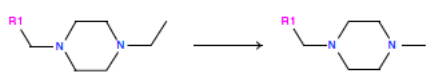
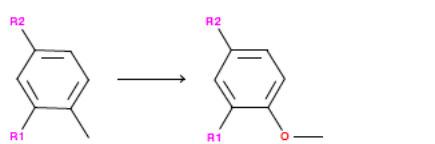
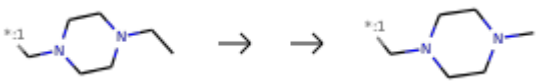
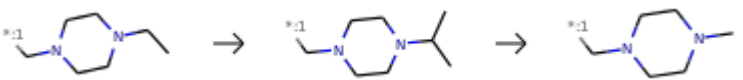
Table Row	Transformation	Fold Change	# Increasing	# Decreasing	delta SlogP
4/1		2.95	6	0	-0.77
4/2		2.47	7	1	-0.67
4/3		2.19	8	1	-0.39
4/4		1.77	17	4	-0.30

Figure 7 Miscellaneous lipophilicity reducing transformations. For these transformations we show the fold reduction in hERG inhibition, the number of examples that increase the IC_{50}, the number that decrease the IC_{50}, and change that this transformation makes in the SlogP model of logP.

Transformation (no "normal" MMP):



Only one variation available:



n=1, Δ_1=0.07

n=2, Δ_2=0.315

Δ = 0.39

Acknowledgments

Roche

Stefanie Bendels

Manfred Kansy

Wolfgang Guba

Olivier Roche

...and all the LUCID users

Others

Greg Landrum (RDKit)

Jameed Hussain (MMP algorithm)