The use of Design of Experiments to develop Efficient Arrays for SAR and Property Exploration

Chris Luscombe,

Computational Chemistry

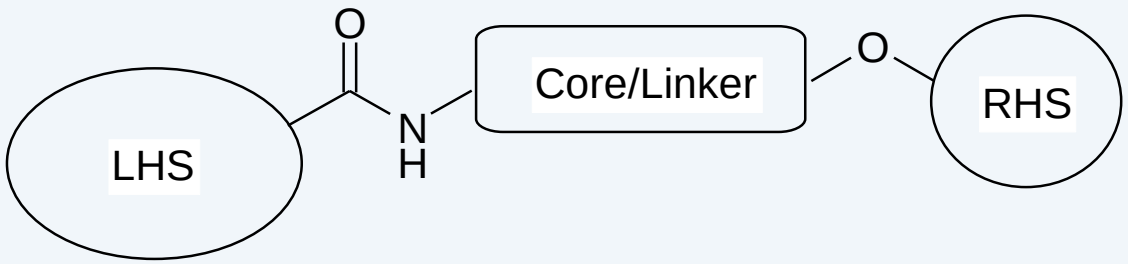
GlaxoSmithKline

Summary of Talk

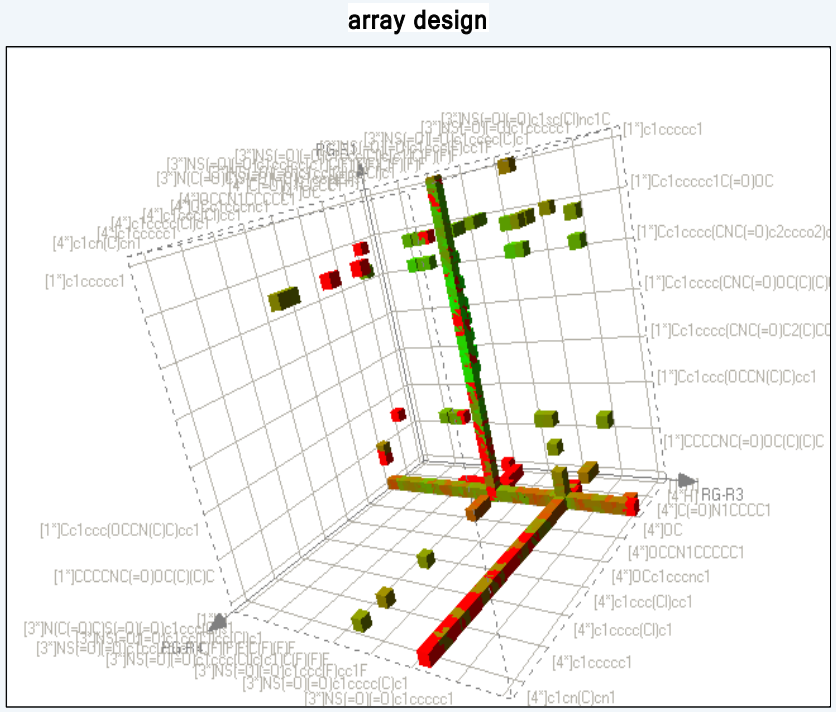
- Traditional approaches
- SAR
- Free-Wilson
- Design of experiments
- Examples
- Learnings
- Conclusions

Objectives of Lead Optimisation

- Design Array experiments to answer SAR questions to enhance potency
- Improve physicochemical properties
- Discover new monomer groups of interest.



How do we traditionally determine SAR

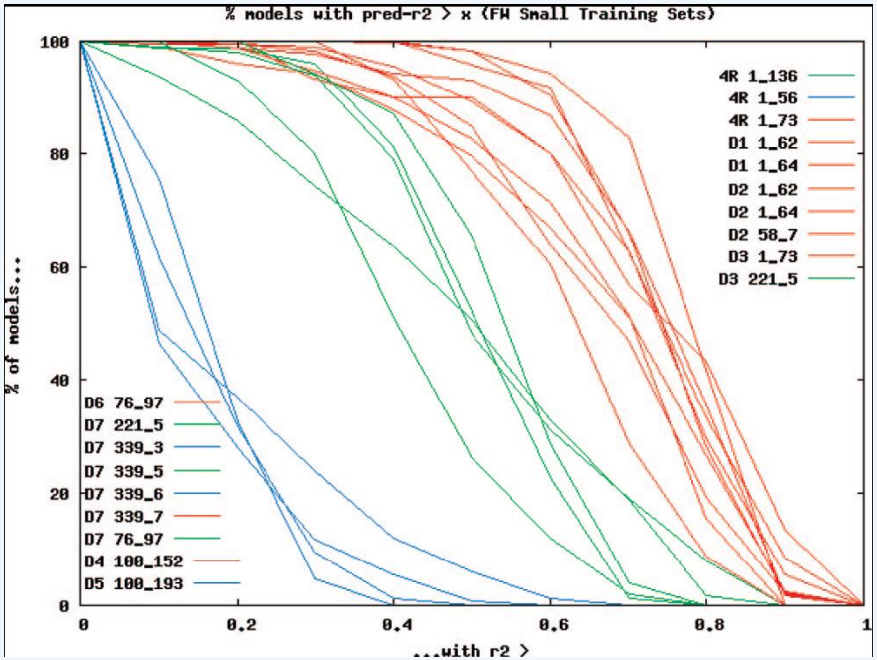


- Optimisation at a single position allows
  - Easy synthesis planning
  - Detailed understanding of SAR
- Assumes FW type additivity
  - Substituent contributions at different positions are independent and additive
- This approach is widely used and very successful

Free Wilson theory R1-Core-R2

- First mathematical technique for quantitative SAR
- Response = effect of Core + effect R1 substituent + effect of R2 substituent
- Assumptions
  - Core makes a constant contribution
  - All contributions are additive
  - No interactions between core and substituent
  - No interaction between substituents
- Can only explore chemical space defined by R-group combinations in the training set

Assessing Additivity Assumptions



Assessment of Additive/Nonadditive Effects in Structure-Activity Relationships: Implications for Iterative Drug Design *J. Med. Chem.* ***2008, 51, 7552–7562*** Yogendra Patel, Valerie J. Gillet, Trevor Howe, Joaquin Pastor, Julen Oyarzabal, and Peter Willett

Design of Experiments (DOE)

- Experimental Design approaches are well established for the optimization of multi-factor experiments, such as reaction conditions.
- Typically these domains utilize ‘continuous’ variables such as temperature, addition rate, time etc
- Can these same techniques be use where each variable is categorical?

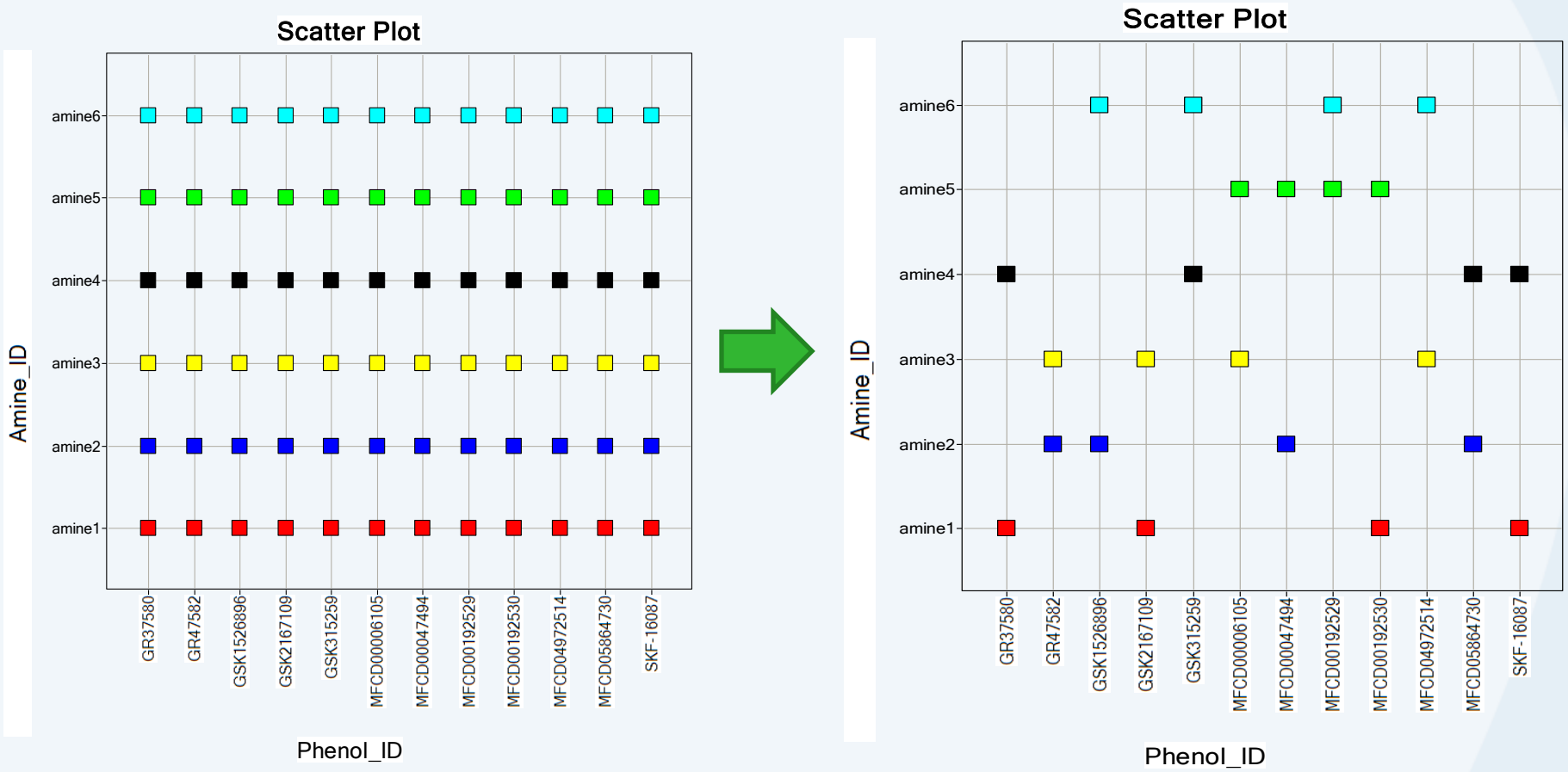
DOE in Medicinal Chemistry?

- We propose that Design of Experiments (DOE) based approaches can be applied to array scenarios where the full (e.g. M x N) array cannot be synthesized for practical reasons.

- By treating each monomer in the array as a categorical factor of the design, a balanced fractional (“Sparse”) array design can be generated.

- This novel approach can be successfully used to understand and exploit the SAR of a late stage optimisation programme

Example of a Sparse Array
1/3rd fraction from an 6 x 12 array



Questions

- Is the fraction selected sufficient to explore the chemistry space?
- Can we adequately assess monomer potential?
- Can we predict the ‘missing’ compounds?
- Is it a practical way to direct chemistry synthesis?
- Is it an efficient process?

- Does it work?

Sparse Array :What are the key steps?

Monomers

- Identify/ select which monomers to incorporate into the design

Design

- Create an experimental design template appropriate to the investigational space define from the monomer numbers

Optimise

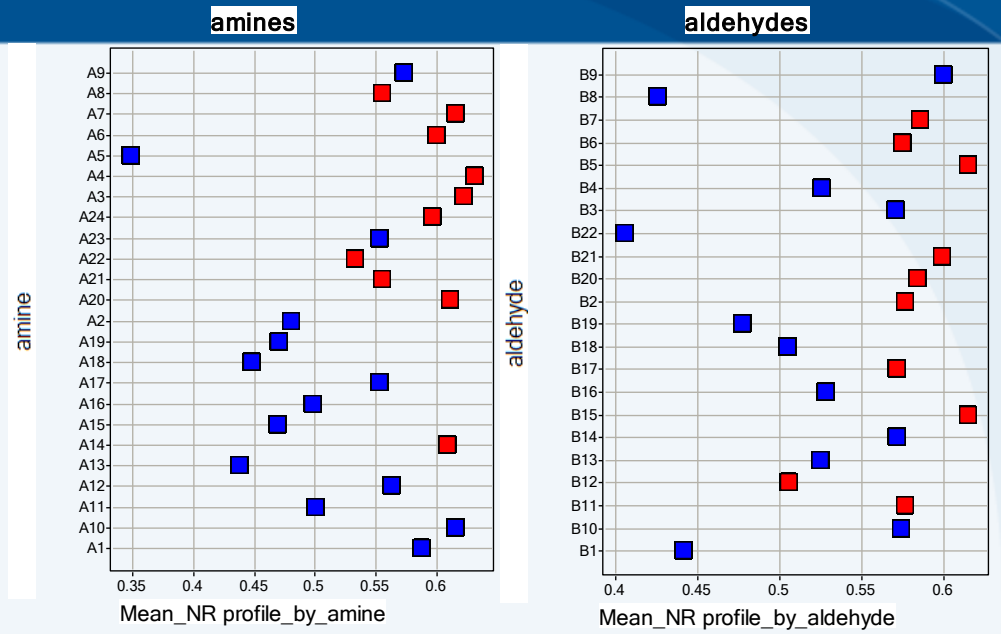
- Allocate monomers into the design so define which compounds to actually synthesize

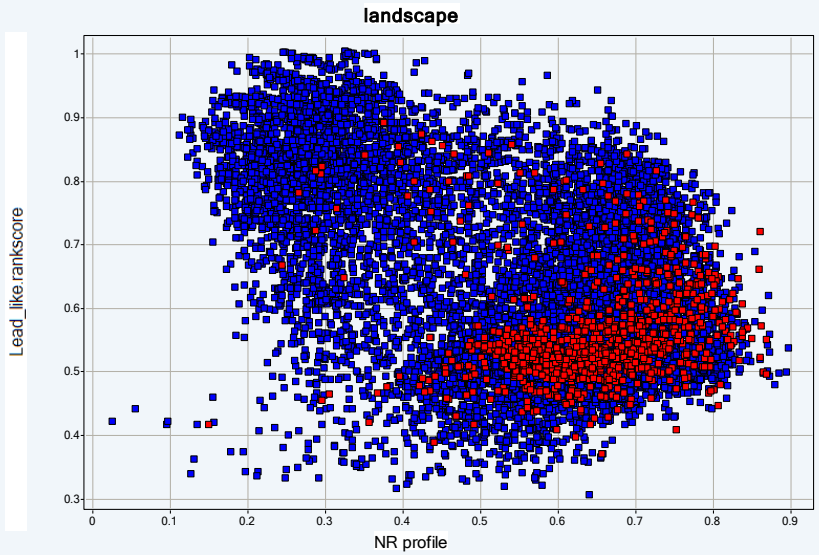
Analyse

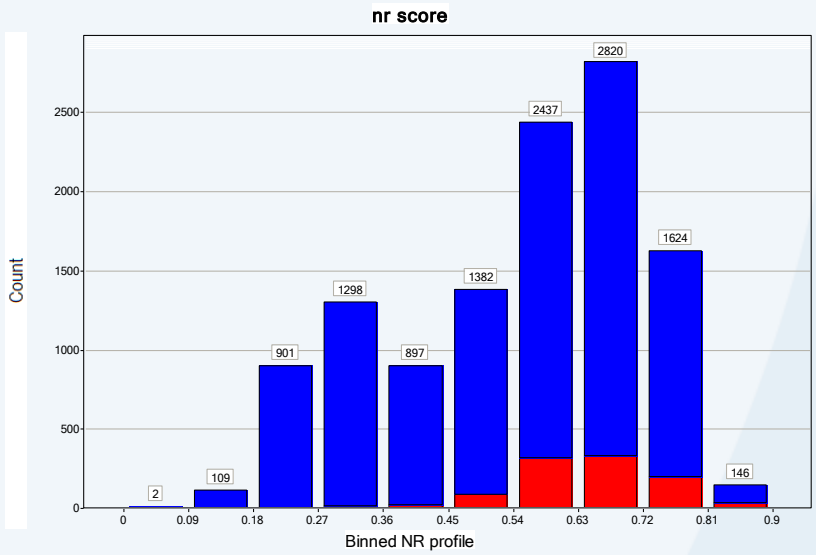
- Measure assay endpoints and build free-Wilson models to understand the SAR

Monomer Selection

- Identify appropriate monomers at each position
- Use diversity, physico-chemical, ADME and scientific rationale to reduce the monomer lists
- Calculate the average desirability score from each monomer across the whole virtual library.
- Select the higher scoring ones to be included in the Final DOE array design







Sparse Array :What are the key steps?

Monomers

- Identify/ select which monomers to incorporate into the design

Design

- Create an experimental design template define from the monomer numbers

Optimise

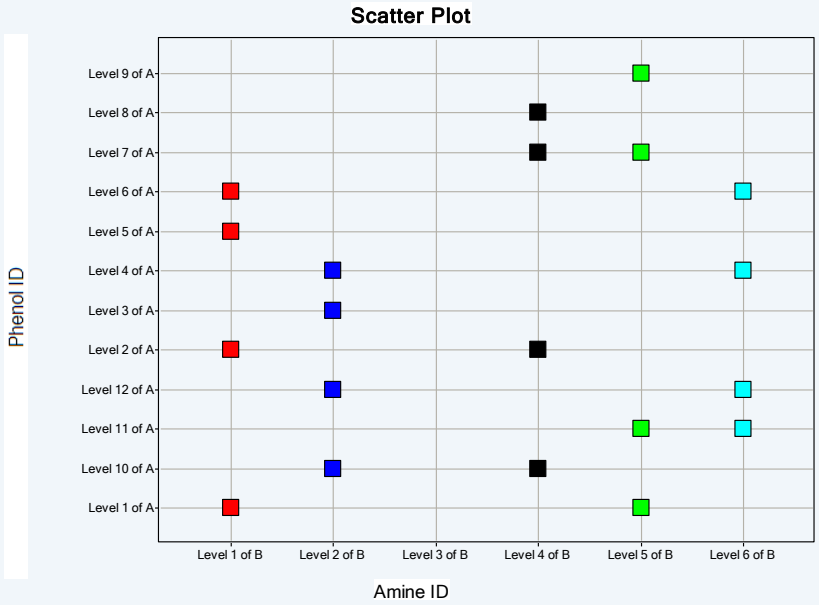
- Allocate monomers into the design so define which compounds to synthesize

Analyse

- Measure assay endpoints and build free-Wilson models to understand the SAR

Design Creation (Sparse arrays)

- Create an in-complete balanced D-Optimal design
  - Even numbers of monomers at each R position
  - D Optimality
  - Force balance



Many software packages around which can generate these types of Experimental Design

Sparse Array :What are the key steps?

Monomers

- Identify/ select which monomers to incorporate into the design

Design

- Create an experimental design template appropriate to the investigational space define from the monomer numbers

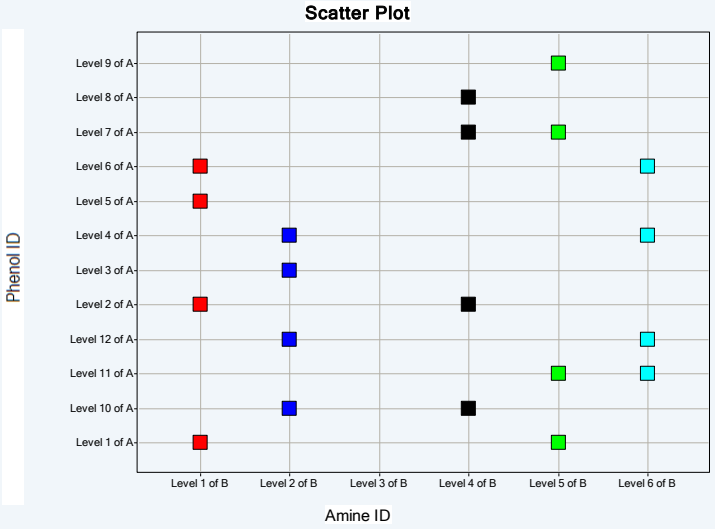
Optimise

- Allocate monomers into the design to define which compounds to actually synthesize

Analyse

- Measure assay endpoints and build free-Wilson models to understand the SAR

Which Monomer at which Position?



Compounds	A:Phenols	B:Amines
1	Phenol 3	Amine 2
2	Phenol 6	Amine 6
3	Phenol 7	Amine 5
4	Phenol 3	Amine 3
5	Phenol 9	Amine 5
6	Phenol 6	Amine 1
7	Phenol 10	Amine 4
8	Phenol 10	Amine 2
9	Phenol 8	Amine 4
10	Phenol 5	Amine 3
11	Phenol 4	Amine 6
12	Phenol 5	Amine 1
13	Phenol 1	Amine 1
14	Phenol 11	Amine 5
15	Phenol 12	Amine 2
16	Phenol 12	Amine 6
17	Phenol 1	Amine 5
18	Phenol 8	Amine 3
19	Phenol 7	Amine 4
20	Phenol 11	Amine 6
21	Phenol 2	Amine 1
22	Phenol 9	Amine 3
23	Phenol 4	Amine 2
24	Phenol 2	Amine 4

- In principle monomers could be allocated in any order, including random, into the DOE array
- GSK use an in-house algorithmic approach to allocate monomers into the define positions in the DOE array so as to optimise the compounds to be synthesized against another property
  - Eg diversity,
  - lead-likeness,
  - logP etc

Sparse Array :What are the key steps?

Monomers

- Identify/ select which monomers to incorporate into the design

Design

- Create an experimental design template appropriate to the investigational space define from the monomer numbers

Optimise

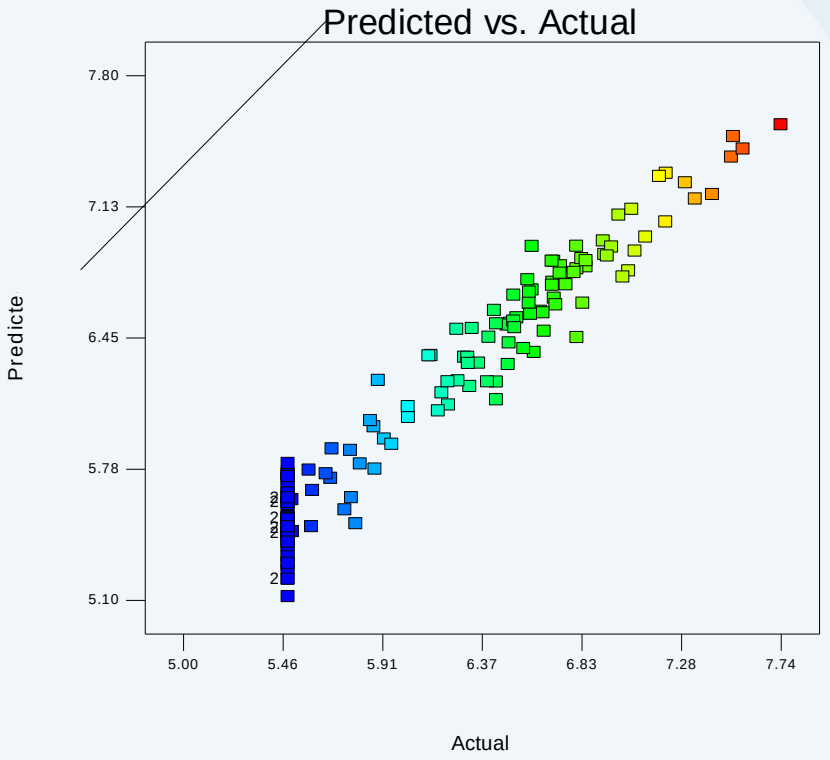
- Allocate monomers into the design so define which compounds to actually synthesize

Analyse

- Measure assay endpoints and build free-Wilson models to understand the SAR

FW analysis of monomer contribution

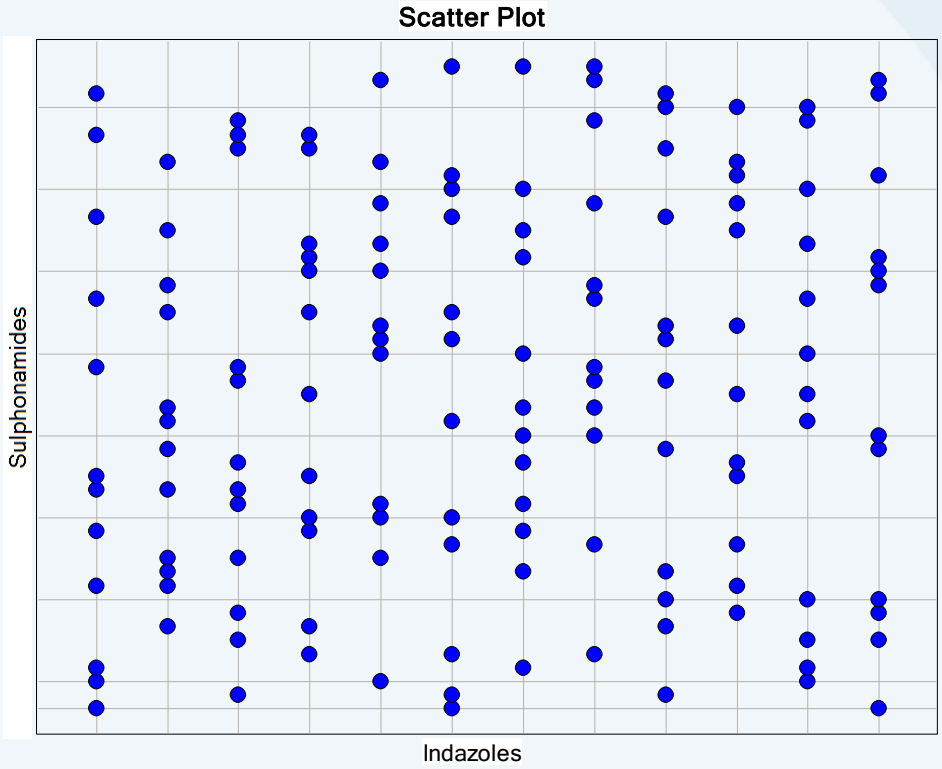
- A Free –Wilson analysis is a regression based approach to establish monomer contributions to a predictive model
- A high degree of fit suggests that the potency profile could be additive in nature.
  - The presence of outliers may imply non-additive behaviour
  - Assess potential interaction terms between monomers if the output appears to be non-additive



Example 1
Sparse array to evaluate defined N x M combinatorial space with a fractional subset

- **Design**
  - 12 Indazoles (R1)
    - Identified using classical SAR approaches
  - 48 sulphonyl chlorides monomers (R2)
    - selected from library using a variety of criteria
      - Lead-likeness score

R2

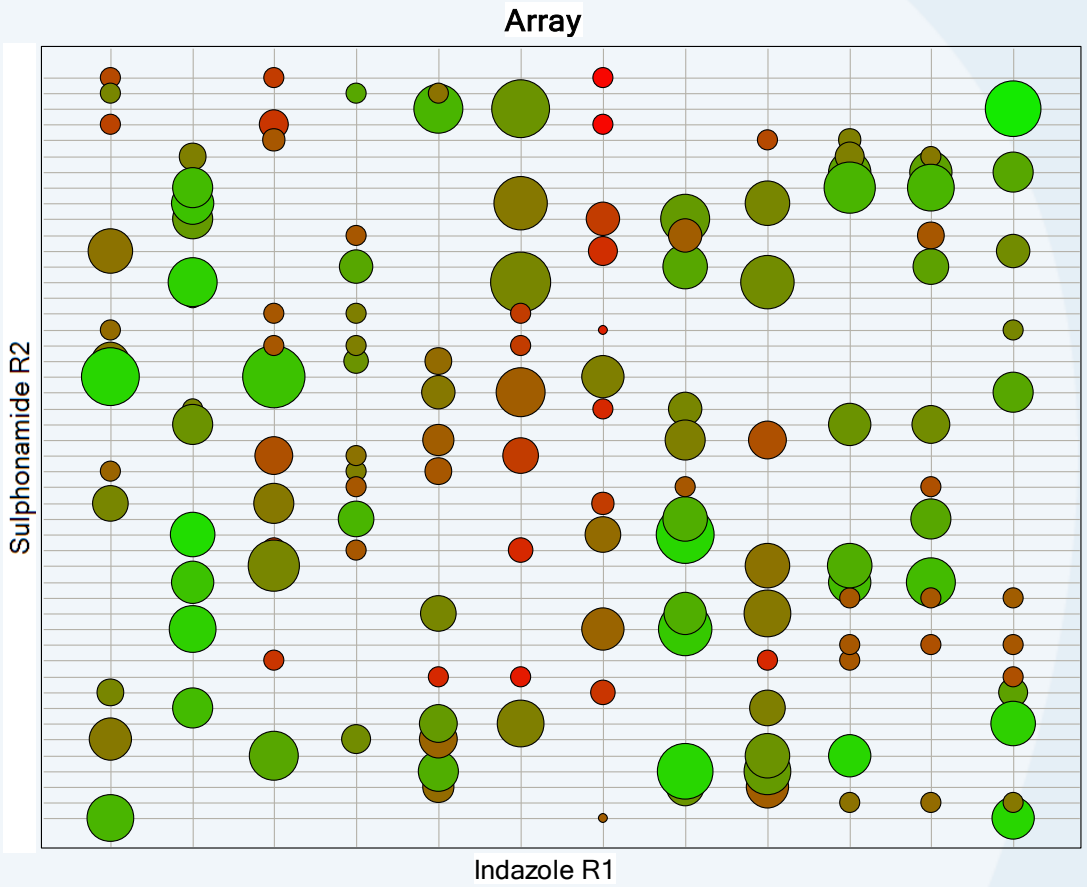


R1

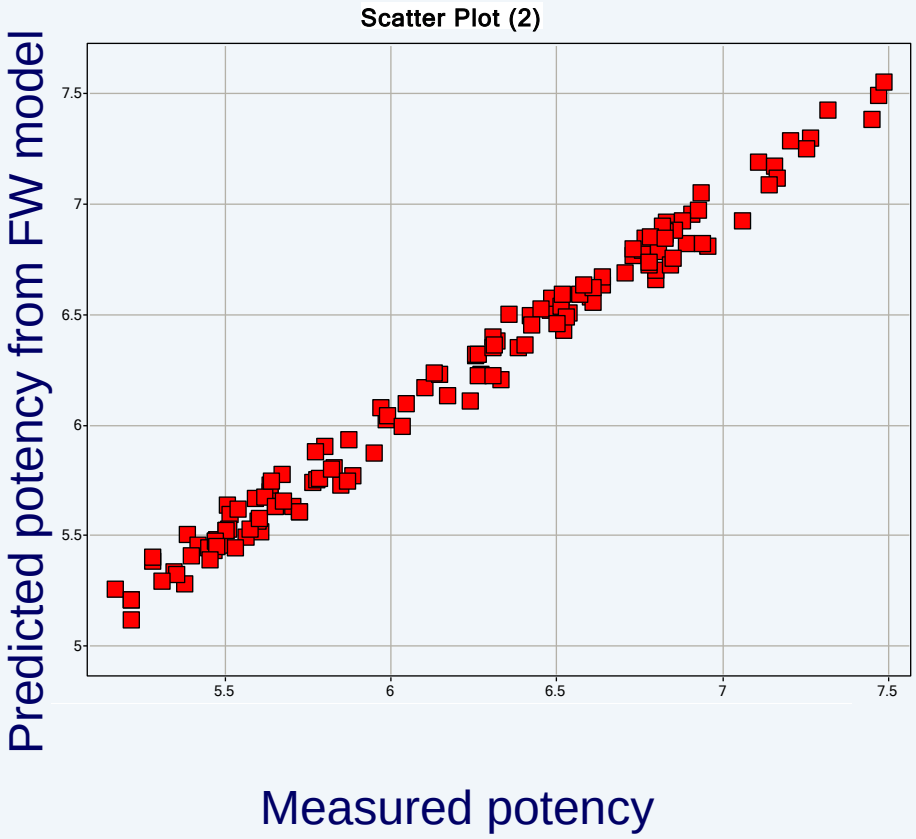
- 12 monomers per R1
- 3 monomers per R2

Measured Potency for the Sparse array

- 142 of 144 compounds from patchwork array were synthesised and tested
- Coloured for potency, sized by ligand efficiency
- Clear that some Indazoles are more promising than others



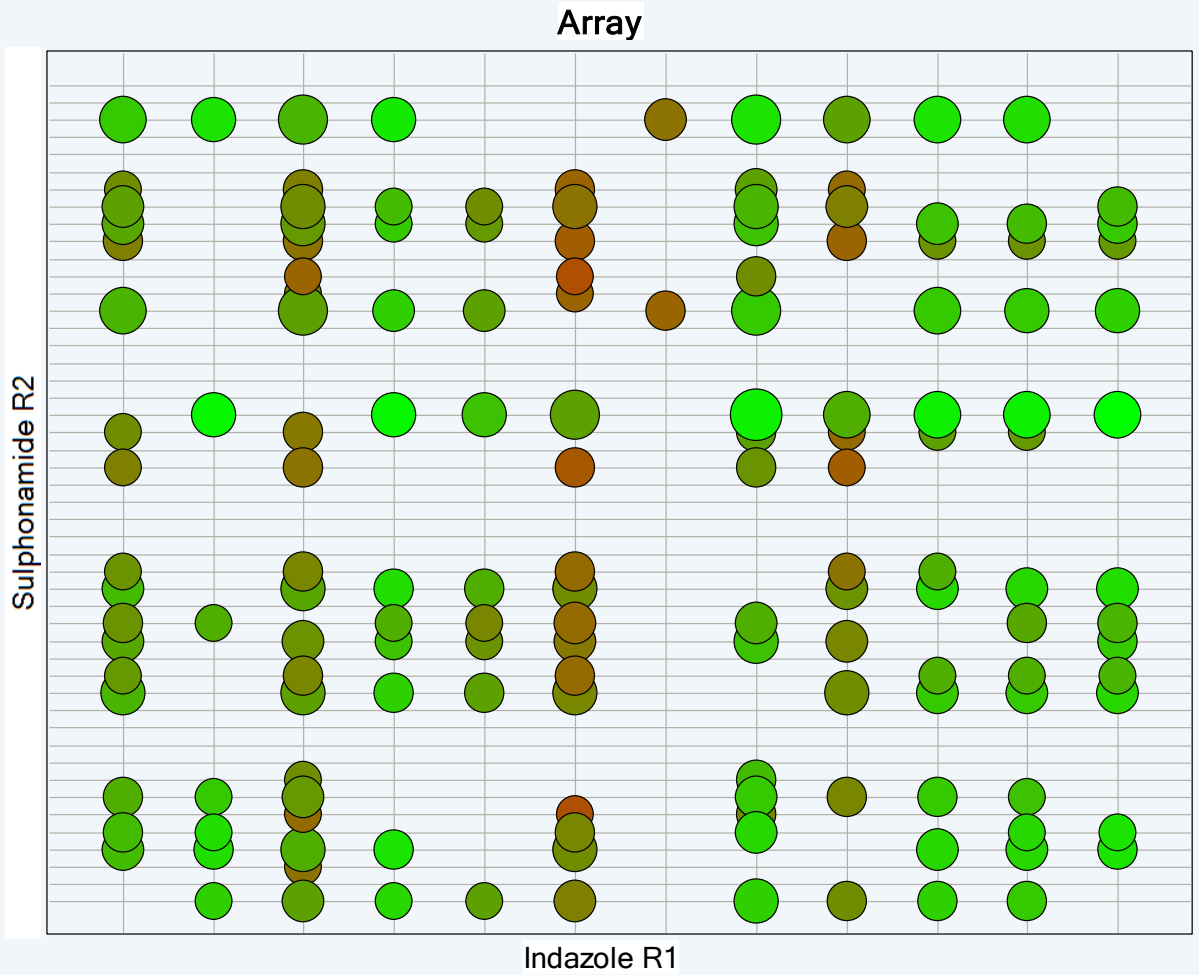
Sparse Array Data Analysis



- Statistical analysis was done to evaluate ‘additivity’
- Free Wilson model: Predicted potencies were plotted against measured potencies
- The FW model show potential excellent additivity with no outliers.

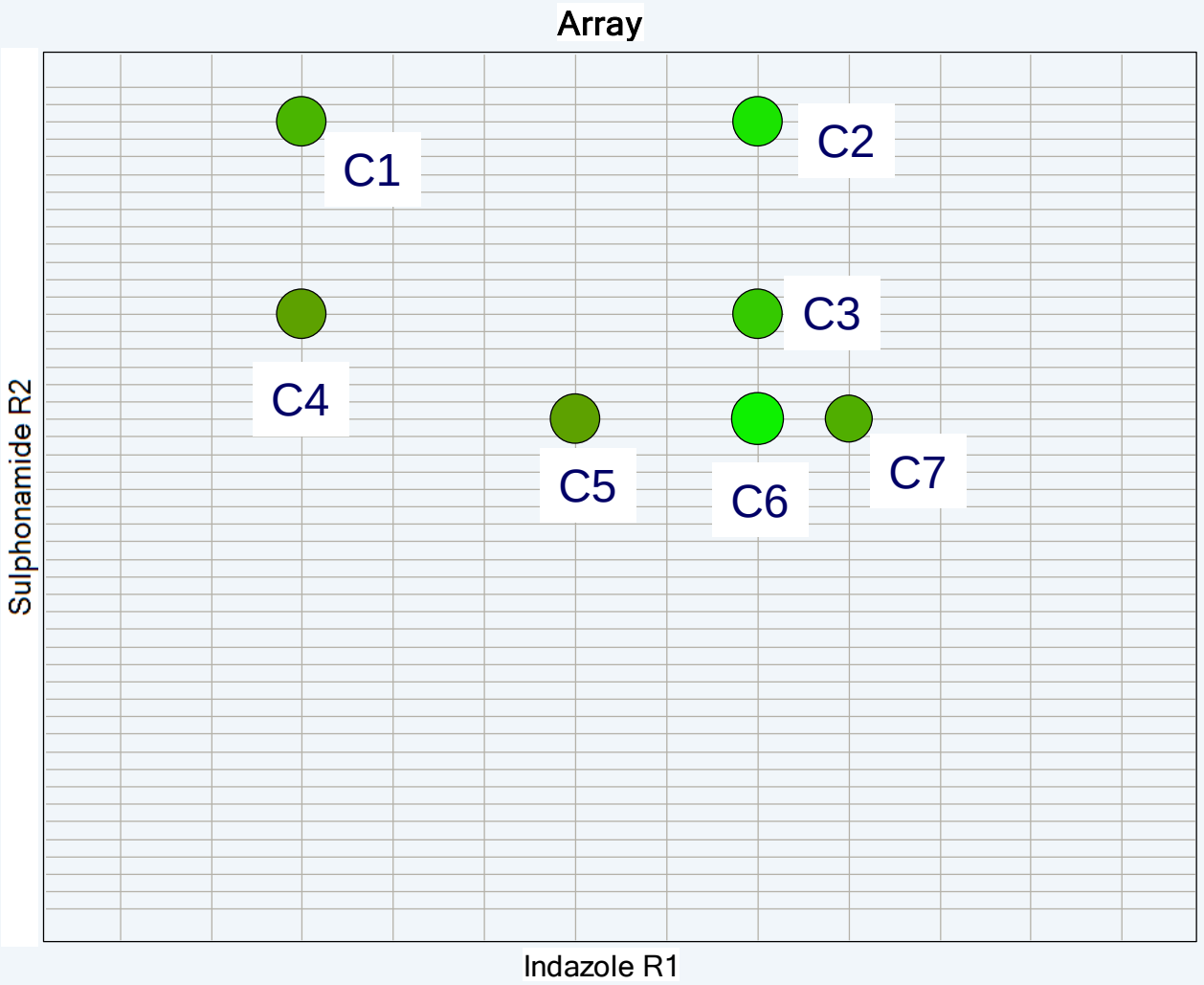
Predicted Potency for the complete array of 576 compounds (Fit and Predict), only Actives (pIC50>6.5 shown)

RG-R2
(48 Variants)



RG-R1 (12 Variants)

Find the predicted most potent compounds that haven't already been synthesized



RG-R2
(48 variants)

RG-R1 (12 variants)

Predicted potent compounds

- All compounds subsequently synthesized had measured potencies within +/- 0.2 pIC50 of the predicted value
- Validated the Additivity assumption
- Identified promising alternatives which were sent for further PK analysis – potential back up to the current pre-candidate

C1
Predicted GTPgS = 7.6
BEI = 16.0
Measured = 7.6

C2
Predicted GTPgS = 7.5
BEI = 13.5
Measured = 7.6

C3
Predicted GTPgS = 7.5
BEI = 14.8
Measured = 7.3

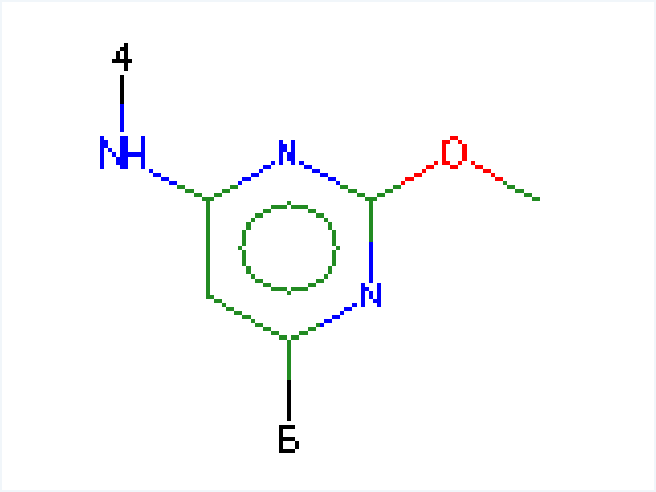
C4
Predicted GTPgS = 7.5
BEI = 14.2
Measured = 7.4

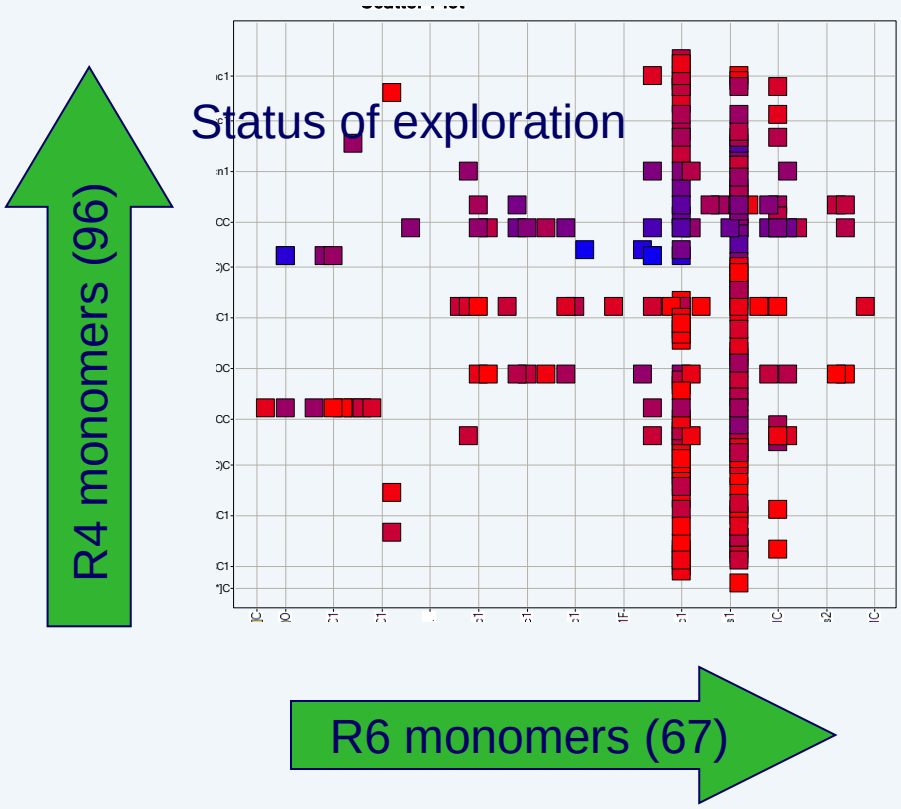
C5
Predicted GTPgS = 7.6
BEI = 15.6
Measured = 7.5

CAT friendly example
Sparse Array Automation

- CAT : Automated array chemistry system
- A particular design (nicknamed the Tetris array) which is ‘array automation’ friendly and thus allows these investigational approaches to be carried out efficiently from a synthetic perspective.

Exploration of Chemical space coverage for a Dual targetting programme

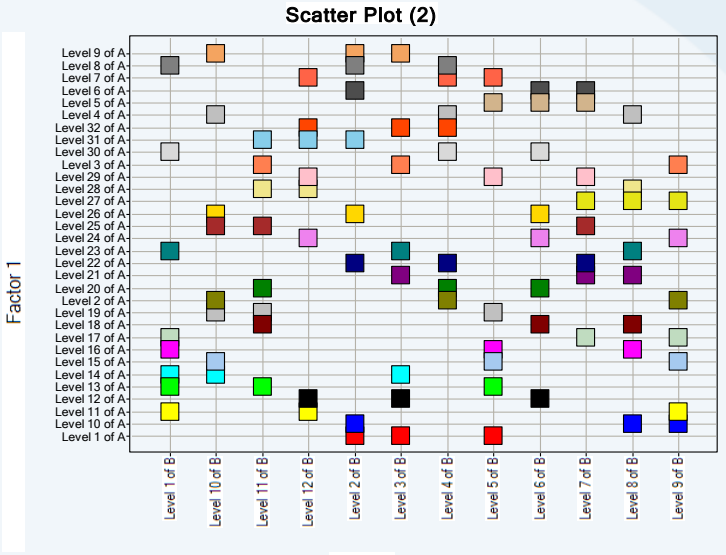




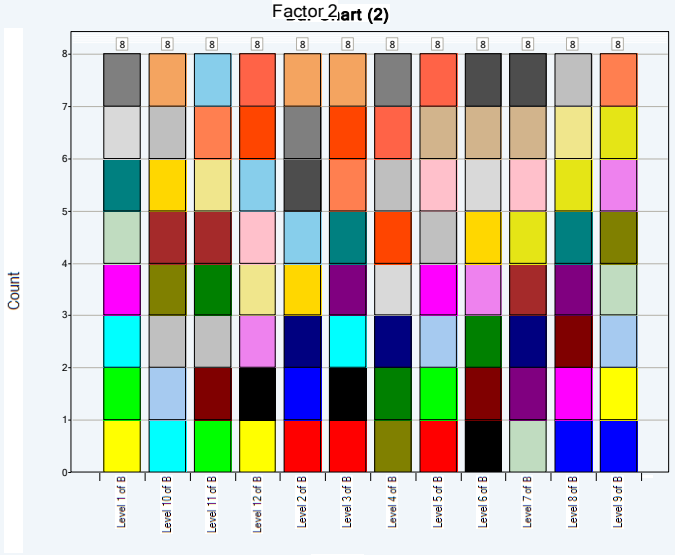
32 R4 and 12 R6 monomers were chosen for inclusion in Sparse array

CAT friendly 8 (from 32) x 12 Tetris Array

- The experimental design chosen is a 8 x12 chosen from a potential 32 x 12 fully enumerated array (384 potential compounds).
  - (¼ fraction)
- Each coloured block represents one of the 32 R4 monomers
  - Each R4 monomer is used 3 times
  - Each R6 monomer is used 12 times



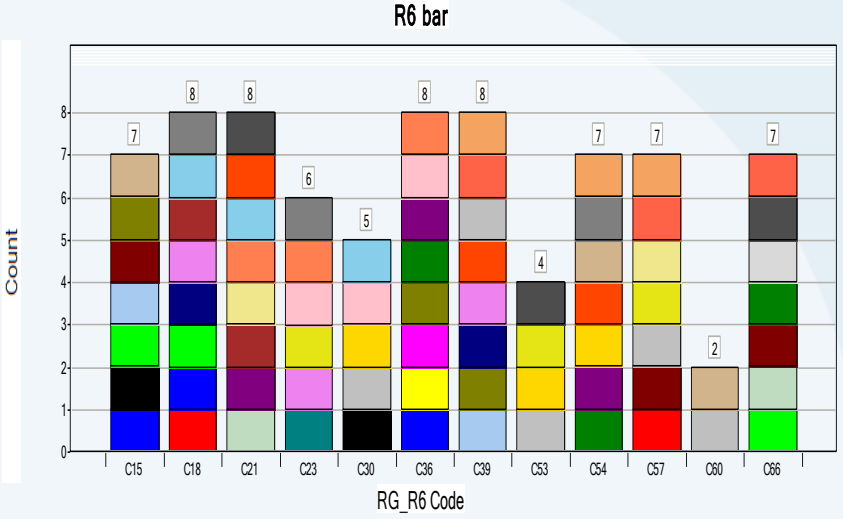
R4 Monomers

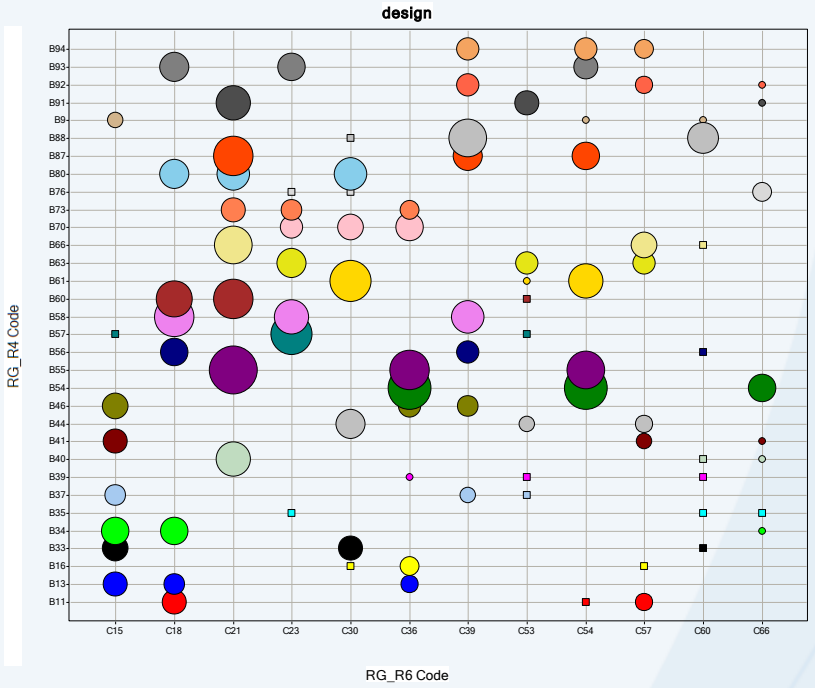


R6 Monomers

Sparse array results

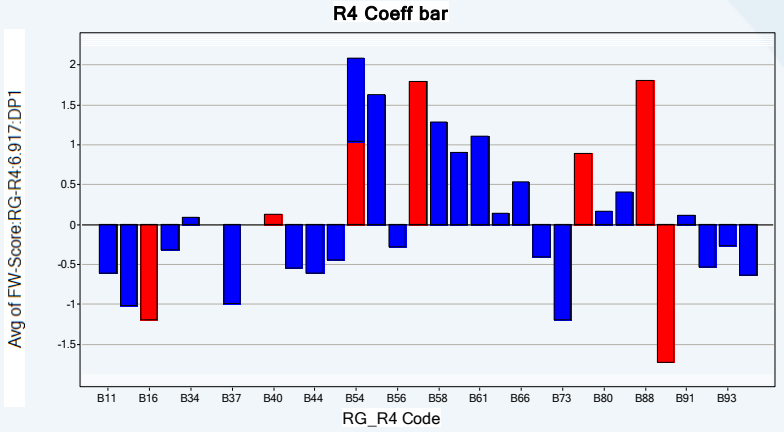
- Using the CAT the synthesis was done efficiently and effectively
  - Synthesis was actually done using 8 linear (1x12) arrays
- For the Sparse array synthesis of 77 of the 96 compounds was achieved and the compounds delivered to screening.
  - This is approximately 75% of full sparse array
  - Only 20% of the fully enumerated array

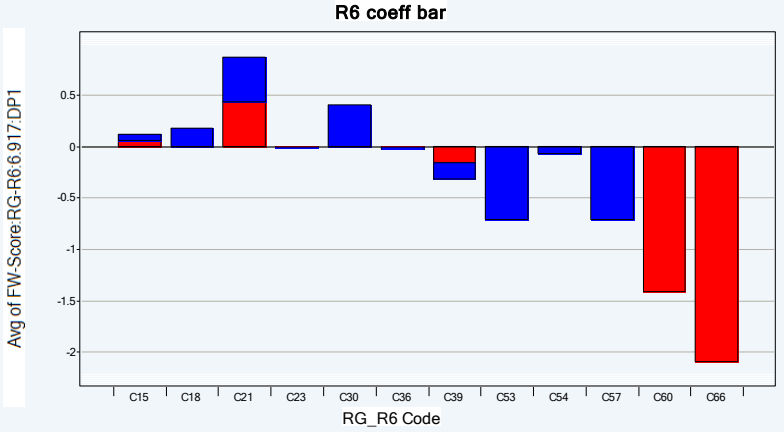




Monomer contribution

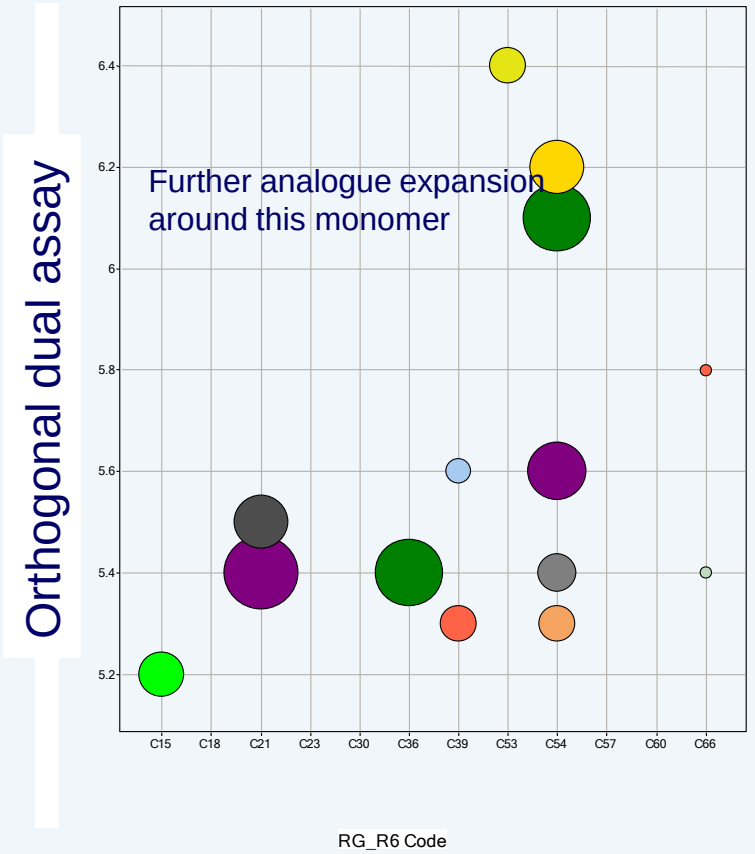
- The Programme team concluded that the chemistry within this area of chemical space was well understood wrt target potency.
- The Programme team predicted potent analogues with targetted physchem profiles for synthesis





Further Developments Dual target

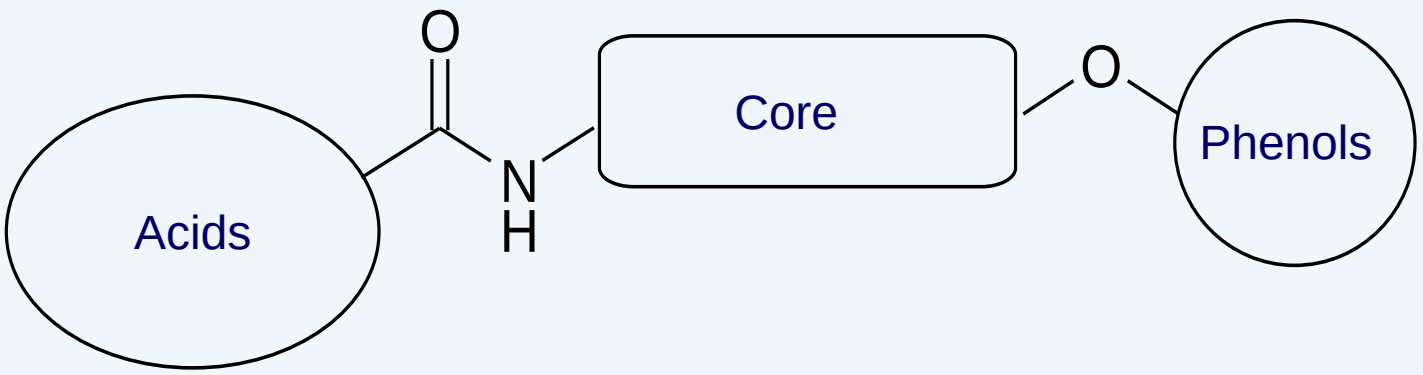
- The monomers chosen in the array were selected to create Primary actives but were not thought likely to have any potential in Secondary target assay
- However, surprisingly 14 compounds were found to be active in the second assay
  - Currently being followed up in the programme team as potential dual antagonists



Sized by Primary potency coloured by R4 group

EXAMPLE 3
EXTENDED 3 RG TETRIS SPARSE ARRAY

3 points of change on the molecule



Extended 3 RG Tetris Sparse array

Cores (A) = 3 (These were used to explore a stereo chemistry question)
Phenols (B) = 4
Acids (C) = 24

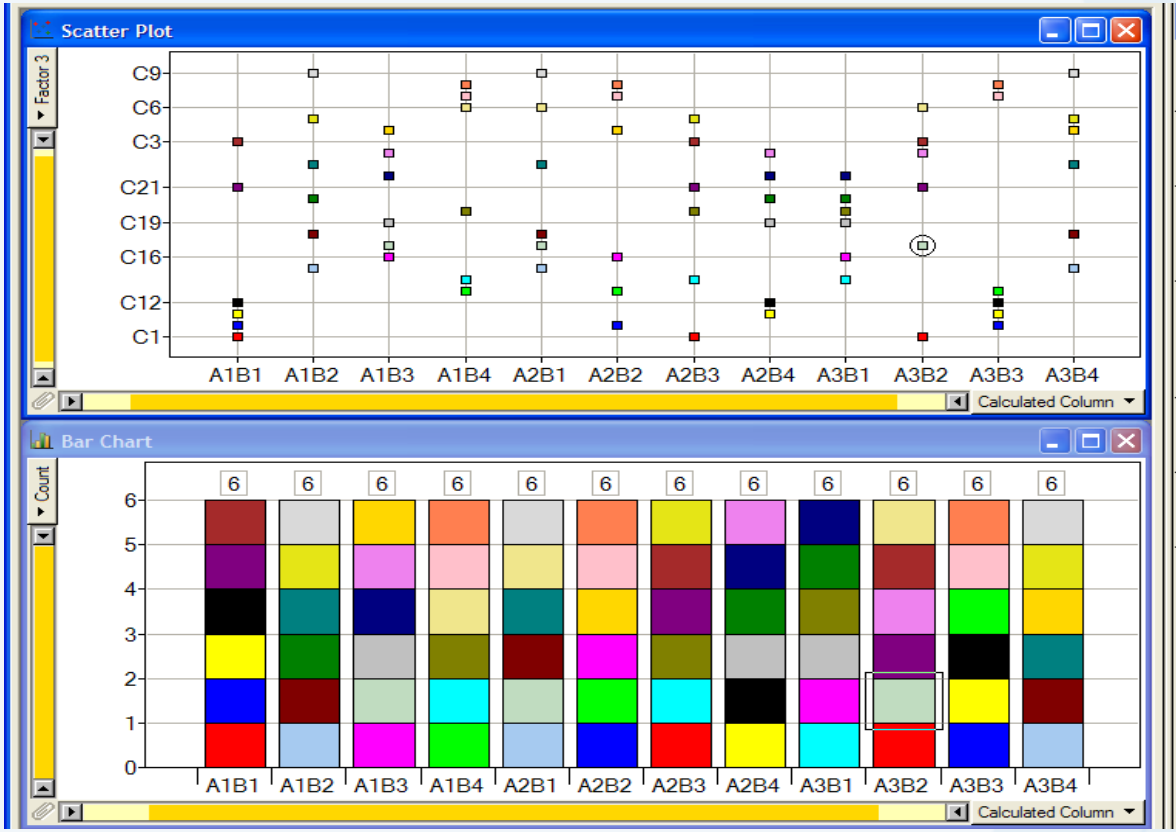
All acids represented 3 times

3x4x24 = 288 compounds

25% of full array synthesised

Distribution ‘balanced’

Extended TETRIS array



Coloured by Acid monomer group

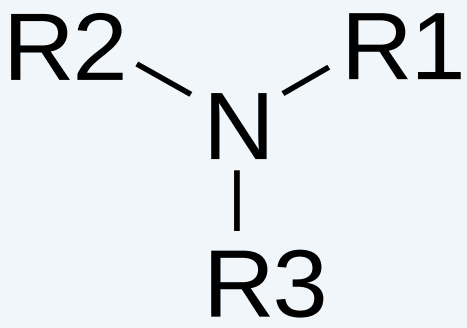
OTHER DESIGN TYPES

Latin Squares: Symmetrical design spaces

Useful for n x n x n problems
where n = number of monomers in each RG position

Eg 6 R1 x 6 R2 x 6 R3

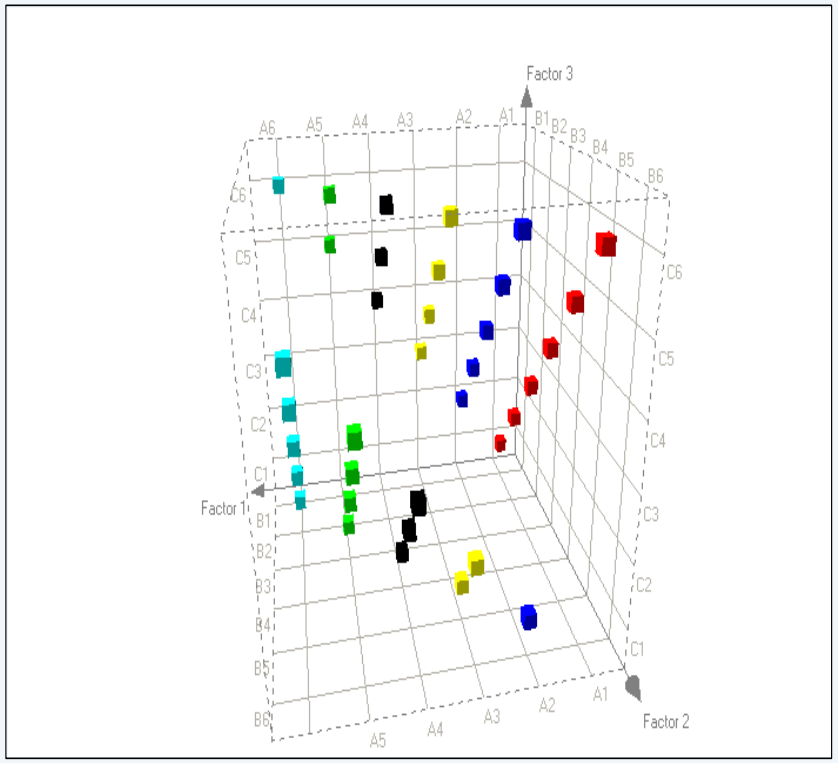
A 1/n fraction is selected



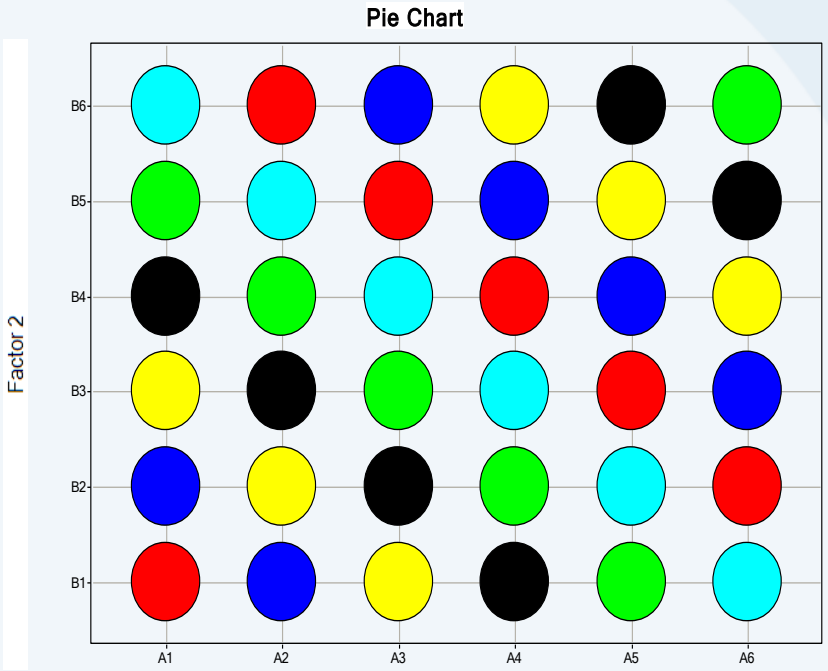
Three RG positions – Latin Squares

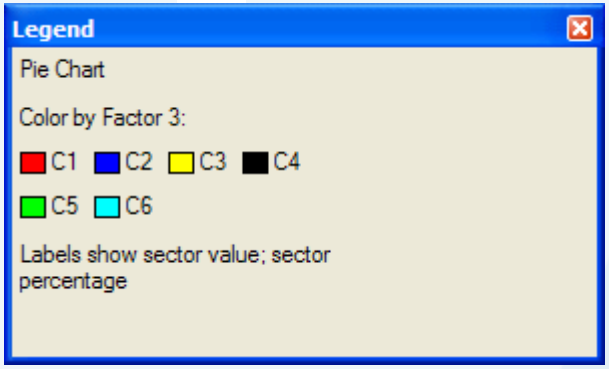
Predictive Array Design: LIPKIN, ROSE,SAR and QSAR in Environmental Research, 2002 Vol 13 (3-4) pp425-432

Scatter Plot



Pie Chart





Each possible pair of monomers is present once and that each monomer is present an equal number of times – defined by the array dimensions.

Pros and Cons of Sparse array approaches

Objective exploration generates an optimal data set for ANOVA / Free-Wilson analysis.

Complete evaluation of potency response within the design space from only a fraction of the possible compounds

Defined endpoint to the work

Excellent data set for QSAR

Chemistry may be more difficult to carry-out

Needs a reasonable resource commitment upfront

Needs majority of compounds chosen in the array to be made and measured for the analysis to be robust

Assumes Additivity (but then so does Linear SAR exploration)

Learnings from experience

- Ideally 3 examples minimum for each monomer within the design, although 2 will work for a robust assay and chemistry
- Need to have confidence in getting some active compounds
  - If all the compounds are inactive its difficult to fit a model!
- Confidence in ability to synthesize compounds
  - Some loss of particular compounds can be tolerated but if whole reactions fail then the array design will be compromised

Summary

- Experimental Design may provide an alternative /complementary strategy which may be suitable in some circumstances
  - E.g. Initial exploration of new monomer space
  - Identification of back up compounds
  - Establish Addivity in the series
- Efficient Lead Optimisation by exploring more than one point of change at the same time on the molecular template
- Can unearth some surprises which may never have been found by traditional processes

- There are different design types for different situations
  - Software is available to create the designs
  - Work well in situations where the bespoke synthesis is contracted out

Acknowledgements

- Tony Cooper
- Heather Hobbs

Medicinal Chemistry

- Nick Barton
- Stephen Pickett
- Darren Green

Computational Chemistry

Thank-You

References to literature to date

- How design Concepts can Improve Experimentation: Mager 1997
  - Use of iterative search techniques to find monomers which have the required levels of particular physico chemical properties to fit into an ideal experimental design
- Statistical Molecular Design of BB for CombiChem: Linusson, Wold etal 1999
  - Selection of BB's using t- scores of the enumerated library as variables and then applying D-Optimal , Space Filling or Cluster based selection strategies
- Predictive Array Design: LIPKIN, ROSE etal, 2002
  - Use of Latin Squares
  - Simulated Annealing for Monomer assignment