



UK-QSAR Autumn Meeting 2014
Uni of Cambridge, Cambridge, UK

Evidence-based medicinal chemistry

...using matched molecular series

Noel O'Boyle and Roger Sayle

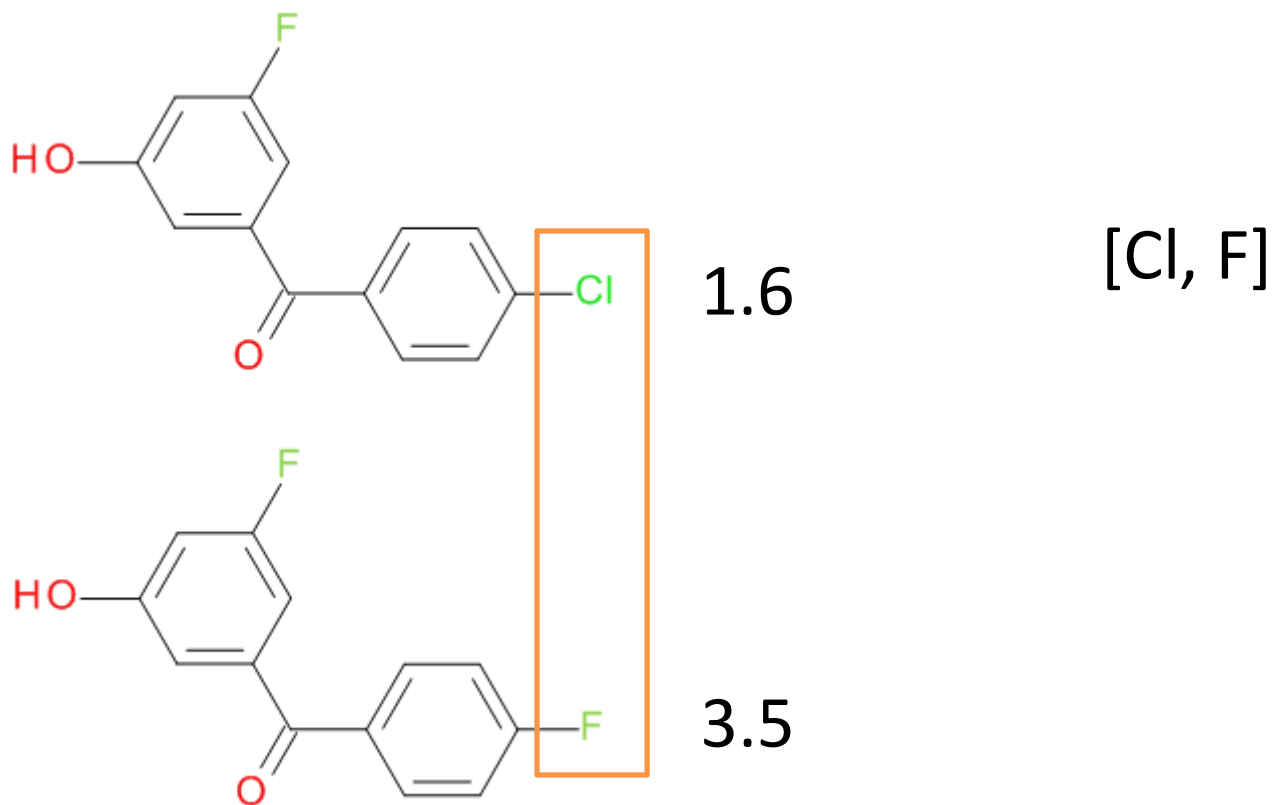
NextMove Software

Jonas Boström

AstraZeneca



MATCHED (MOLECULAR) PAIRS



Coined by Kenny and Sadowski in 2005*

Easier to predict **differences** in the values of a property than it is to predict the value itself

* Chemoinformatics in drug discovery, Wiley, 271–285.



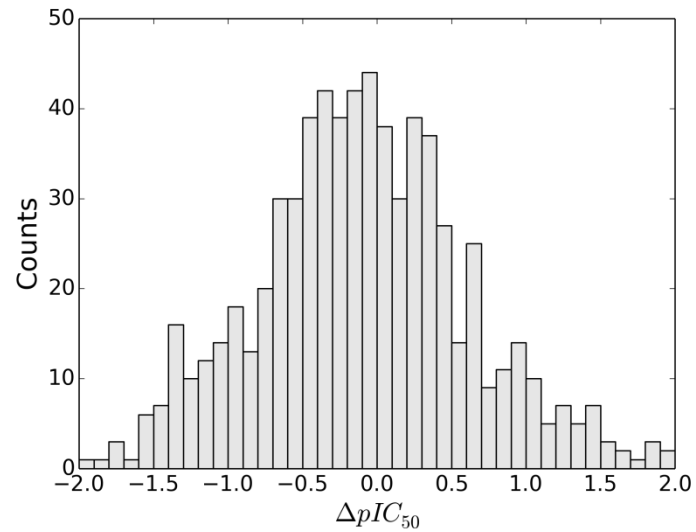
MATCHED PAIR USAGE

- **Successfully** used for:
 - Rationalising and predicting physicochemical property changes
 - Finding bioisosteres
- **Not very successful** in improving activity
 - Activity changes dependent on binding environment
- Hajduk, Sauer. *J. Med. Chem.* **2008**, *51*, 553
 - Data from 30 protein targets at Abbott
 - Most R group transformations led to potency changes normally distributed around 0



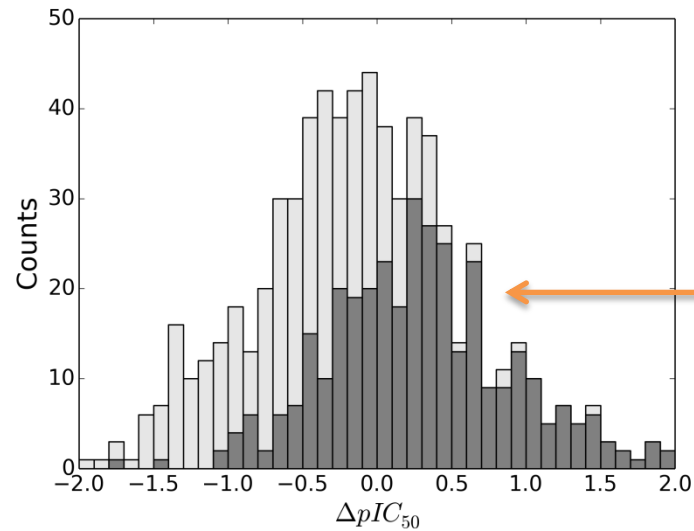
MATCHED PAIRS AND ACTIVITY

$pIC_{50}(CC) - pIC_{50}(CCCC)$



MATCHED PAIRS AND ACTIVITY

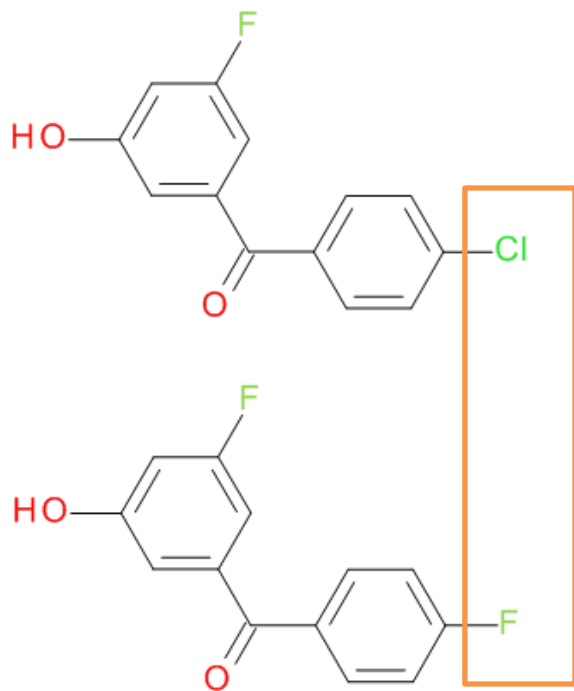
$pIC_{50}(CC) - pIC_{50}(CCCC)$



For those cases where:
[CCC > CCCC]



MATCHED SERIES OF LENGTH 2 = MATCHED PAIR

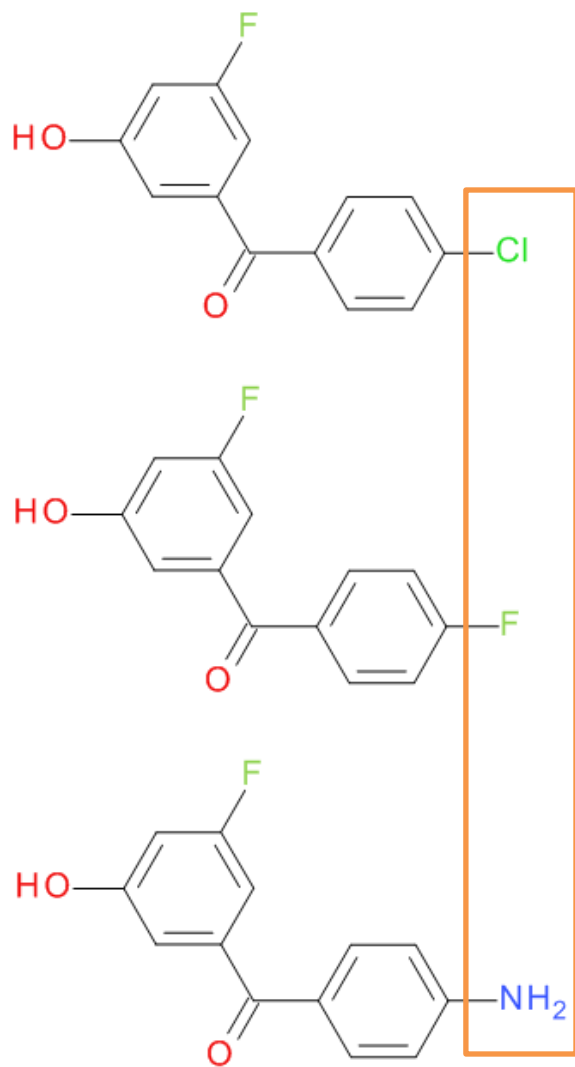


[Cl, F]

“Matching molecular series” introduced by Wawer and Bajorath, *J. Med. Chem.* **2011**, 54, 2944



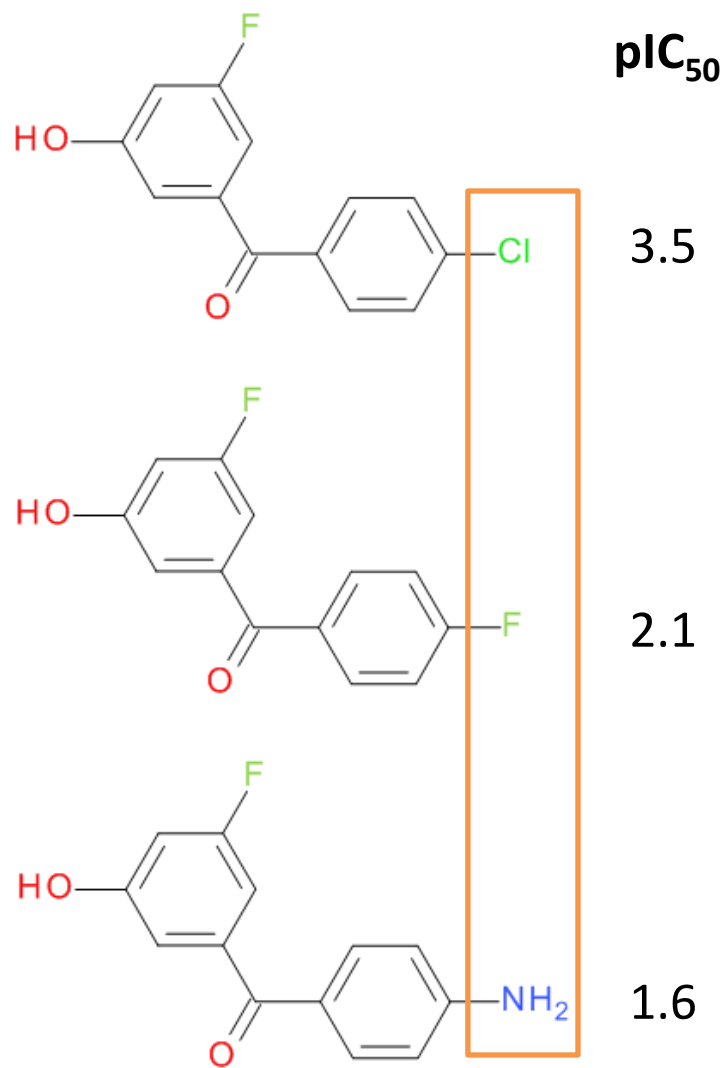
MATCHED SERIES OF LENGTH 3



[Cl, F, NH₂]



ORDERED MATCHED SERIES OF LENGTH 3



[Cl > F > NH₂]



PREFERRED ORDERS: HALIDES (N=2)

For an ordered matched series (i.e. $A > B > C > \dots$), there are $N!$ ways of arranging the R Groups:

Series	Observations*
F > H	9761
H > F	8685

Would expect 9223 for each assuming the order is random

– We can calculate **enrichment**

*Dataset is ChEMBL19 IC₅₀ data for binding assays (transformed to pIC₅₀ values)



PREFERRED ORDERS: HALIDES (N=2)

For an ordered matched series (i.e. $A > B > C > \dots$), there are $N!$ ways of arranging the R Groups:

Series	Enrichment	Observations
F > H	1.06*	9761
H > F	0.94*	8685

Would expect 9223 for each assuming the order is random

– We can calculate **enrichment**

*Significant at 0.05 level according to binomial test after correcting for multiple testing (Bonferroni with N-1)



PREFERRED ORDERS: HALIDES (N=3)

Series	Enrichment	Observations
Cl > F > H	1.90*	1478
H > F > Cl	1.08	838
F > Cl > H	0.86*	673
F > H > Cl	0.78*	607
Cl > H > F	0.76*	589
H > Cl > F	0.63*	490



PREFERRED ORDERS: HALIDES (N=4)

Series	Enrichment	Observations
Br > Cl > F > H	5.43*	263
Cl > Br > F > H	3.22*	156
H > F > Cl > Br	1.59*	77
Br > Cl > H > F	1.43	69
F > Cl > Br > H	1.40	68
Cl > Br > H > F	0.85	41
...
H > F > Br > Cl	0.76	37
...
H > Br > F > Cl	0.50*	24
Cl > H > F > Br	0.48*	23
Cl > F > H > Br	0.45*	22
H > Cl > F > Br	0.43*	21
Br > F > H > Cl	0.41*	20
F > H > Br > Cl	0.41*	20
H > Cl > Br > F	0.41*	20
F > Br > H > Cl	0.35*	17
Br > H > F > Cl	0.23*	11

N=2: Max = 1.06, Min = 0.94

N=3: Max = 1.90, Min = 0.63

N=4: Max = 5.43, Min = 0.232

Longer series exhibit greater preferences

If [H>F>Cl] is observed, will Br increase activity further?
 149 observations of [H>F>Cl]
 but only 11 where [Br>H>F>Cl]



FIND R GROUPS THAT INCREASE ACTIVITY



In-house

Query

A > B



MATSY

A > B > C

C > A > B

D > A > B > C

D > A > C > B

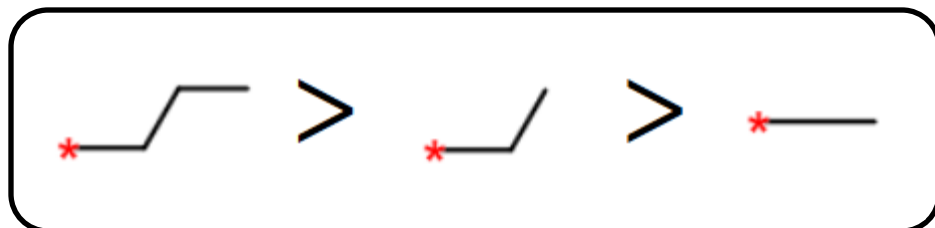
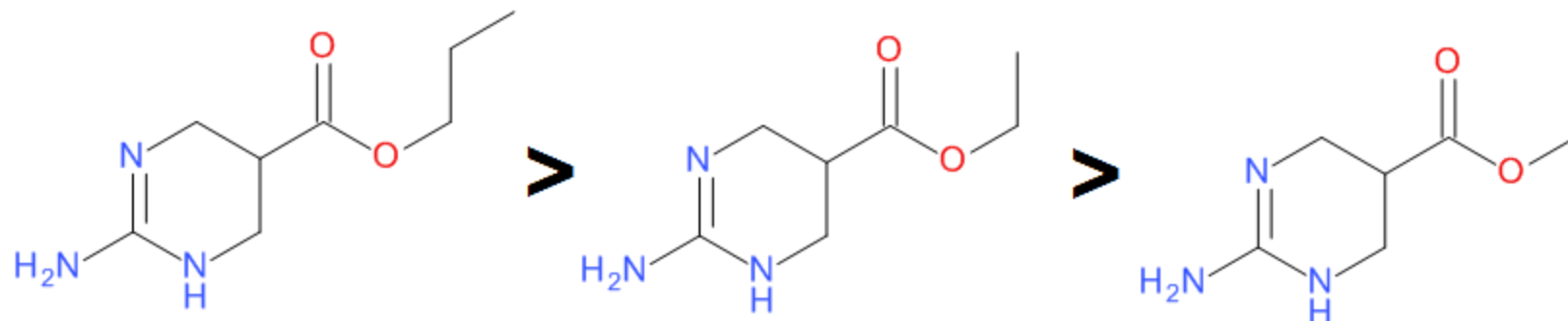
E > D > A > B

...

R Group	Observations	Obs that increase activity	% that increase activity
D	3	3	100
E	1	1	100
C	4	1	25
...

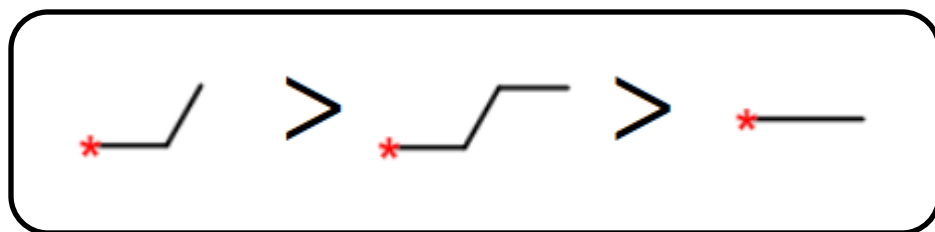
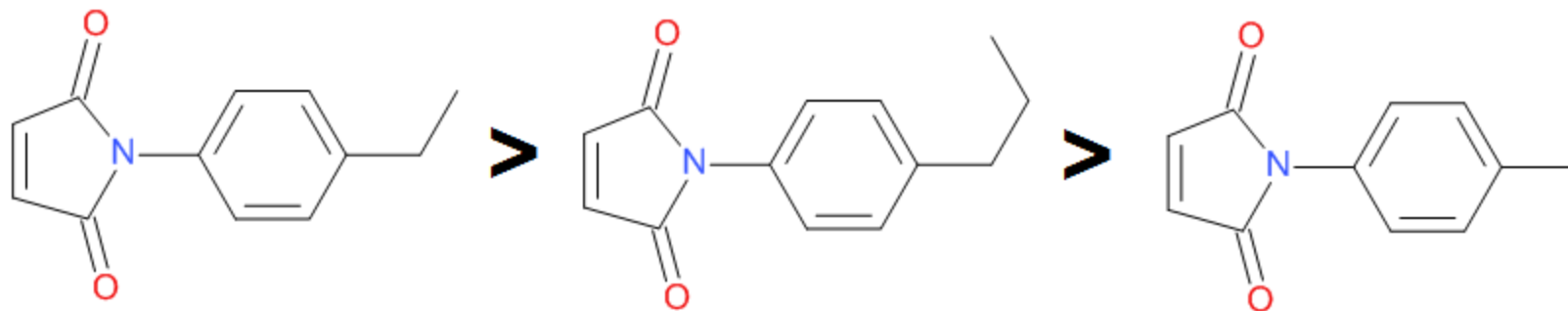


EXAMPLE



	◇ % ▾	Counts ▾	ΔLogP ◇
	90	21	+3.3
	72	60	+1.7
	69	32	+2.8
	63	27	+1.6
	60	40	-0.1

EXAMPLE II

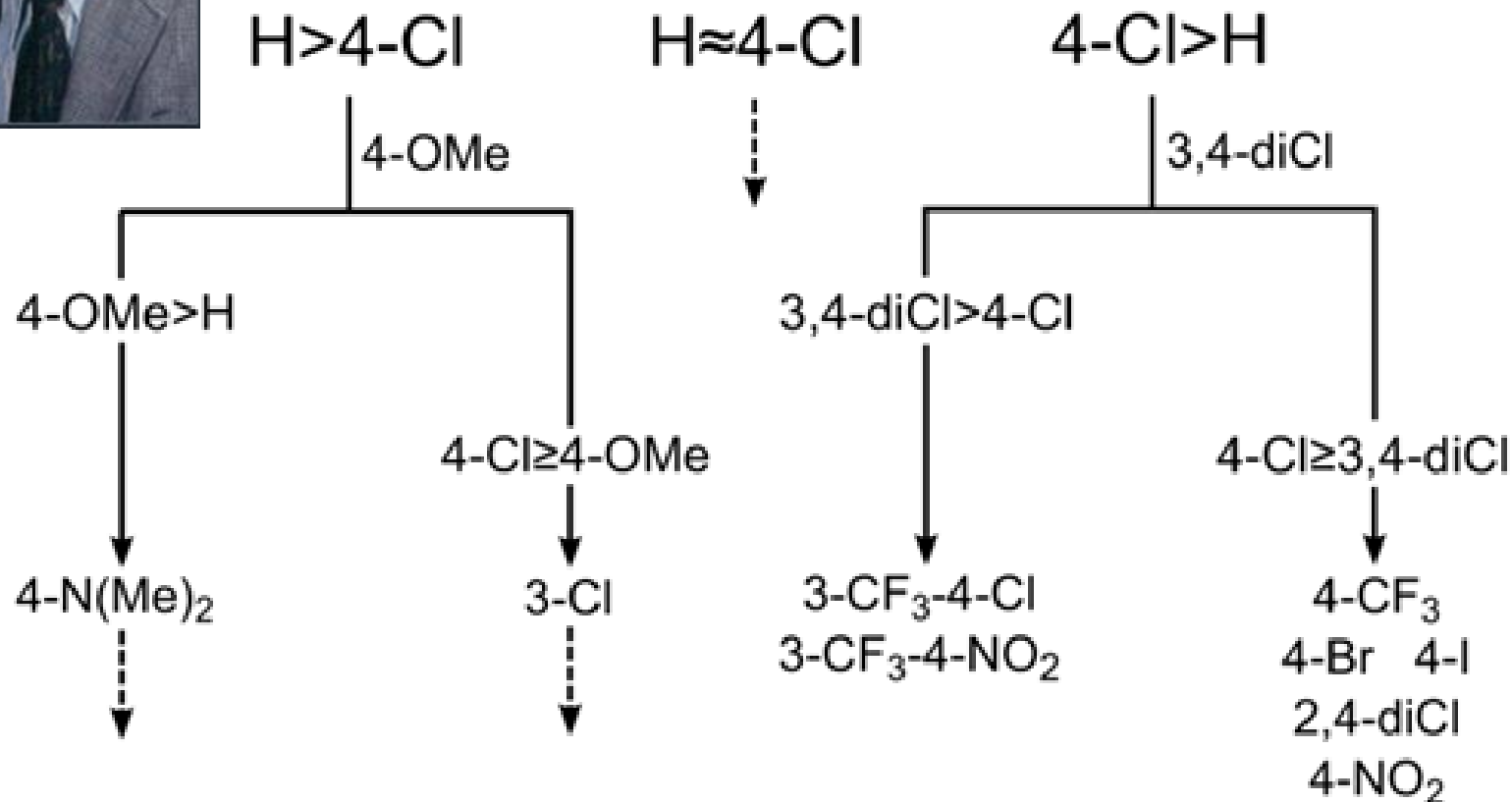


⬆ % ⬇ Counts ⬆ ΔLogP ⬆
>

	38	21	-0.8
	37	27	+0.9
	33	111	+0.3
	33	27	+1.0
	33	21	-1.6



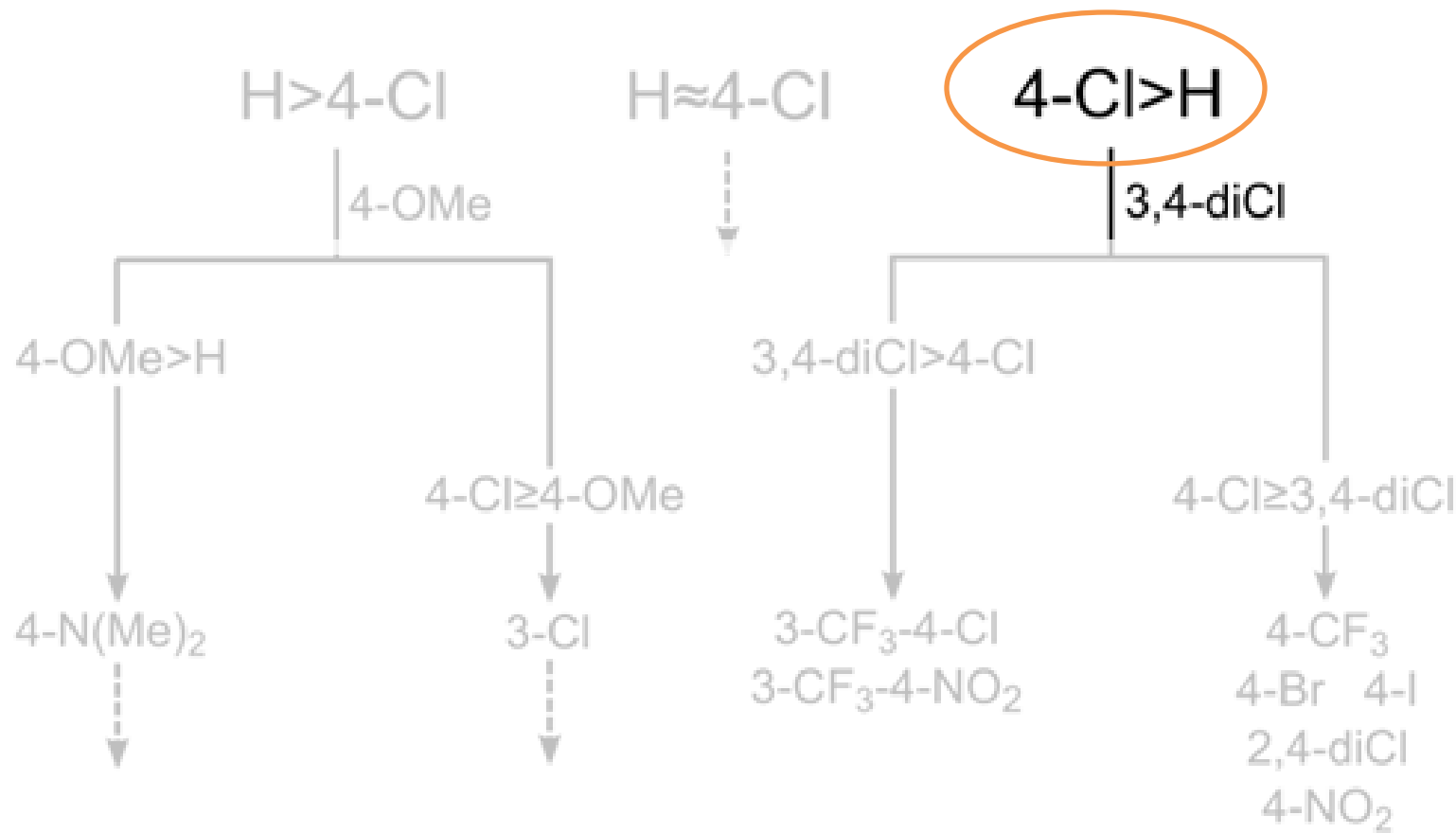
TOPLISS DECISION TREE



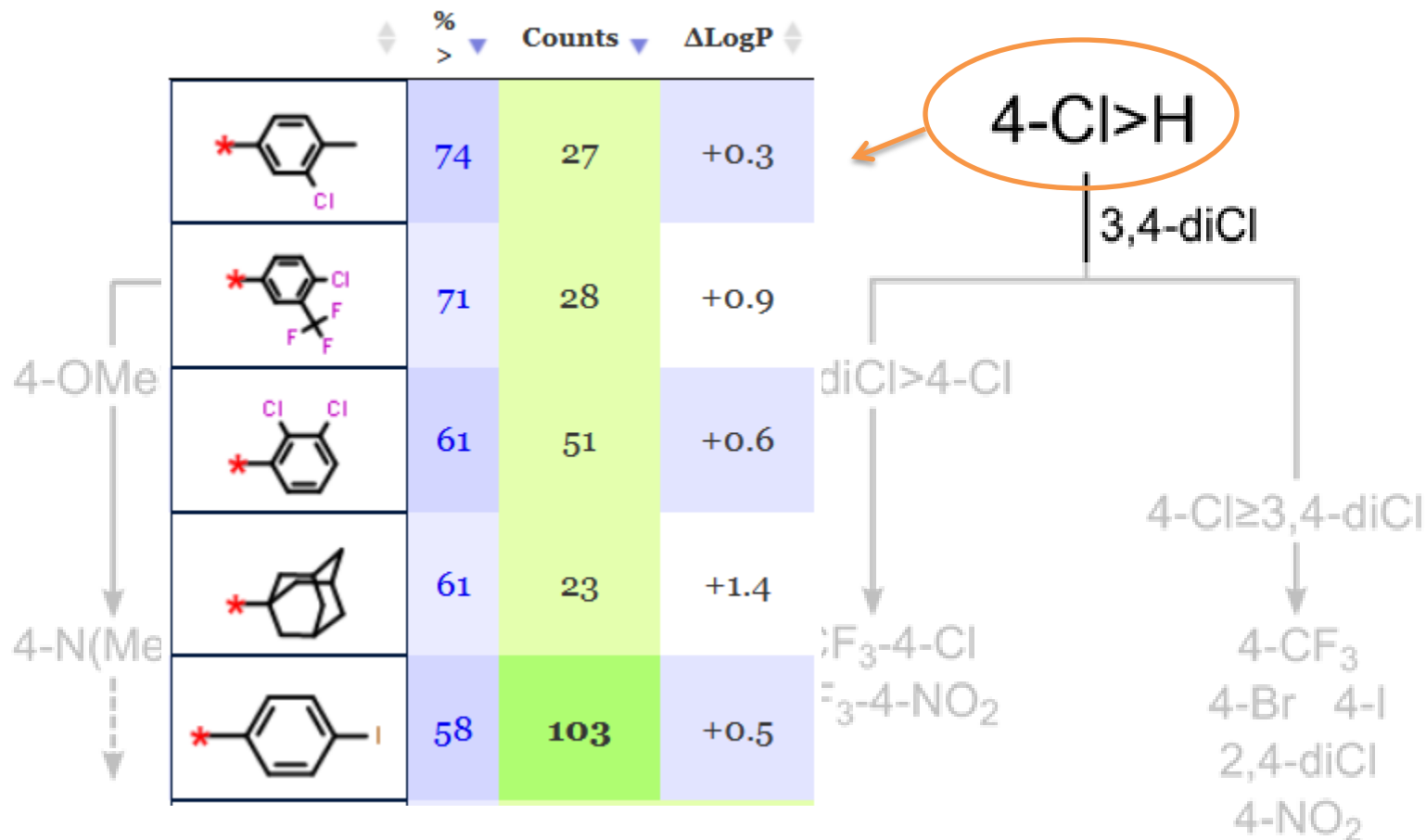
Topliss, J. G. Utilization of Operational Schemes for Analog Synthesis in Drug Design. *J. Med. Chem.* **1972**, *15*, 1006–1011.



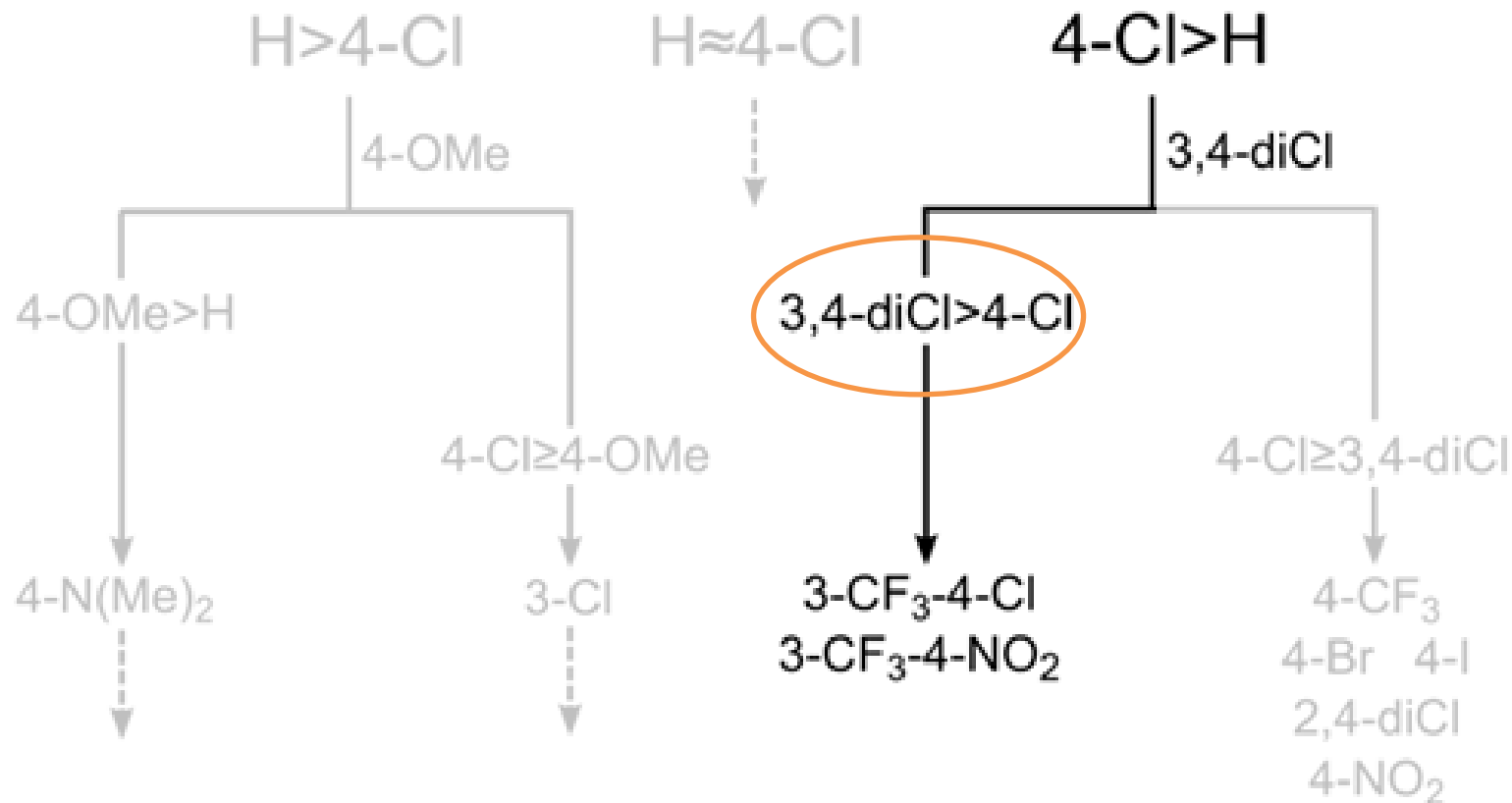
TOPLISS DECISION TREE



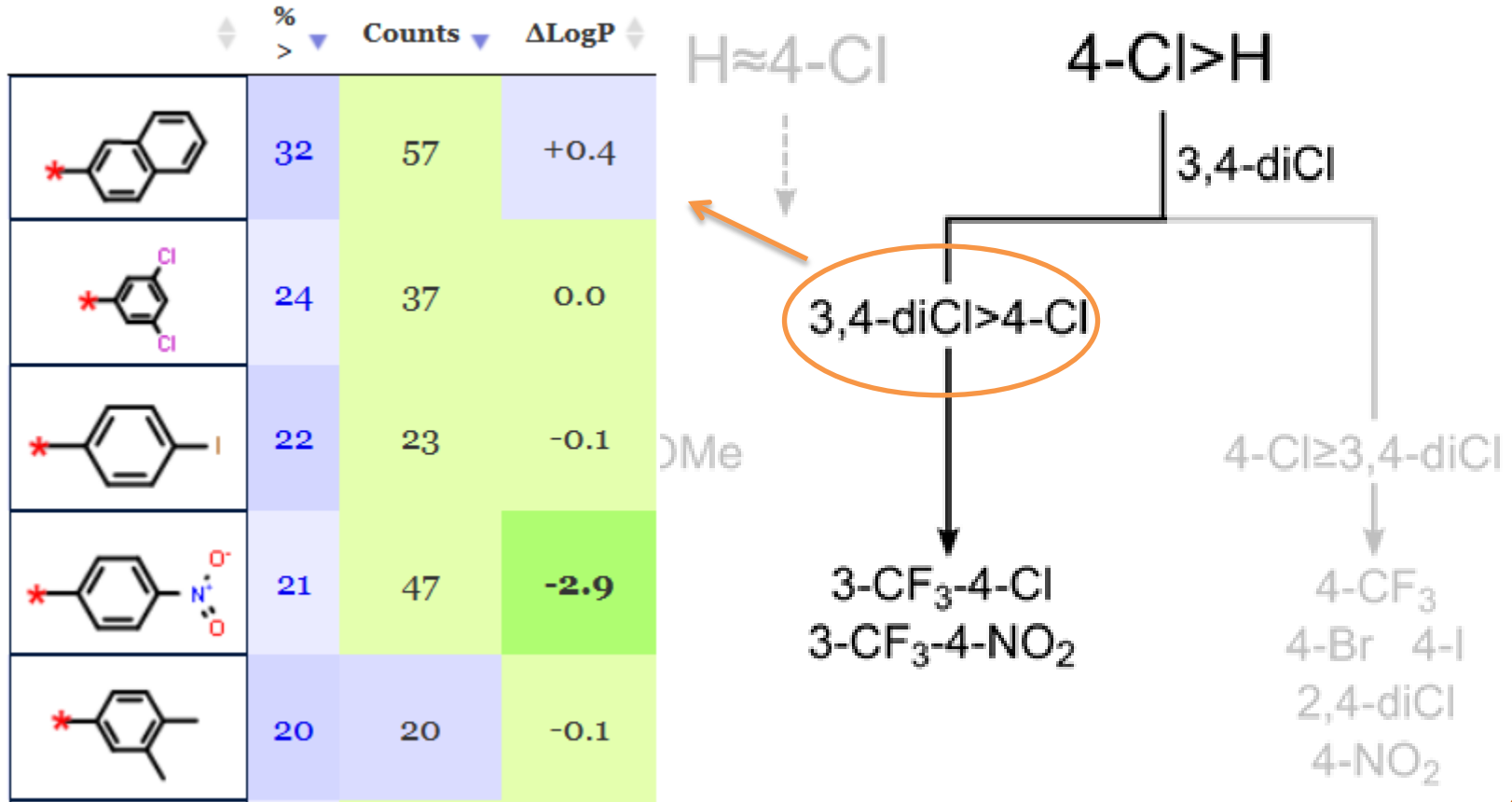
TOPLISS DECISION TREE




TOPLISS DECISION TREE



TOPLISS DECISION TREE

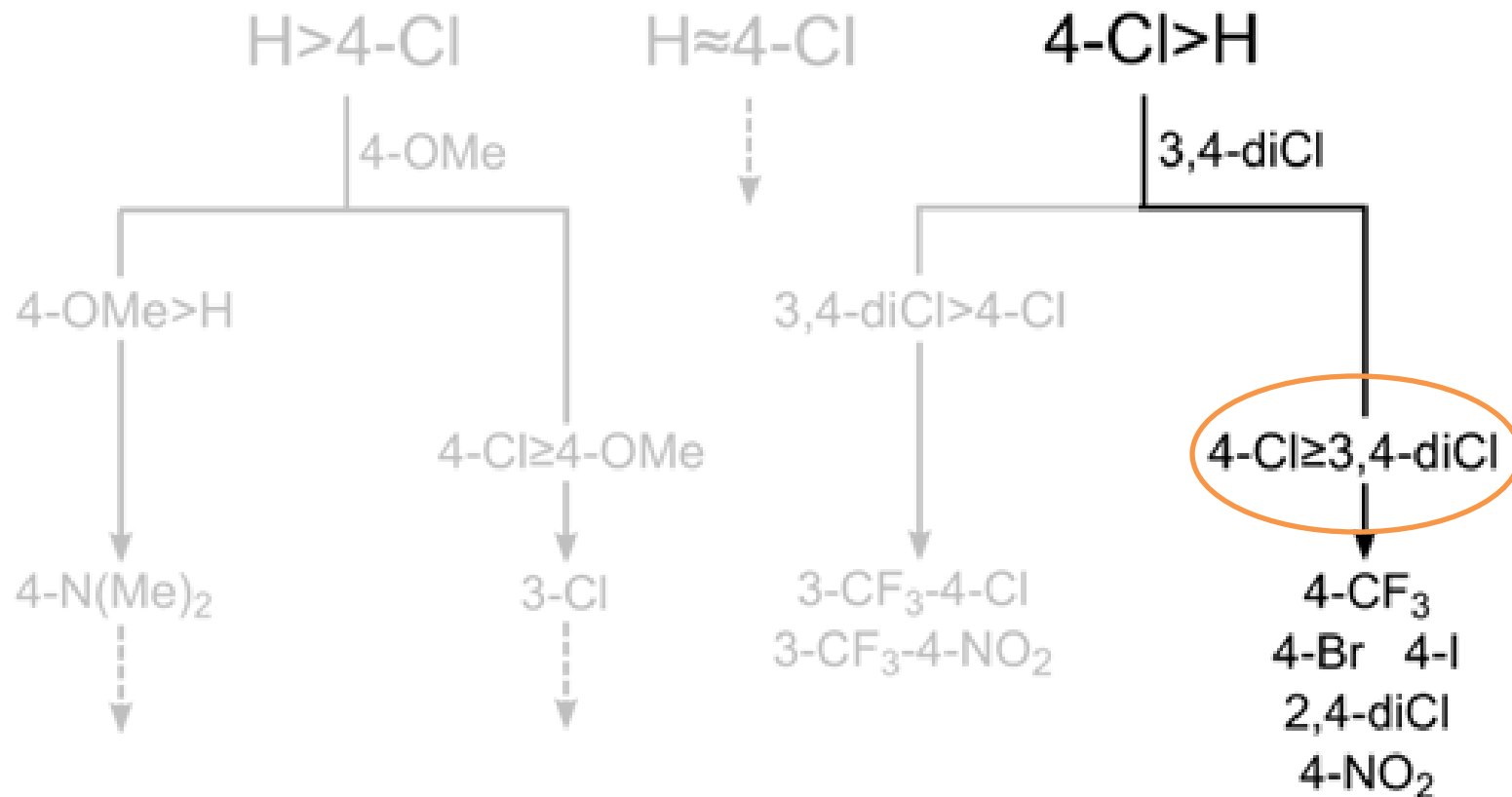


(1st if lower cutoff)

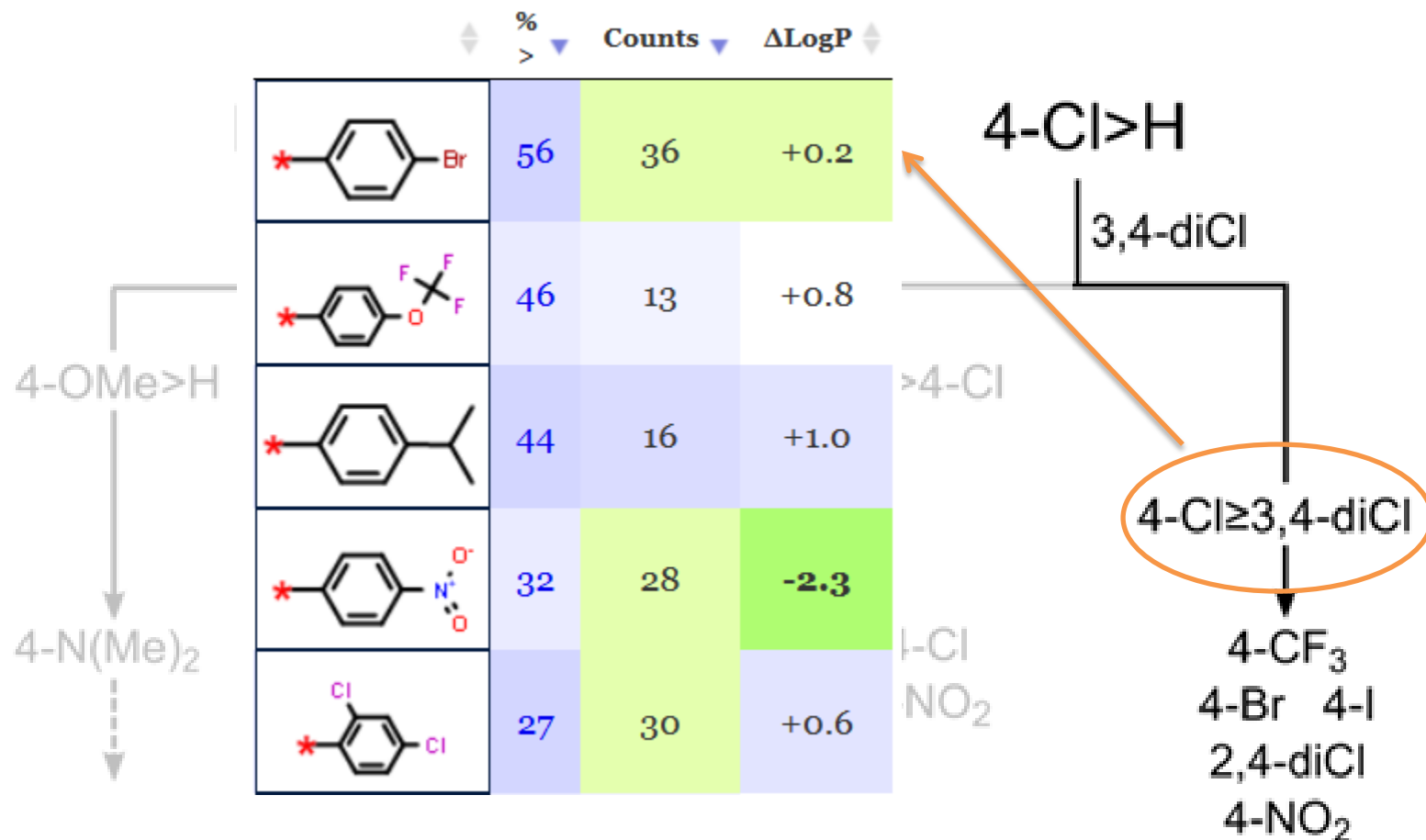
	33	5.00	15	+0.3
---	----	------	----	------



TOPLISS DECISION TREE



TOPLISS DECISION TREE

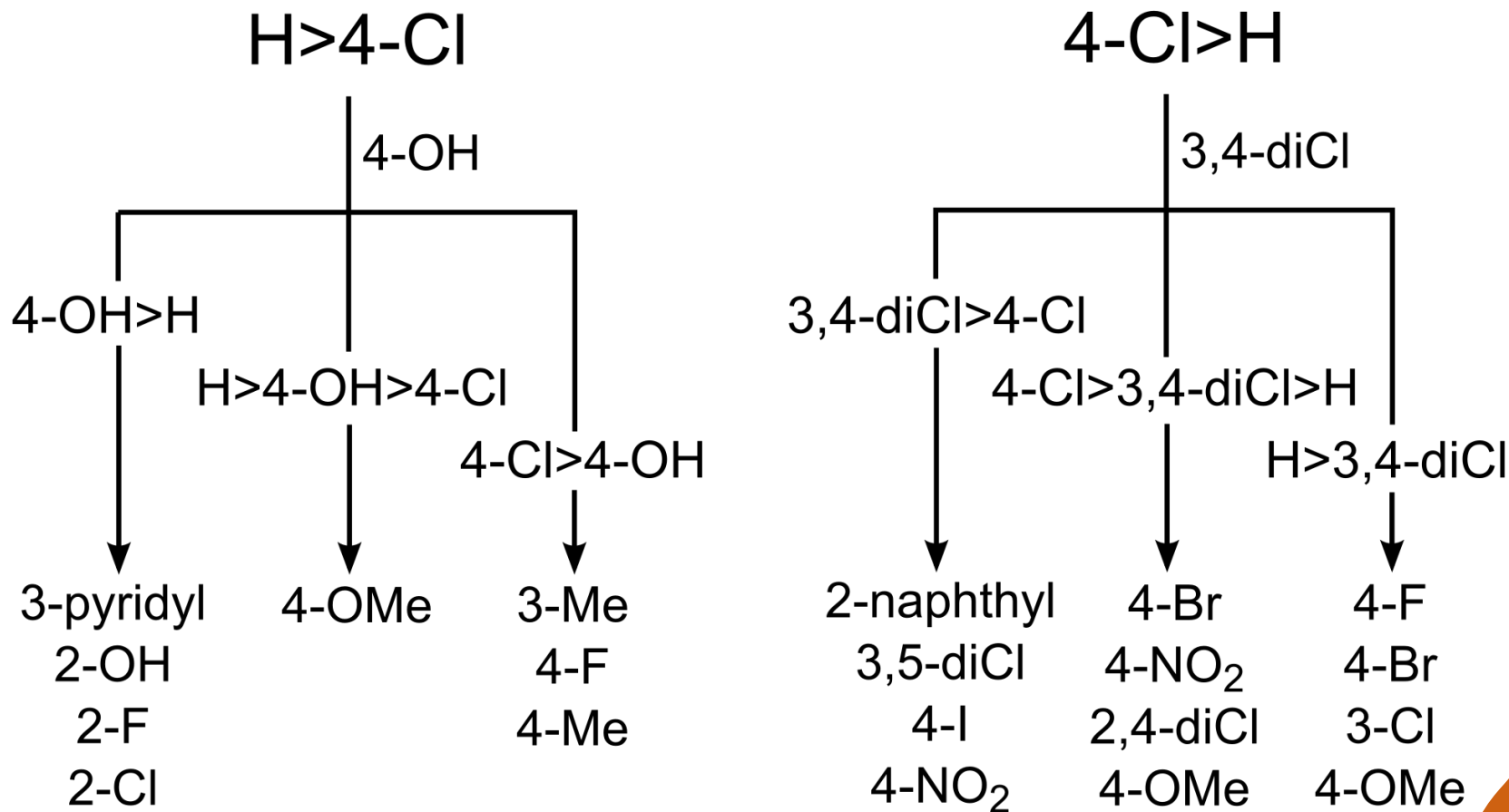


(20th)

	11	28	+0.3
---	----	----	------



MATSY DECISION TREE (ONE OF MANY)



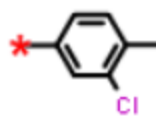
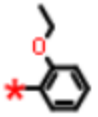
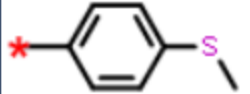
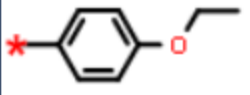
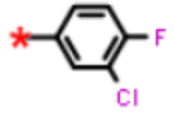
MODIFYING THE PREDICTIONS FOR

4-Cl > H

	% >	Counts
	67	30
	47	30
	46	24
	44	25
	42	77

Kinases

Target-specific

	% >	Counts	ΔLogP
	63	27	+0.3
	55	20	-0.4
	49	63	0.0
	48	46	-0.4
	48	46	+0.1

$\Delta\text{LiPE} > 0$

Incorporate metrics

DRAG-AND-DROP INTERFACE TO MATSY

1/2 3 4 5 6 7 8 9 10 11 12 **Ph 1** Ph 2 Ph 3 Ph 4 Ph 5/6 Custom

Stronger binding < ChEMBL19 pIC50 > Weaker binding

? > [] > Clc1ccc(Cl)cc1 > Clc1ccc(cc1) > []

2277

	Counts	ΔLogP
<chem>Clc1ccc(Cl)cc1</chem>	54 391	+0.6
<chem>c1ccc2ccccc2c1</chem>	54 37	+2.0
<chem>CCCC1=CC=CC=C1</chem>	53 30	+1.1
<chem>Fc1ccc(Cl)cc1</chem>	52 46	+0.1
<chem>BrC1=CC=CC=C1</chem>	50 521	+0.2
<chem>Fc1ccc(Cl)cc1</chem>	50 32	+0.1
<chem>S1=CC=CC=C1</chem>	49 63	0.0
<chem>COCC1=CC=CC=C1</chem>	48 46	-0.4
<chem>Clc1ccc(Cl)cc1</chem>	48 25	+0.3
<chem>Oc1ccc(O)cc1</chem>	48 21	-2.2

Showing 11 to 20 of 111 entries

Previous Next

EVIDENCE-BASED MEDICINAL CHEMISTRY

- Take advantage of the wealth of experience contained in 57k med chem papers
- Synthetic decisions backed by **experimental data**
 - Can drill-down into the data, look at targets, scaffolds
 - Can restrict experimental data used to particular targets, use in-house data rather than ChEMBL
- Does not explain **why**, only that it happens
 - Lots of interesting scientific questions (hint hint)



Evidence-based medicinal chemistry

...using matched molecular series

noel@nextmovesoftware.com



Using Matched Molecular Series as a
Predictive Tool To Optimize Biological
Activity

J. Med. Chem. **2014**, *57*, 2704.

Slides (soon!) at
<http://slideshare.net/nextmovesoftware>

