
Medicinal Chemist's Relationship with Additivity: Are we Taking the Fundamentals for Granted?

J. Guy Breitenbucher

*Streamlining Drug Discovery Conference
San Francisco, CA
Oct. 25, 2018*



IND

Institute for Neurodegenerative Diseases

Additivity as the Birth of QSAR

Journal of Medicinal Chemistry

© Copyright 1964 by the American Chemical Society

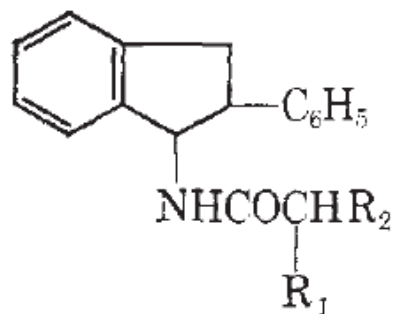
VOLUME 7, NUMBER 4

JULY 6, 1964

A Mathematical Contribution to Structure-Activity Studies

SPENCER M. FREE, JR., AND JAMES W. WILSON

Research and Development Division, Smith Kline and French Laboratories, Philadelphia, Pennsylvania



R_2	R_1		Average
	H	CH_2	
$N(CH_3)_2$	2.13	1.64	1.885
$N(C_2H_5)_2$	1.28	0.85	1.065
	<u>1.705</u>	<u>1.245</u>	<u>1.475</u>

Red annotations: $-0.8 \downarrow$ next to the $N(CH_3)_2$ row, $-0.8 \downarrow$ next to the $N(C_2H_5)_2$ row, and -0.5 between the two columns of R_1 values.

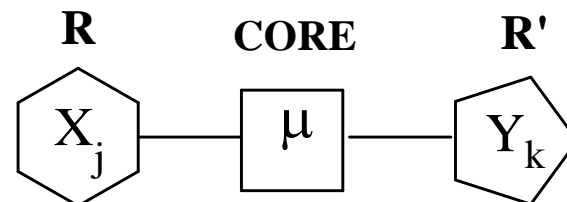
$$\text{Response} = \text{Average} + X_i + Y_j$$

“The proposed models should not be criticized as ignoring the combination of several substituents that produce a biological response in excess of the additive estimation. Such results will appear in some analog series.”

Free-Wilson Modifications

➤ *Fujita-Ban (1971)*

- $-\log A = \mu + X_j + Y_k$
 - Modified Free-Wilson by using log units and making μ the activity of unsubstituted core, X and Y are the activity contribution of the substituents in two different positions on the core.
 - X and Y are determined empirically for a specific biological system and position on the core.



➤ *Bocek and Kopecky (1965)*

- $A = \text{Avg} + X_j + Y_k + e_j e_k$
 - SAR was not always additive so they introduced the concept of an interaction term.

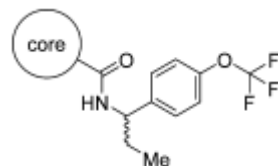
$$e_j e_k = pK_i - (\mu + X_j + X_k)$$

We Usually Assume Additivity by Inference

- *We analyzed 500 SAR papers (JMC, BOMCL, ACSMCL) and found only 4 where the additive assumption was checked!!*
- *Most papers follow a standard progression of linear SAR:*

Assuming Additivity by Inference

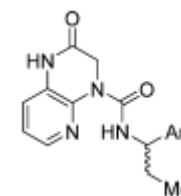
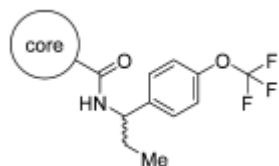
➤ We analyzed 500 SAR papers (JMC, BOMCL, ACSMCL) and found only 4 where the additive assumption was checked!!



compd	core	anticipated binding mode ^a	PDE2A IC ₅₀ (nM) ^b
4a		(A)	53 (47–59)
5		(A)	480 (350–650)
6		(A) or (B)	78 (61–98)
7		(A) or (B)	21 (18–26)
8		(C)	6500 (4200–9800)
9		(C) or (D)	19000 (15000–24000)
10		(A) or (B)	70000 (40000–120000)

Assuming Additivity by Inference

➤ We analyzed 500 SAR papers (JMC, BOMCL, ACSMCL) and found only 4 where the additive assumption was checked!!

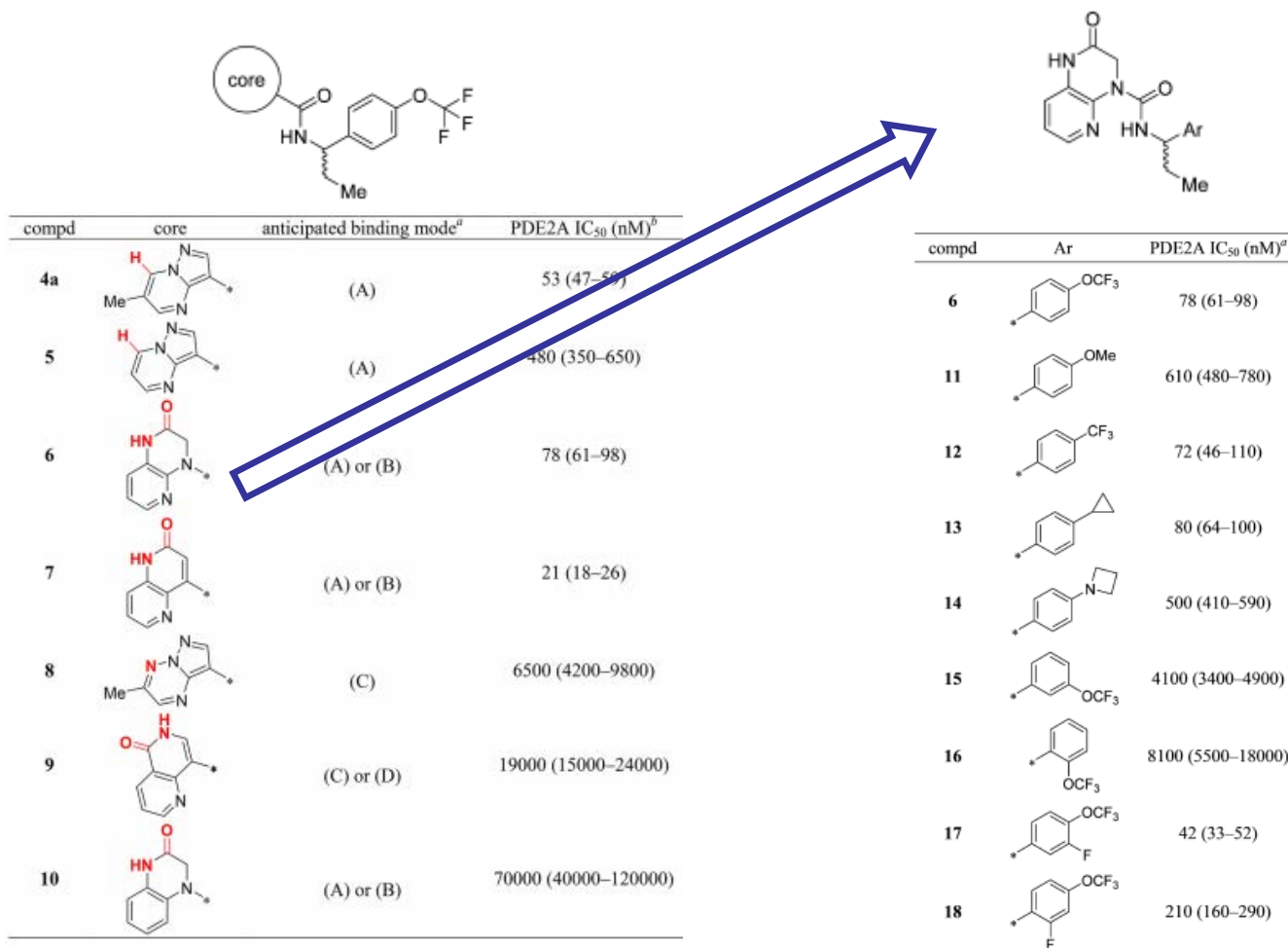


compd	core	anticipated binding mode ^d	PDE2A IC ₅₀ (nM) ^b
4a		(A)	53 (47–59)
5		(A)	480 (350–650)
6		(A) or (B)	78 (61–98)
7		(A) or (B)	21 (18–26)
8		(C)	6500 (4200–9800)
9		(C) or (D)	19000 (15000–24000)
10		(A) or (B)	70000 (40000–120000)

compd	Ar	PDE2A IC ₅₀ (nM) ^e
6		78 (61–98)
11		610 (480–780)
12		72 (46–110)
13		80 (64–100)
14		500 (410–590)
15		4100 (3400–4900)
16		8100 (5500–18000)
17		42 (33–52)
18		210 (160–290)

Assuming Additivity by Inference

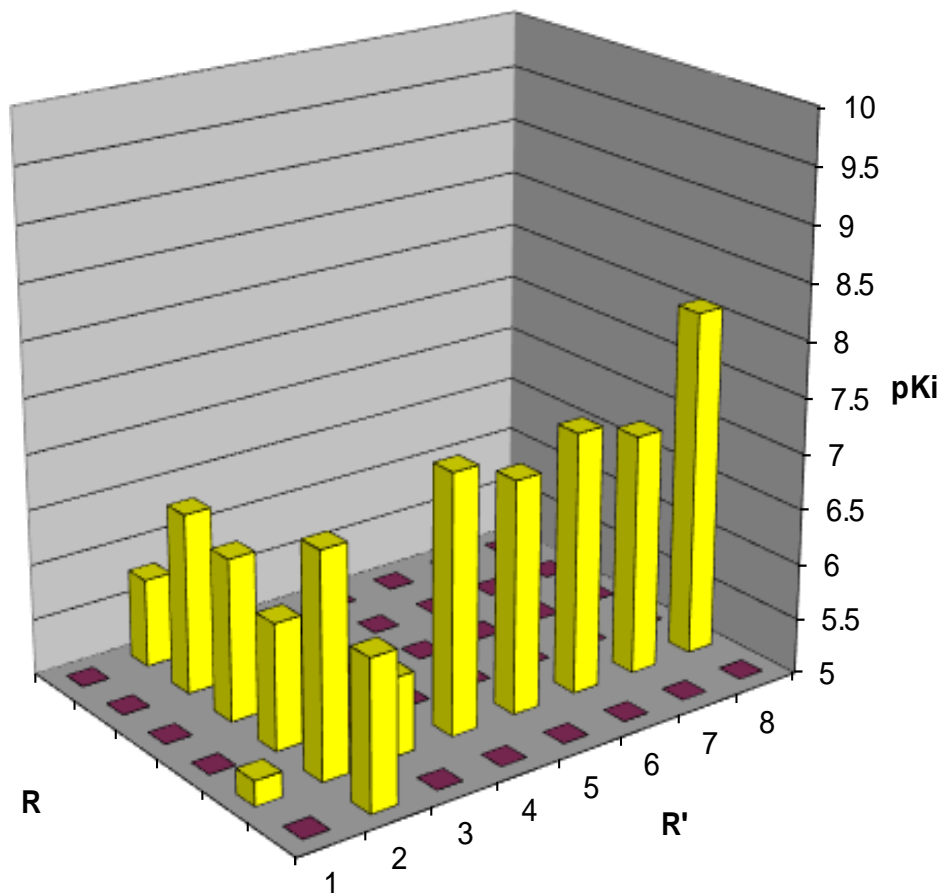
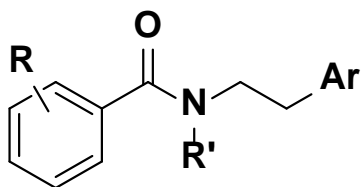
➤ We analyzed 500 SAR papers (JMC, BOMCL, ACSMCL) and found only 4 where the additive assumption was checked!!



Assuming Additivity by Inference

- *We analyzed 500 SAR papers (JMC, BOMCL, ACSMCL) and found only 4 where the additive assumption was checked!!*
- *I think chemists are aware of the possibility that non-additivity exists, but they are so comfortable with the assumption that they don't think it is important to check.*
- *I will contend that checking additivity not only prevents chemists from missing good compounds but also highlights critical changes in SAR.*

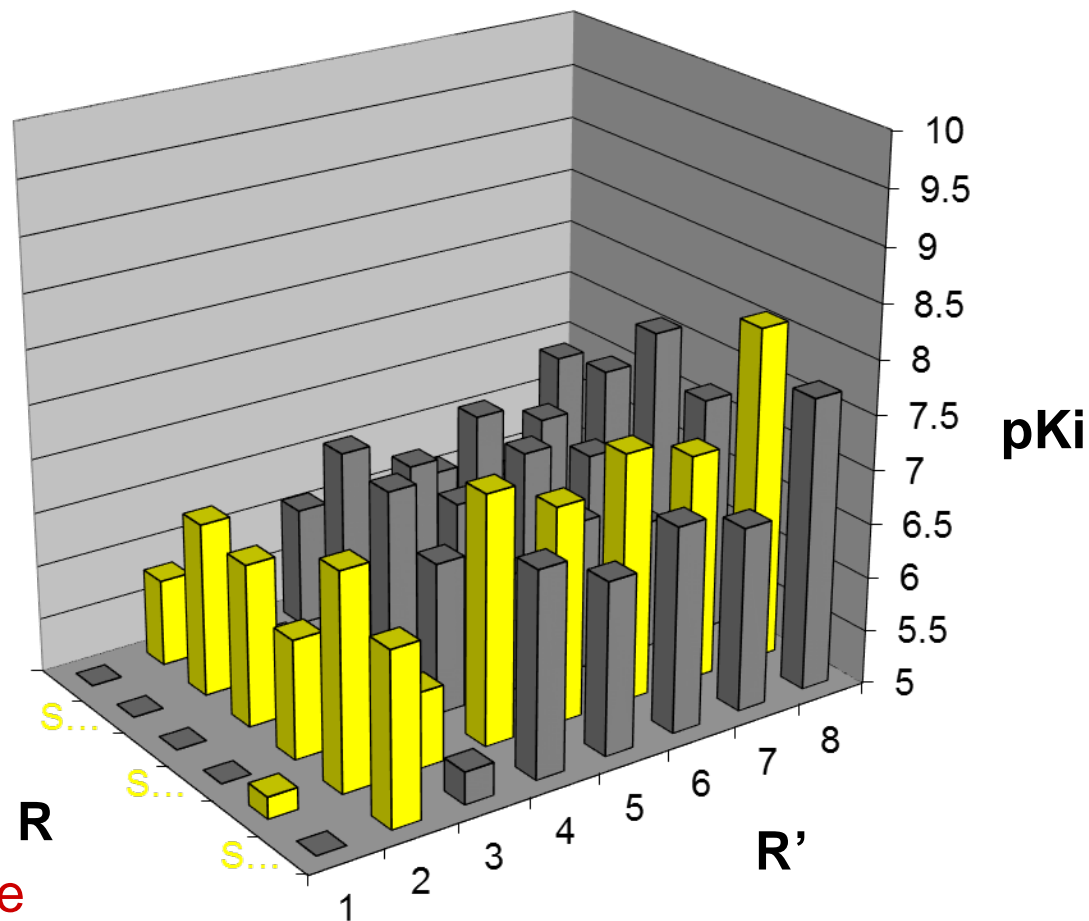
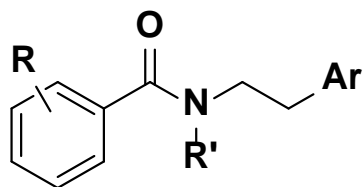
What are we assuming with linear SAR?



JNJ SAR of Orexin Antagonists

What are we assuming with linear SAR?

This is a visual quantitative representation of assuming additivity.

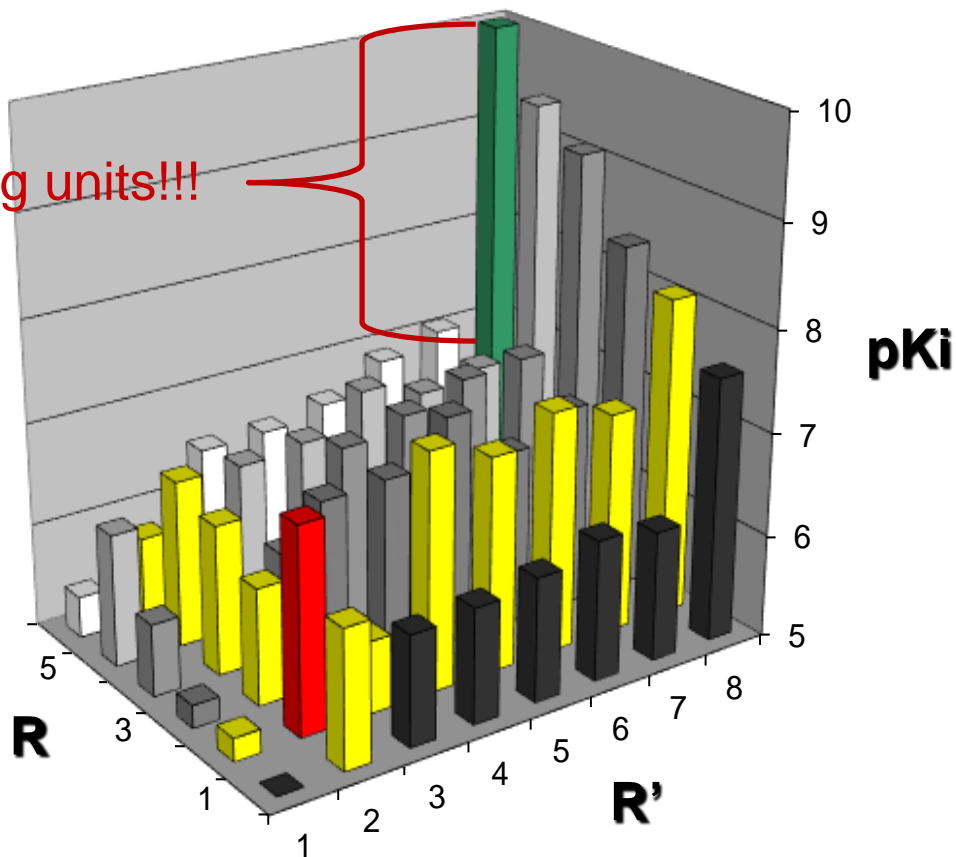
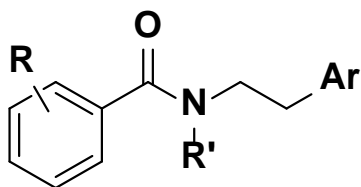


$$pK_i = \mu + X_j + Y_k : \text{Additive}$$

SAR of Orexin Antagonists

What if your data looks like this?

Non-Additivity $e_j e_k = 3 \log \text{ units!!!}$



*2 other Orexin libraries showed a similar lack of additivity.
...in 500 medchem papers we only checked additivity 4 times!!!*

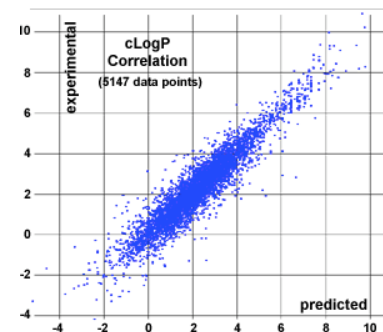
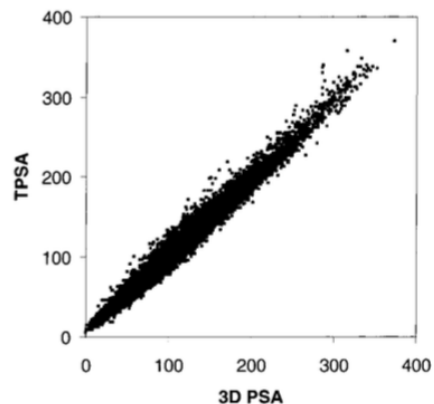
Why do we assume additivity?

It starts from the fundamentals of our chemistry training (Anchoring effect).

- Thermodynamics: $\Delta H = \sum \Delta H_{(\text{bonds broken})} - \sum \Delta H_{(\text{bonds formed})}$ or $\Delta S = \Delta S_{\text{trans}} + \Delta S_{\text{rot}} + \Delta S_{\text{vib}}$
- clogP: adding contributions of every atom based on its atom type
- tPSA = \sum Polar fragments

Table 1. Atomic Contributions (\AA^2) to PSA

atom type ^a	PSA contrib	atom type ^a	PSA contrib
[N](-*)(-*)-*	3.24	[nH](c:*)	15.79
[N](-*)=*	12.36	[n+](c:*)(:*)	4.10
[N]#*	23.79	[n+](-*)(:*)	3.88
[N](-*)(=*)=**b	11.68	[nH+](c:*)	14.14
[N](=*)#**c	13.60	[O](-*)-*	9.23
[N](-*)(-*)-*-1 ^d	3.01	[O](-*)-*-1 ^d	12.53
[NH](-*)-*	12.03	[O]=*	17.07
[NH]1-1*-1 ^d	21.94	[OH]1-*	20.23
[NH]1-*	23.85	[O-]1-*	23.06
[NH2]1-*	26.02	[o](c:*)	13.14
[N+](-*)(-*)(-*)(-*)-*	0.00	[S](-*)-*	25.30
[N+](-*)(-*)(-*)=*	3.01	[S]1=*	32.09
[N+](-*)#**e	4.36	[S](-*)(-*)(=*)	19.21
[NH+](-*)(-*)(-*)-*	4.44	[S](-*)(-*)(=*)=*	8.38
[NH+](-*)(=*)=*	13.97	[SH]1-*	38.80
[NH2+](-*)(-*)-*	16.61	[s](c:*)	28.24
[NH2+]1=*	25.59	[s](=*)(c:*)	21.70
[NH3+]1=*	27.64	[P](-*)(-*)(-*)-*	13.59
[n](c:*)	12.89	[P](-*)(-*)(=*)	34.14
[n](c:*)(c:*)	4.41	[P](-*)(-*)(-*)(-*)=*	9.81
[n](-*)(c:*)	4.93	[PH](-*)(-*)(-*)=*	23.47
[n](=*)(c:*)*/	8.39		

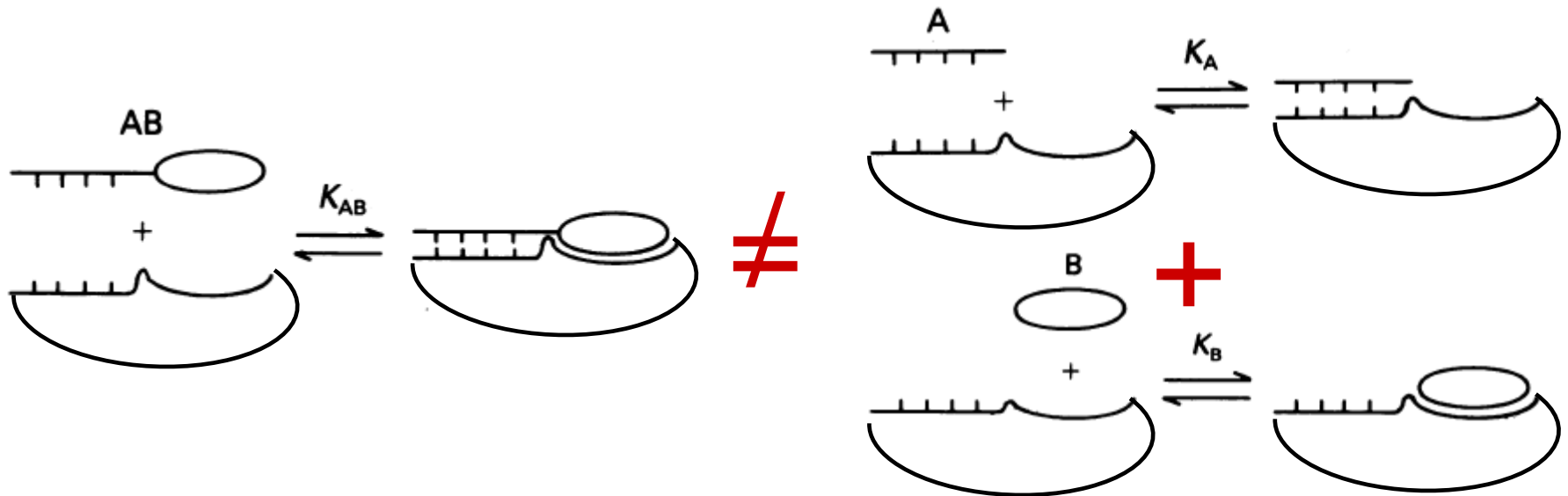


Ligand Protein Interactions are not Bulk Physical Properties

How would you respond if I said that a phenyl ring contributed the same amount of energy to all ligand-protein binding events, regardless of context?

Yet that is exactly how we treat a phenyl ring when we calculate physical descriptors?
(clogP, TPSA, MW, MolVol, ΔH of formation).

...and yet we expect that simple physical properties can predict biology data?

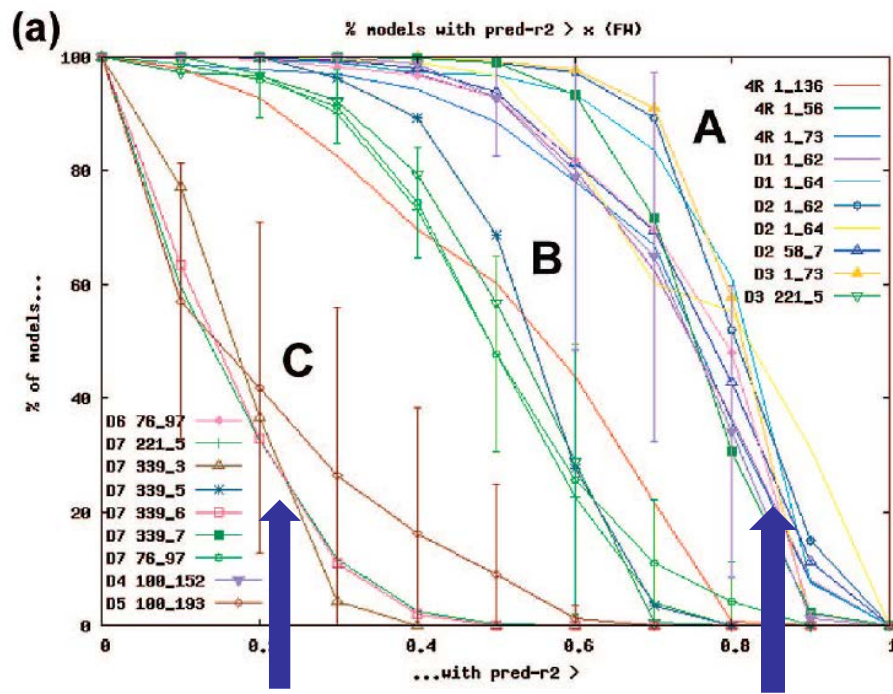
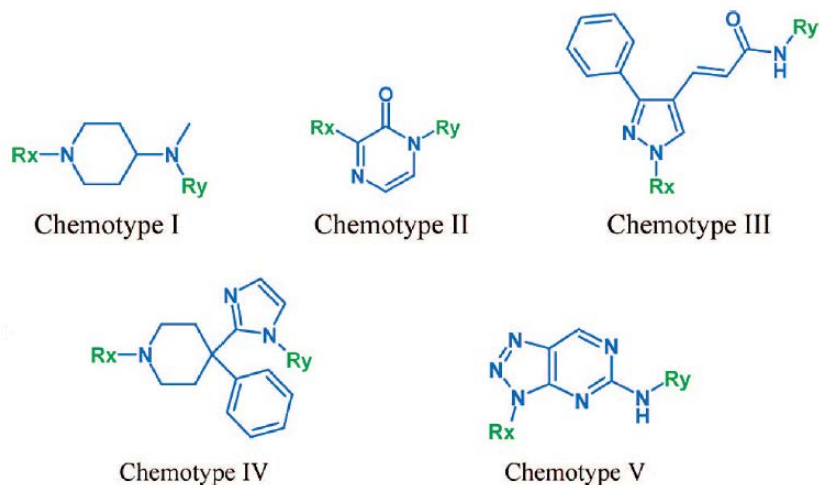


Entropy, fit, and strain all play key roles in our thinking about additivity.

How general is additivity in SAR?

- Janssen data mining exercise.
 - 19 SAR data sets
 - Sets had 40-178 members
 - At least 2 variable groups
 - Range of data values >2 log units
 - Sets were 80-66% complete

$$y = C + \sum_{i=1}^n a_i X_i + \sum_{j=1}^m b_j Y_j$$

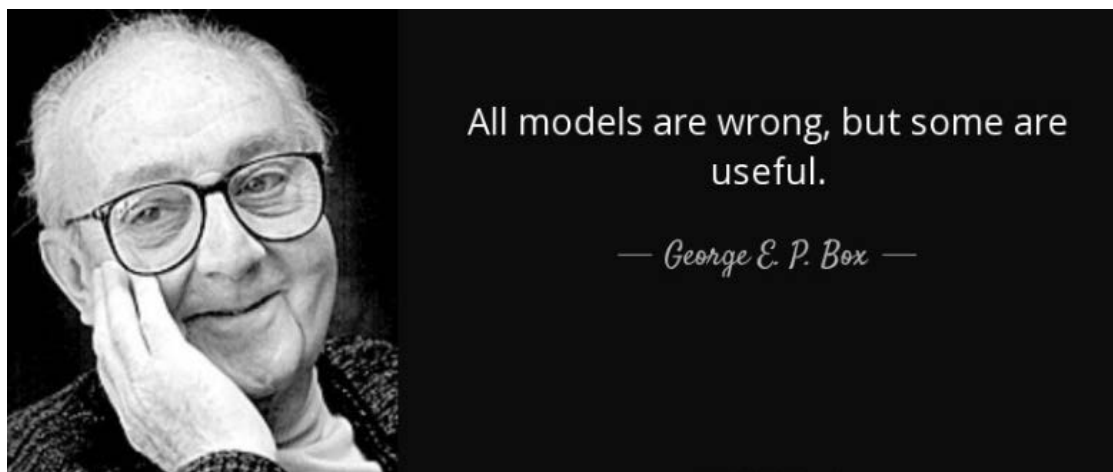


4 Clearly non-additive

10 Mostly additive

Where does this leave us?

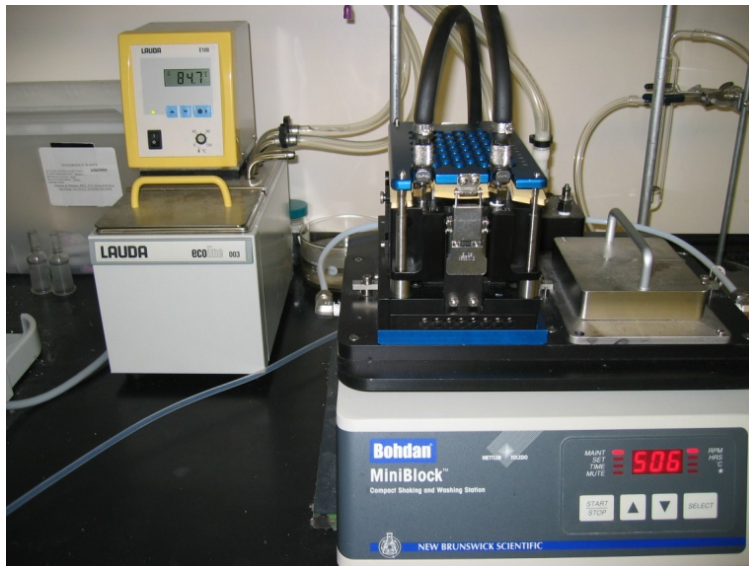
- Additivity is an attractive assumption and useful when true.
- However, we know that for ligand-protein SAR the assumption of additivity is tenuous at best.
- ...but we never seem to actually check when we do SAR?



- Is there a way to conduct and analyze our SAR differently?

Matrix Synthesis is Practical

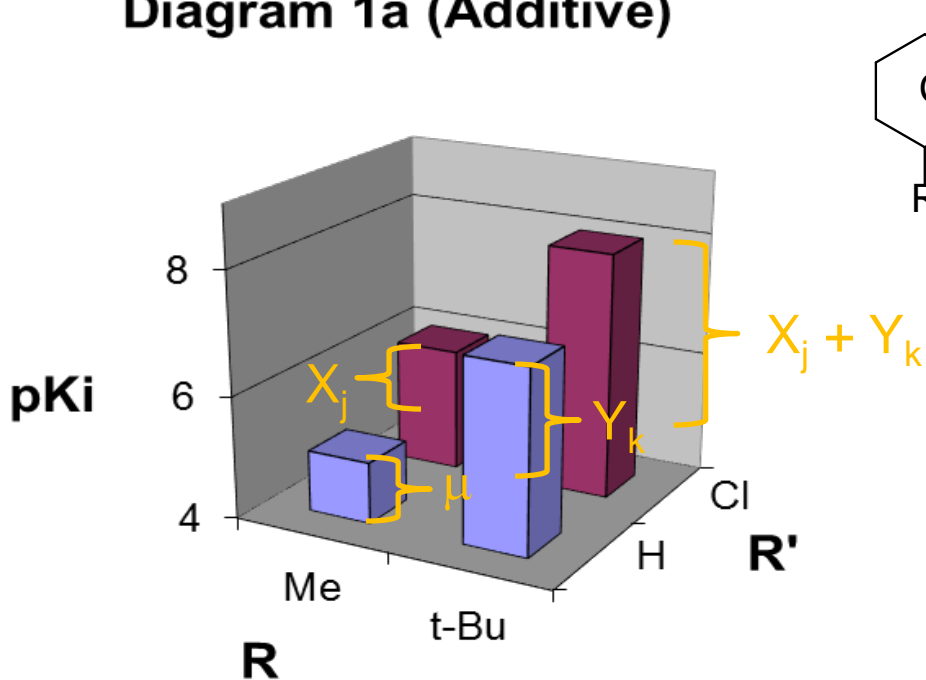
Since 1990s we have the tools to make all the compounds in a matrix.



Yet we still don't do it 95% of the time!

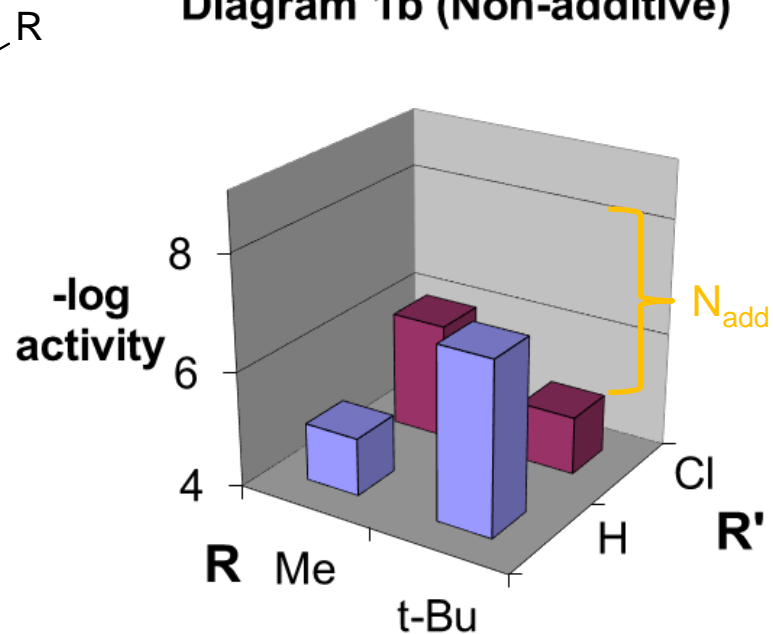
Visualizing Additivity in SAR

Diagram 1a (Additive)



$$pKi = \mu + X_j + Y_k : \text{Additive}$$

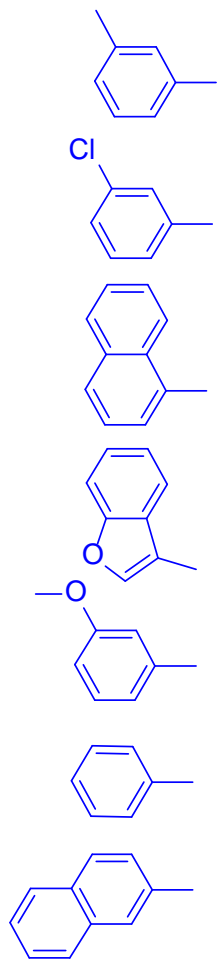
Diagram 1b (Non-additive)



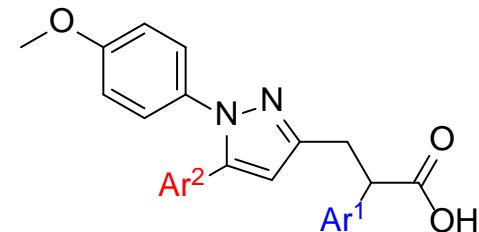
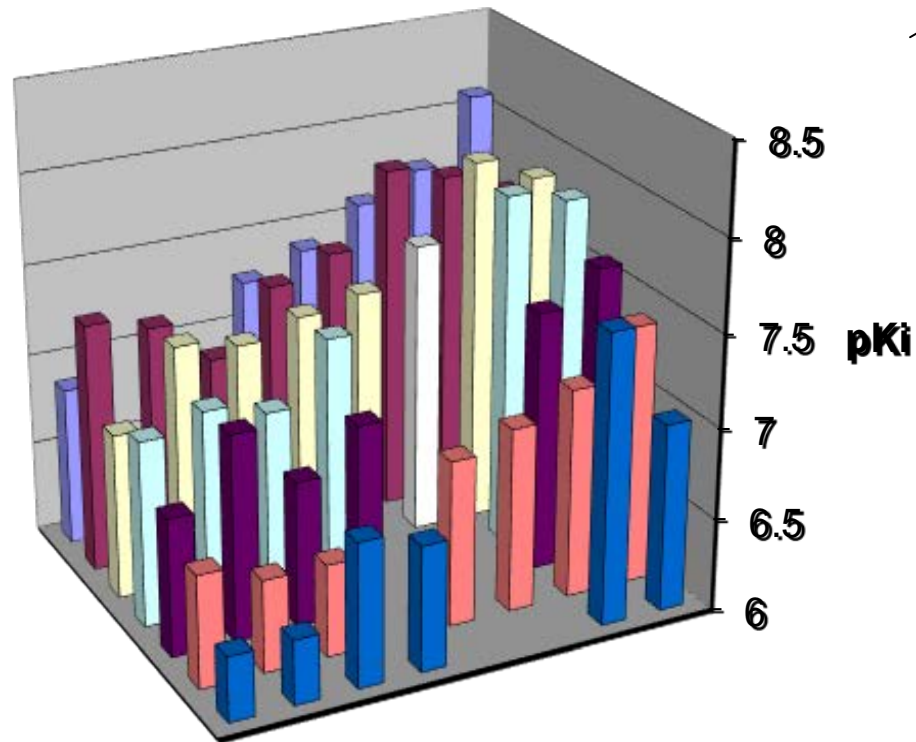
$$pKi - (\mu + X_j + Y_k) = e_j e_k = N_{\text{add}} = \text{non-zero}$$

Therefore $e_x e_y$ is a term that can be calculated for a system and is equal to the degree of non-additivity in the system.

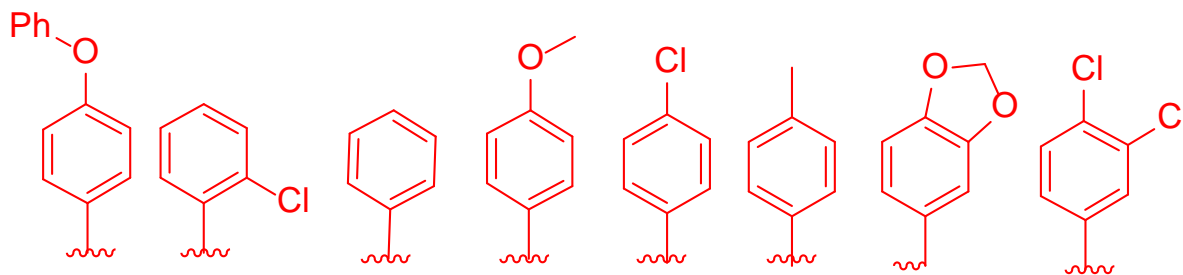
Library of CCK₁ Antagonists



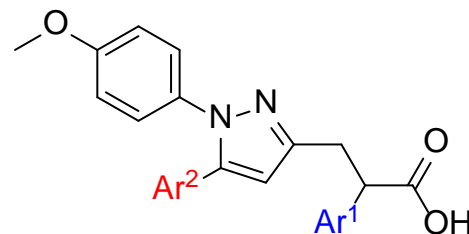
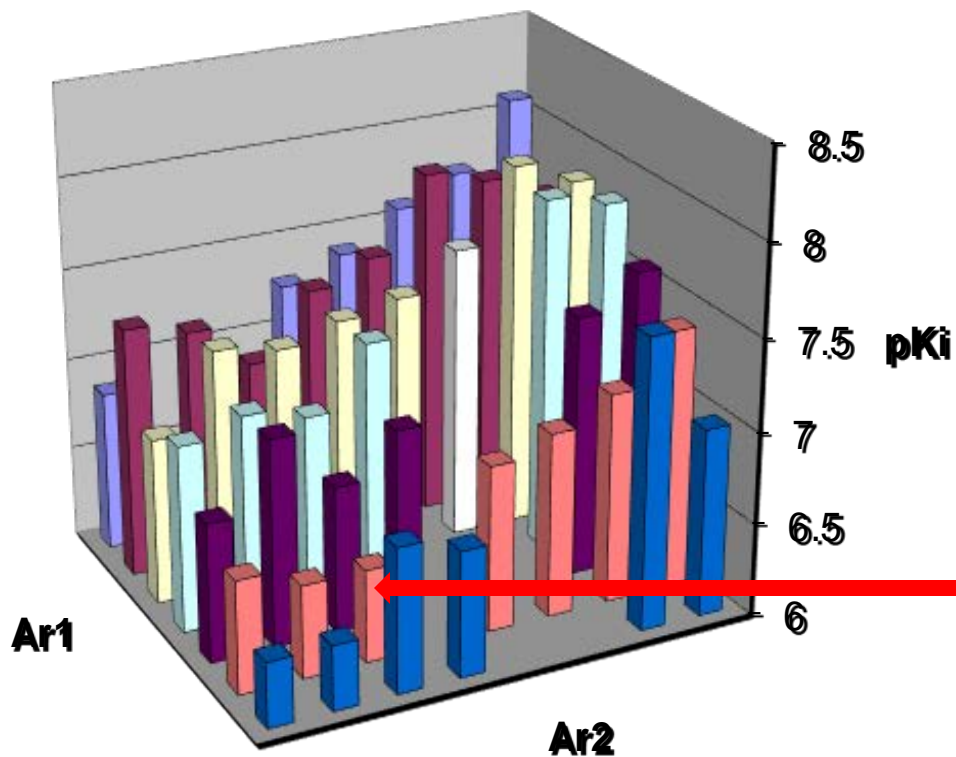
Ar1



Ar2



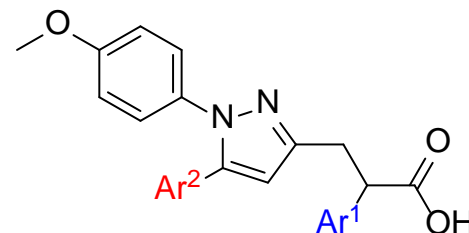
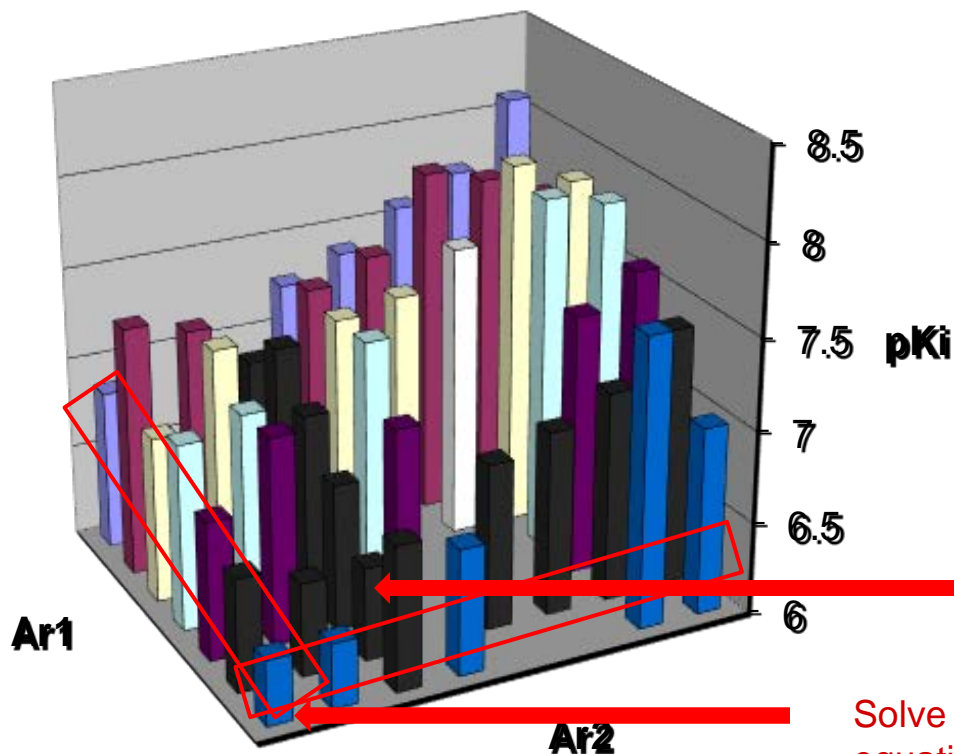
How to determine additivity?



$$ApK_{i(jk)} = \mu + X_j(Ar^1_j) + Y_k(Ar^2_k)$$

$$\mu = 6.5$$

How to do the Math?



$$pK_{i(jk)} = \mu + X_j(\text{Ar}^1_j) + Y_k(\text{Ar}^2_k)$$

$$\mu = 6.5 \quad (X_0 = 0; Y_0 = 0)$$

Solve for X_j and Y_k by best fit regression using the 42 equations generated.

Examples:

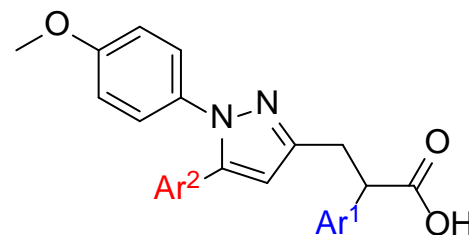
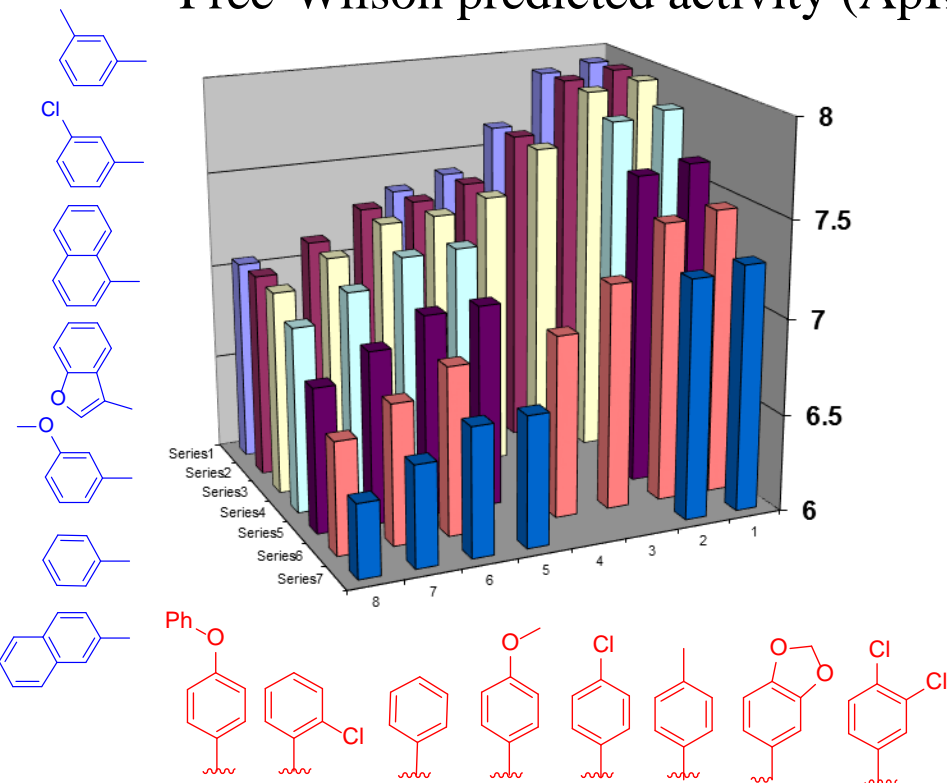
6 + 7 = 13 Variables
6 x 7 = 42 Equations

6.3 = 6.5 + X1 + Y1	6.3 = 6.5 + X1 + Y1
6.7 = 6.5 + X2 + Y1	6.3 = 6.5 + X1 + Y2
7.1 = 6.5 + X3 + Y1	6.7 = 6.5 + X1 + Y3
7.0 = 6.5 + X4 + Y1	NA = 6.5 + X1 + Y4
7.4 = 6.5 + X5 + Y1	NA = 6.5 + X1 + Y5
7.0 = 6.5 + X6 + Y1	7.6 = 6.5 + X1 + Y6
	7.0 = 6.5 + X1 + Y7

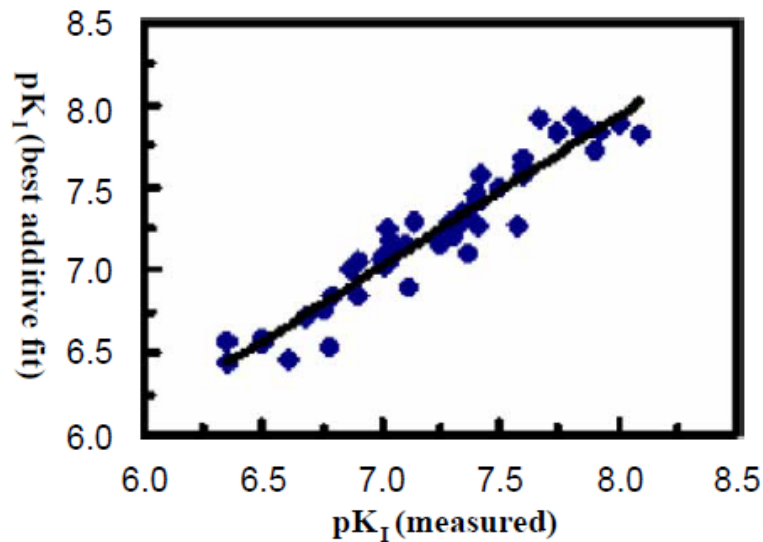
ect...

Additive Analysis using Free Wilson

Free-Wilson predicted activity (ApK_i)



Additive Correlation

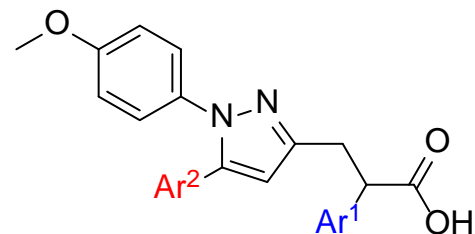
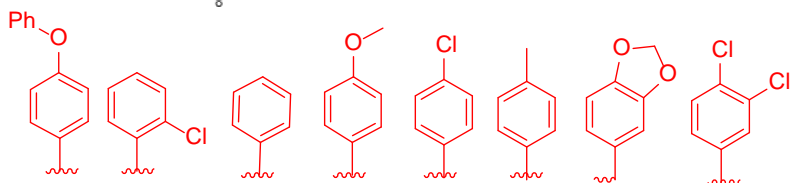
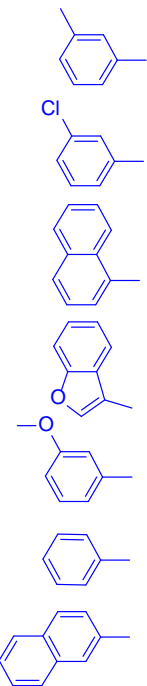
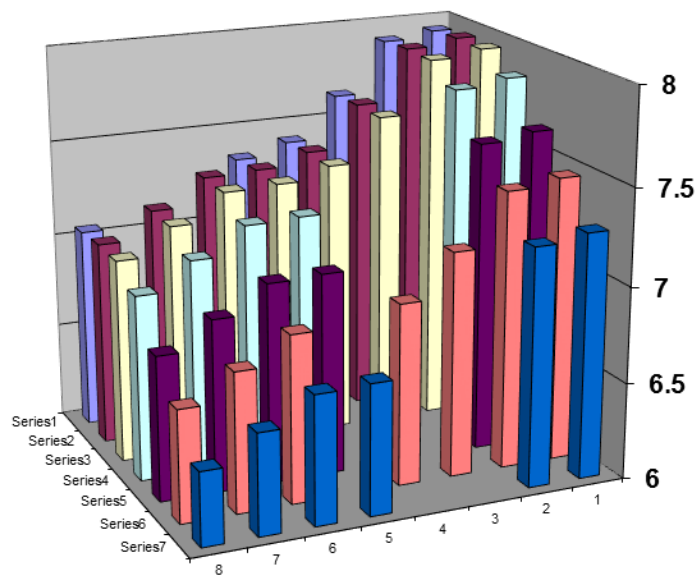


Through linear regression we find the best possible value for each X_j and Y_k with the actual data. Then plug those values back in to get the ApK_i for each compound = $\mu + X_j + Y_k$

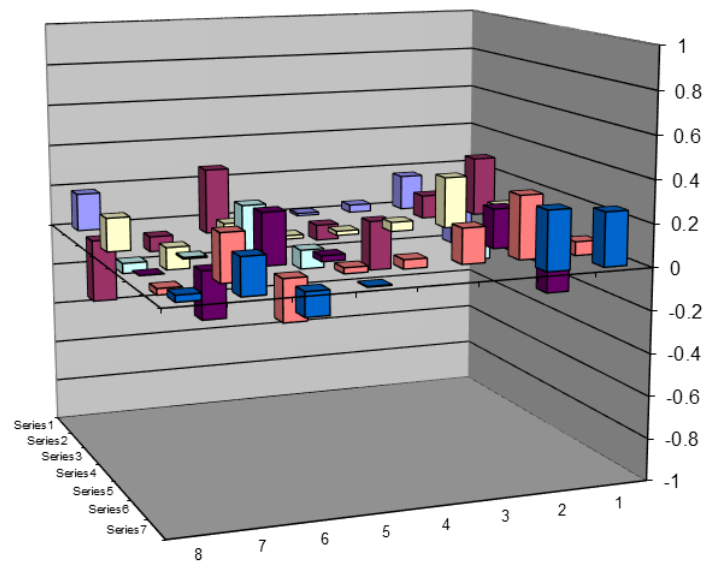
In this case the additive model works!

Additive Analysis using Free Wilson

Free-Wilson predicted activity (A_{pKi})

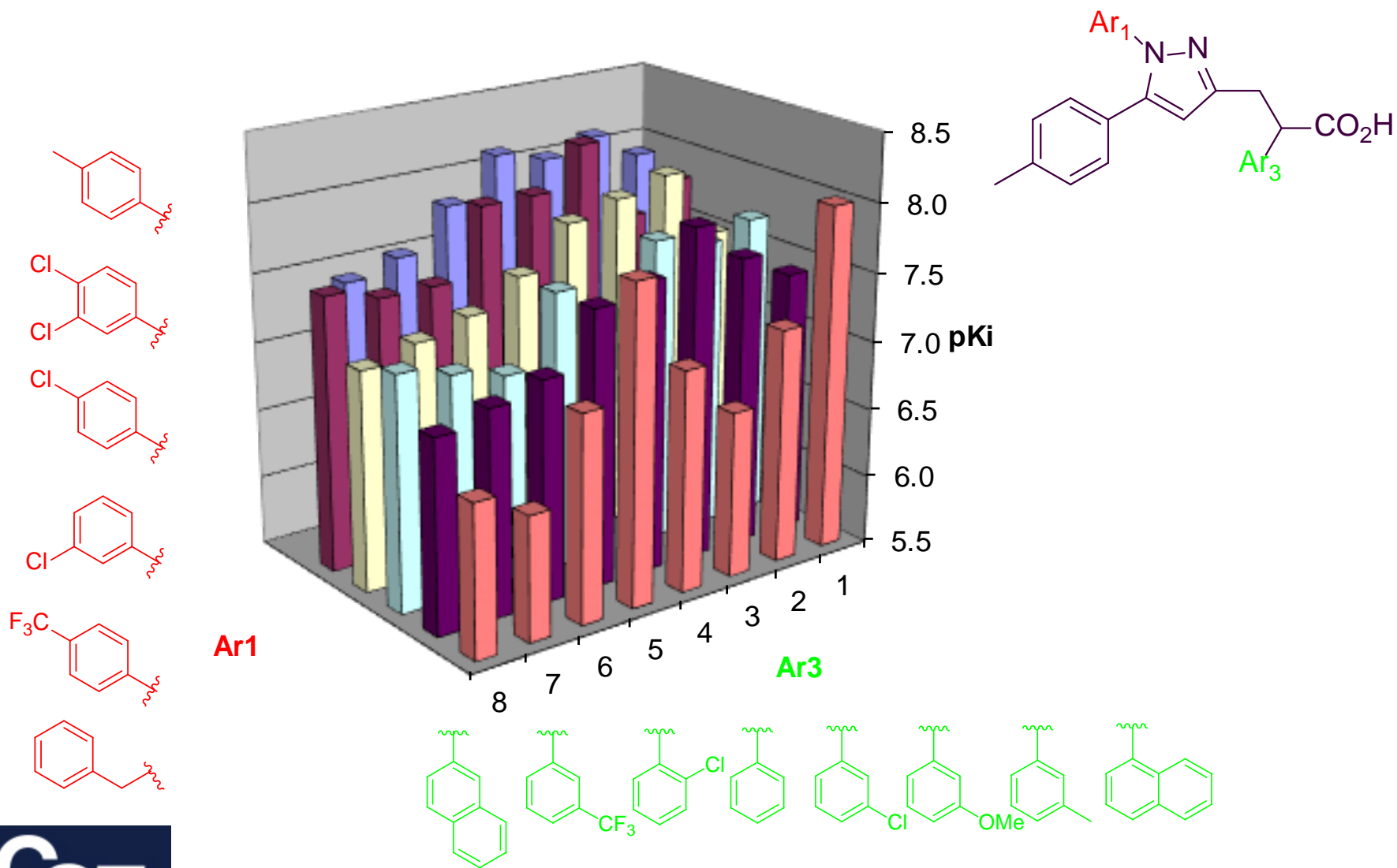


Deviation from prediction

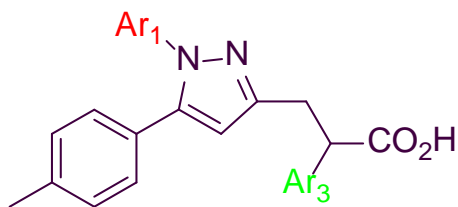


$$\text{Deviation from additivity} = D_{\text{add}} = pKi - A_{pKi} = e_x e_y$$

Library 2

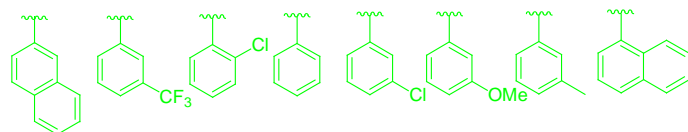
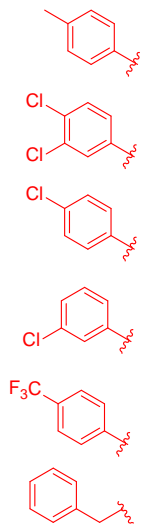
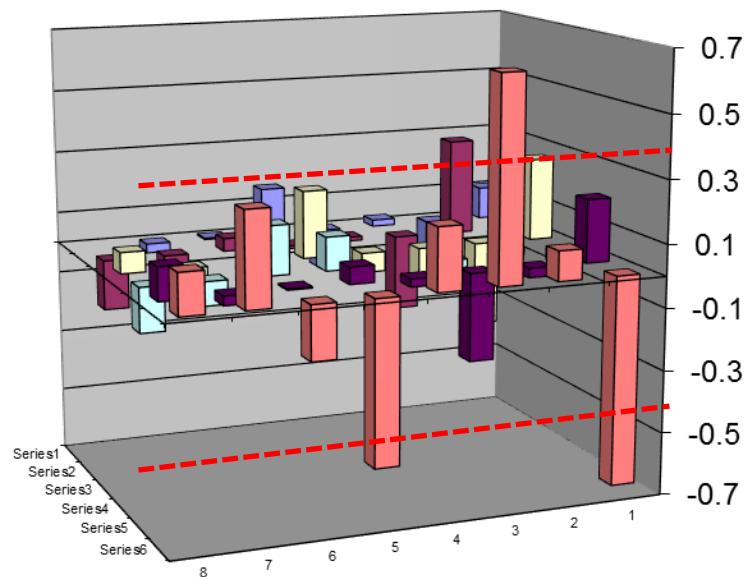
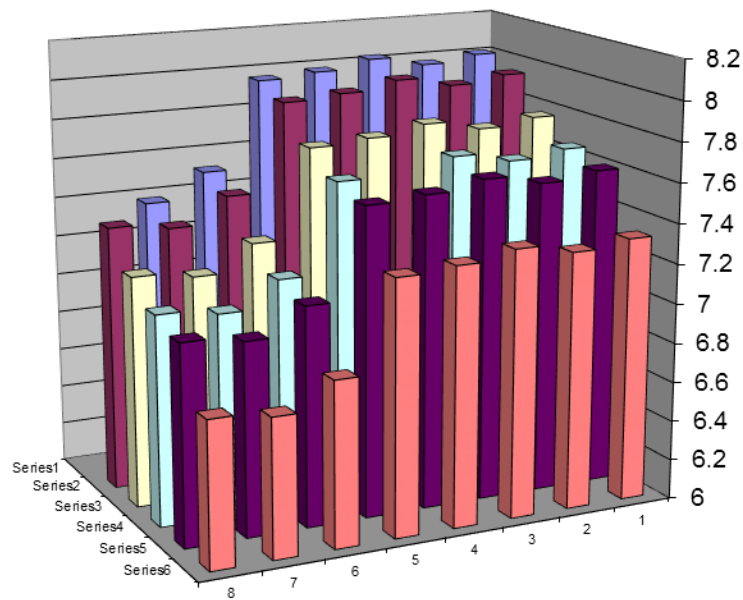


Additivity Analysis of Second Library



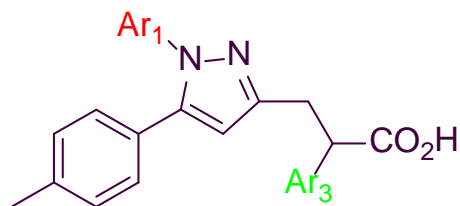
Free-Wilson Predicted activity

Deviation from Predicted



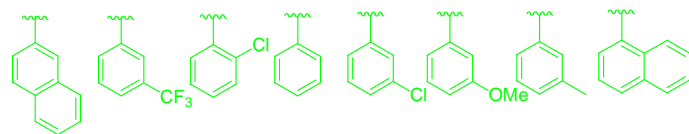
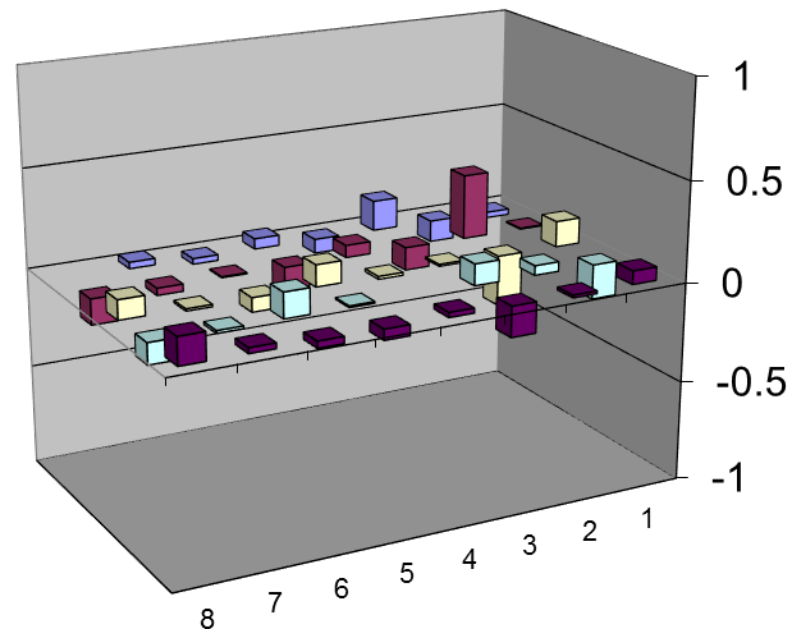
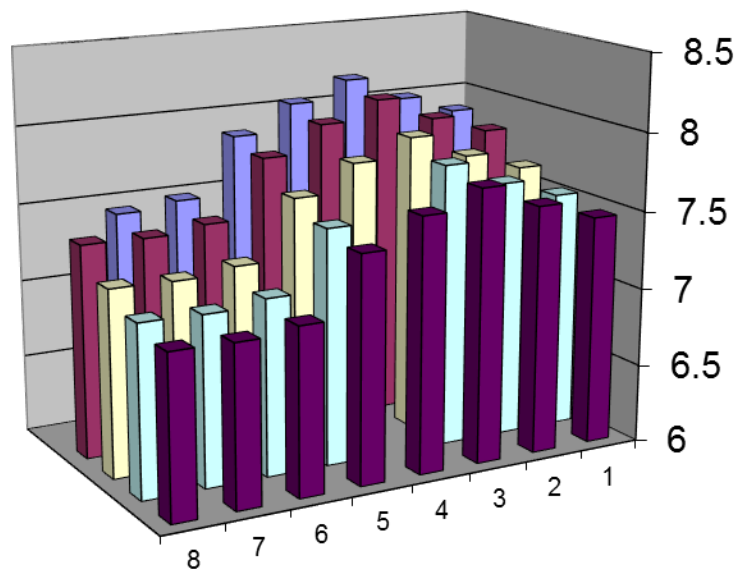
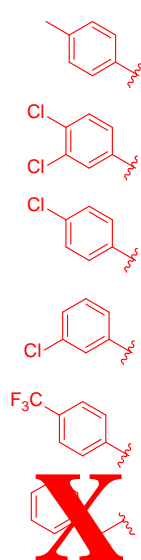
For this assay $\sigma = 0.2$
 $p = 95\%$ when $D_{add} > 2\sigma = 0.4$

Analysis of Second Library



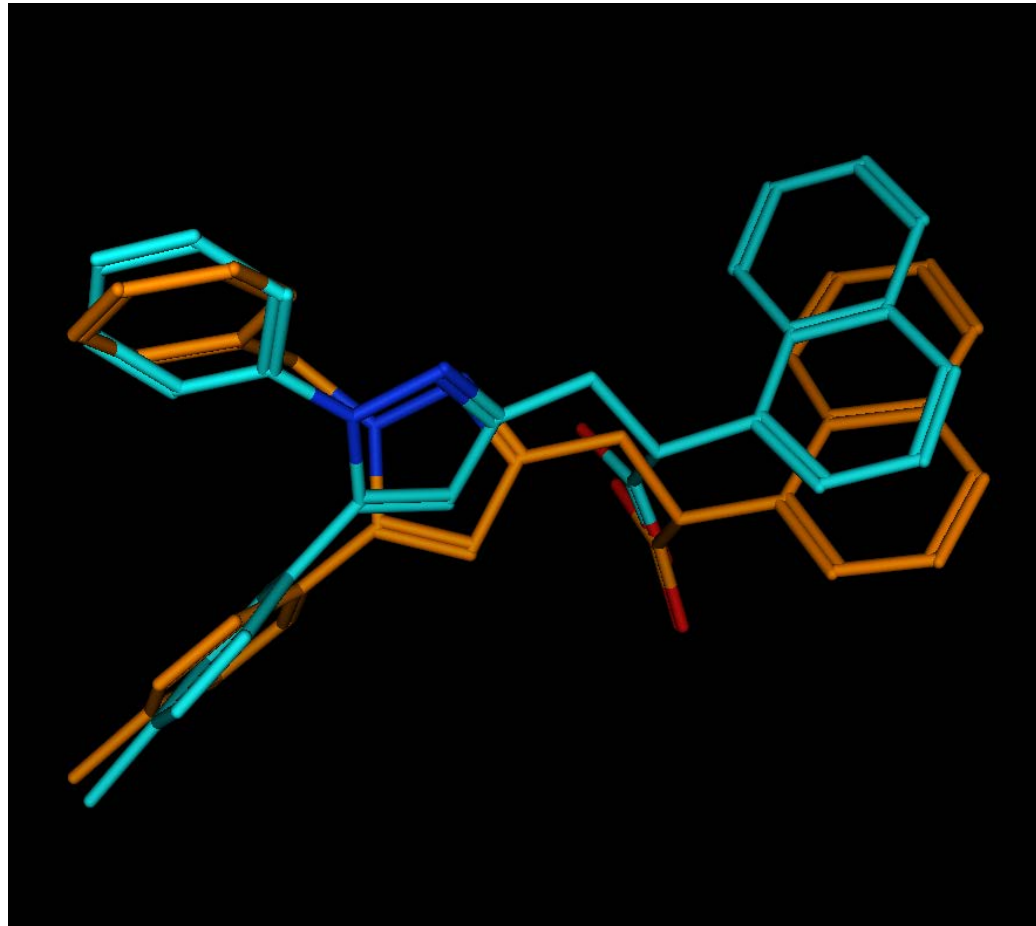
Free-Wilson Predicted activity

Deviation From Predicted

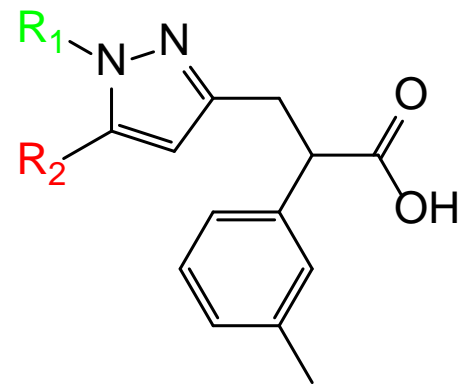
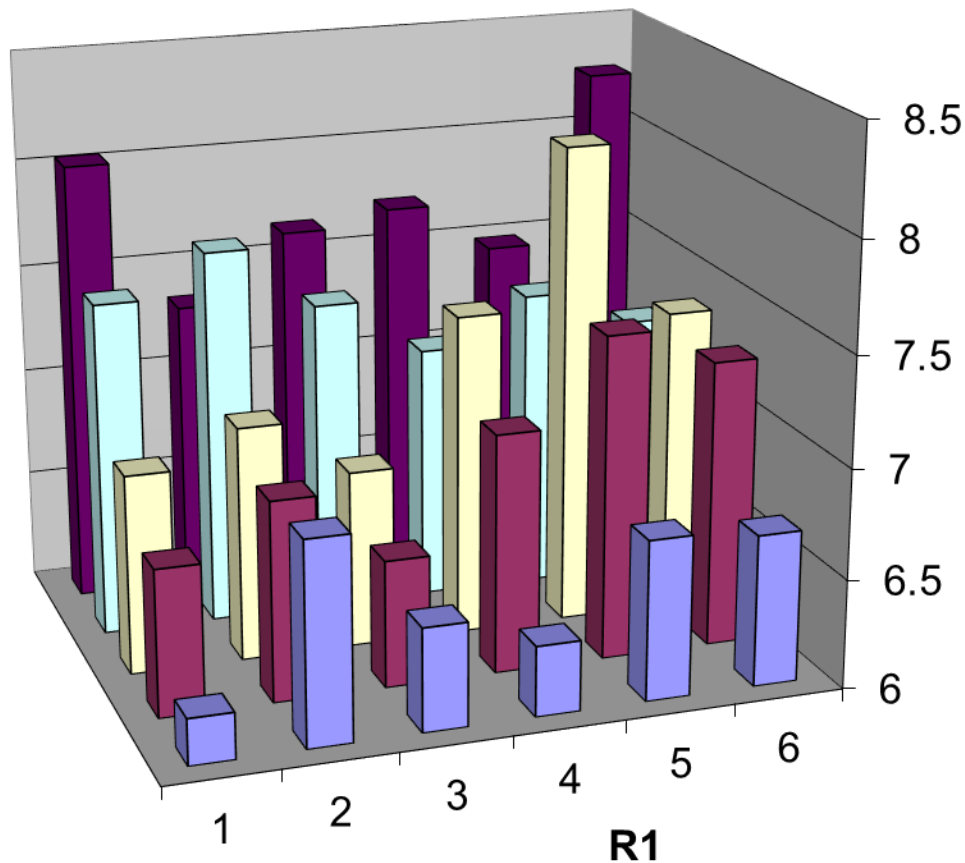
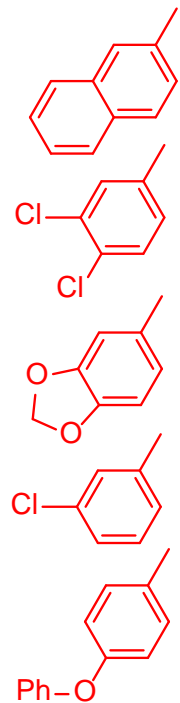


One small change in Ar₁ and you have to start all over understanding Ar₃!

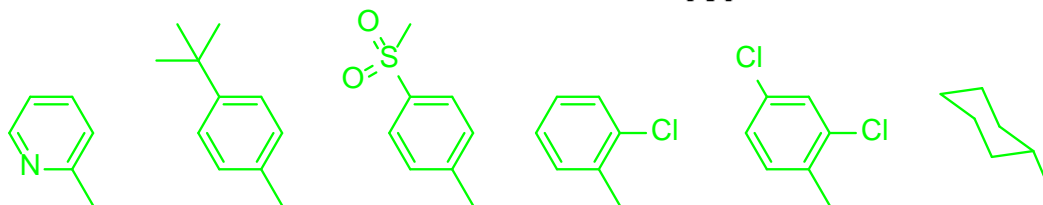
Possible change in binding mode



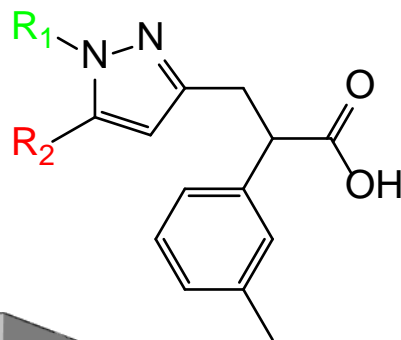
Library Three



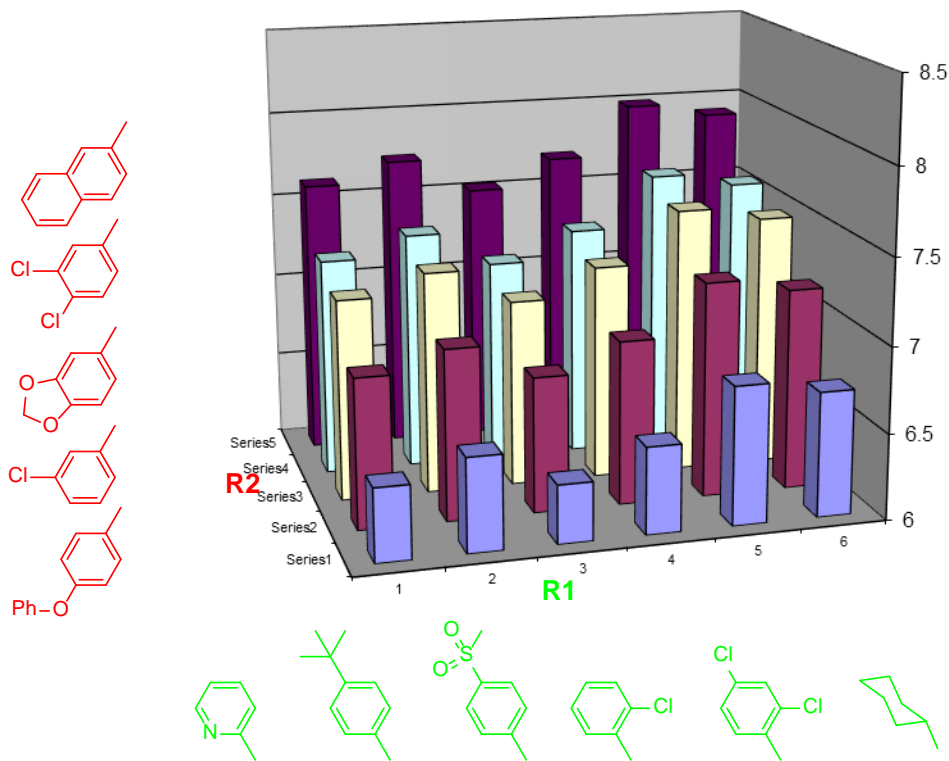
pKi



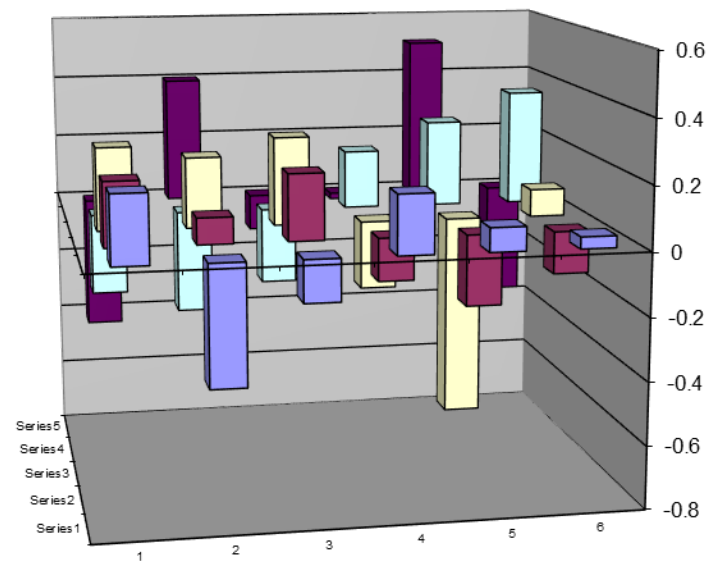
Analysis of Library Three



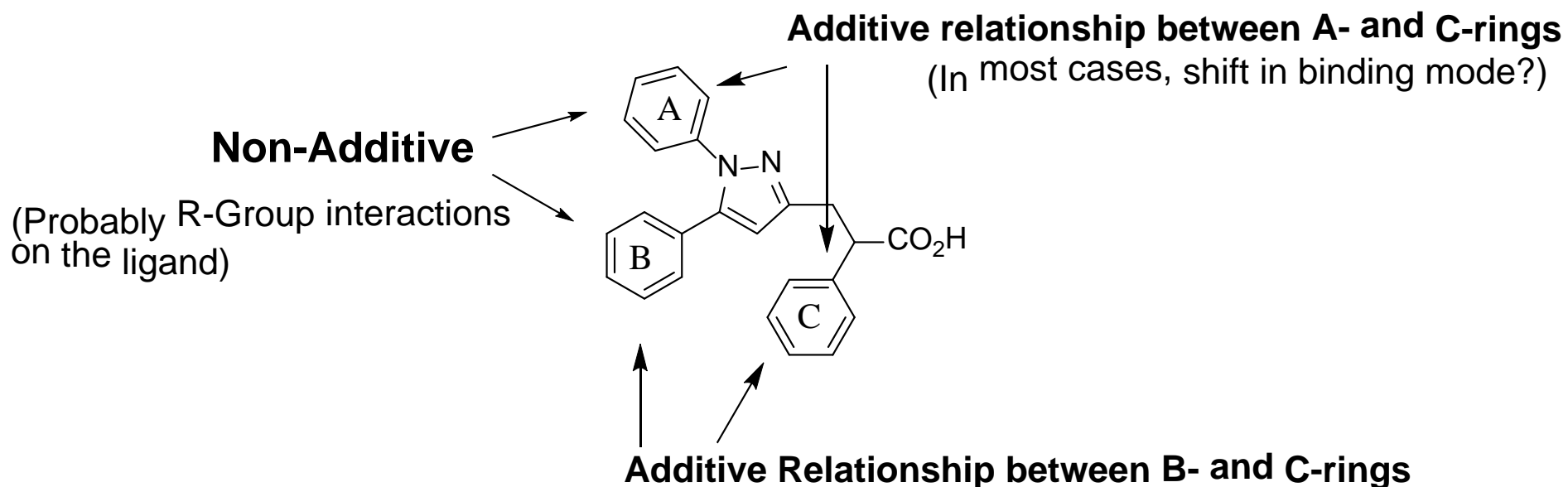
Free-Wilson predicted activity



Deviation from prediction



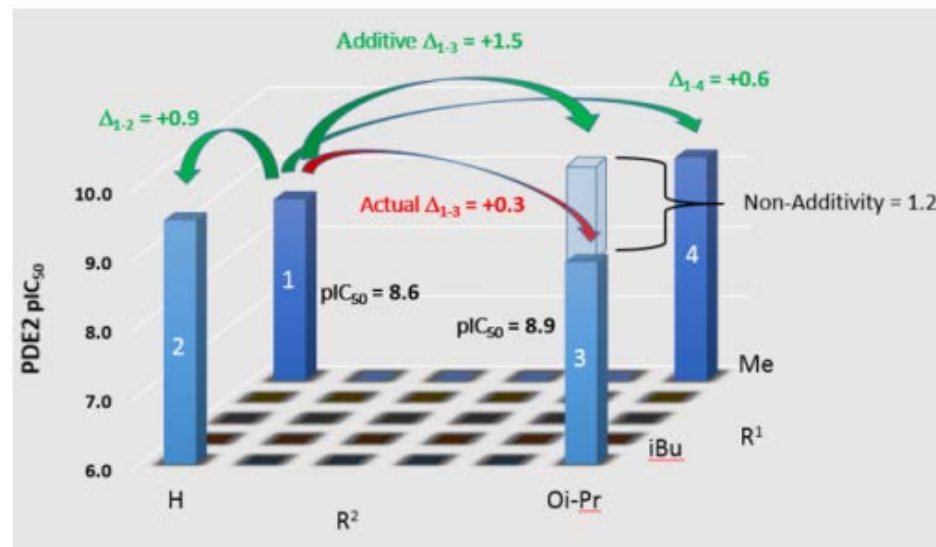
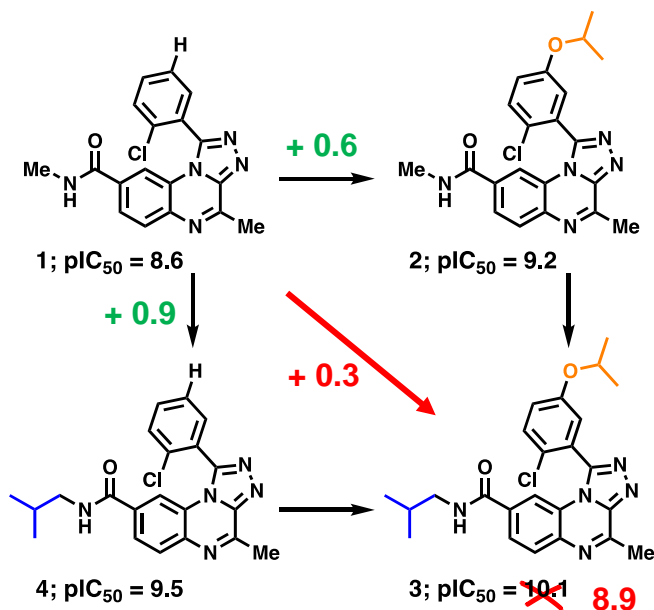
Lessons Learned from 3 Libraries



Even within a single series there are non-additive parts and additive parts of the SAR!

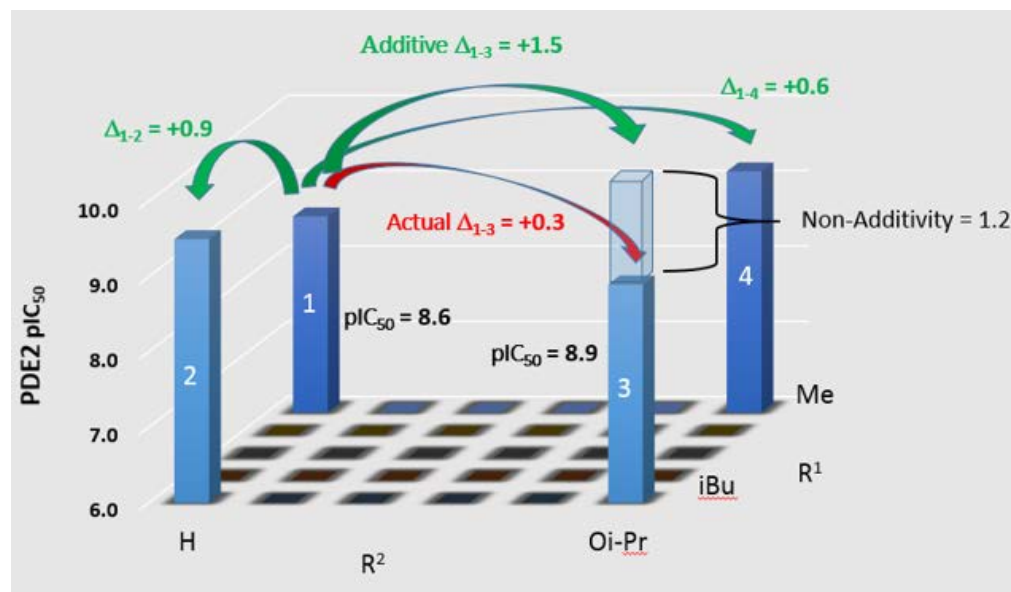
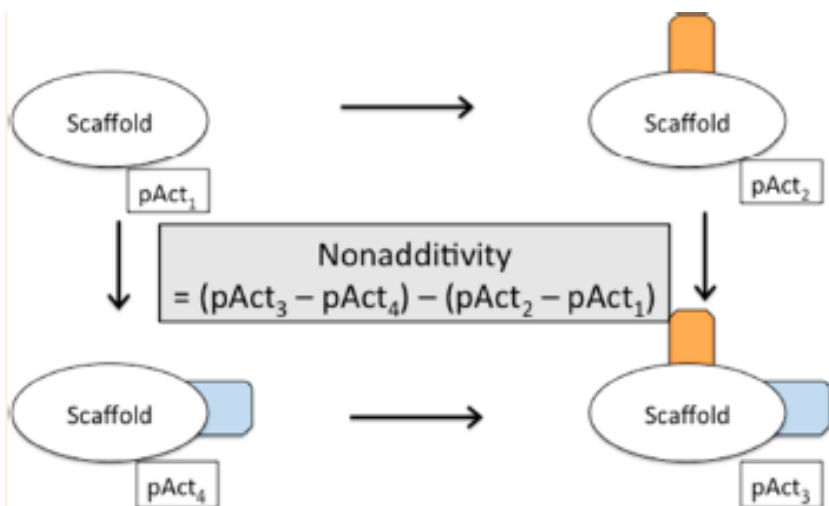
PDE2 Case Study

Phosphodiesterase 2 (PDE2) inhibitors for the treatment of memory disorders



- Evidence for non-additive SAR early (< 10 analogs)
- Can this data provide insights into ligand-protein interactions?

New Mathematical Method Solved Problems



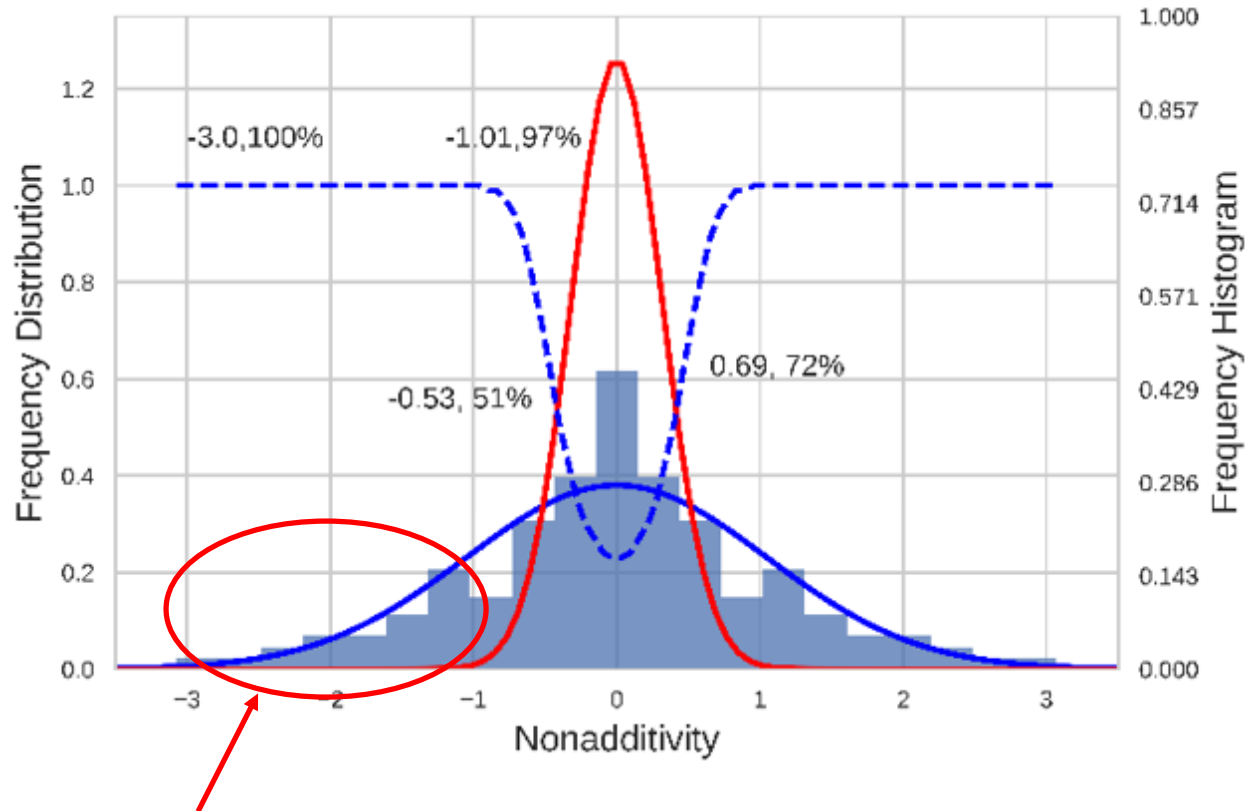
This method allows determination of additivity in sparse matrix's where you might have missing data.

$$\text{Non-additivity} = (pA_3 - pA_4) - (pA_2 - pA_1) = e_X e_Y = pKi - (\mu + X_j + Y_k)$$

It's all the same equation!

However in the Kramer case one is looking at each individual set of 4 compounds. In the linear regression method additivity is measured against a theoretical average of all compounds.

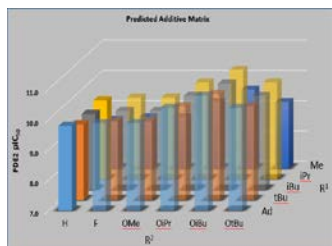
Non-Additivity Analysis



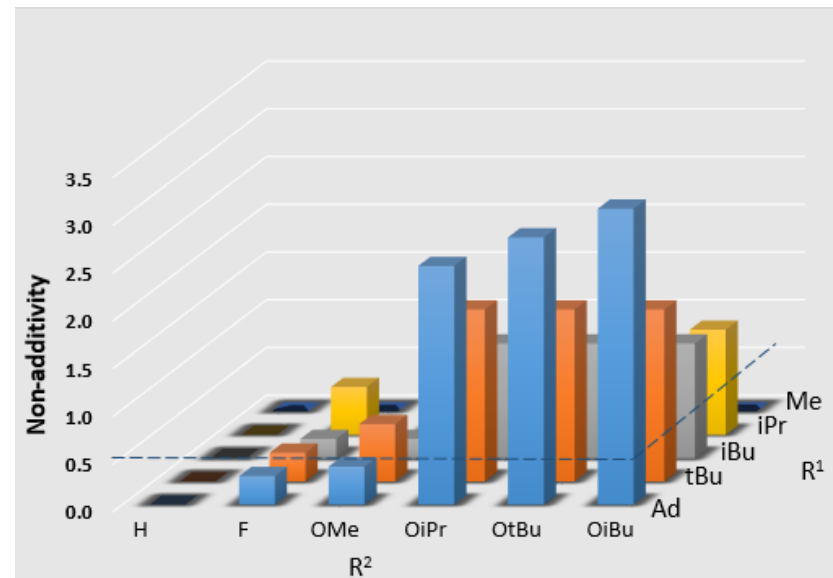
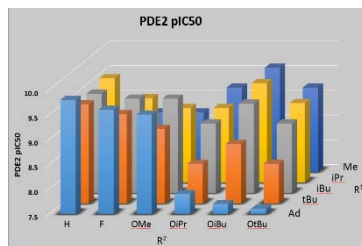
26% of the matrix is non-additive ($p = 97\%$)
Largest D_{add} is 3 log units from additive!
They happen when $R_1 = \text{Large}$ and $R_2 = \text{Large}$

Systematic Deviation from Additivity

Predicted Data



Actual Data

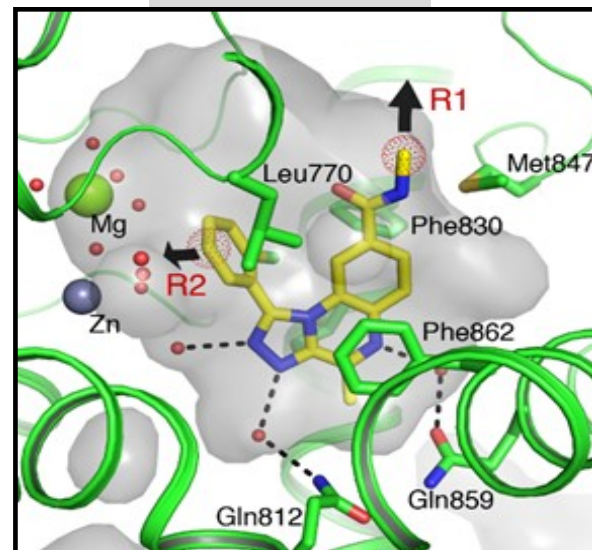
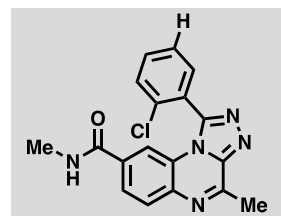
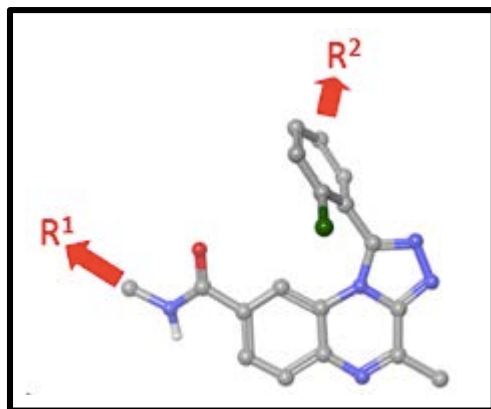


54% of the data is non-additive (99% confidence)

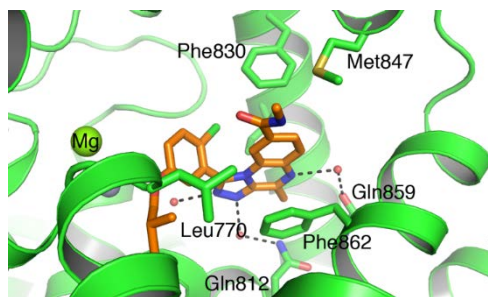
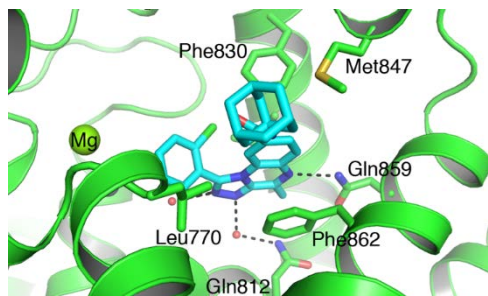
The Cause of Non-Additive SAR

Common reasons for non-additive SAR:

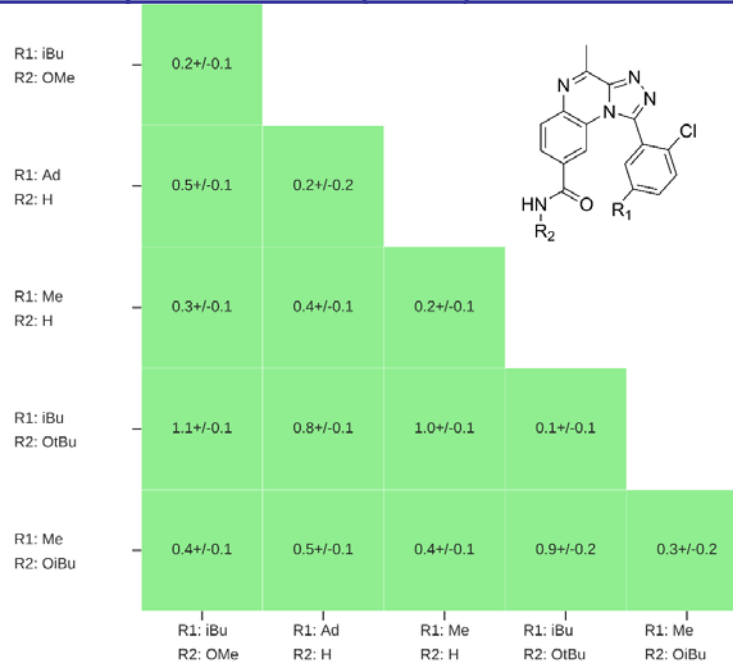
- Interactions between substituents
- Different binding modes



Crystallography and Structural Characterization

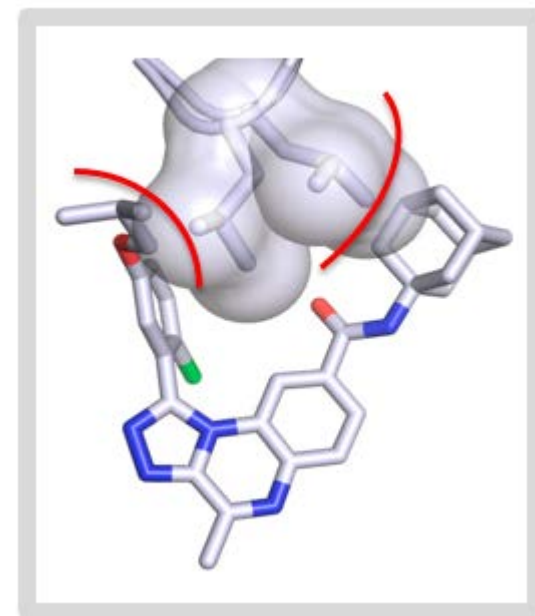
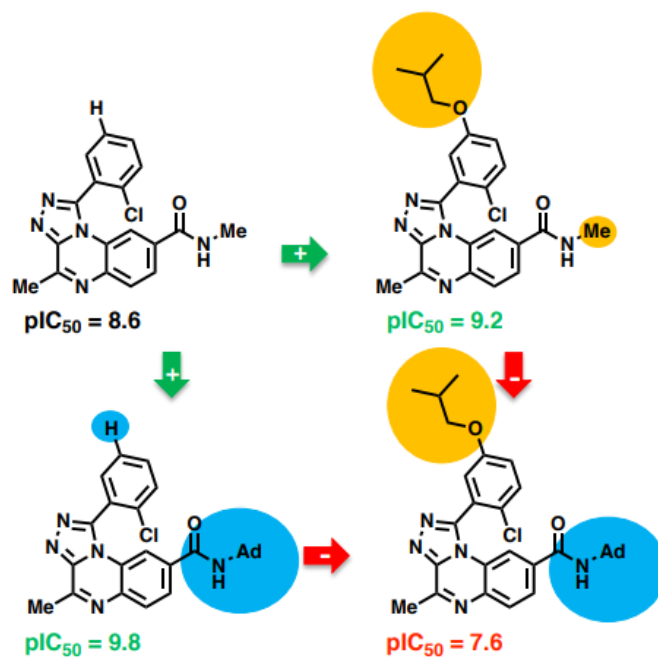
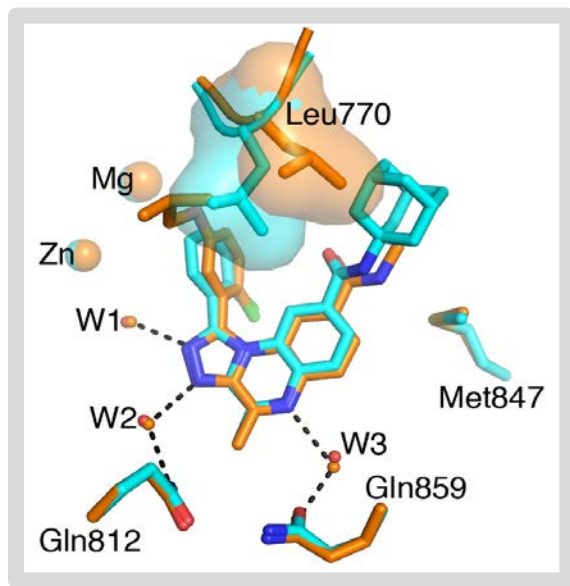


Root-mean-square deviation (RMSD) values between structures



- All compounds adopt a similar binding mode (RMSD values $\leq 1 \text{ \AA}$)
- Ligand rearrangement in the binding site is not the explanation for non-additivity

Protein Conformational Changes!



Leu770 works as a molecular divider that distributes ligand binding volume between R_1 and R_2 pockets.

First documented case of large non-additive SAR effects being confirmed as resulting from changes in protein conformation.

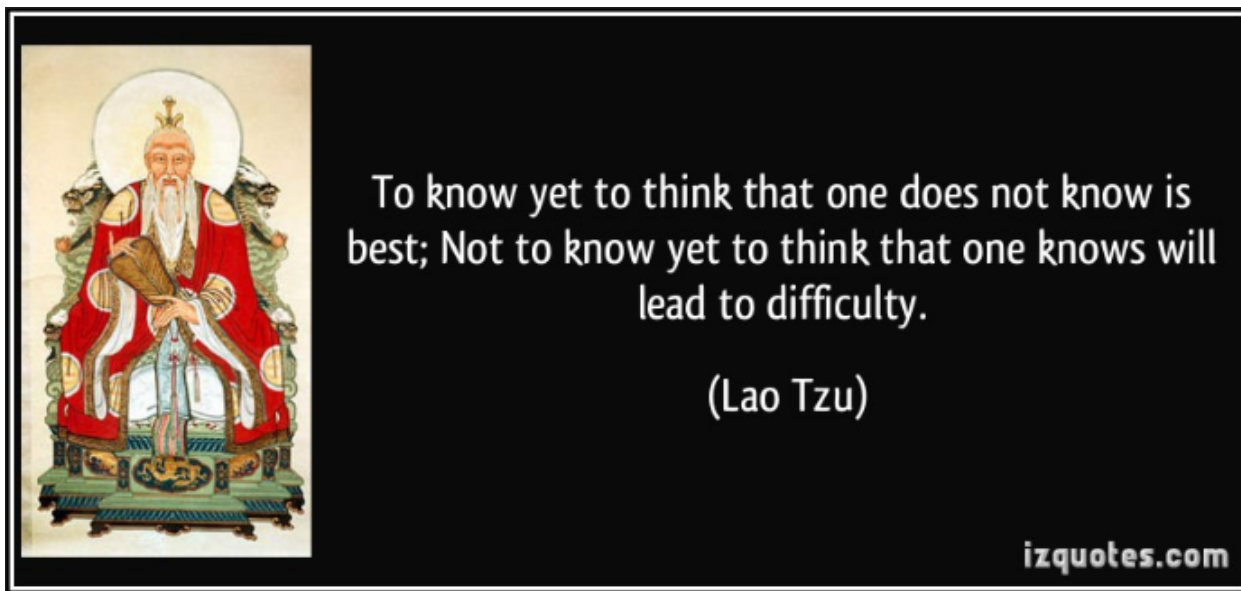
Conclusion

- We too often assume additive SAR when we should be checking.
- Non-additive effects can be very large (>3 log units).
- The presence of non-additive effects can lead to understanding of changes in binding modes or ligand conformation.

- Doesn't the first step in building a QSAR model have to be determining additivity in the data set?
 - When non-additivity exists can there be a meaningful correlation of biological effects with additive bulk physical properties (cLogP, tPSA, ect.)?

Where can we go?

- *Additivity in ADME properties?*
- *Additivity can predict changes in binding mode, how about changes in function (i.e. agonist/antagonist switch)?*
- *In phenotypic screening can non-additivity suggest change in mechanism?*



Acknowledgements

Laurent Gomez

Johnson and Johnson

Mike Hack
Kristen Arienti
Kelly McClure
Clark Sehon
Liming Huang

Dart NeuroScience

Markus Metz
William Sinko
Rui Xu
Brandon Selfridge
William Vernier
Kiev Ly
Richard Truong

Biology teams (JNJ and Dart) and Structural Biology team (Dart)