

**The effects of
structural changes on
molecular properties:
Matched molecular pair
analysis**

Andrew Leach

Computational Chemistry

AstraZeneca, UK

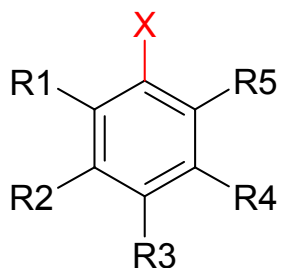
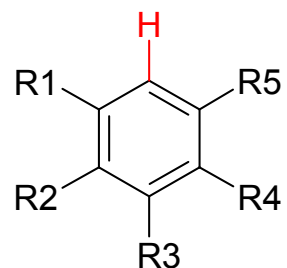
Overview

- What are matched pairs?
- Why use matched pairs?
- Worked Examples

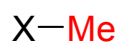
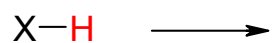
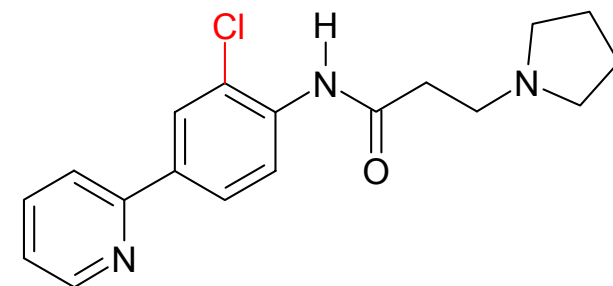
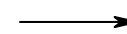
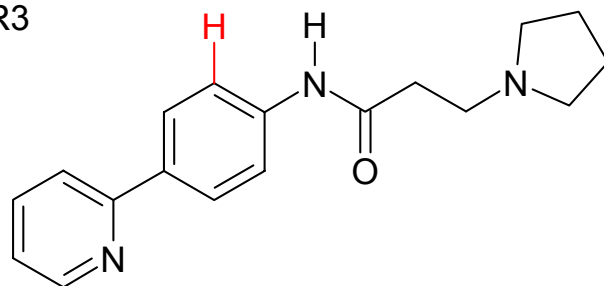
What are matched molecular pairs?

- Molecules that differ only by a (preferably small) structural change
- Remainder of the molecule is exactly the same

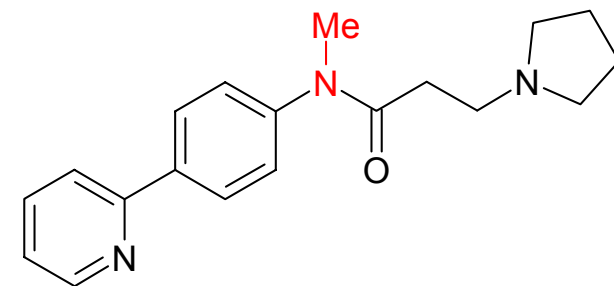
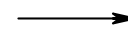
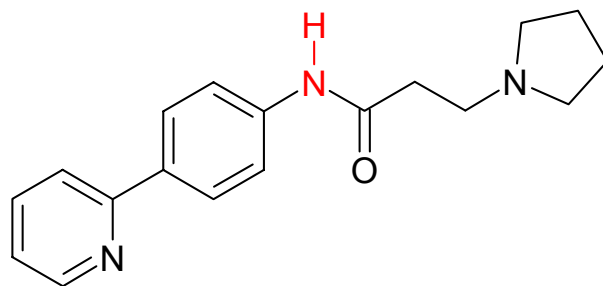
Examples of Structural Changes



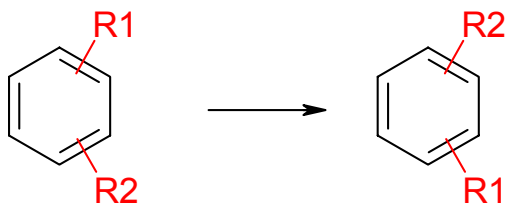
X = F, Cl, Br, Me, OMe, NO₂, CN etc.



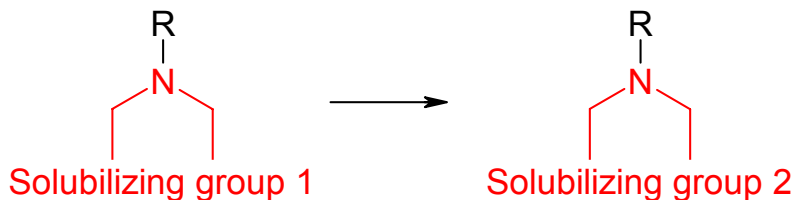
X = N in amine, amide, sulfonamide
O in hydroxyl, phenol, acid



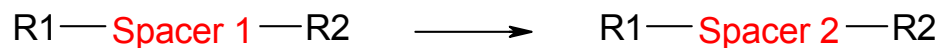
Examples of Structural Changes



Regioisomers



Morpholine, piperidine, pyrrolidine, piperazine etc.



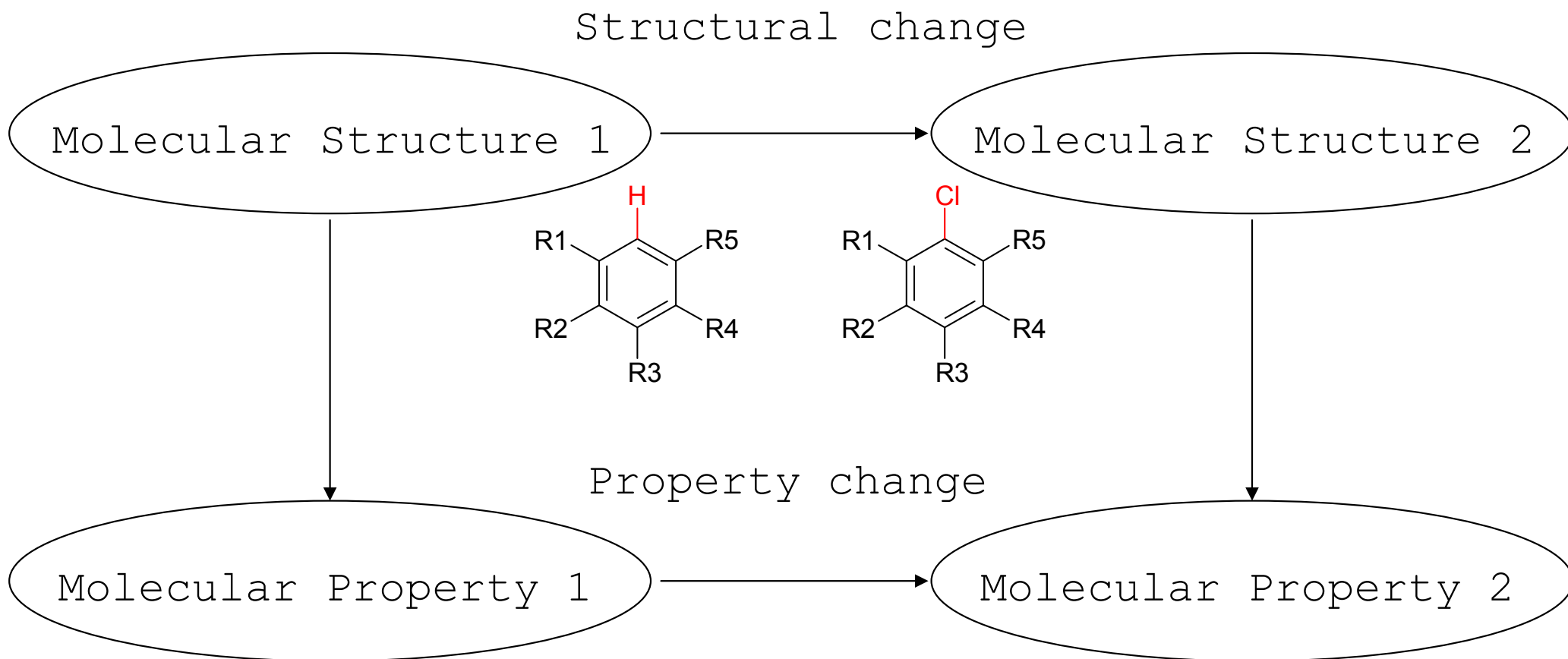
Spacer = amide, sulfonamide, ether, amine, alkene etc.

Enantiomers

Why Use Matched Pairs?

- Common to look at matched pairs for one or two cases of interest
- Key concept - locate all sets of matched pairs that fit the outline e.g. every set of pairs of molecules that differ only by addition of Cl to an aromatic ring
- Find the corresponding mean change in a property of interest
- If the set is structurally diverse and sufficiently large this mean gives the intrinsic effect of making the structural change

Matched Molecular Pairs



Mean change in solubility = -0.7 log units

Advantages of Matched Pairs

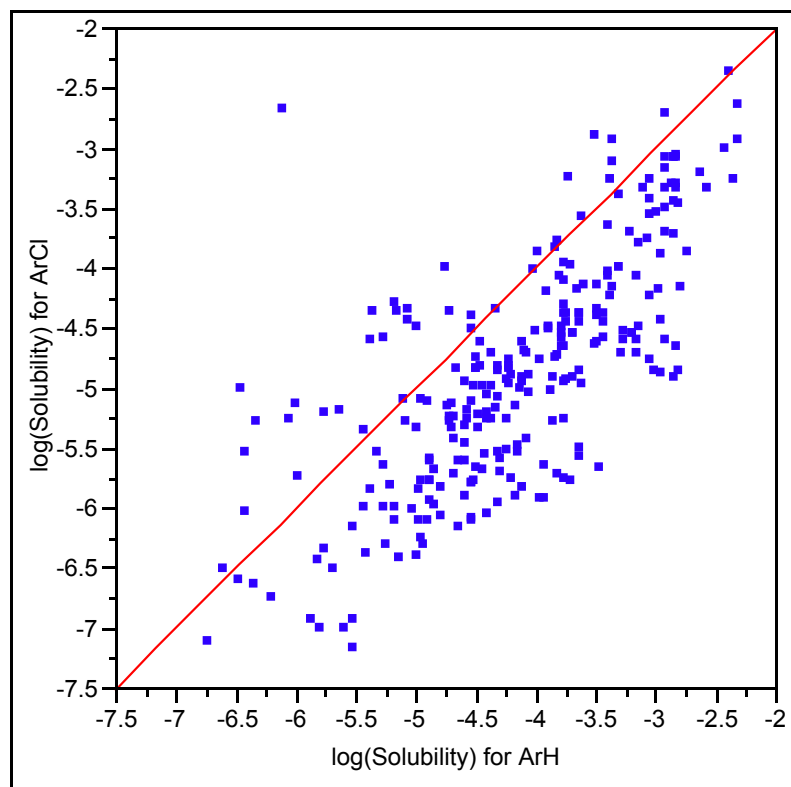
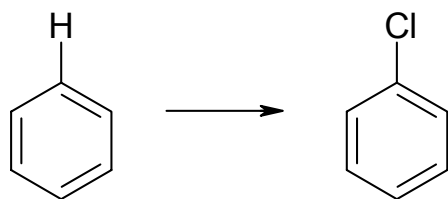
- Uses only real data
- Any property can be used
- Looking at changes in structure – easier to do and in line with design process
- Identify outliers and quantify the variation

Issues with Matched Pairs

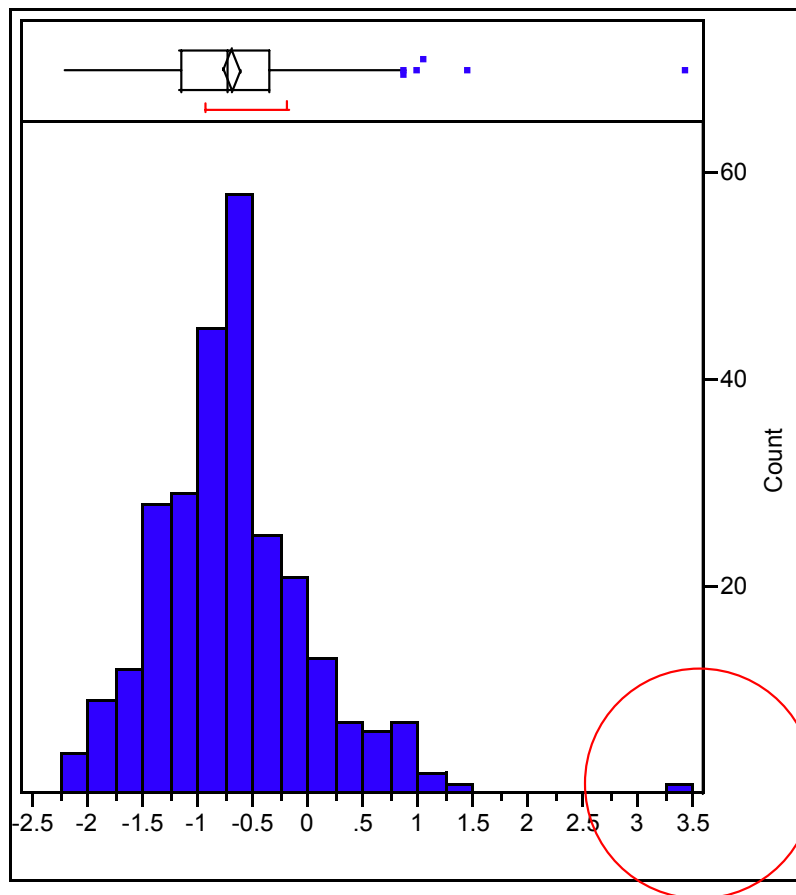
- Generality only as general as the dataset
 - keep transparent
- Statistical difference
 - is there an effect?
 - is there enough data?

Worked Examples 1: Solubility effects of aromatic substituents

After identifying pairs, add data
Example: addition of Cl to an aromatic ring



Worked Examples 1: Solubility effects of aromatic substituents



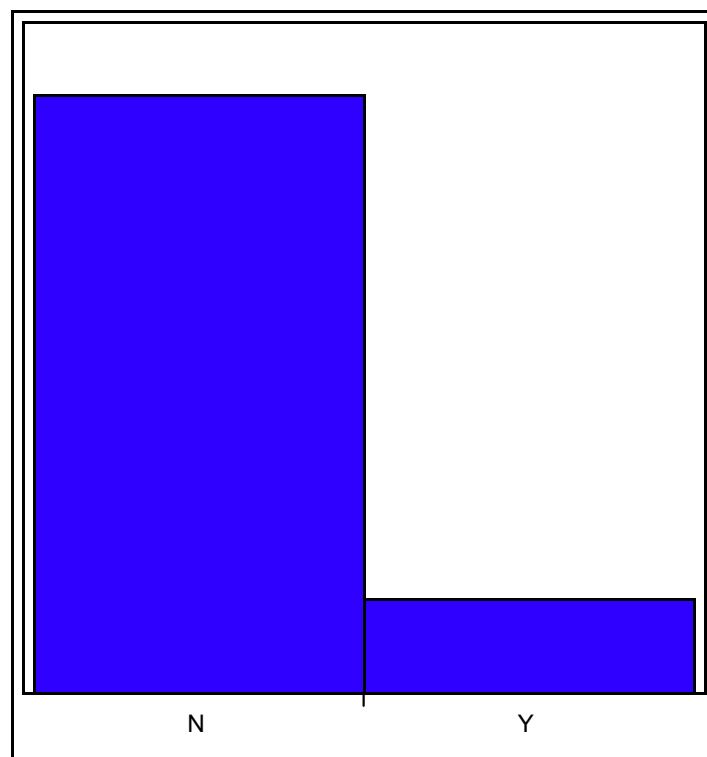
Assay measures
solubility from
solids

Mean difference in $\log(\text{Solubility}) = -0.7$

Standard deviation = 0.7

Standard error in the mean = 0.04

Worked Examples 1: Solubility effects of aromatic substituents



% of times ArCl is more soluble than ArH = 14 %

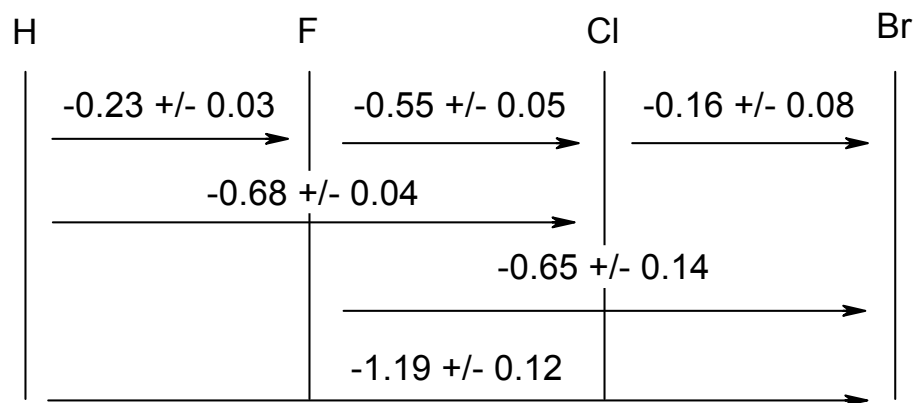
% of times ArCl is less soluble than ArH = 86 %

Worked Examples 1: Solubility effects of aromatic substituents

Substituent	Mean of change in log(solubility)	Standard deviation	Standard error	% showing increase	% showing decrease
F	-0.23	0.71	0.03	34	66
Cl	-0.68	0.70	0.04	14	86
Br	-1.19	0.69	0.12	3	97
CF ₃	-0.80	0.72	0.09	18	82
Me	-0.26	0.68	0.06	30	70
CN	-0.23	0.84	0.09	34	66
OMe	-0.06	0.71	0.05	47	53
SO ₂ Me	+0.22	0.69	0.15	70	30

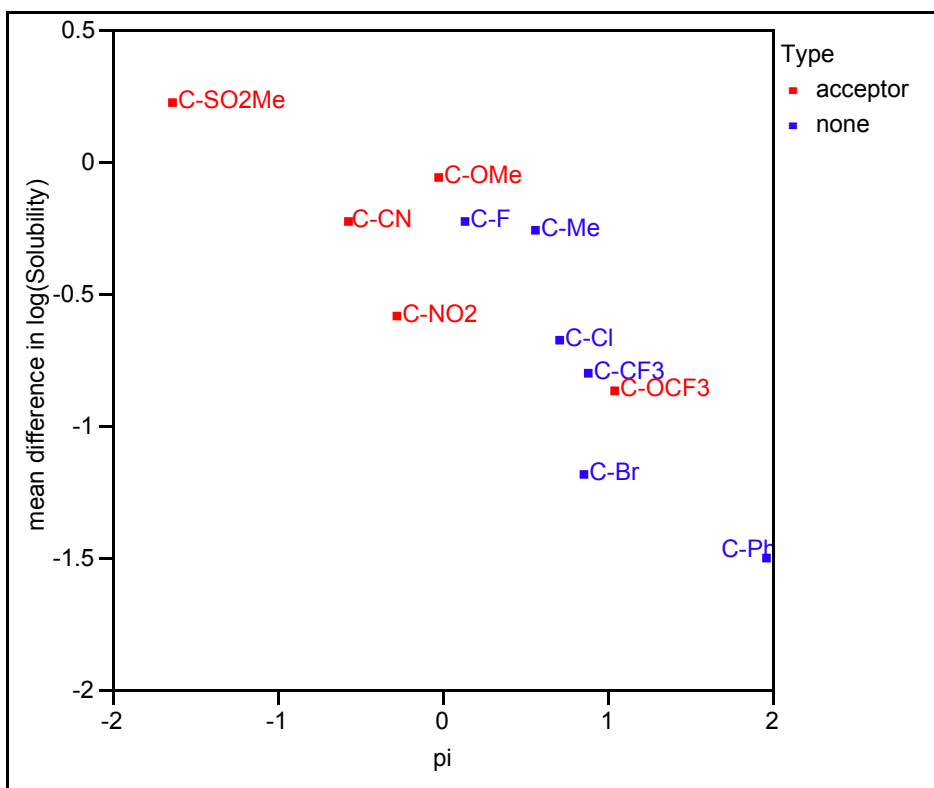
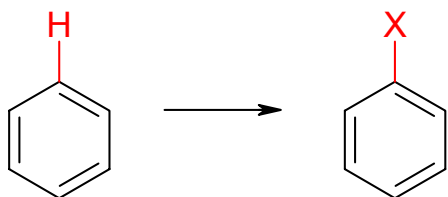
Worked Examples 1: Solubility effects of aromatic substituents

Substituent interchanges rather than additions:
Ar-H is a good reference state



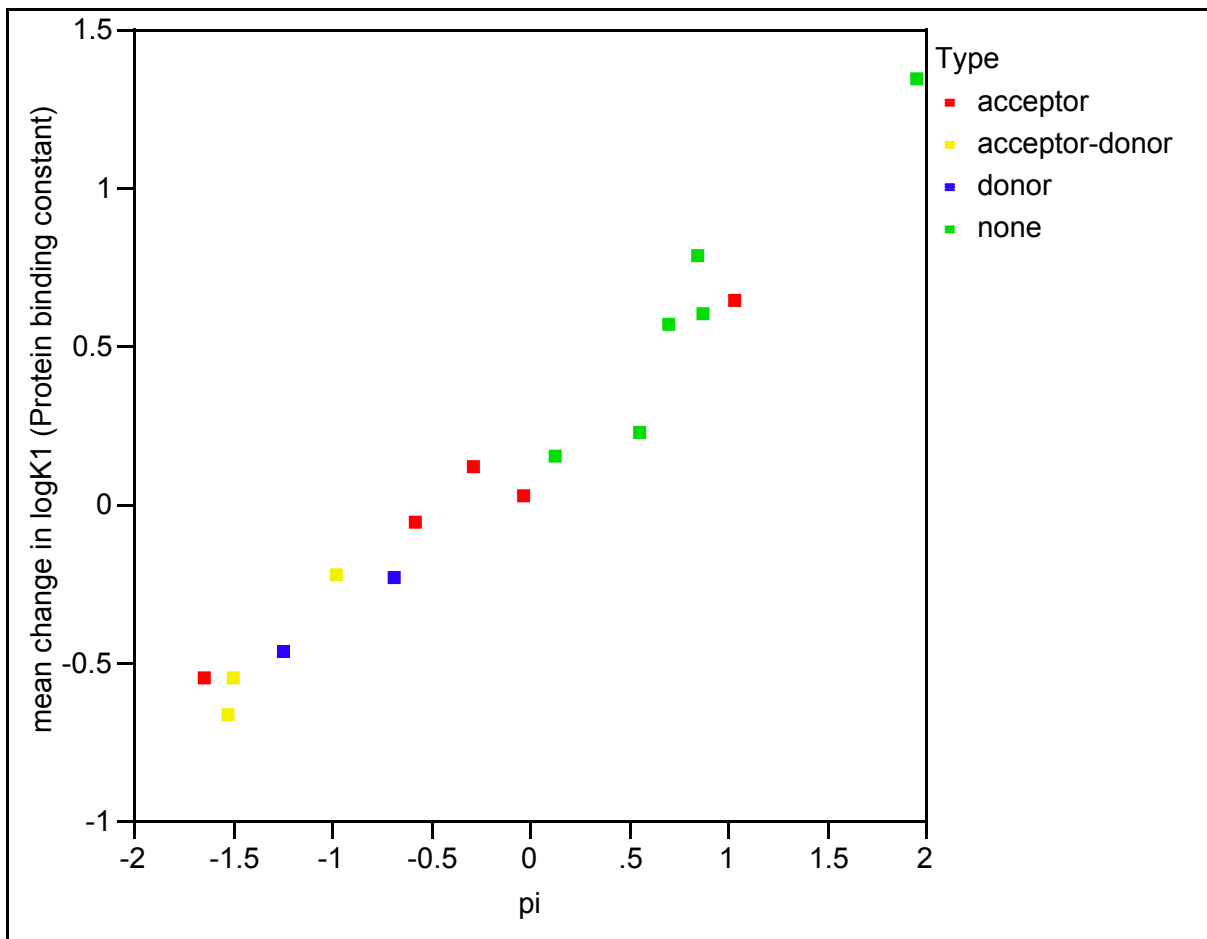
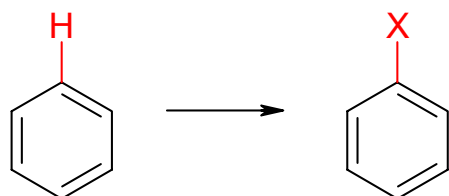
Worked Examples 1: Solubility effects of aromatic substituents

Role of lipophilicity



π = average change in logP for addition of X

Worked Examples 2: plasma protein binding effects of aromatic substituents



π = average change in logP for addition of X

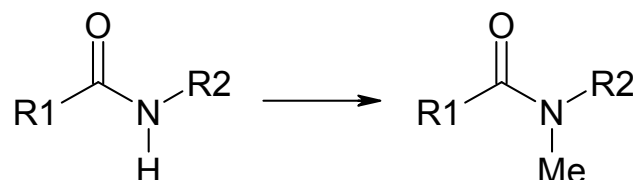
Protein binding in rat plasma (albumin + others)

Other properties

- Physical properties (solubility, lipophilicity, protein binding)
- Potency against biological target (enzyme or cell)
- Pharmacokinetic (in vitro or in vivo)

Worked example 3: Methylation of amides

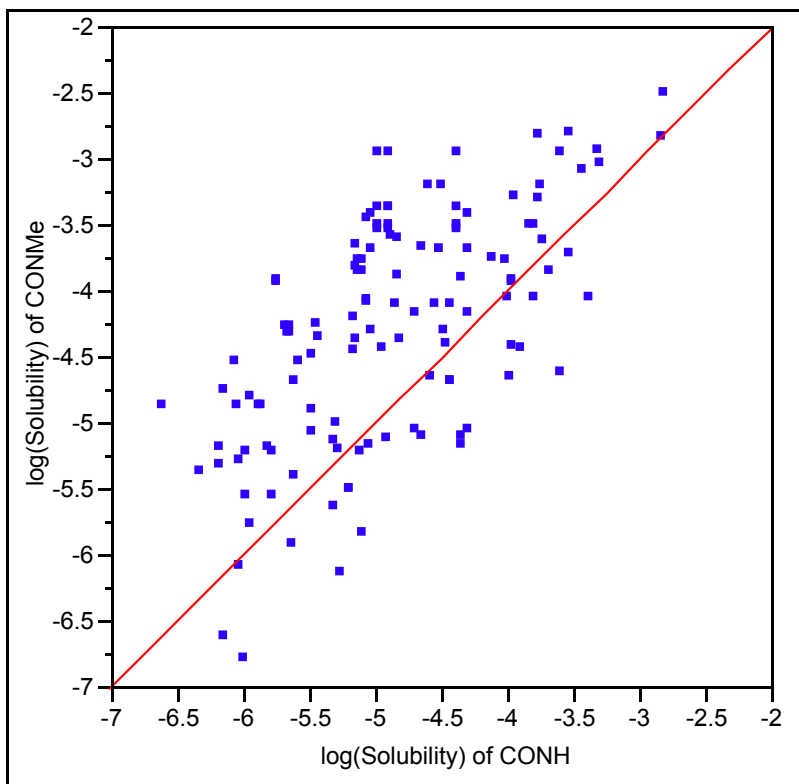
- Project observation that methylation of amides INCREASED solubility in one or two cases



- Matched pairs is an ideal method to test whether this is a general phenomenon or specific to those few compounds

Worked example 3: Methylation of amides

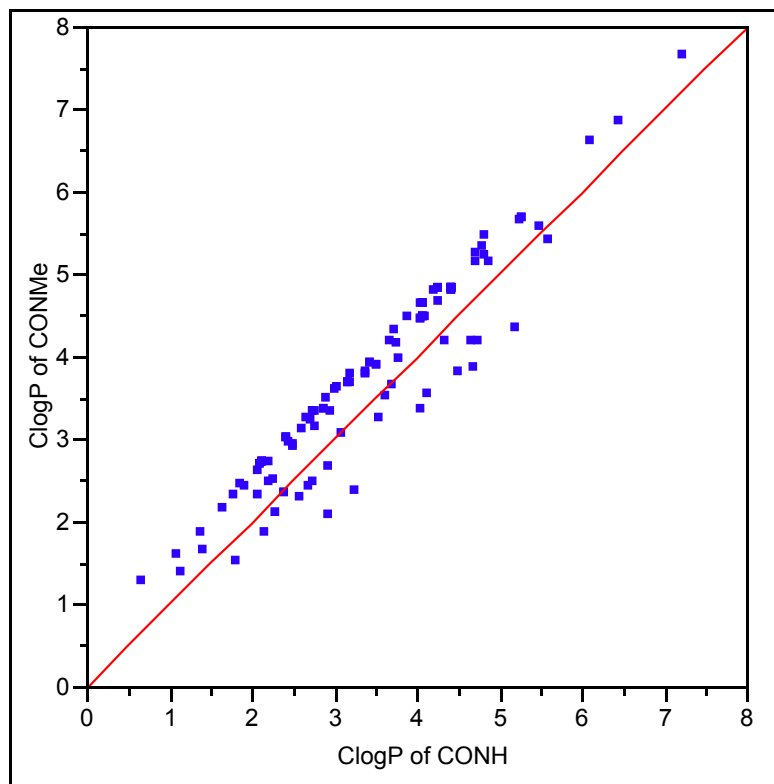
Survey of whole company database of solubility



Mean change = +0.61
SD = 0.72
SE in mean = 0.06
% of cases where
CONMe is more soluble
than CONH = 77 %

Worked example 3: Methylation of amides

Not lipophilicity driven?



Mean change = +0.34

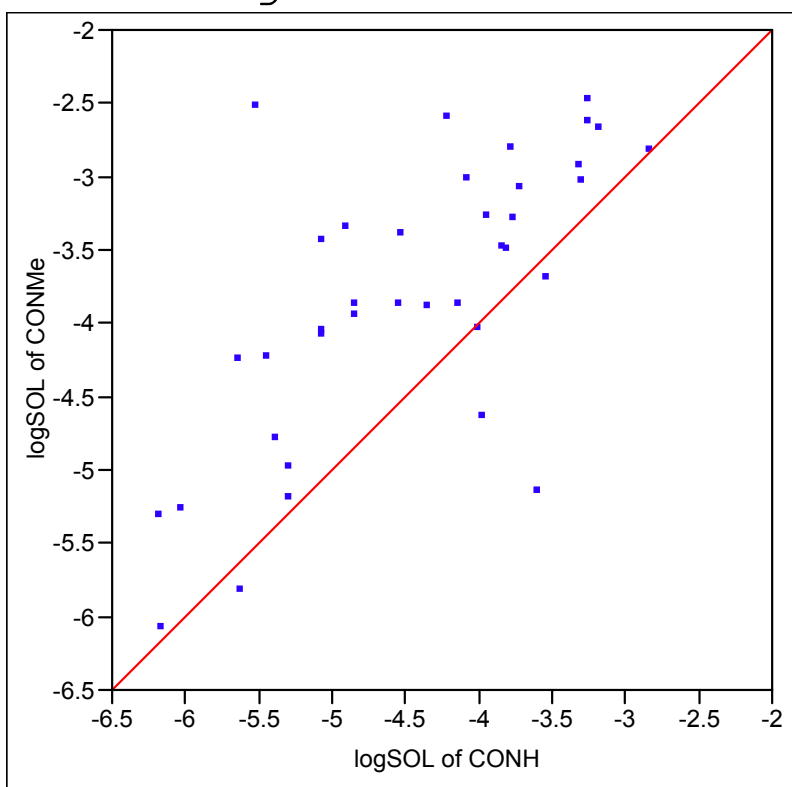
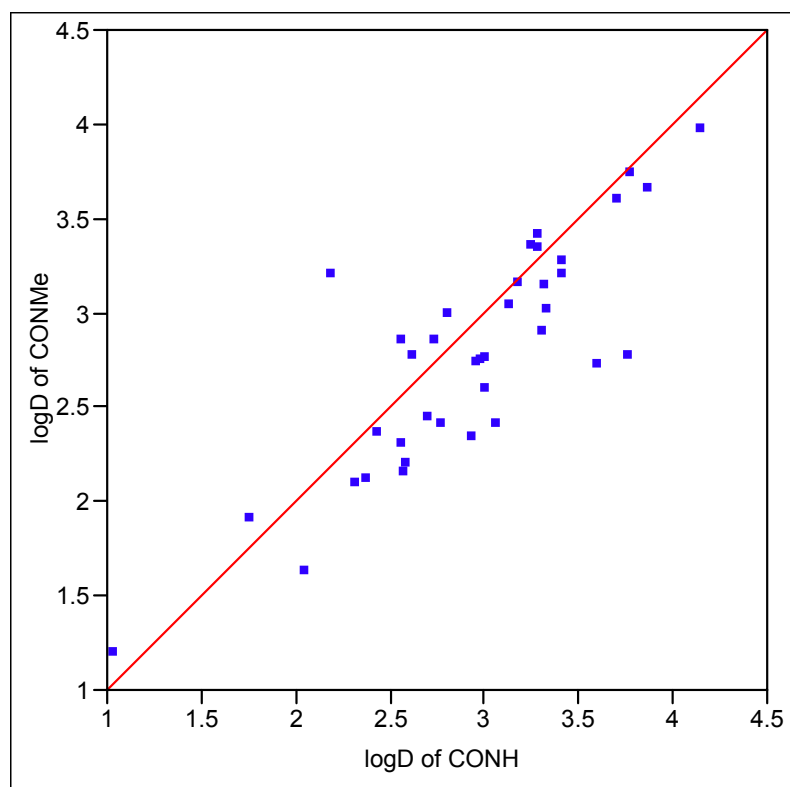
SD = 0.37

SE in mean = 0.03

% of cases where CONMe
is more lipophilic than
CONH = 82 %

Worked example 3: Methylation of amides

Not lipophilicity driven? – Pairs with
measured solubility and logD



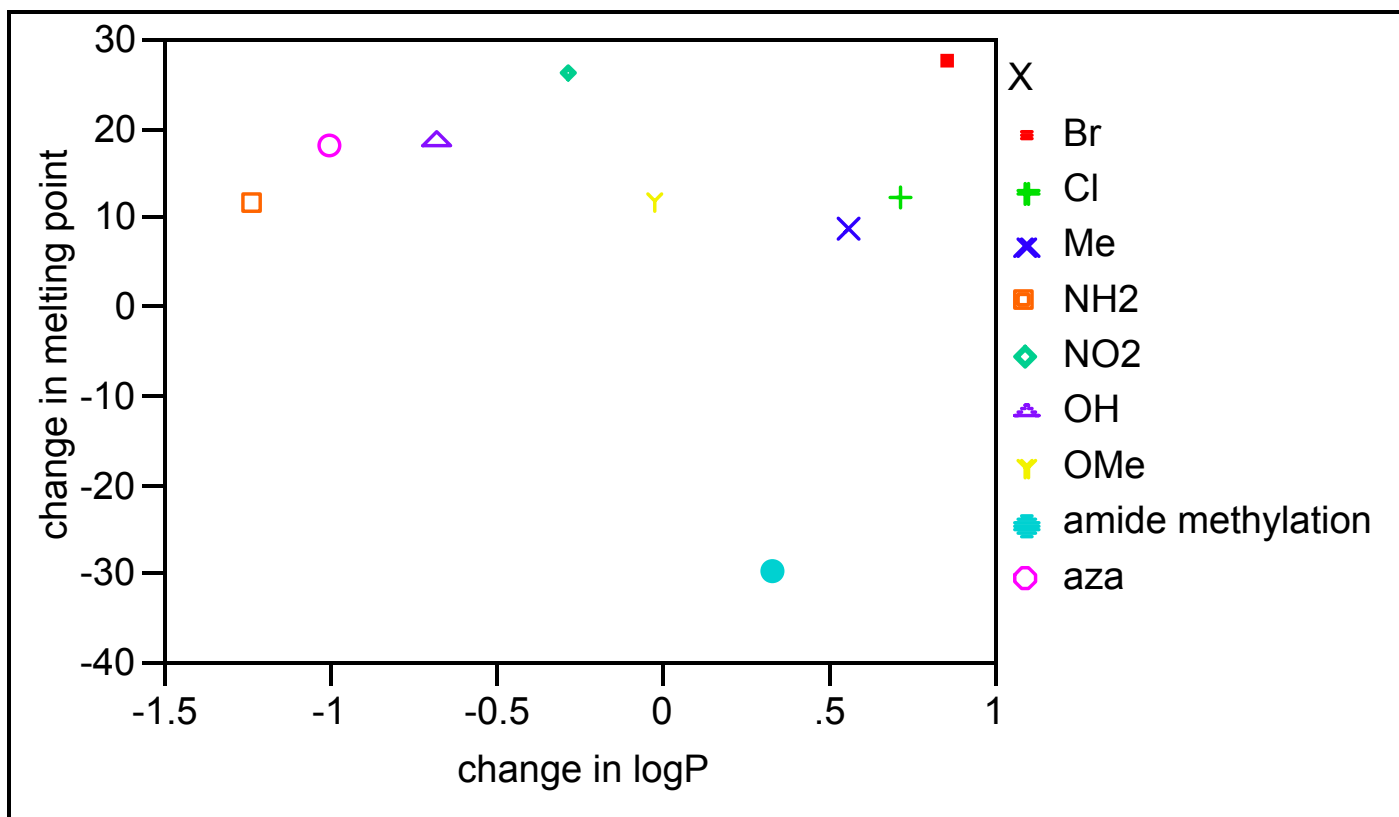
Mean change in $\log(\text{Solubility}) = +0.64$

Mean change in $\log D = -0.16$; in $\text{clogP} = +0.03$

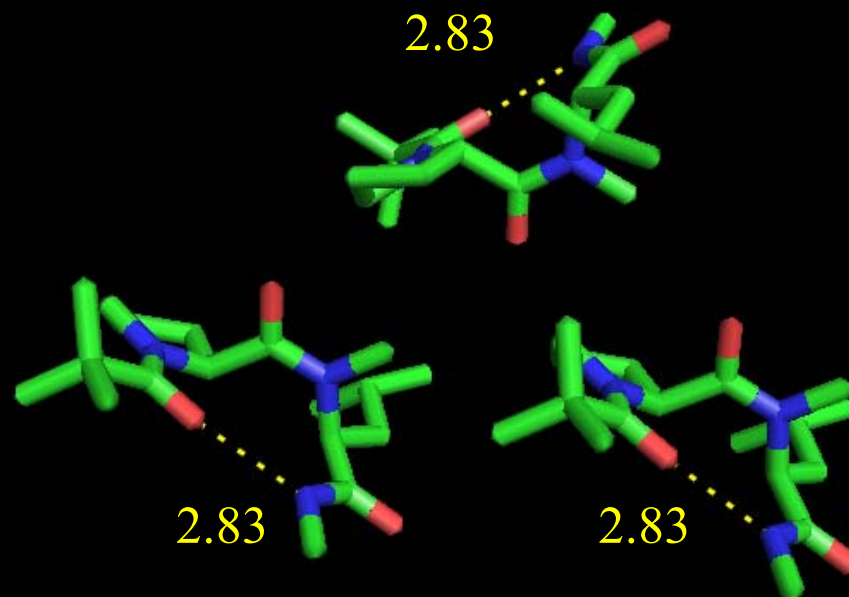
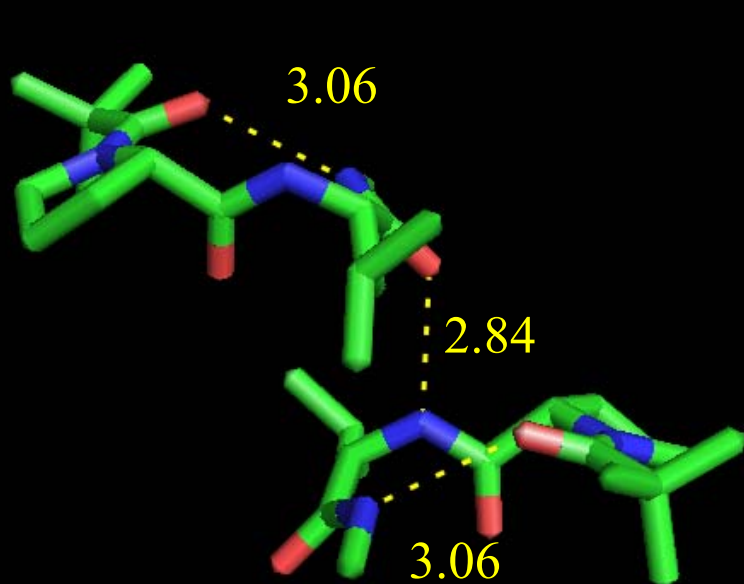
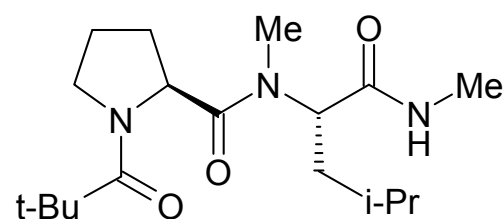
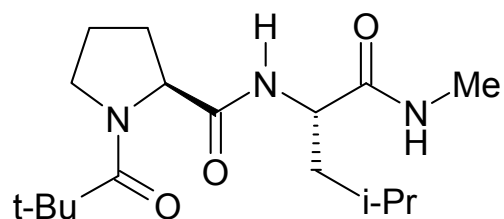
SD in $\log D = 0.34$; in $\text{clogP} = 0.44$

SE in mean for $\log D = 0.06$; for $\text{clogP} = 0.07$

An aside - the solid state and melting points



An aside - the solid state and melting points



Conclusions

- Matched molecular pairs provide chemical insight
- Guide molecular design
- Leverage information in databases
- Can be interpretable - insight
- Need not be interpretable!
- Can analyze wide variety of databases
- Can price up molecular shopping list

Acknowledgments

- Peter Kenny, Huw Jones, Dave Cosgrove, Jeff Morris and other Alderley Park computational chemistry colleagues
- Nicola Colclough, Matt Wood, Linette Ruston, Phil MacFaul and the rest of the Alderley Park Physical Chemistry Group