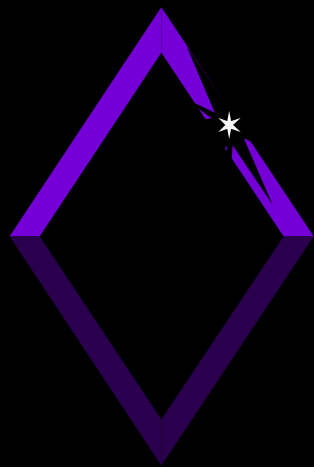


Expectations from High Throughput Screening



Jeff Morris

Paula Kitts

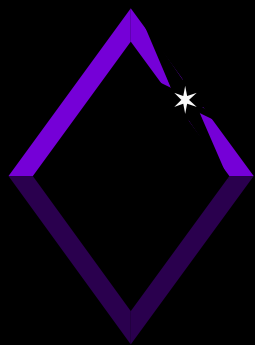
Richard Button

Zeneca Pharmaceuticals

Alderley Park

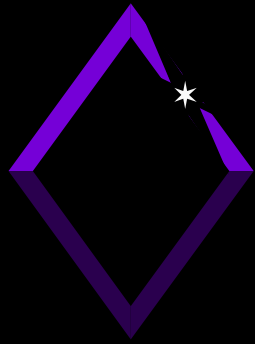
Macclesfield UK

ZENECA



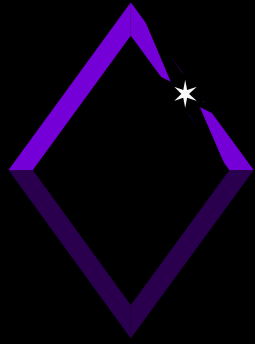
Content

- ◆ HTS in Zeneca
- ◆ Biological Expectations
- ◆ Chemical Expectations
- ◆ slightly grubby SAR



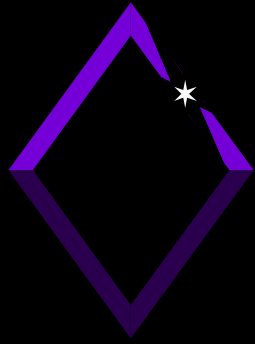
HTS in Zeneca

- ◆ Screen entire collection each time
- ◆ No replicates
- ◆ ~5K selected compounds for dose response.
- ◆ Computational tools in hands of Project Chemists.
 - ◆ Perceptions
 - ◆ HTS
 - ◆ Collection



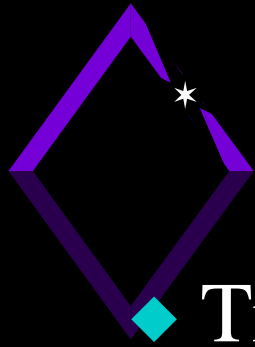
What happens if we don't find anything?

- ◆ Have we done it right?
 - ◆ what have we missed?
 - ◆ how many false positives?
 - ◆ screening concentration?
 - ◆ cut off correct?
- ◆ Intuition is not reliable guide



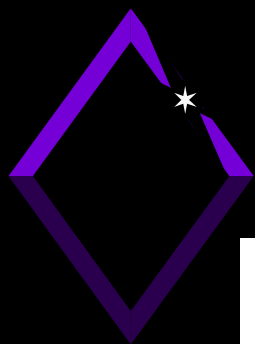
Simulate HTS?

- ◆ Calculate a set of IC50's based on random numbers
- ◆ Calculate associated % response
- ◆ Add noise
- ◆ This is our HTS output

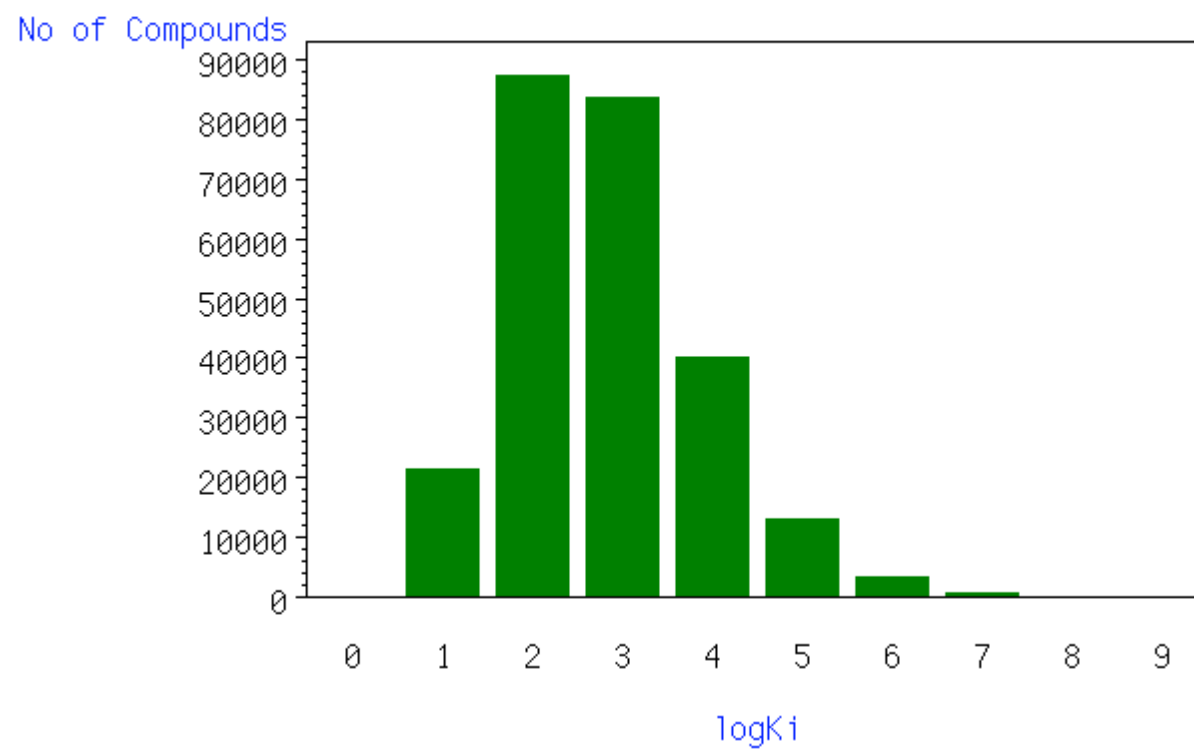


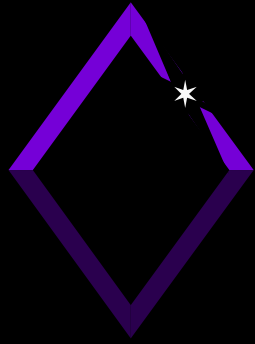
Distribution of 500,000 IC₅₀

- ◆ Tried quite a few distributions
 - ◆ properties
 - ◆ most compounds not active
 - ◆ no compounds better than 1nM
- ◆ Log transformation of uniform random numbers
 - ◆ small number of actives
- ◆ Random numbers from gamma distribution
 - ◆ larger hit list



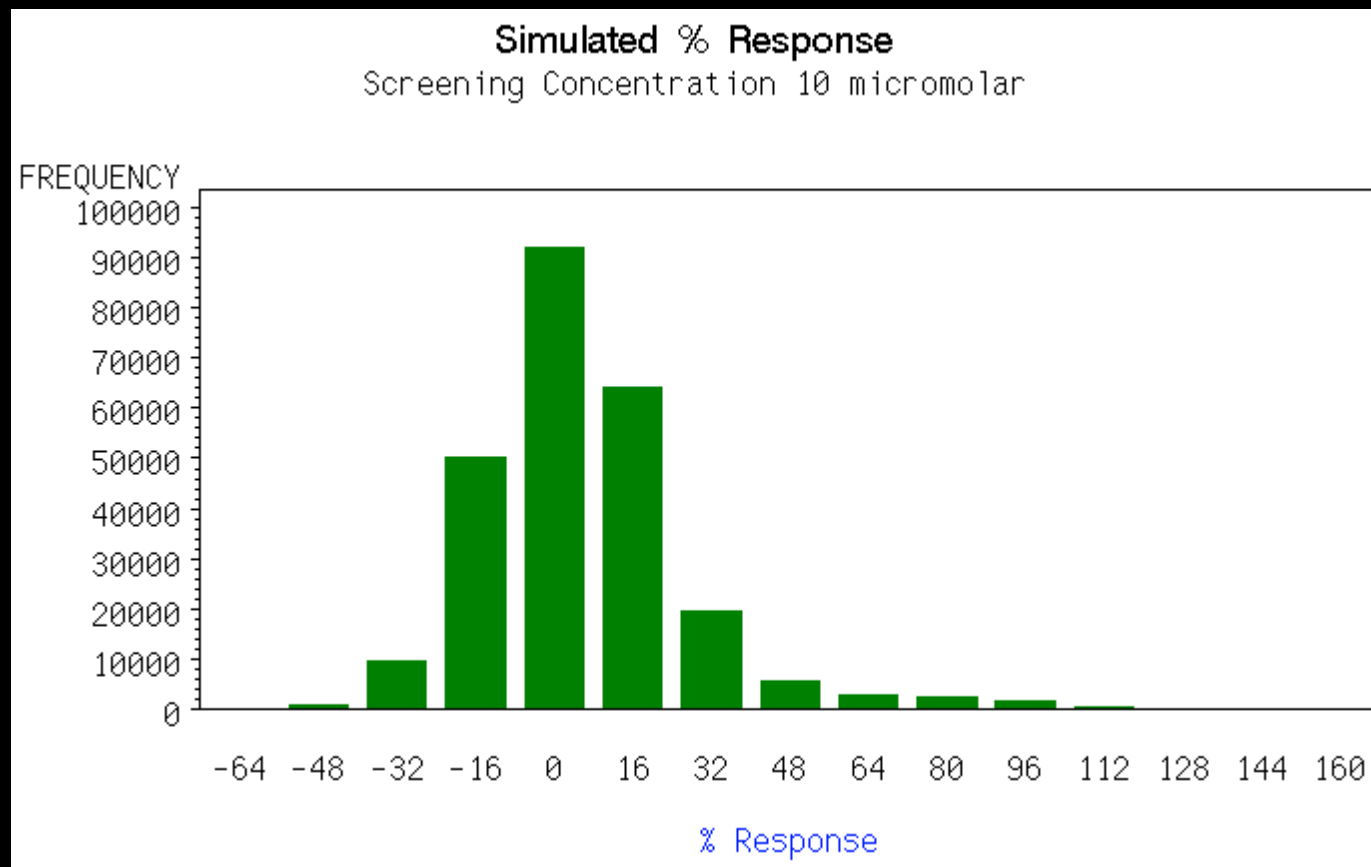
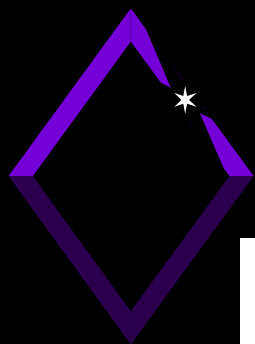
Simulated Distribution logKi

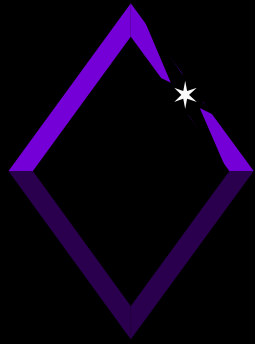




Calculate % response

- ◆ Using IC50 calculate % response
 - ◆ 3 10 and 100 μ M
- ◆ Add noise
 - ◆ Normally distributed noise added.
 - ◆ average sd set to 8-15%
 - ◆ this is our HTS hit list





False positives/False negatives?

◆ Compounds grouped by IC50:

< 100 μ M

Never want

10-100 μ M

Maybe

1-10 μ M

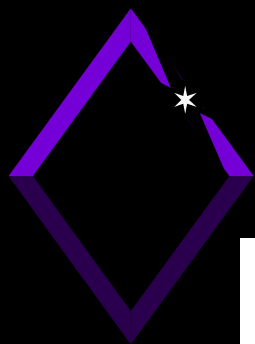
Probably

<1 μ M.

Always

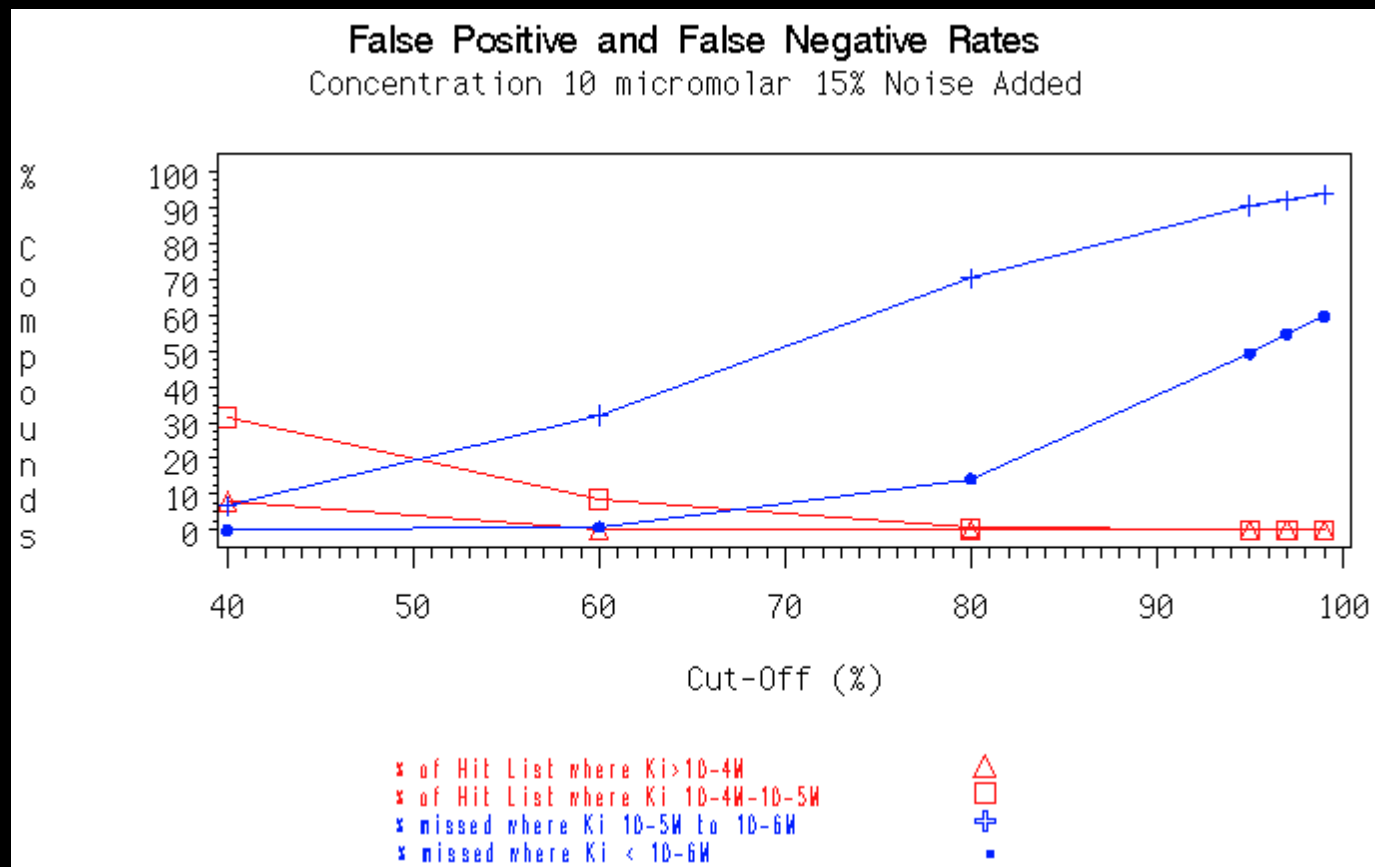
How many of each group do we find?

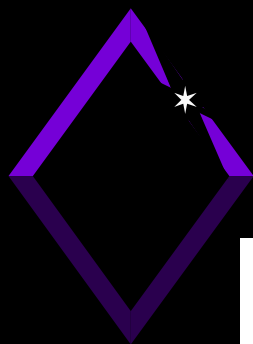
How many do we miss?



How many active compounds do we miss?

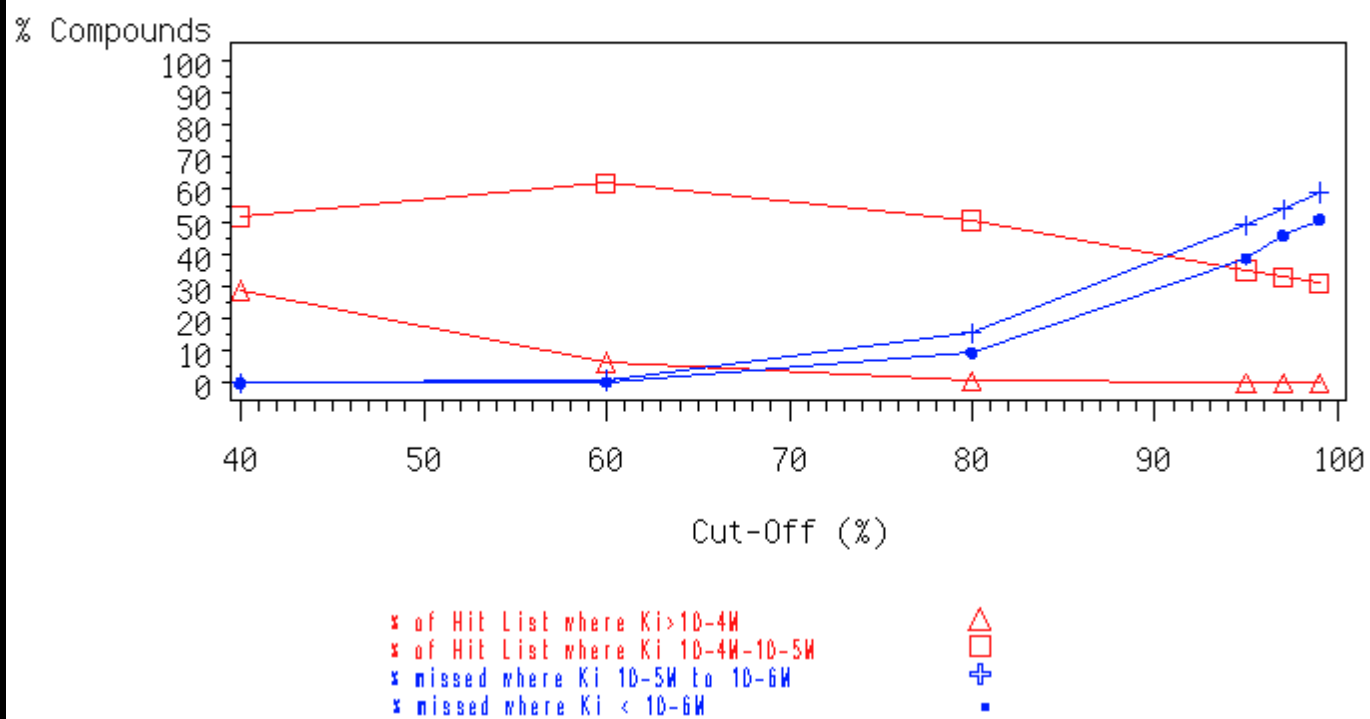
How many 'false positives' contaminate hit list?

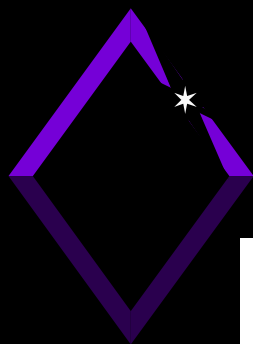




False Positive and False Negative Rates

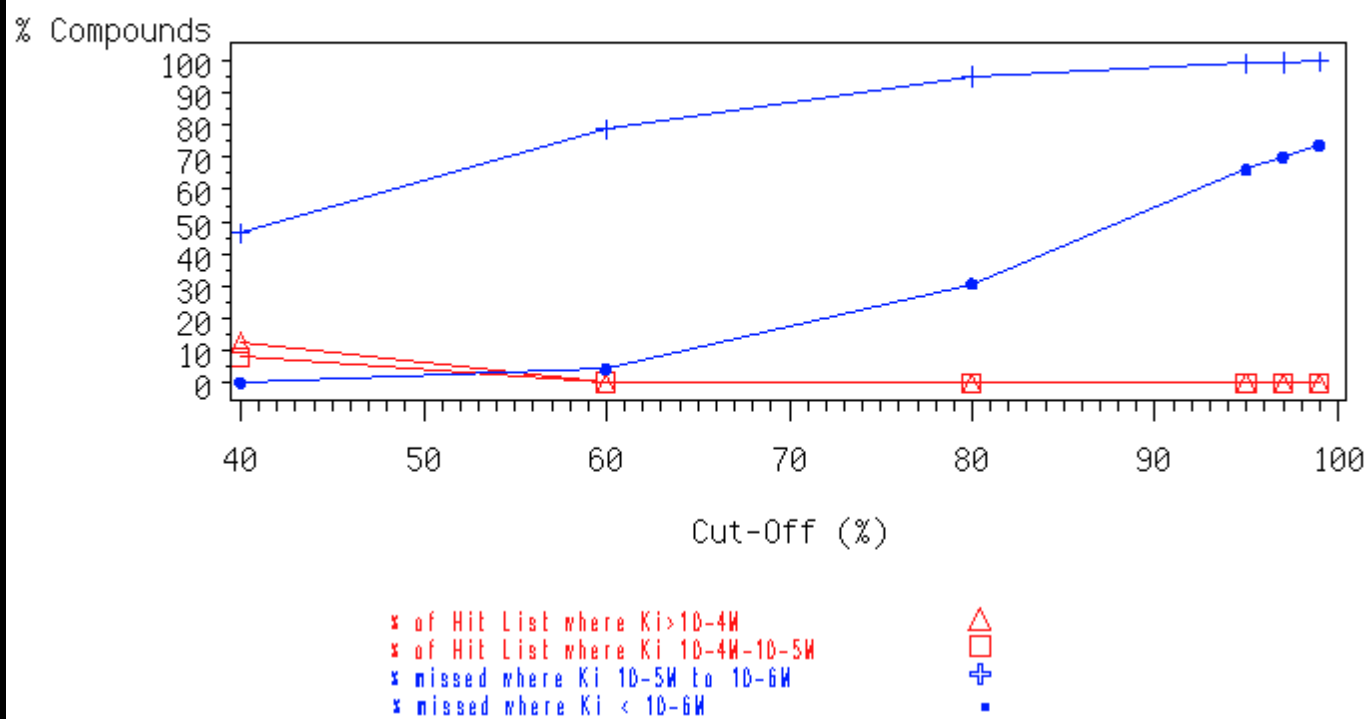
Concentration 100 micromolar 15% Noise Added

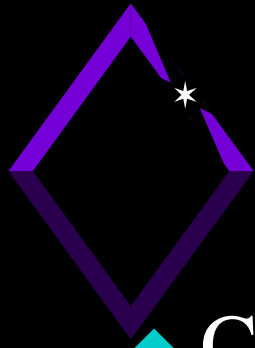




False Positive and False Negative Rates

Concentration 3 micromolar 15% Noise Added





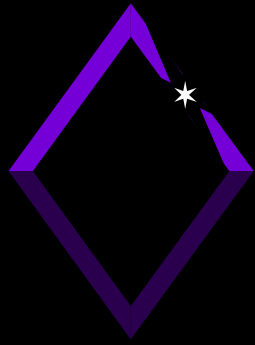
Limitations of Simulations

◆ Cons

- ◆ Simplest possible situation
 - ◆ perfect data 'clean' noise
- ◆ No systematic effects
 - ◆ solubility, screening technology, cell free systems only
- ◆ Every compound equally likely to be active
 - ◆ neighbourhood behaviour?

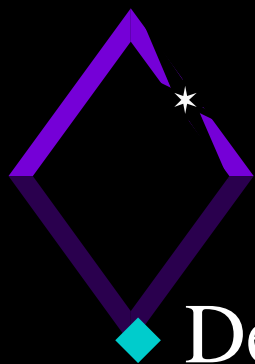
◆ Pros

- ◆ Simple
- ◆ Framework to try out ideas



Summary Biological Expectations

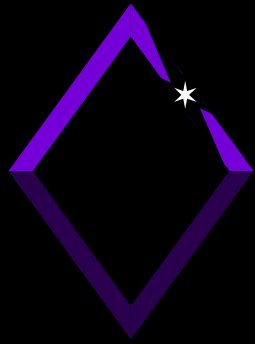
- ◆ ‘Testing everything’ finding everything
- ◆ ‘We must not miss anything’ = hit list list contaminated with ‘false positives’
- ◆ Framework for looking at these processes.
 - ◆ Suppose you have 2 screens one is cytotoxic could you find compounds of interest?



Chemical Composition of Hitlist

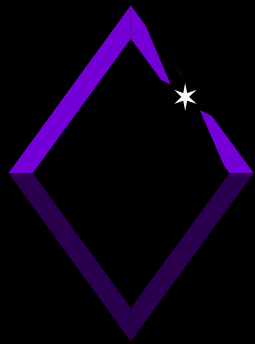
◆ Democracy rules!

- ◆ On average expect to see representation of collection
- ◆ Collection 13% acids 2% diacids
hit list 13% acids 2% diacids.
- ◆ Structural series
- ◆ Chemical 'dross'
 - ◆ dyes, biguanides, cations, alkylating agents, acylating agents



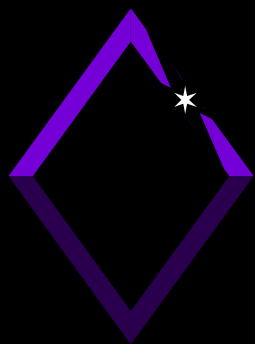
Unselective Compounds

- ◆ Define unselective as hitting more than 3 screens
 - ◆ Over 80% of compounds active in HTS hit 1 or 2 screens.
- ◆ On average unselectives (17% collection) expected to make up 35% of any hitlist
 - ◆ Chemical 'dross' concentrated in the unselectives!



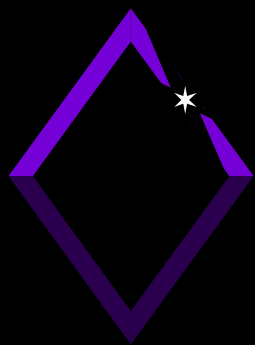
slightly grubby SAR

- ◆ Actives from HTS (no replicates)
- ◆ How many chemical families?
- ◆ How many similar tested how many active.
- ◆ How much chemical dross.



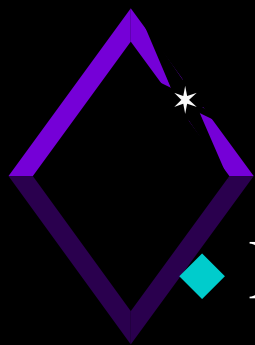
Clustering

- ◆ Work with groups of related structures
 - ◆ Clustering based on Daylight Fingerprints
 - ◆ Sphere exclusion method
 - ◆ fast
 - ◆ 'clean' clusters.
 - ◆ 2-D because synthetic routes important.
 - ◆ Neighbourhood behaviour
 - ◆ based on Daylight fingerprints
 - ◆ remove toxophores/reactives/undesireables.
 - ◆ Set of Smarts definitions, mwt cut-offs, CLOGP cutoffs



Conclusions

- ◆ Starting to understand the process.
 - ◆ Projects starting to understand the process.
- ◆ Computational tools all in the hands of Medicinal Chemists.
 - ◆ Builds in prejudice.



Expectation from HTS

- ◆ Expect large proportion weak actives
 - ◆ if cut offs set low.
- ◆ Expect representation of collection
 - ◆ including 'dross' if it's there
- ◆ Expect to see 'unselective' compounds forming large chunks of hitlists
- ◆ Hit lists entirely in line with expectation
 - ◆ It's not poor HTS data!