

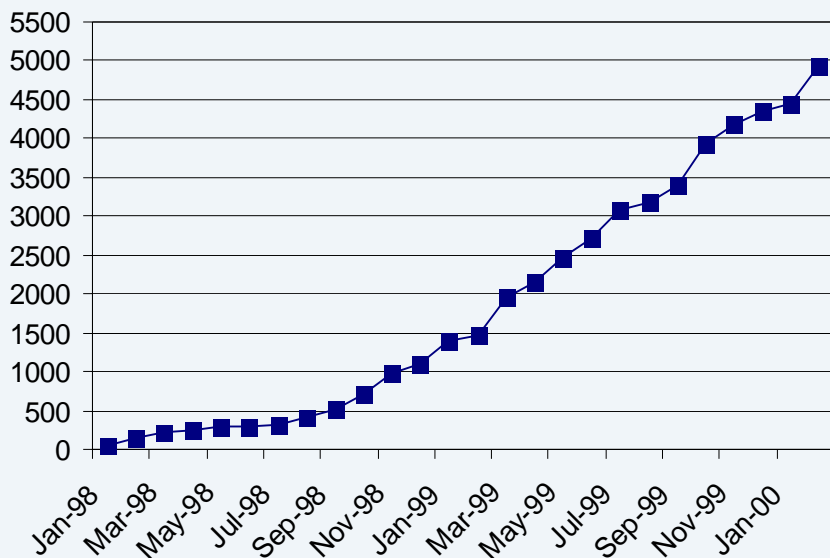


# The SAR Toolkit

Francis Atkinson and Gianpaolo Bravi  
MDR CSC

# The need for improved tools for SAR analysis

- Lead-Op programs can generate huge amounts of data...



- Program ran for two years
- Three series optimised simultaneously
- nearly 5k molecules synthesised

- Data analysis using existing tools very time demanding
- Subtle trends and outliers alike can be missed
- 'Historical' SAR can be neglected

# What should we provide?

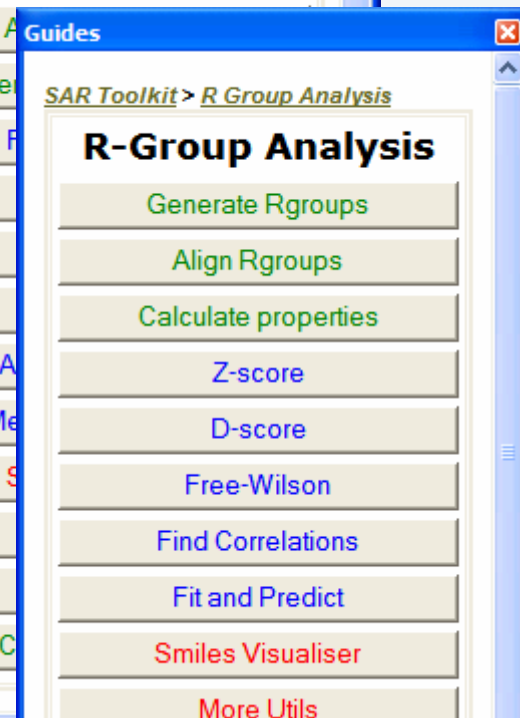
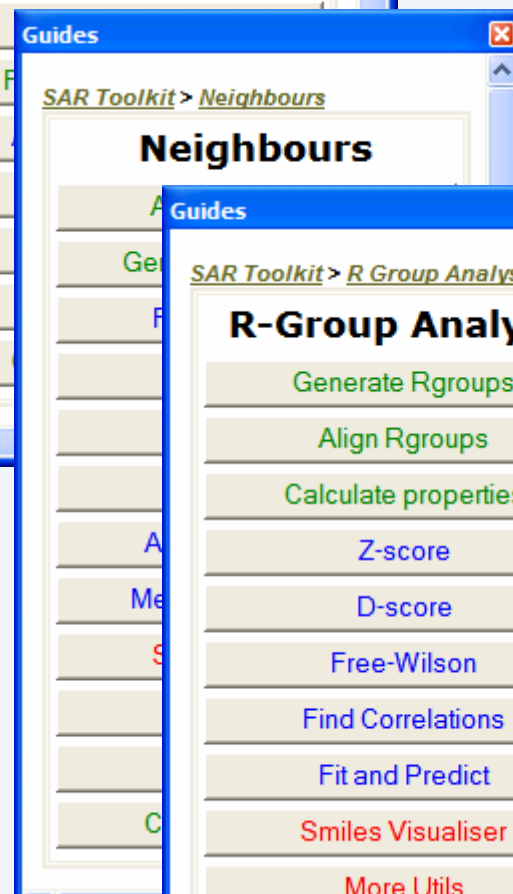
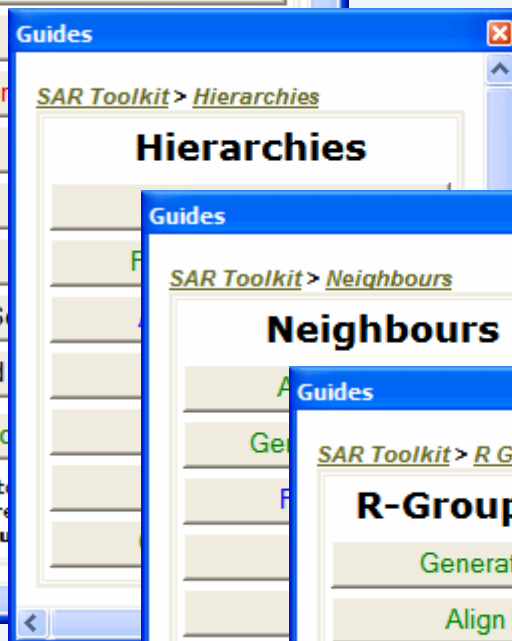
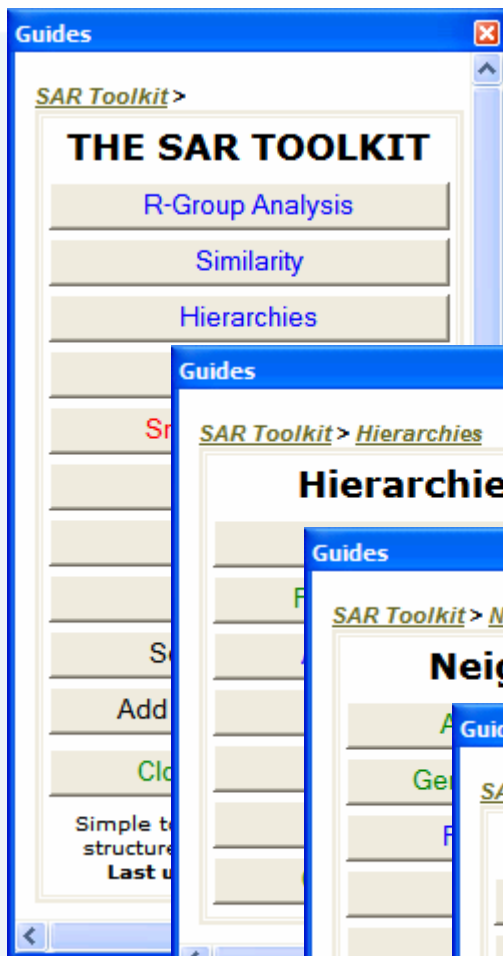
- Tools suitable for Medicinal Chemists
- Tools for Exploratory Data Analysis
  - “Maximize insight into a data set”
  - “Uncover underlying structure”
  - SAR *not* QSAR
    - Model building can come later
- Tools that...
  - Retrieve, collate and format biological data
  - Organise data for easy browsing
  - Highlight trends and outliers
  - Facilitate discussion of SAR and dissemination of conclusions

# Spotfire chosen as a platform

- Accessible to all within GSK
  - Already popular within MedChem community
- Allows rapid development & deployment of tools
  - Embedded web-browser
  - Client-side scripting with JavaScript
  - Easy integration of Unix-based tools...
    - Daylight Toolkit
    - Scripting with Perl *etc.*

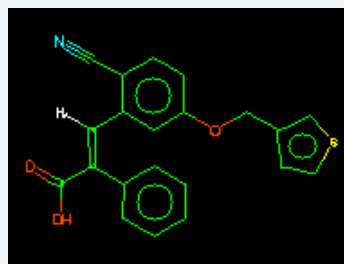
# Outline

- The SAR Toolkit
  - Hierarchies
  - Nearest Neighbours
    - Similarity
    - Single-point changes
  - R-groups
    - Fragmentation
    - D-score analysis

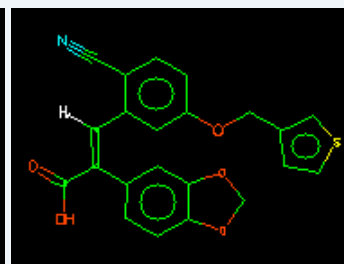


# Introduction to Hierarchies

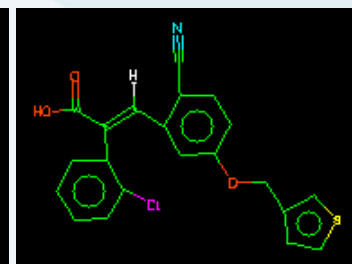
- Take 9 molecules...
- It may not be hard to analyse these molecules
- But ...



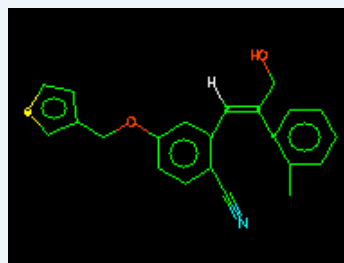
7.18



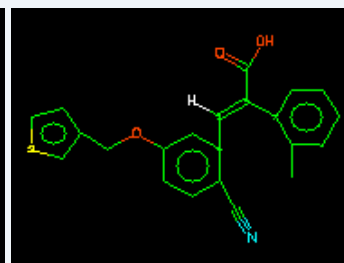
6.69



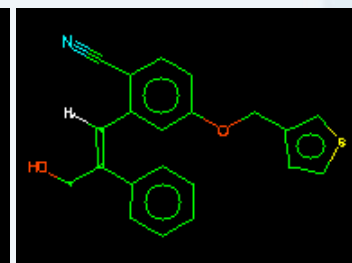
6.69



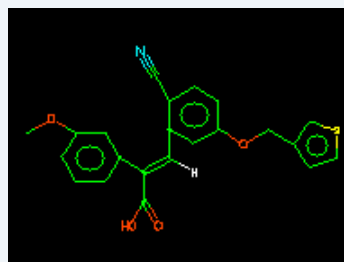
6.52



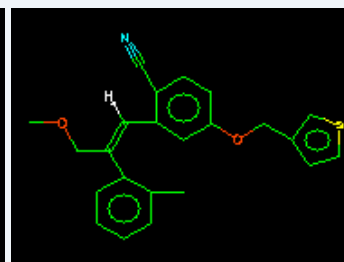
6.52



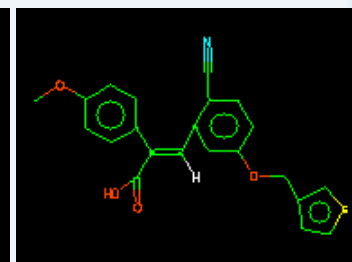
6.15



5.95



4

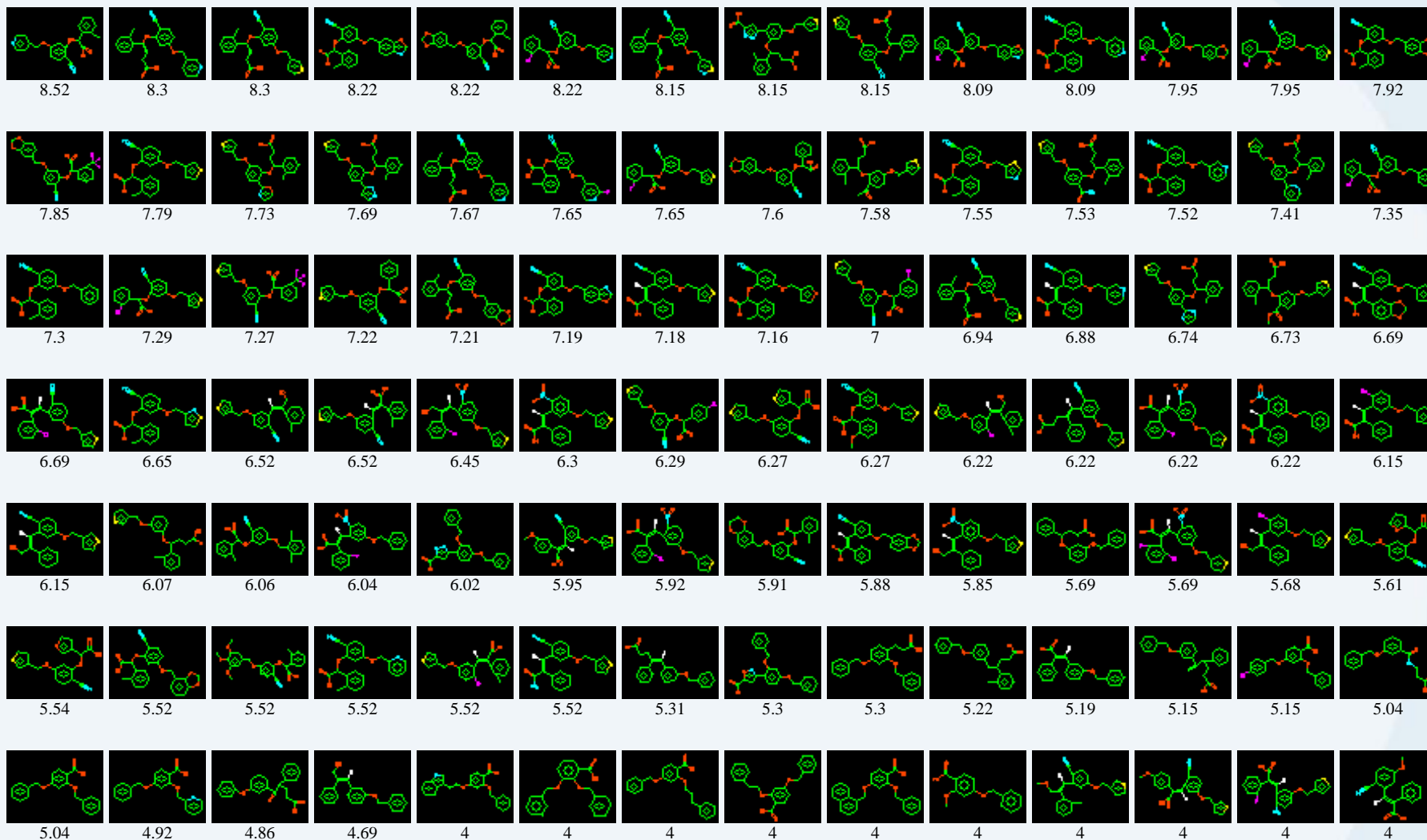


4

- $ET_A$  antagonists
- Non-GSK set (from Aureus)

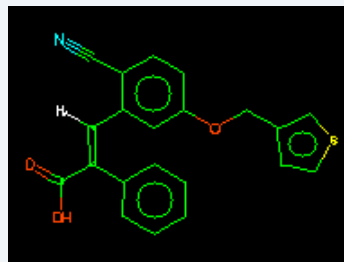
# ET<sub>A</sub> antagonists

- 98 members in set (after removing duplicates)

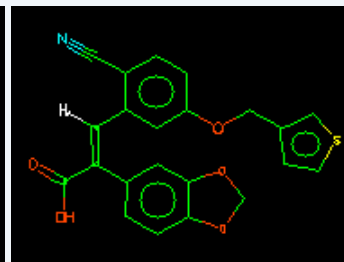


# Introduction to Hierarchies

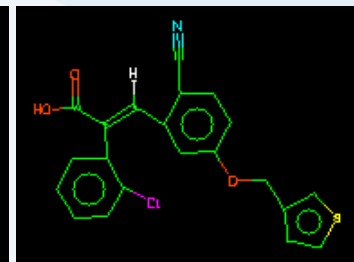
- Things may get easier if we rearrange molecules into a structural hierarchy



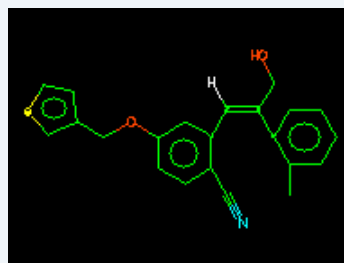
7.18



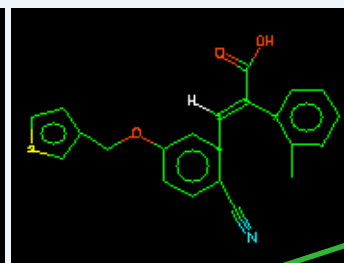
6.69



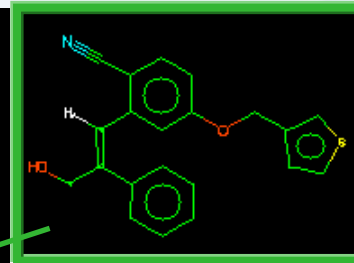
6.69



6.52

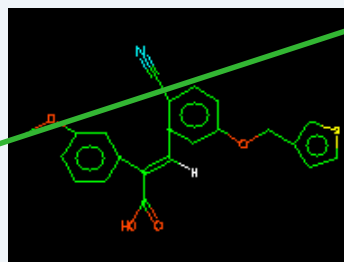


6.52

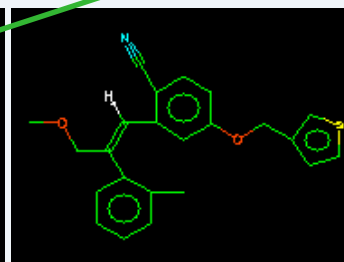


6.15

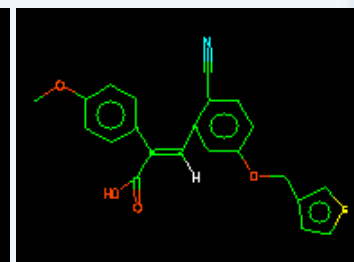
Substructure of all the others: becomes the Parent



5.95



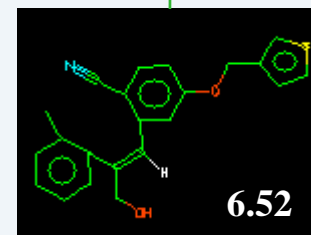
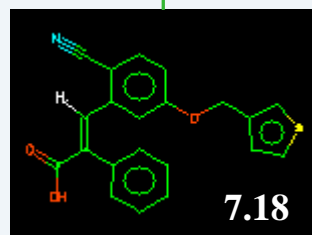
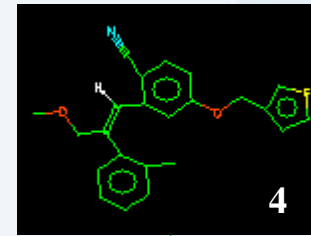
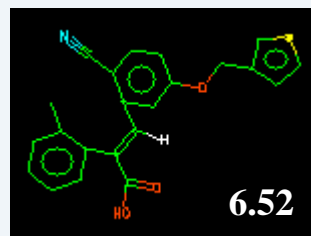
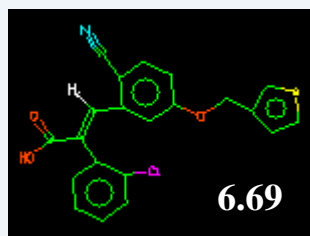
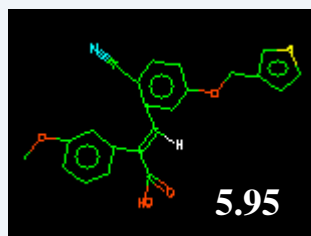
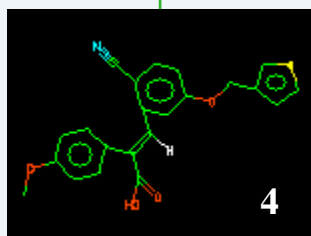
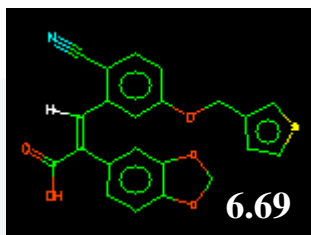
4



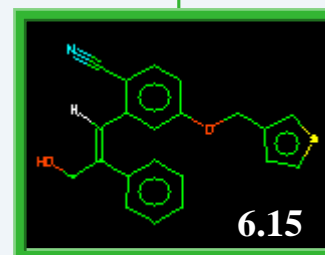
4

ET<sub>A</sub> antagonists, non-GSK set (source Aureus)

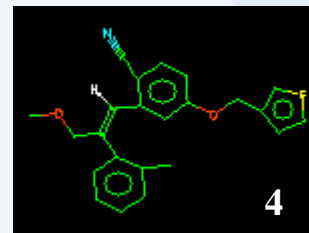
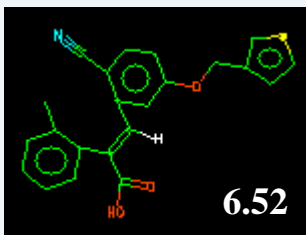
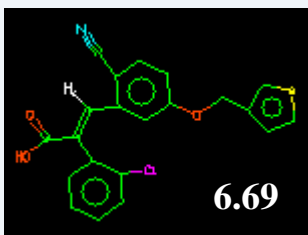
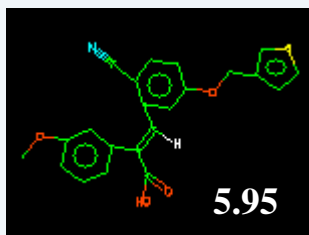
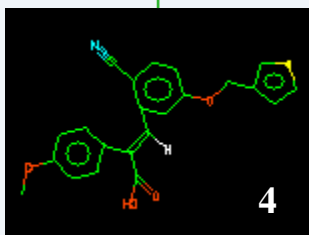
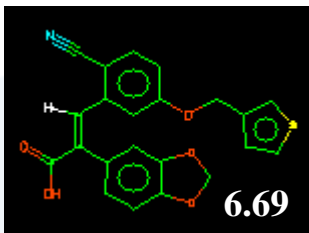
# Children



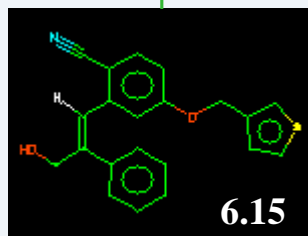
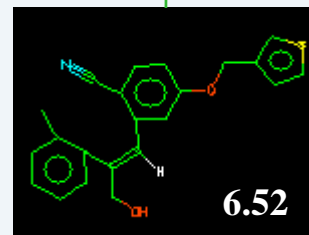
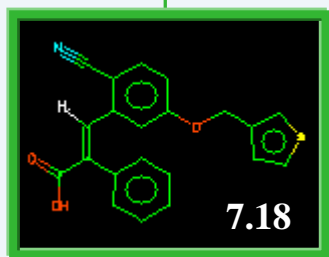
# Parent



# Children

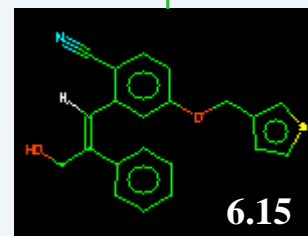
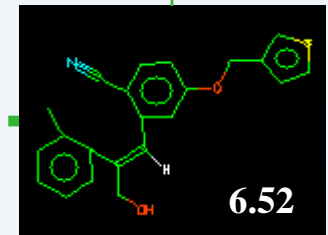
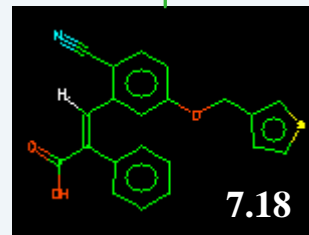
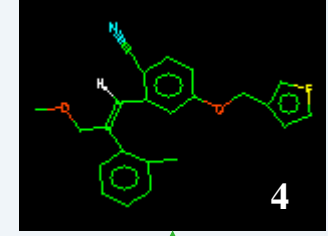
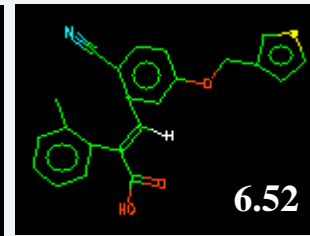
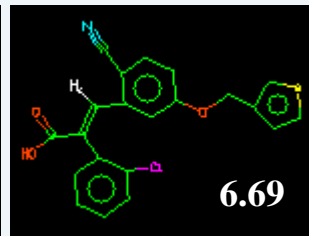
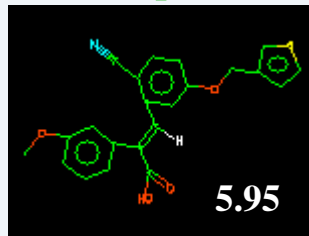
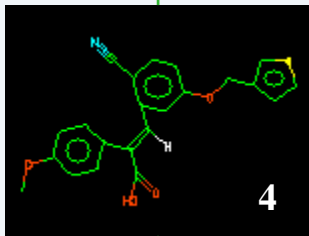
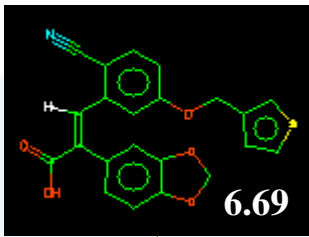


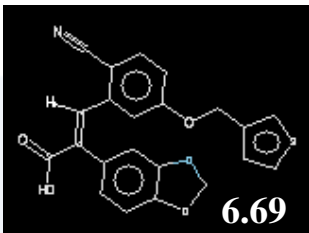
# Parent



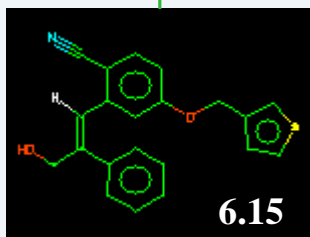
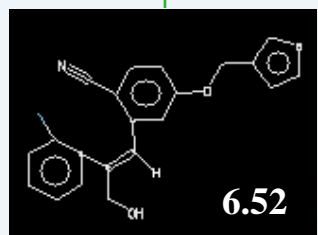
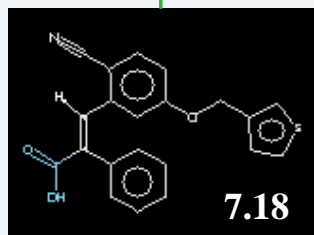
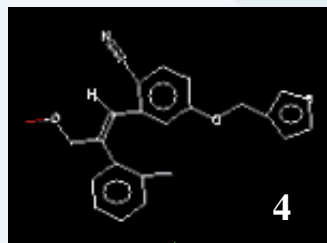
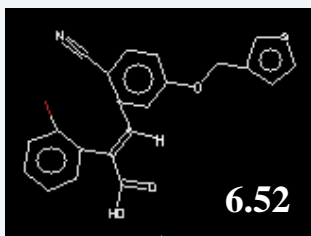
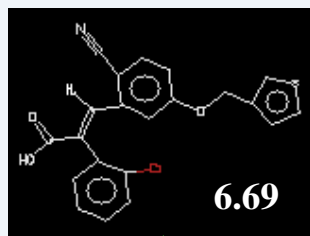
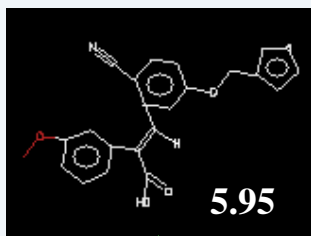
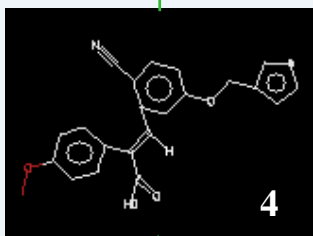
- Children may be Parents of a sub-hierarchy

- Children may have multiple Parents





- Highlight regions where Children differ from Parent
- Coloured according to activity difference

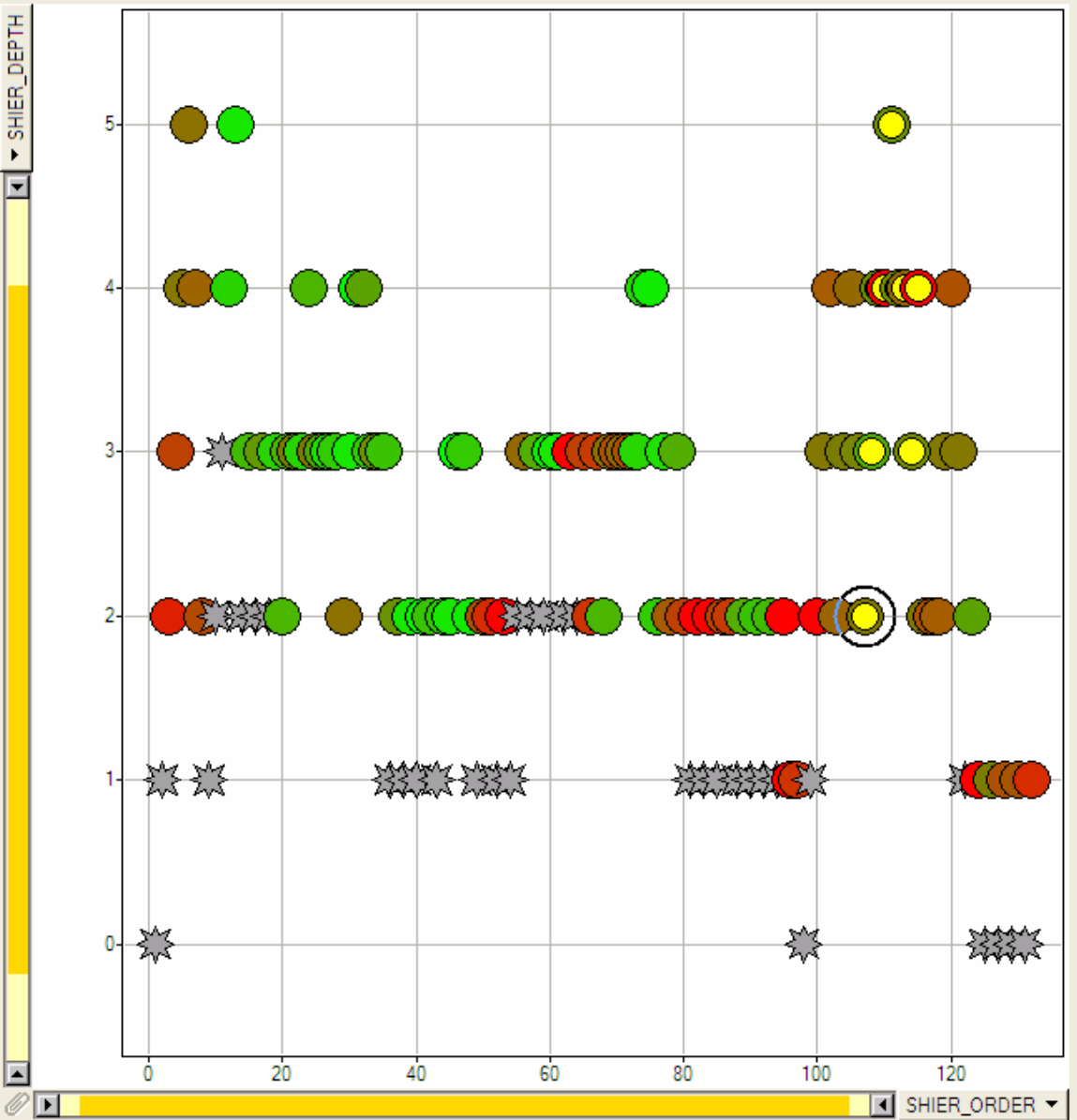
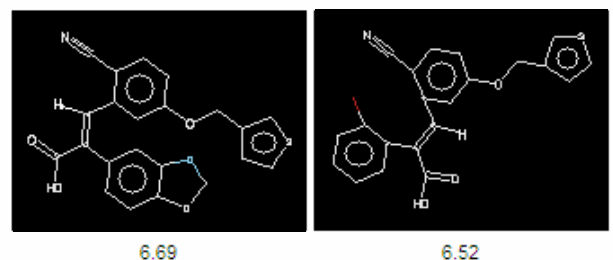
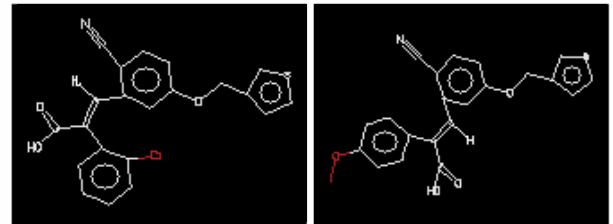
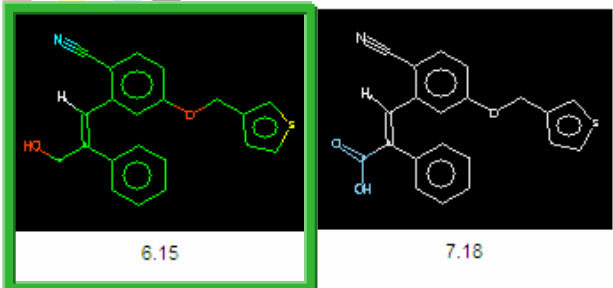




Guides

Smiles Visualiser

Mode: Cluster ? Cluster  
 Column: SHIER\_MEMBERS\_SMILES ?  
[Save Settings](#), [Load Settings](#) ?  
 Toggle: [Options](#), [Active](#), [Highlighted](#), [Selections](#), [Colouring](#) ?  
 Cluster Records: [Copy To Clipboard](#) [Export To Word](#) ?  
 Lower Higher ?



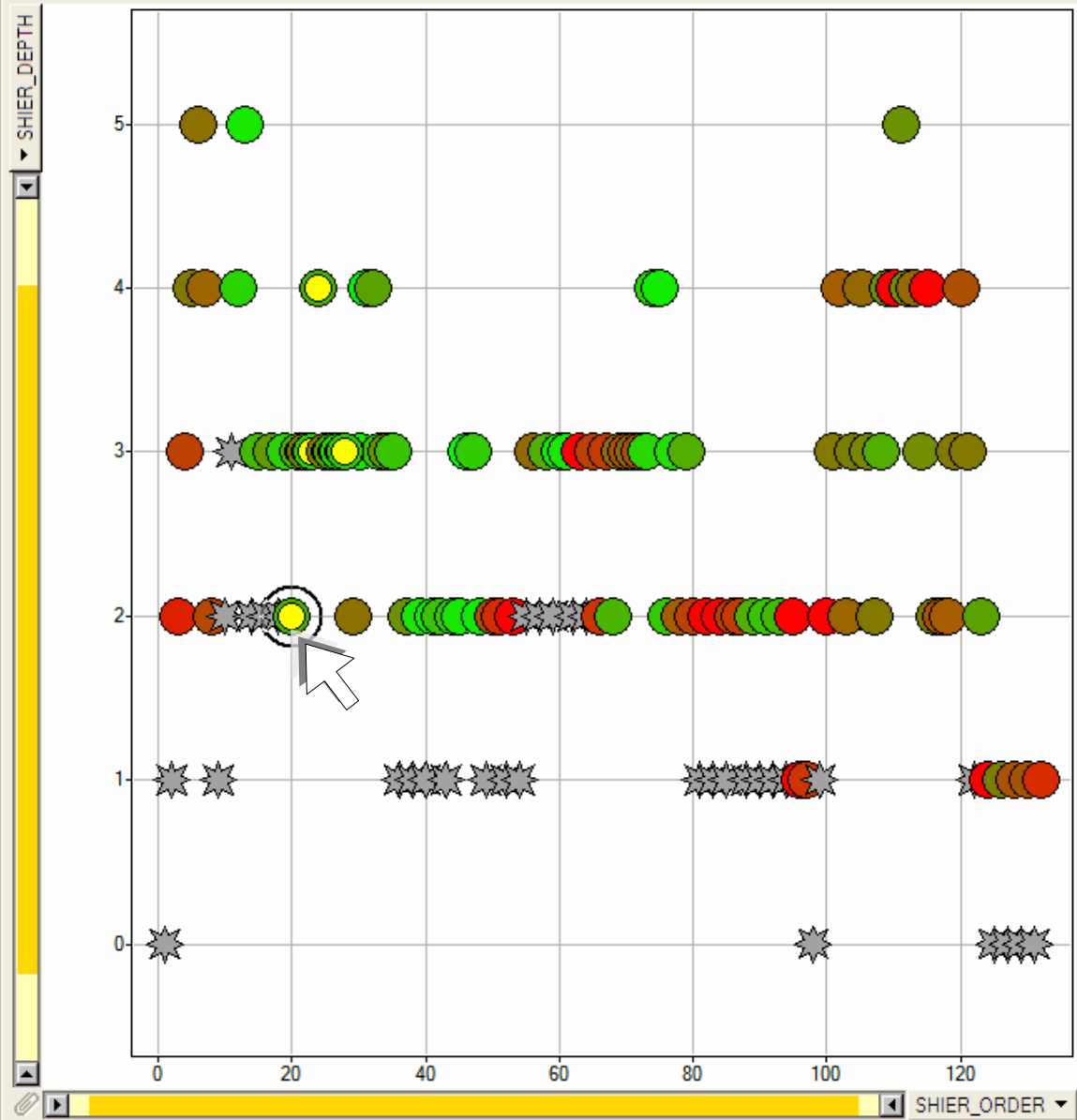
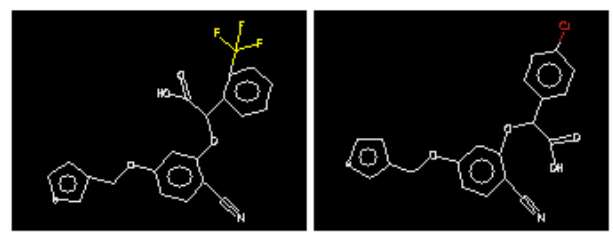
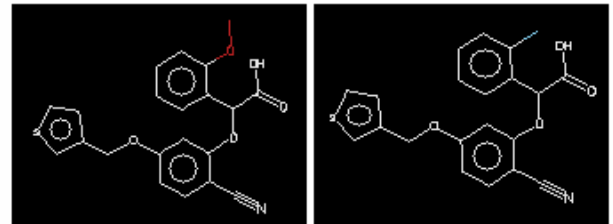
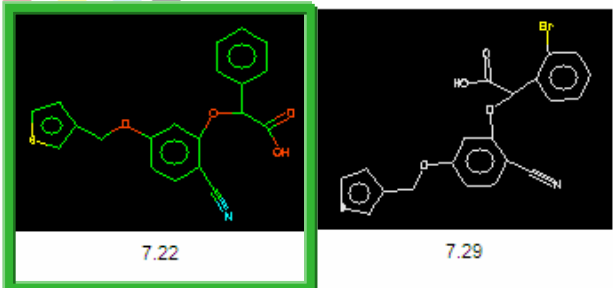
Scatter Plot Hierarchy



Guides

Smiles Visualiser

Mode: Cluster ? Cluster  
 Column: SHIER\_MEMBERS\_SMILES ?  
[Save Settings](#), [Load Settings](#) ?  
 Toggle: [Options](#), [Active](#), [Highlighted](#), [Selections](#), [Colouring](#) ?  
 Cluster Records: [Copy To Clipboard](#) [Export To Word](#) ?  
 Lower Higher ?

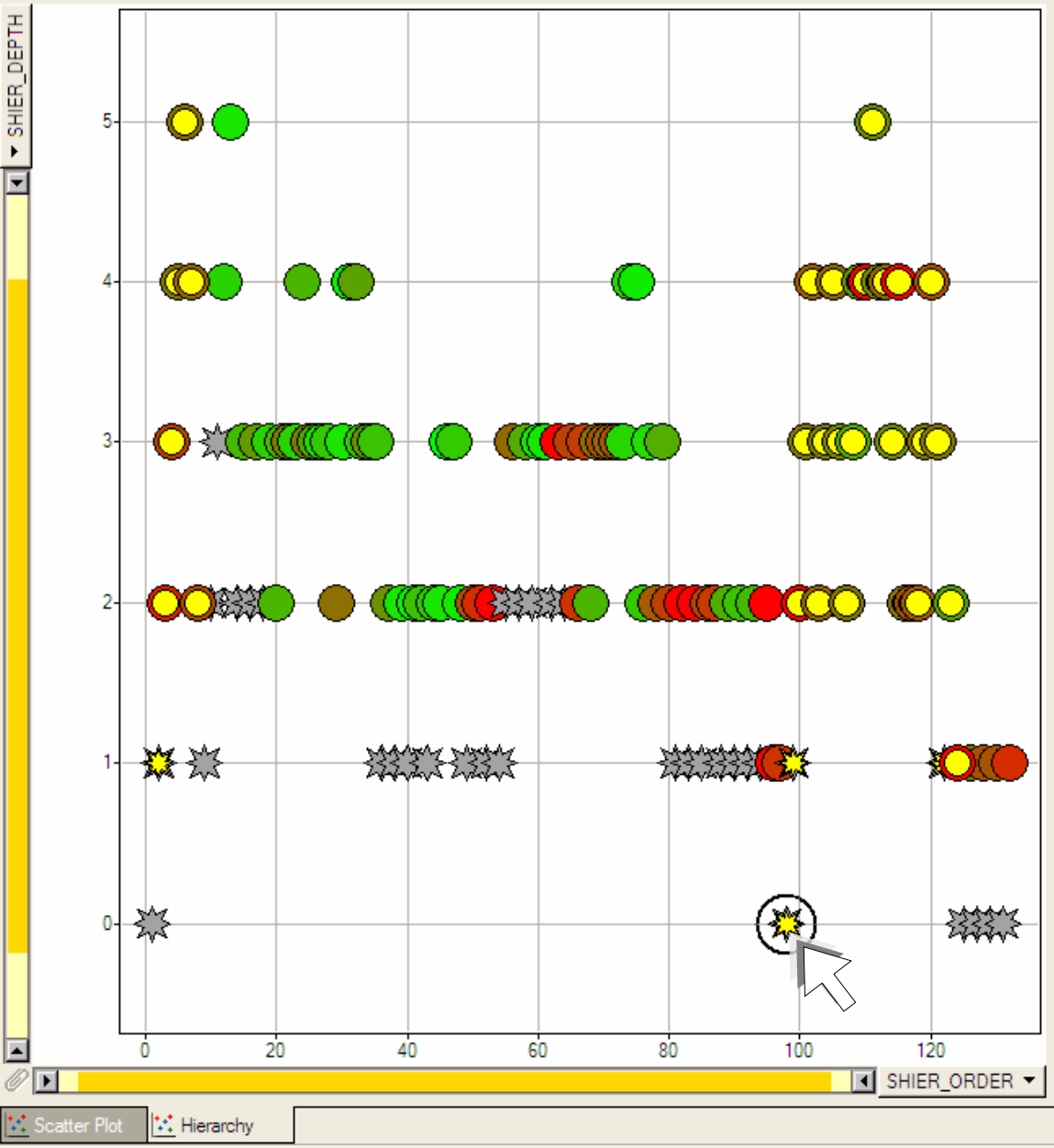
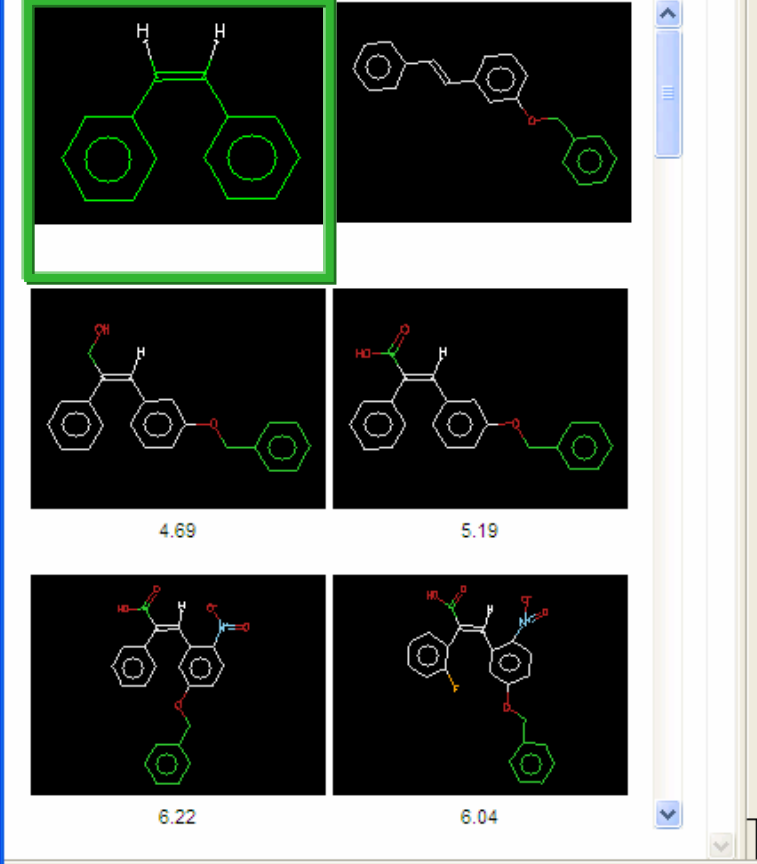


Scatter Plot Hierarchy

Guides

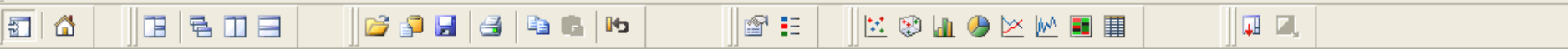
Smiles Visualiser

Mode: Cluster ? Cluster  
Column: SHIER\_MEMBERS\_SMILES ?  
[Save Settings](#), [Load Settings](#) ?  
Toggle: [Options](#), [Active](#), [Highlighted](#), [Selections](#), [Colouring](#) ?  
Cluster Records: [Copy To Clipboard](#) [Export To Word](#) ?  
Lower Higher ?



# Introduction to Nearest Neighbours

- Neighbouring molecules have 'similar' structures
  - They are generally expected to have similar activities
  - When they don't, the structural differences can be very informative
- Nearest Neighbours can be defined in various ways...
- Using similarity metric with molecular fingerprint
  - Various metric / fingerprint combinations possible
    - *E.g.* Daylight fingerprint & Tanimoto coefficient
    - Also 2D pharmacophores, Reduced Graphs
  - Neighbours are defined by a similarity threshold
    - Values *ca.* 0.8 commonly used with Daylight FP / Tanimoto combination
  - Neighbouring molecules can differ in various places
- Using 'single-point changes'
  - Neighbouring molecules can only differ in one position
  - Degree of difference is measured by change in number of atoms
  - Neighbours are defined by a threshold value for this difference



Guides

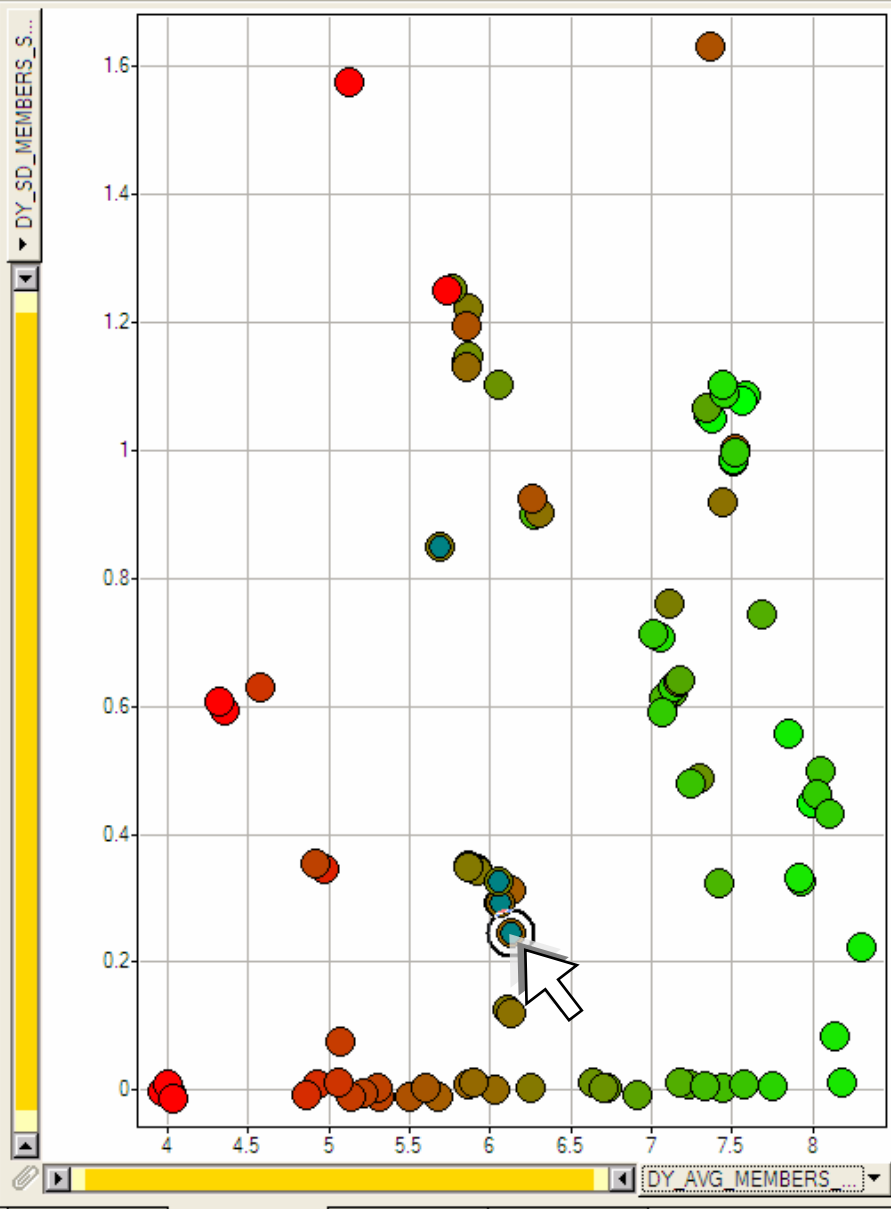
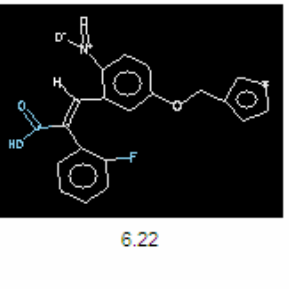
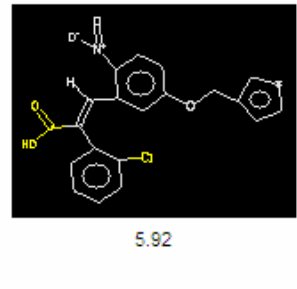
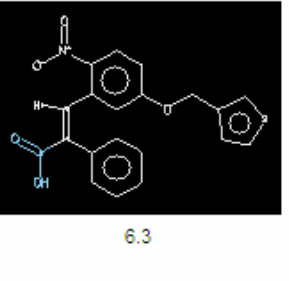
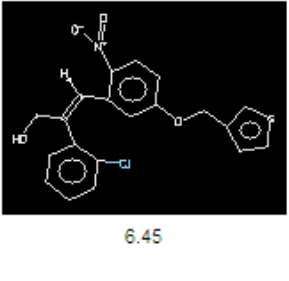
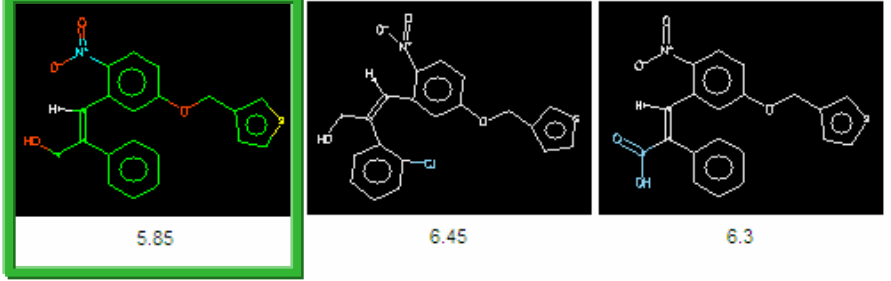
Smiles Visualiser

Mode: Cluster ? Cluster Column: DY\_MEMBERS\_SMILES\_>=0.85 ?

Min similarity 0.85 ? Save Settings, Load Settings ?

Toggle: Options, Active, Highlighted, Selections, Colouring ?

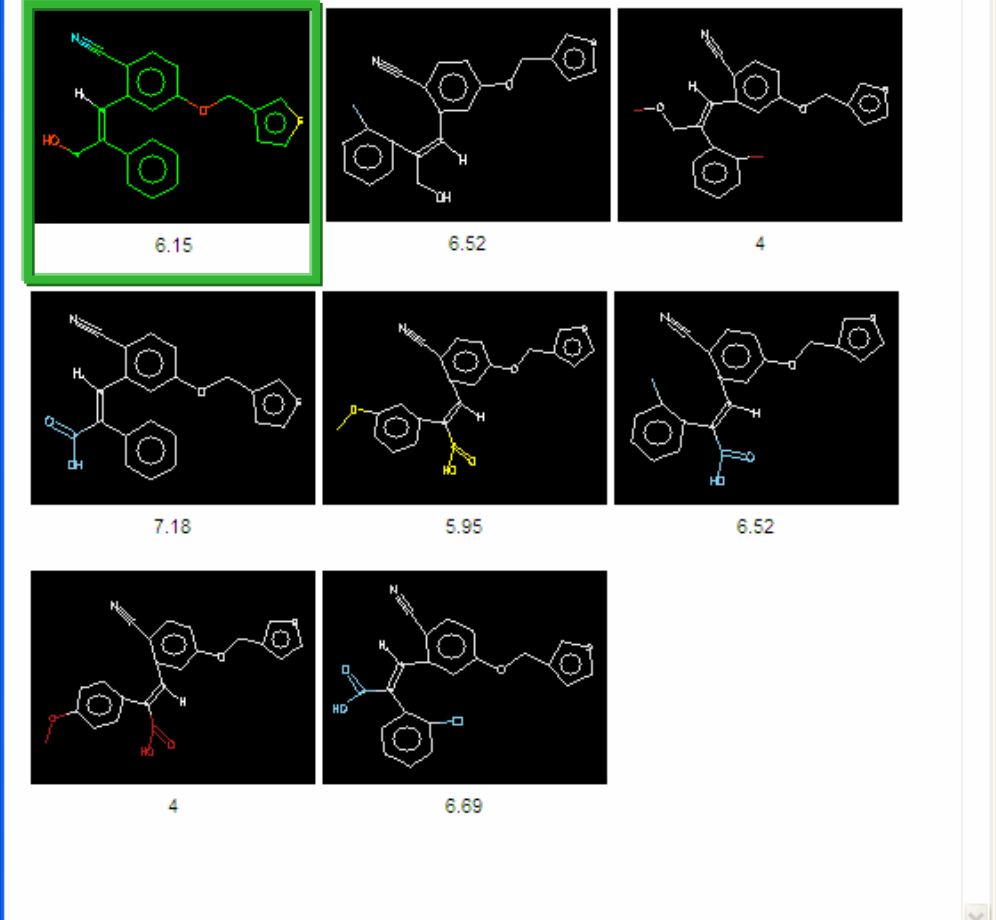
Cluster Records: Copy To Clipboard Export To Word ? Lower Higher ?



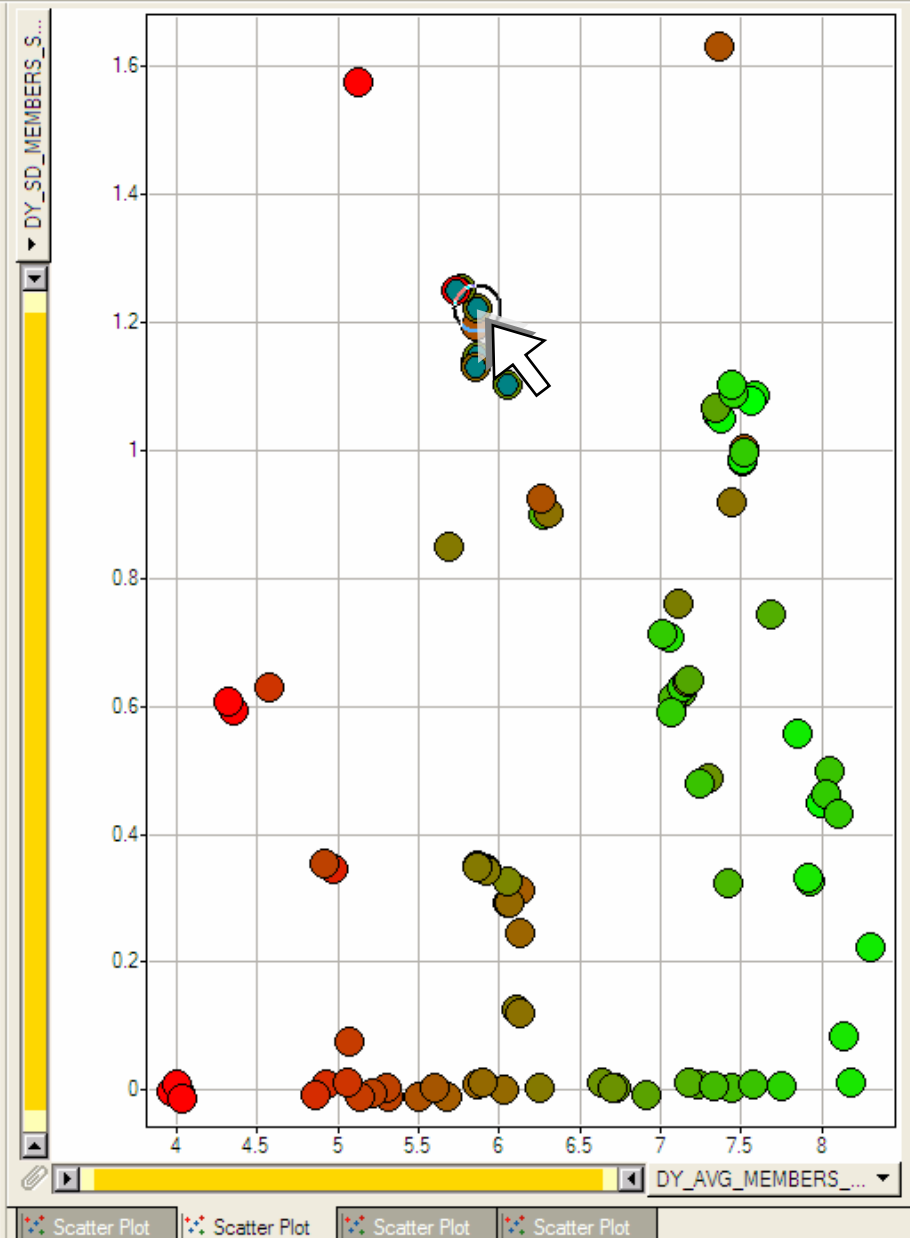
Guides

Smiles Visualiser

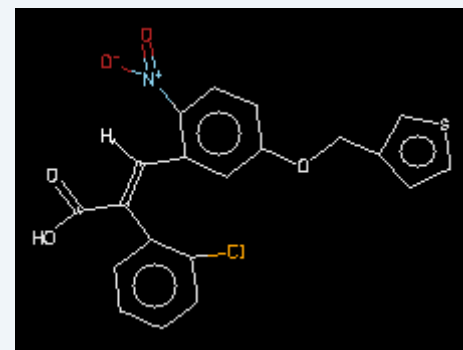
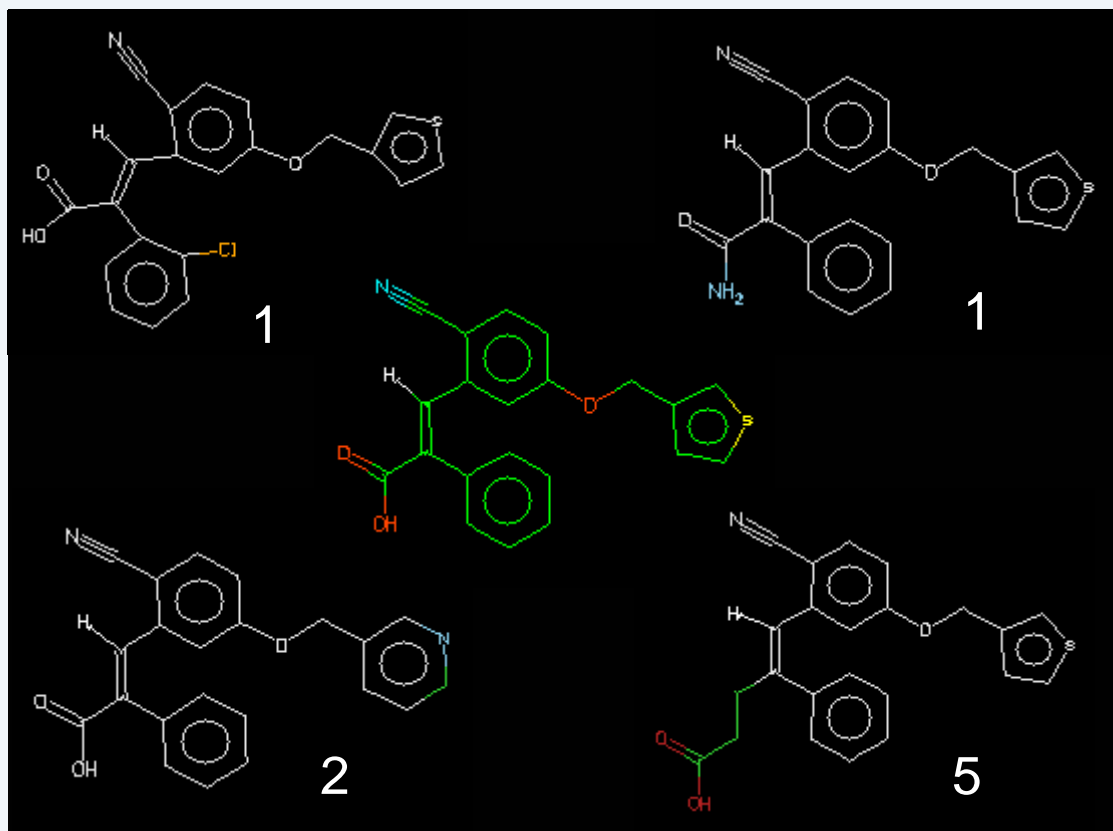
Mode: Cluster ? Cluster Column: DY\_MEMBERS\_SMILES\_>=0.85 ?  
Min similarity: 0.85 ? Save Settings, Load Settings ?  
Toggle: Options, Active, Highlighted, Selections, Colouring ?  
Cluster Records: Copy To Clipboard Export To Word ? Lower Higher ?



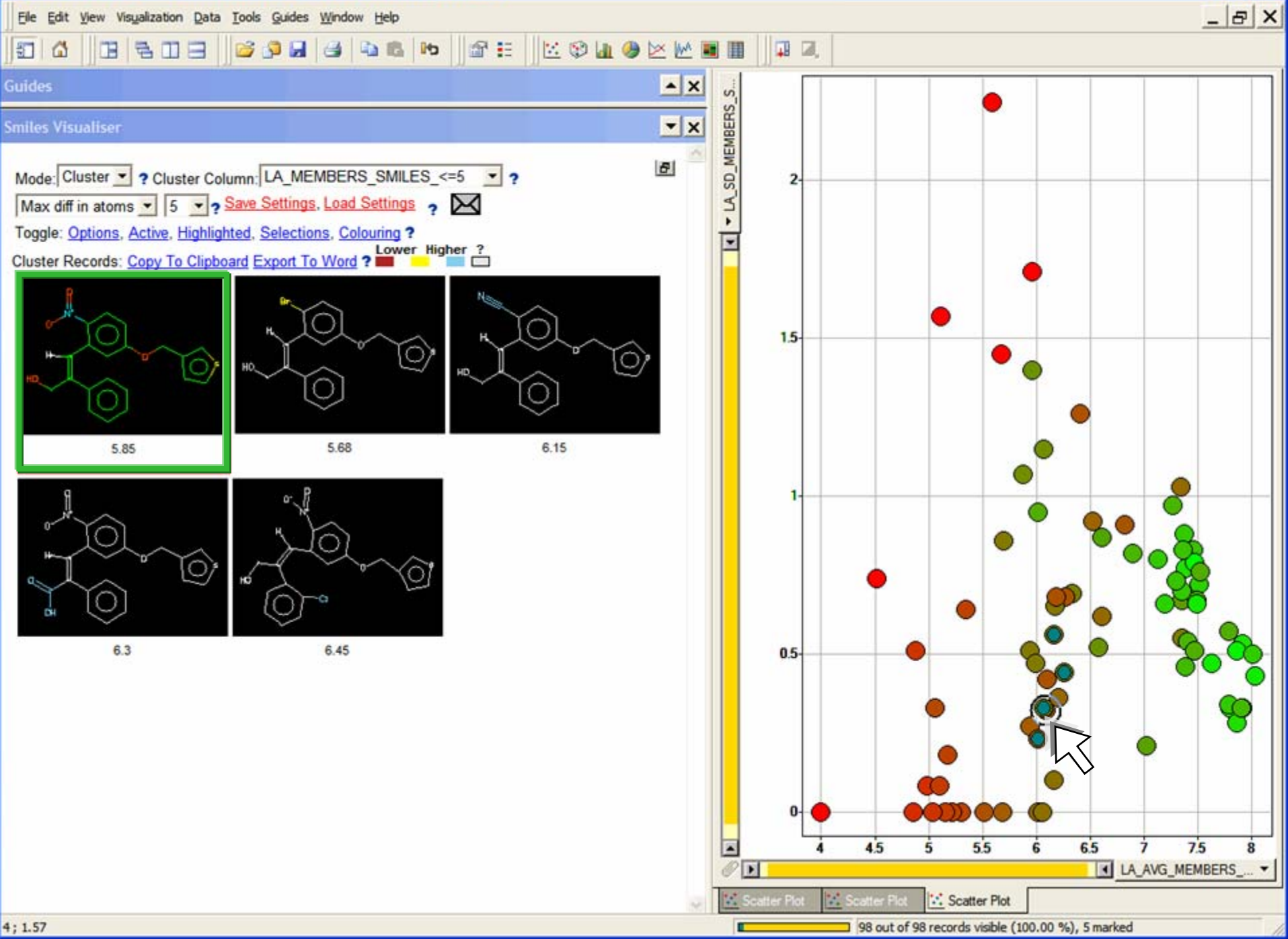
A grid of chemical structures (SMILES) displayed on a black background. The structures are arranged in three rows and three columns. The first structure in the top row is highlighted with a green border and labeled '6.15'. Other structures are labeled with numbers: 6.52, 4, 7.18, 5.95, 6.52, 4, and 6.69. The structures show various chemical groups including nitriles, hydroxyl groups, and aromatic rings.



# Single-point changes: 'Look-Alike'



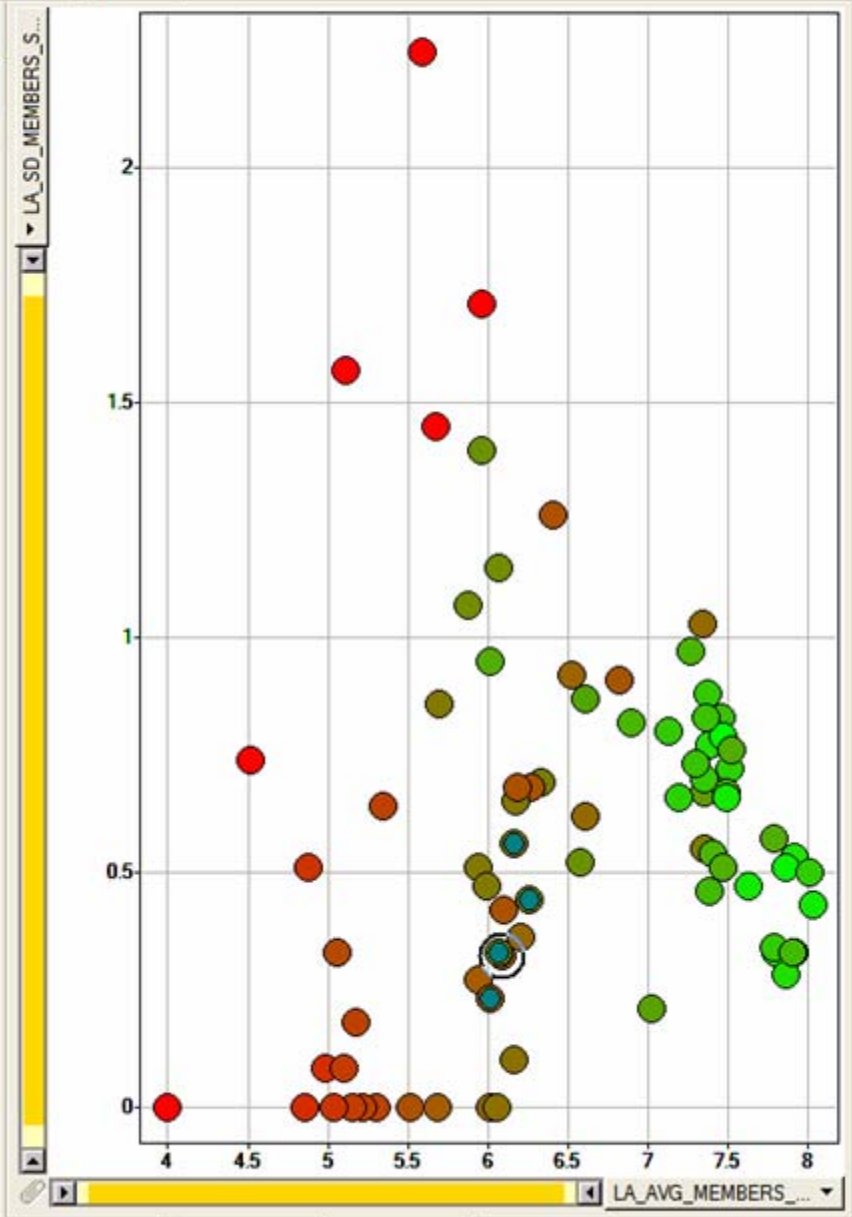
DIFFERENCE IS NOT LOCALISED THEREFORE MOLECULES ARE NOT PAIRED



Guides

Smiles Visualiser

Mode: Cluster ? Cluster Column: LA\_MEMBERS\_SMILES\_<=5 ?  
Max diff in atoms 5 ? Save Settings, Load Settings ?  
Toggle: Options, Active, Highlighted, Selections, Colouring ?  
Cluster Records: Copy To Clipboard Export To Word ? Lower Higher ?



Guides

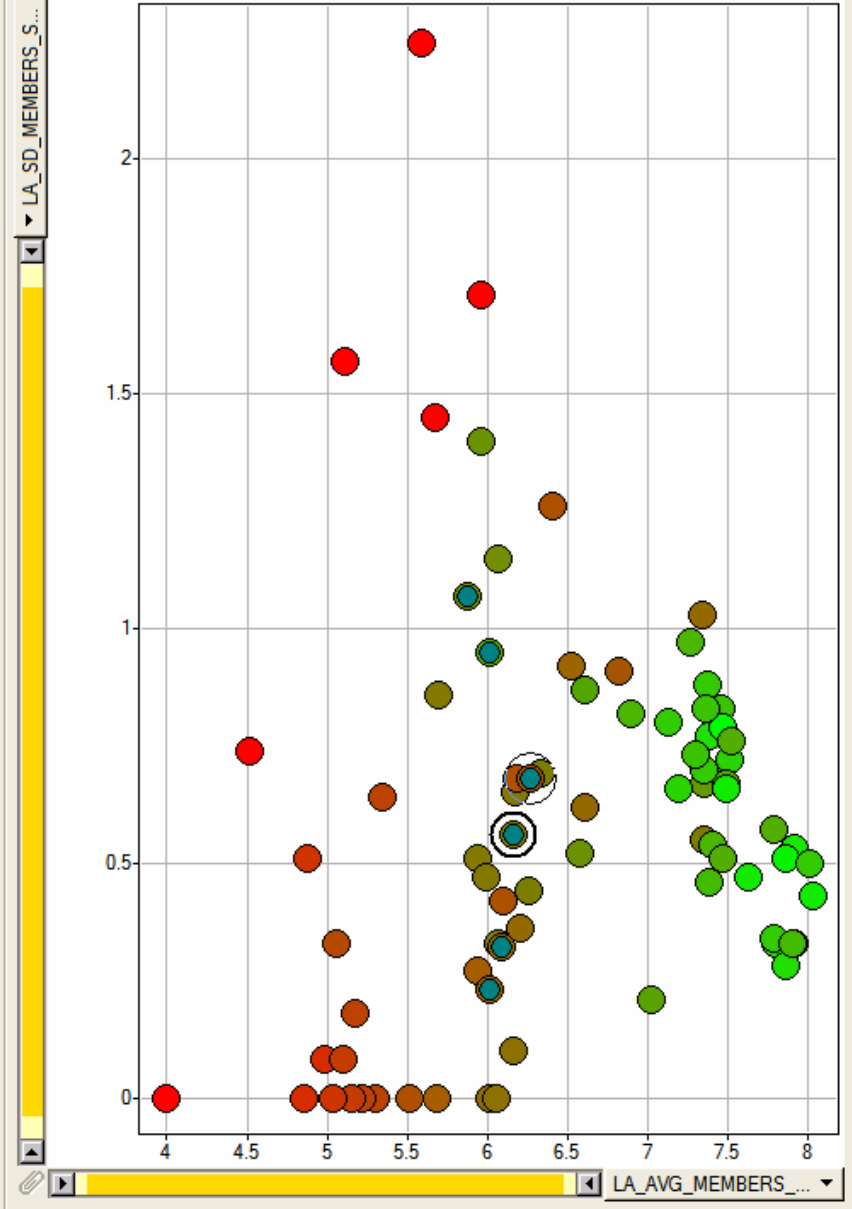
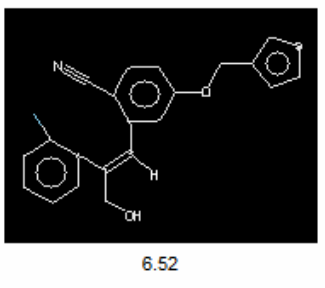
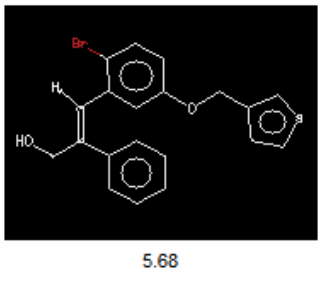
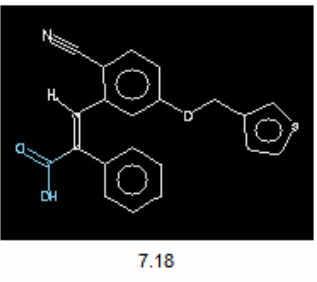
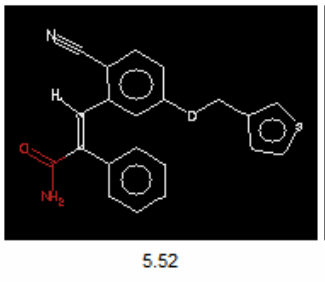
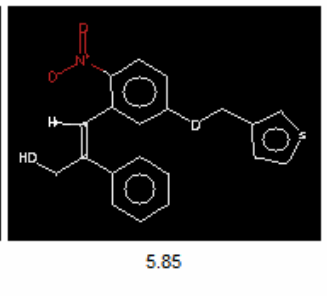
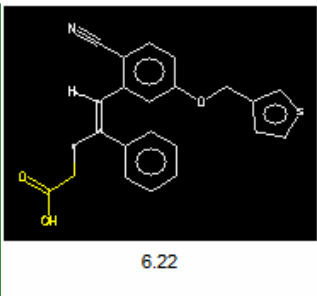
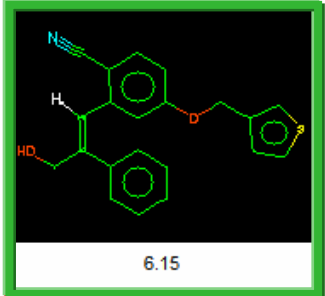
Smiles Visualiser

Mode: Cluster ? Cluster Column: LA\_MEMBERS\_SMILES\_<=5 ?

Max diff in atoms: 5 ? Save Settings, Load Settings ?

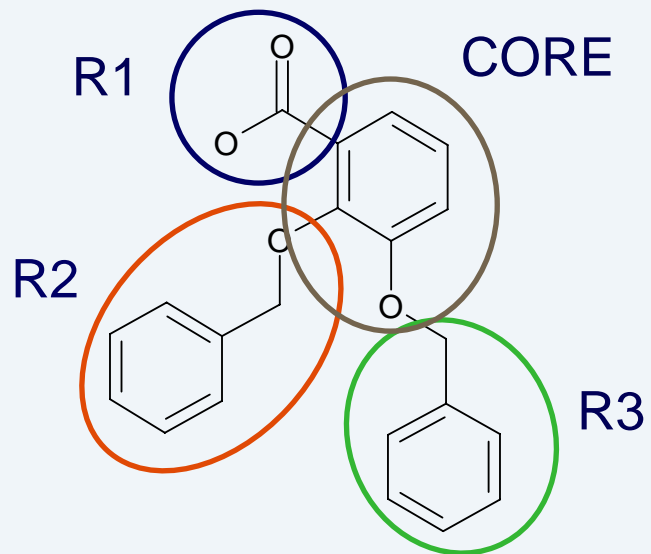
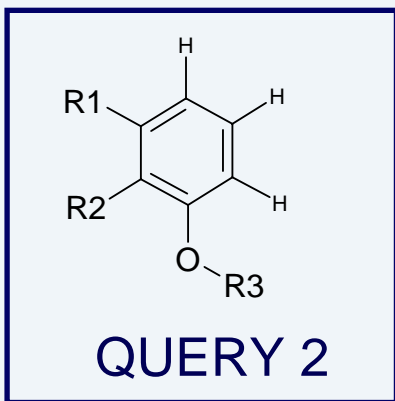
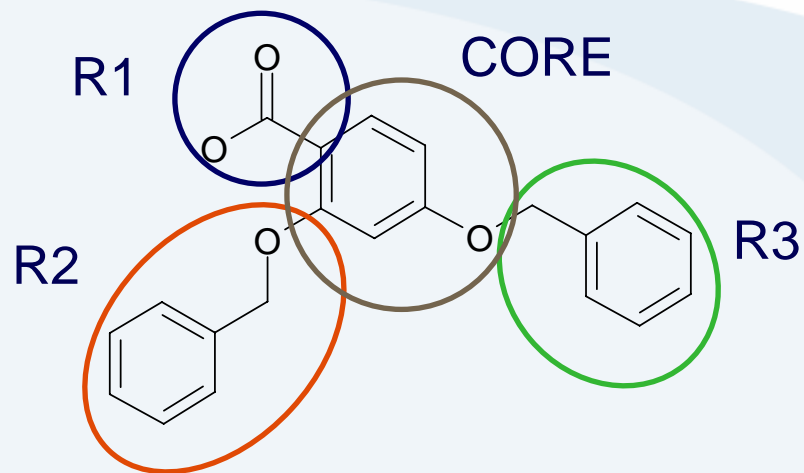
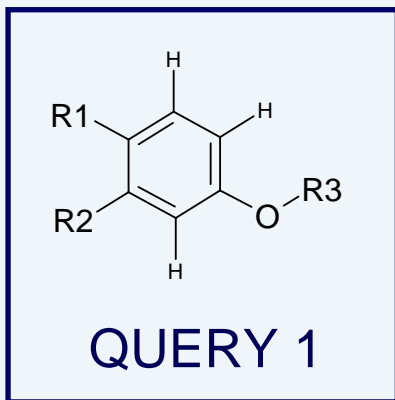
Toggle: Options, Active, Highlighted, Selections, Colouring ?

Cluster Records: Copy To Clipboard Export To Word ? Lower Higher ?

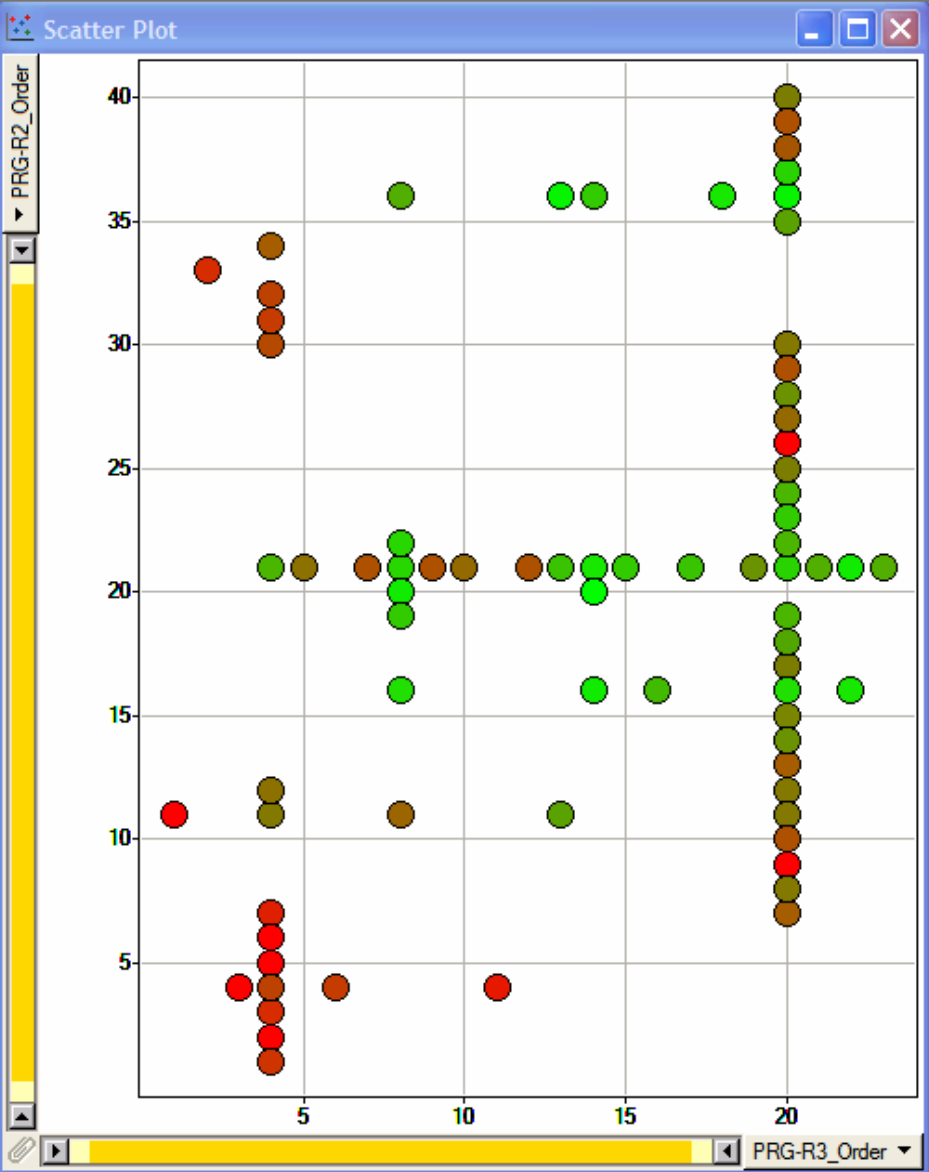
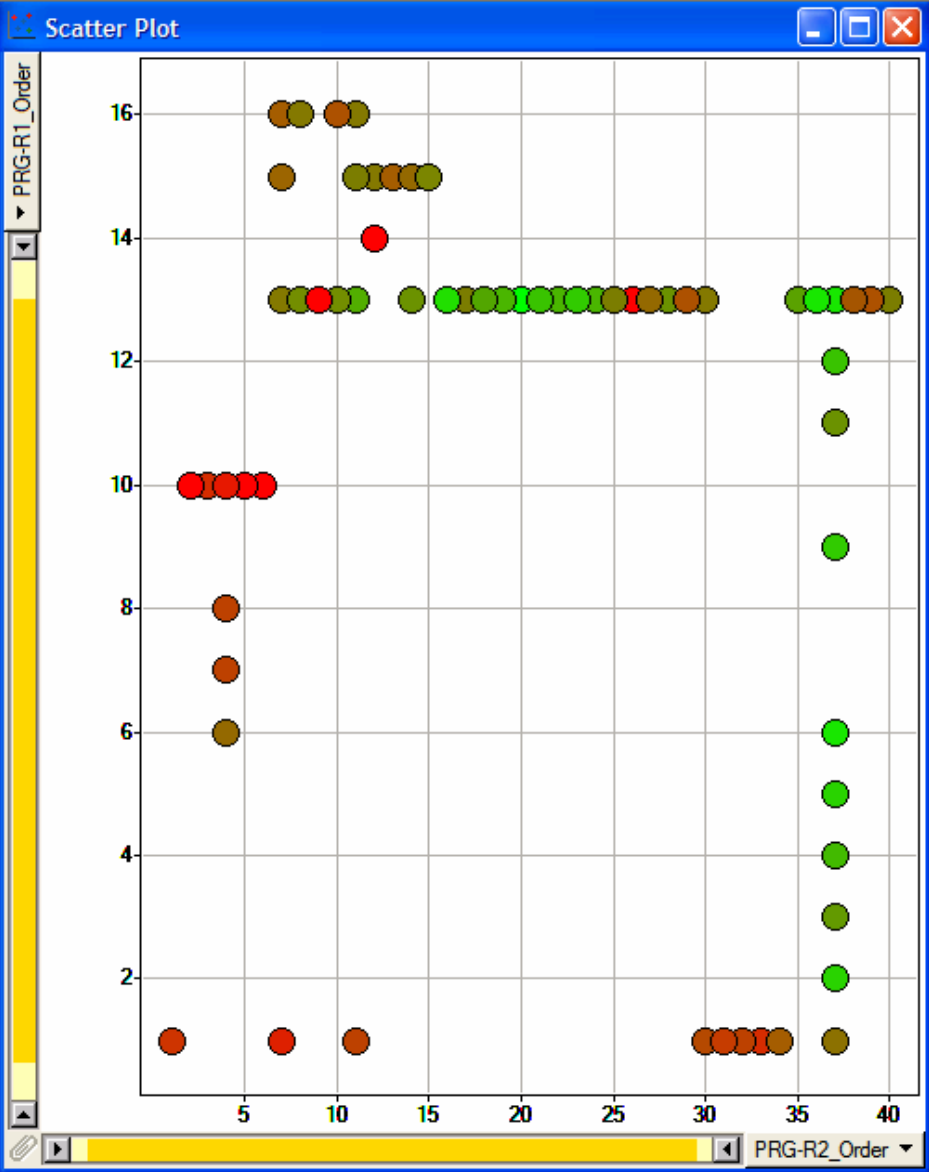


# R-group Analysis

- Widely used by medicinal chemists
  - Simple and intuitive
  - Efficient way of summarising data
  - Focus on points of diversity
- Developed fragmentation tool
  - Embedded in Spotfire
  - Based on Daylight toolkit
  - Highly interactive and versatile
    - Multiple queries can be specified in one run
    - Fragmentations can also be performed sequentially
    - R-groups and cores can be further fragmented
    - Simple rules to overcome symmetry problems
    - Ambiguous compounds are flagged
- R-groups can be analysed...
  - By inspection
    - Various plots available in Spotfire
  - Semi-quantitatively
    - D-scores
    - Z-scores
    - Free-Wilson

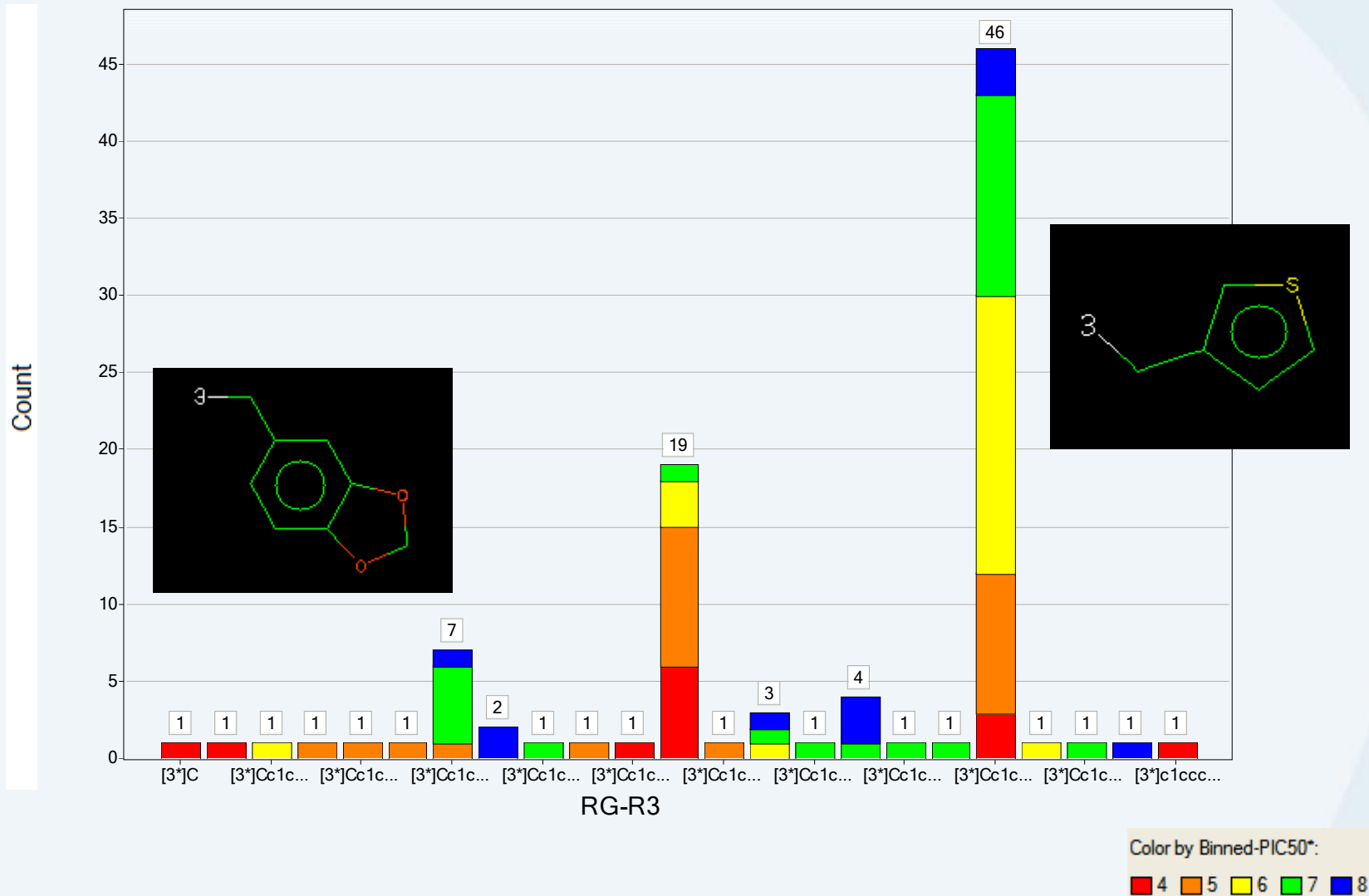






# Introduction to D-scores

Bar Chart



Guides

Smiles Visualiser

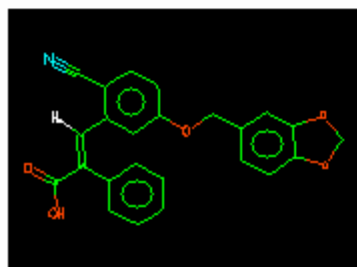
Mode: **Normal** ? SMILES: **SMILES** ?

Select any column?  ? [Save Settings](#), [Load Settings](#) ?

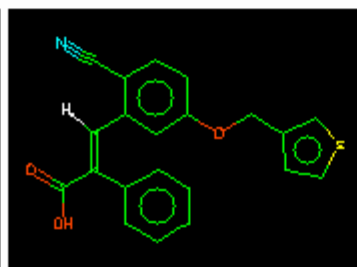


Toggle: [Options](#), [Active](#), [Highlighted](#), [Selections](#), [Colouring](#) ?

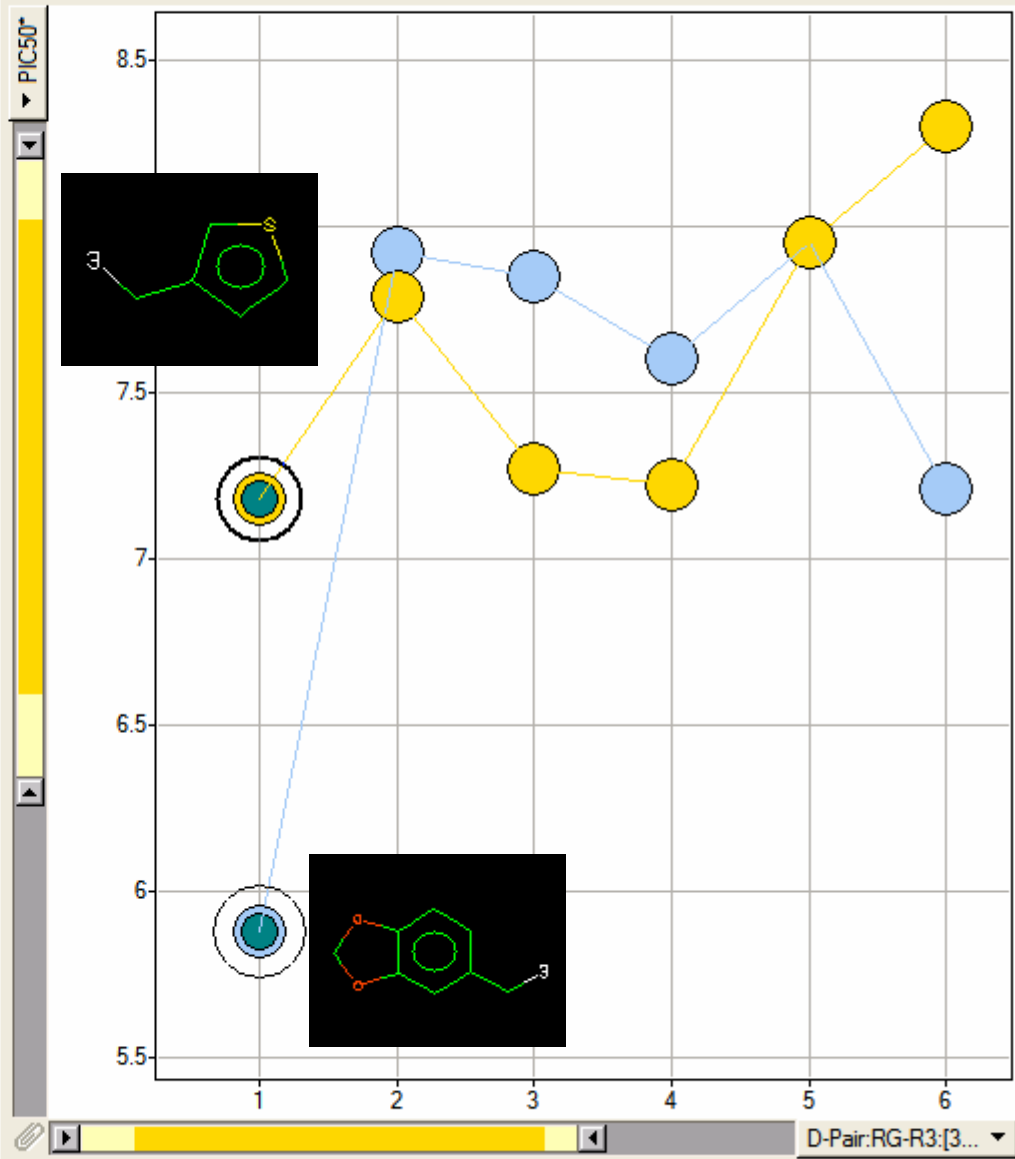
Marked records: [Copy To Clipboard](#) [Export To Word](#) ?



5.88



7.18



Guides

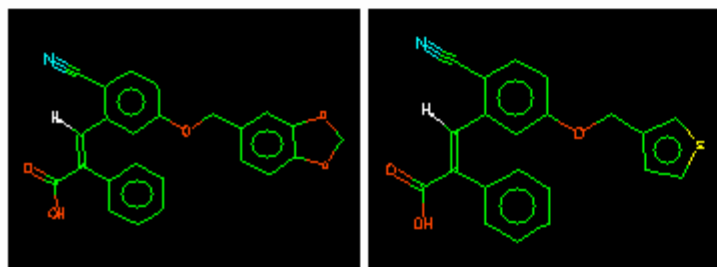
Smiles Visualiser

Mode: Normal ? SMILES: SMILES ?

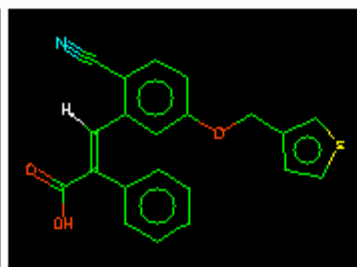
Select any column?  ? [Save Settings](#), [Load Settings](#) ?

Toggle: [Options](#), [Active](#), [Highlighted](#), [Selections](#), [Colouring](#) ?

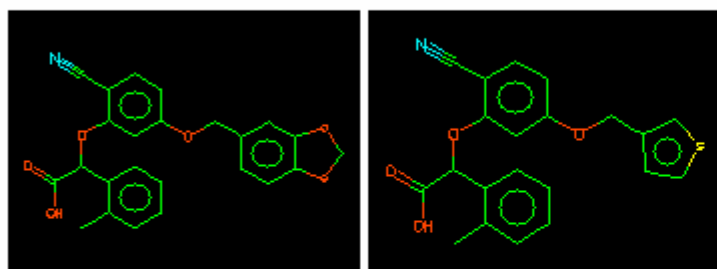
Marked records: [Copy To Clipboard](#) [Export To Word](#) ?



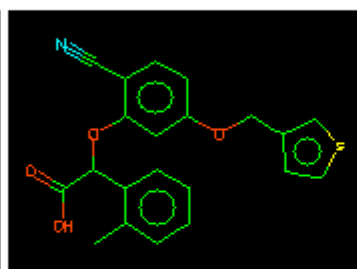
5.88



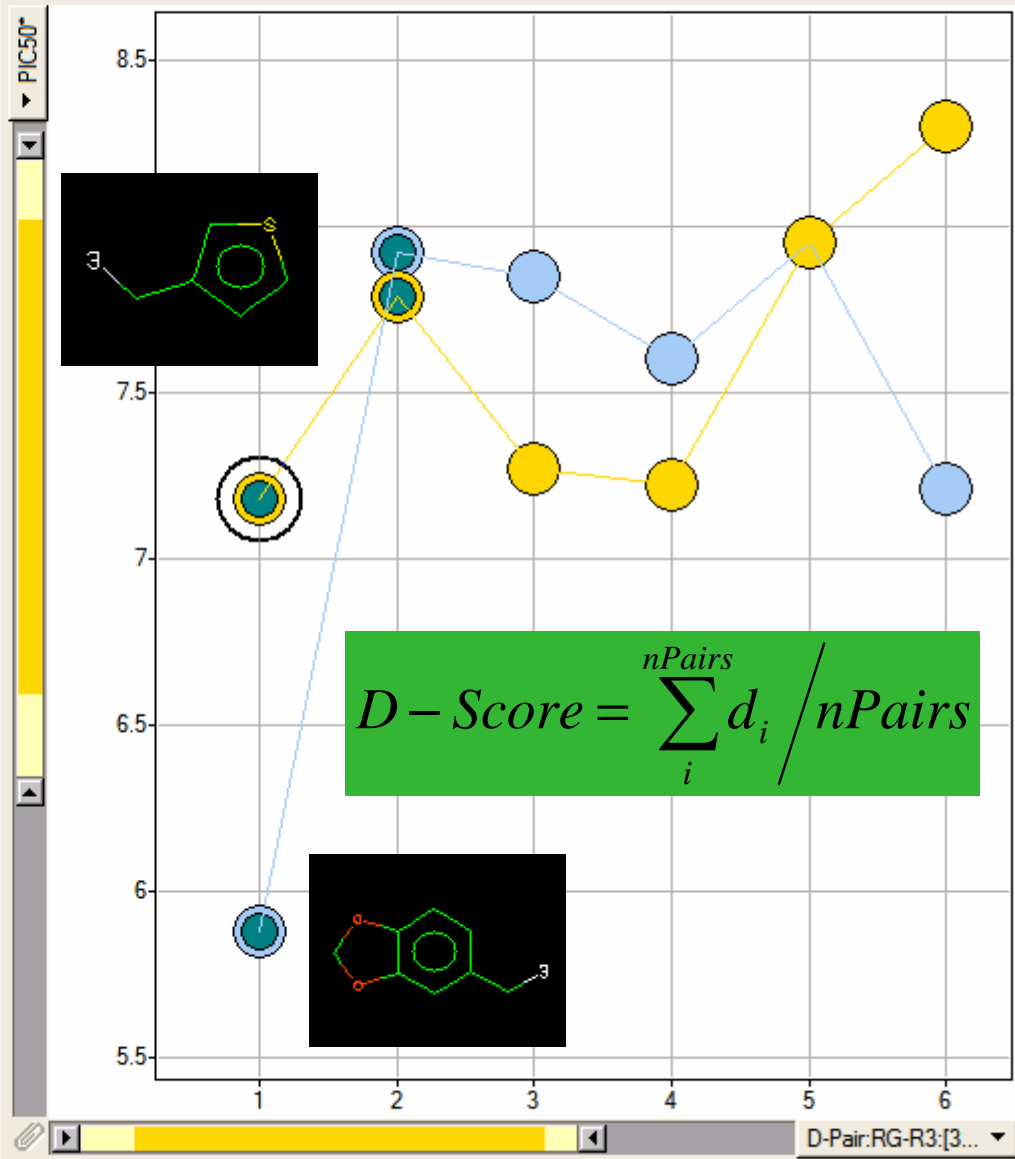
7.18



7.92



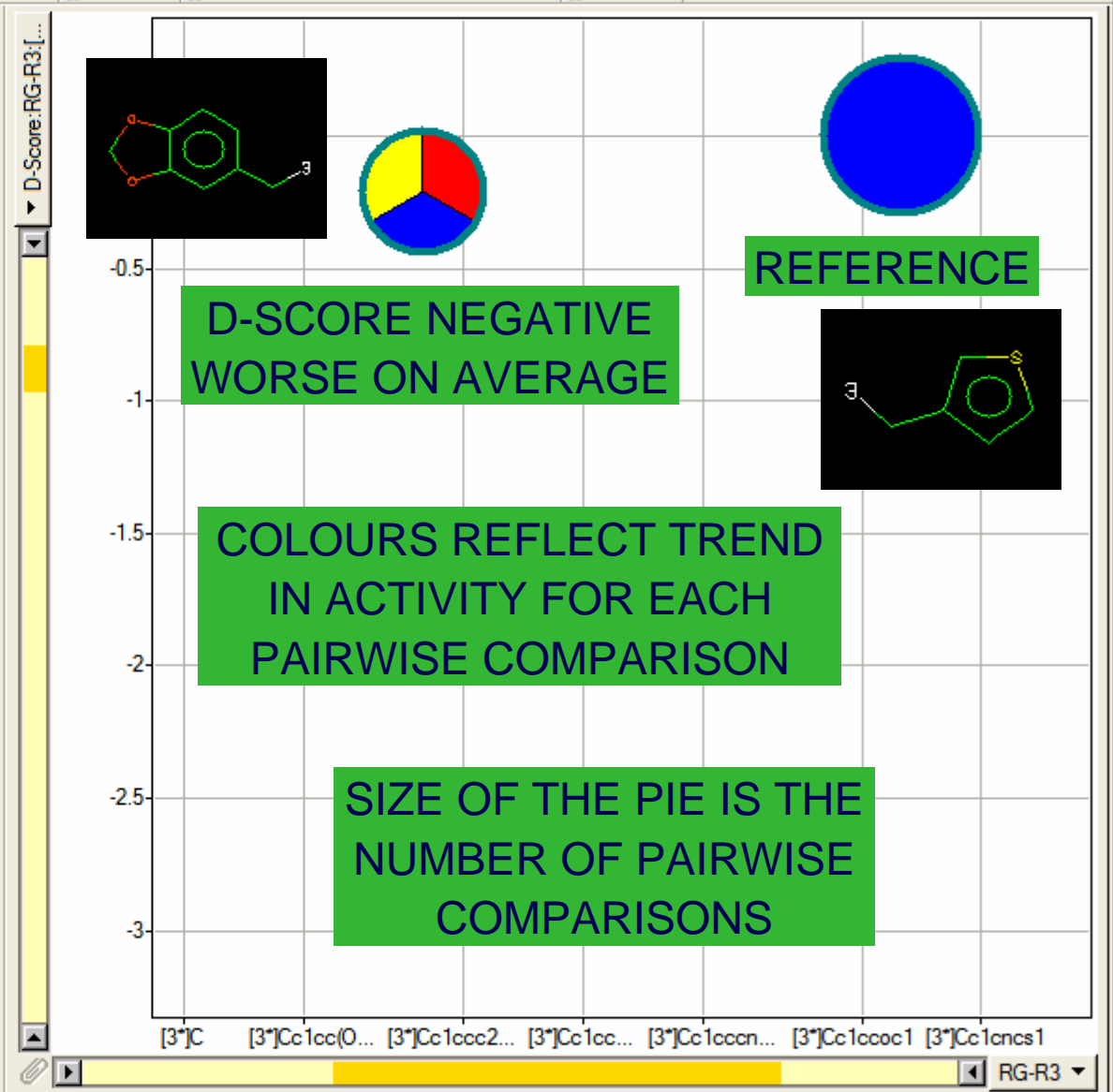
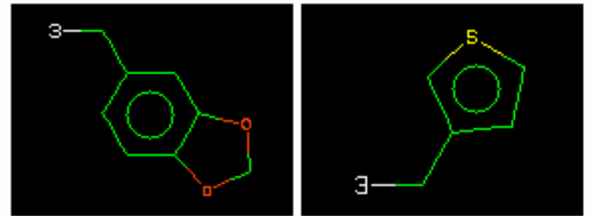
7.79



Guides

Smiles Visualiser

Mode: Normal ? SMILES: RG-R3 ?  
Select any column?  ? [Save Settings](#), [Load Settings](#) ?  
Toggle: [Options](#), [Active](#), [Highlighted](#), [Selections](#), [Colouring](#) ?  
Marked records: [Copy To Clipboard](#) [Export To Word](#) ?



Guides

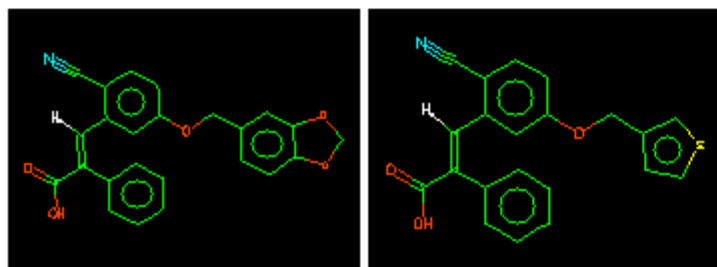
Smiles Visualiser

Mode: Normal ? SMILES: SMILES ?

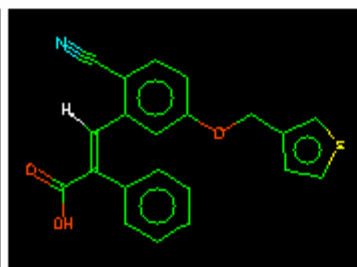
Select any column?  ? [Save Settings](#), [Load Settings](#) ?

Toggle: [Options](#), [Active](#), [Highlighted](#), [Selections](#), [Colouring](#) ?

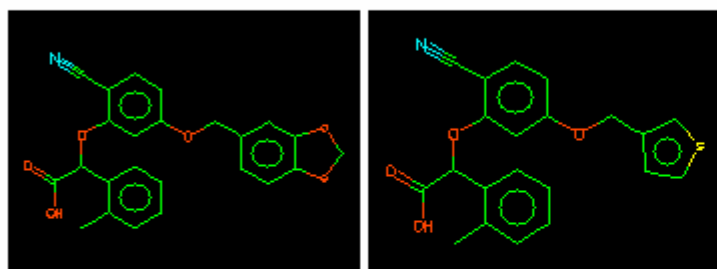
Marked records: [Copy To Clipboard](#) [Export To Word](#) ?



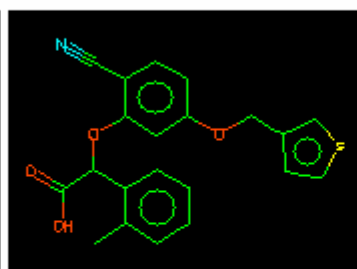
5.88



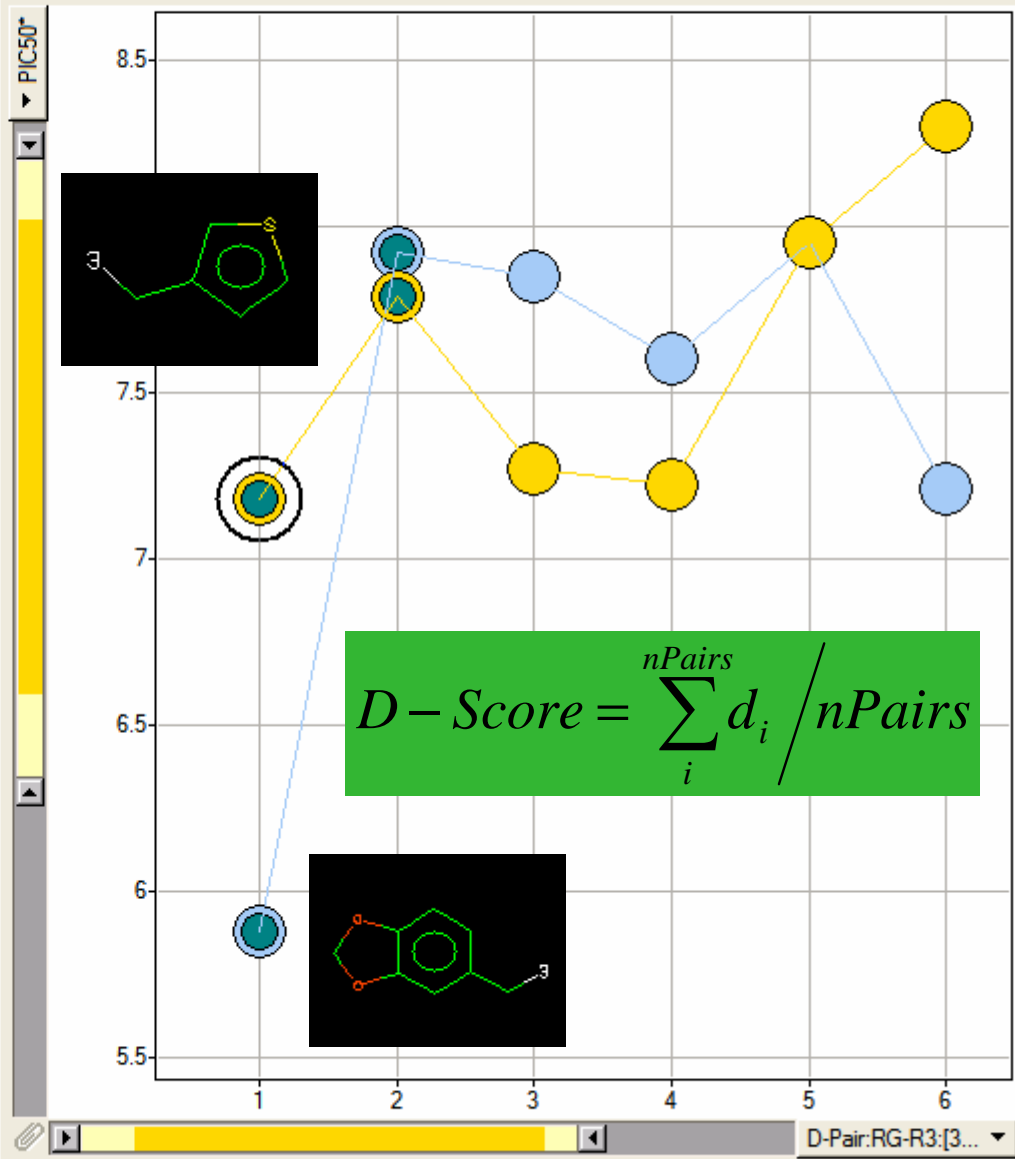
7.18



7.92



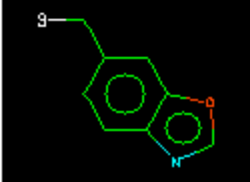

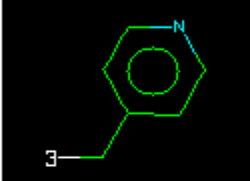
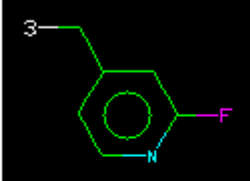

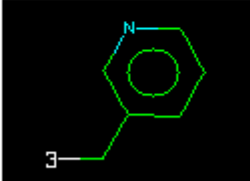
7.79

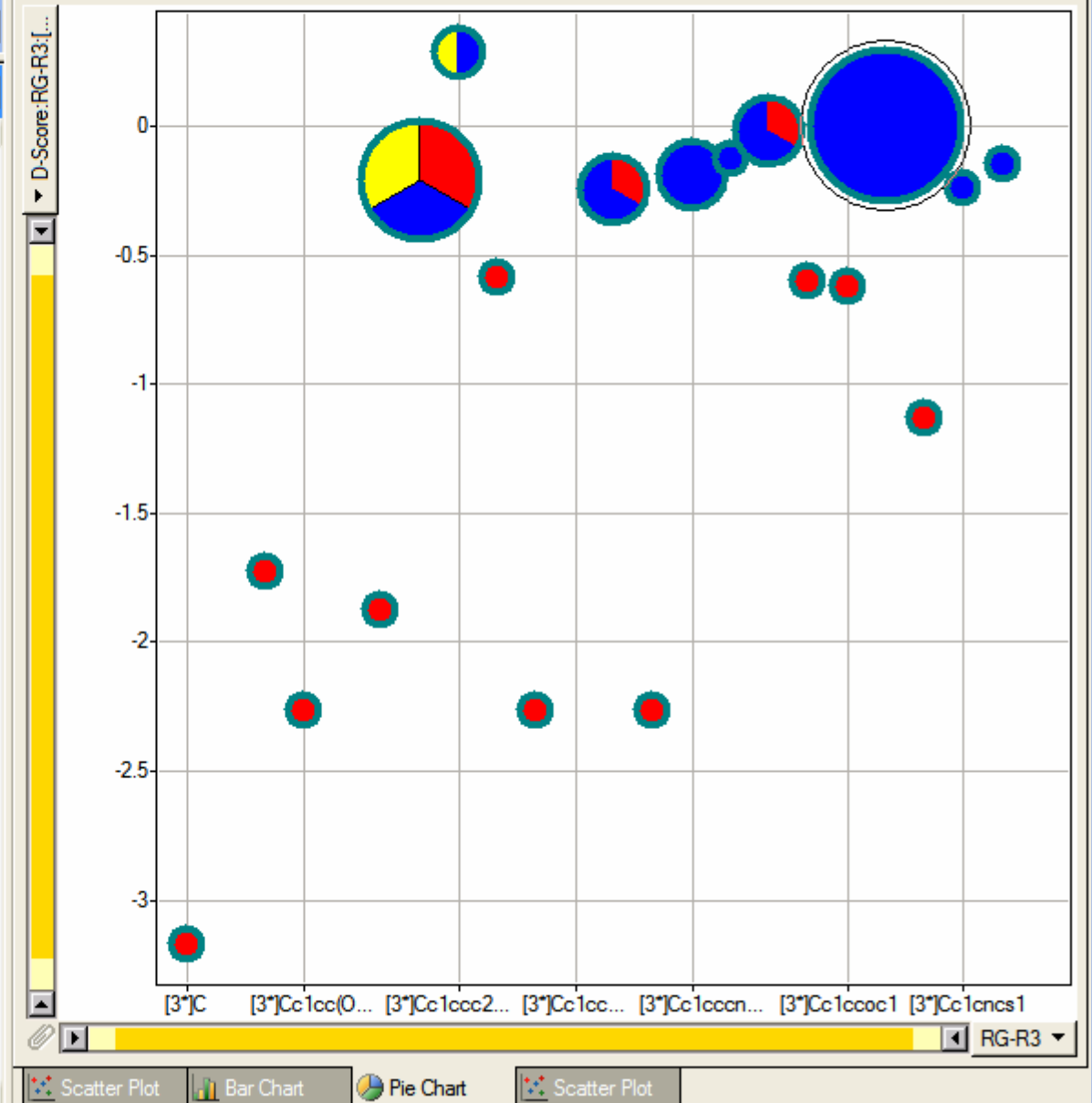


Guides

Smiles Visualiser

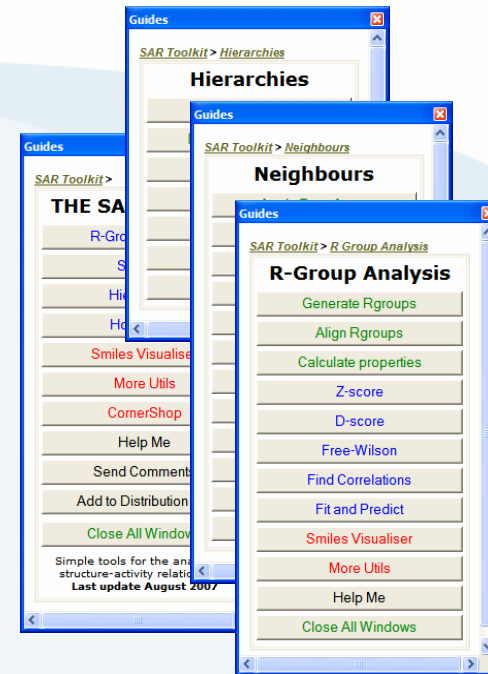
Mode: Normal ? SMILES: RG-R3 ?  
Select any column?  ? [Save Settings](#), [Load Settings](#) ?  
Toggle: [Options](#), [Active](#), [Highlighted](#), [Selections](#), [Colouring](#) ?  
Marked records: [Copy To Clipboard](#) [Export To Word](#) ?

 0.28	 0
 -0.02	 -0.13
 -0.15	 -0.19



# The SAR Toolkit - Summary

- A variety of tools
  - In house and from public domain
  - Complementary to each other
- Integrated with Spotfire
  - Accessible to everyone
  - Highly interactive and simple to use
- Help SAR discussions
  - Easy navigation and visualisation
  - Effective plots for SAR inspection
  - Neighbours, pairwise comparisons can be easily identified and saved
- Currently used by medicinal and computational chemists



# Acknowledgements

- Gianpaolo Bravi
- Francis Atkinson
- Daniel Lowe
- Shane Weaver
- Nicola Richmond
- Gavin Harper
- Martin Saunders
- Colin Edge
- Iain Mclay
- Ian Wall
- Stefan Senger
- Tony Cooper
- Pam Brown
- Tim Ritchie
- Graham Simpson
- Martin Swarbrick