

# Visualising Structured Compound Data in an Unstructured Way

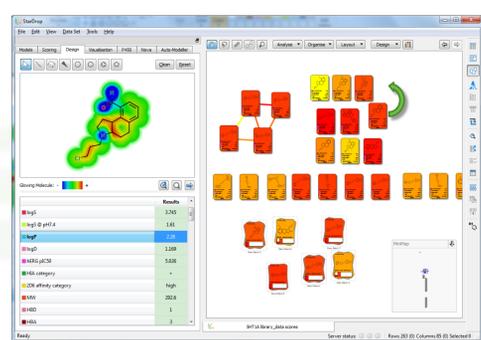
James Chisholm, Chris Leeding, Ed Champness, Hector Garcia Martinez, Alex Elliott, Matthew Segall  
Optibrium Ltd., Cambridge, UK

## Introduction

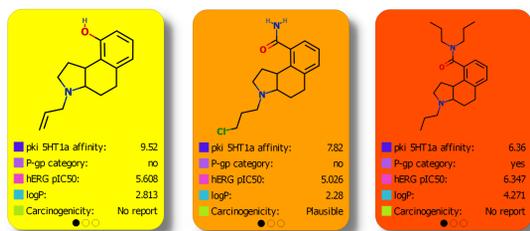
When we view compound data sets, we almost always find ourselves looking at a 'chemical spreadsheet' or a 'form view' of compounds and data in a long list. However, compounds have much more complex relationships that can't be captured in a single, sequential order. We present an alternative view, implemented in the StarDrop™ software [1], in which compound structure and associated data are presented on cards that can be moved and organised freely. This allows chemists to explore their data in a flexible and interactive way and to organise their data based on their perspective and thought processes. Algorithms, such as clustering or matched pair algorithms, can also be applied to organise compounds to draw out interesting patterns or features.

## Cards

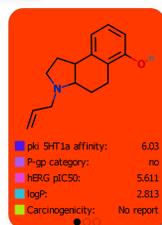
Each card represents a single compound in a data set. Properties can be selected and placed beside a chemical diagram to produce something similar to a 'baseball card' with key compound data. Cards can be easily moved, for example to compare compounds of interest side-by-side. Such simultaneous visualisation of aligned structures together with



properties presents a picture from which a chemist can consider the relationships between compounds and the effect structural changes may have on properties. Such a view may stimulate ideas, facilitate discussions, say in a group setting, and trigger 'why questions' more readily than a spreadsheet or a form view. The view can be panned and scaled to investigate relationships between larger groups of compounds.



Move cards to compare structures and properties

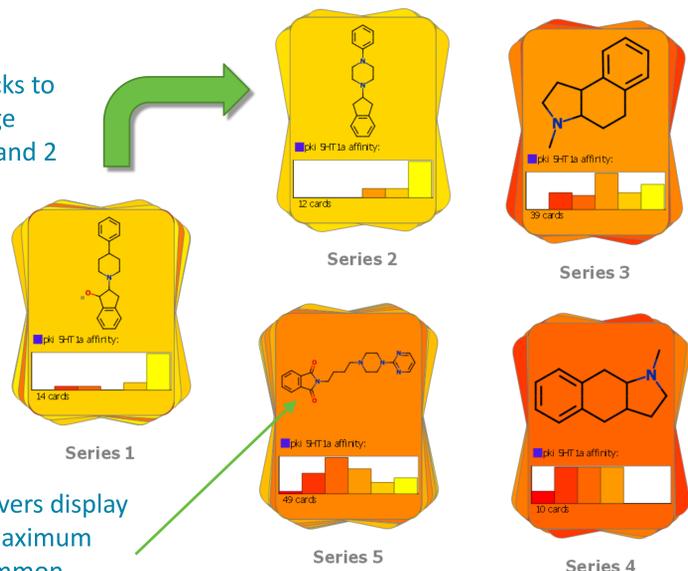


Compare 'baseball cards' with a side-by-side arrangement

## Card Stacks

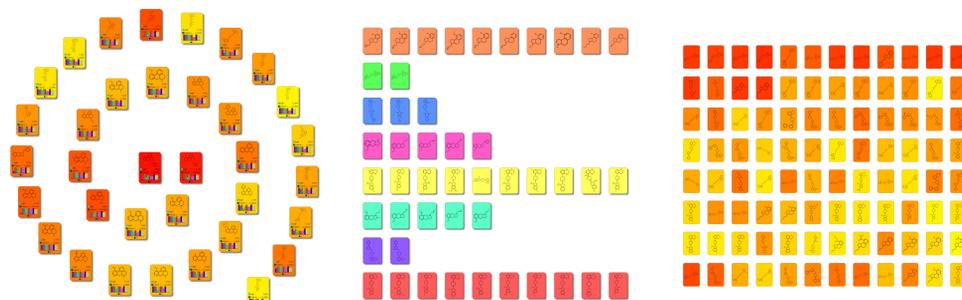
A common task is to assign compounds to particular groups. Following the card analogy, this can be achieved by creating 'stacks'. Stacks can be created manually by dragging one card on top of another or automatically based on properties. In a hit-triaging scenario we may wish to group compounds into different series. The figure below shows an example where cards have been stacked automatically using a maximum common substructure clustering algorithm. The algorithm's assignment can be refined based on a chemist's perspective. For example, we may wish to merge series 1 and 2 shown in the figure below. Stack 'covers' display selected summary information on the compounds in the stack, e.g. a common substructure and distribution of target affinities, as shown below.

Drag stacks to merge series 1 and 2



Stack covers display the maximum common substructure

## Automated Layouts



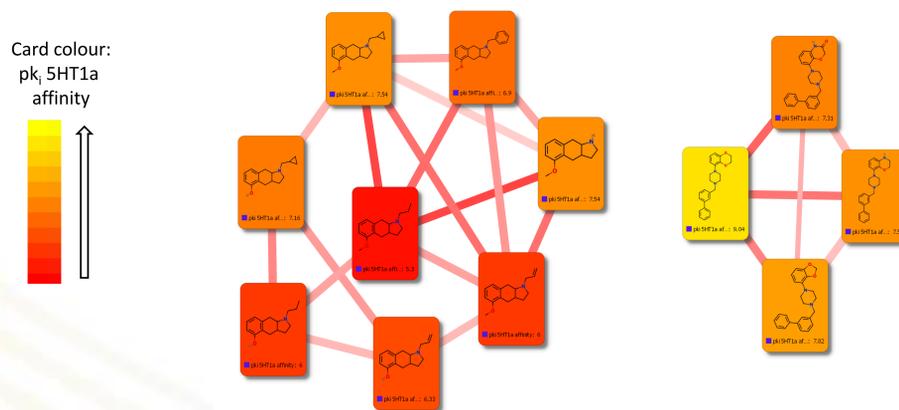
Similarity spiral showing the neighbourhood of a compound

Each row contains compounds with the same scaffold

96 well plate layout

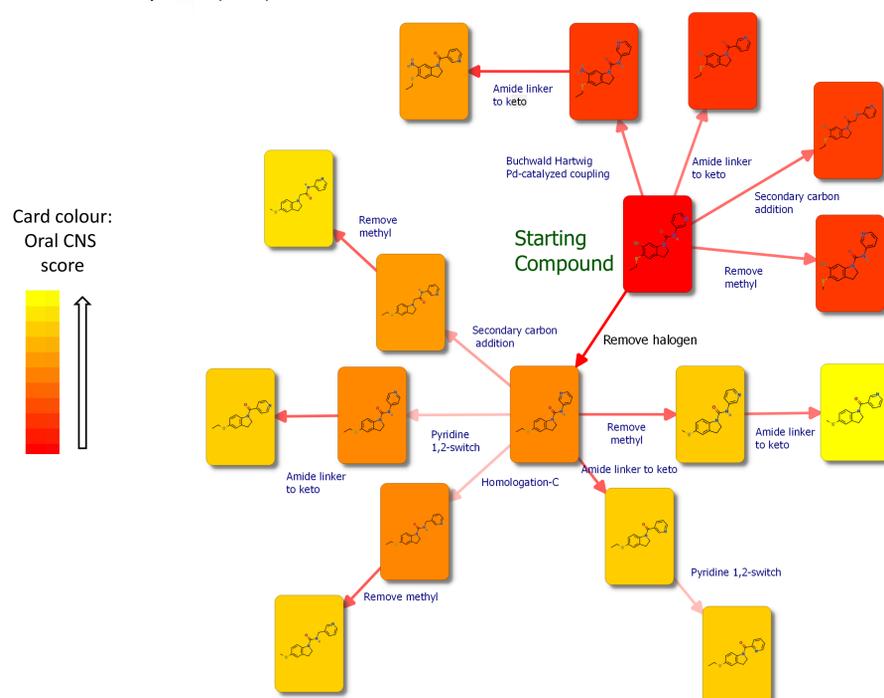
## Matched Pairs

Relationships between compounds can be indicated by linking cards. Cards can be linked manually or automatically and a network layout applied. One application is to present the results of a matched pairs analysis, as shown in the figure below. In this case, links are added automatically to indicate matched pairs. The links are coloured to indicate the change in 5HT1a affinity (pKi) associated with each structural modification, where deeper red links indicate greater changes. The disconnected networks identify different series.



## Idea Flow

Networks can also be used to show the flow of ideas through the course of a project using directed links to show 'parent - child' relationships. The figure below shows the evolution from a starting compound. In this case, the compounds have been generated automatically through the application of common structural transformations, as implemented in StarDrop's Nova™ module [2]. A selection criterion has been applied to steer the generation process to select compounds with higher score for suitable properties for an orally dosed compound for a central nervous system (CNS) indication.



## Conclusions

Rather than imposing a strict linear organisation, a card view allows compounds to be organised freely to produce many different views of compound data. Compounds can be arranged by a chemist in any way that best represents their way of thinking about their data. Automatic layouts can be used in combination with manual intervention to investigate the relationships between compounds and their properties and quickly target high quality series and strategies for further optimisation.

## References

- [1] [www.optibrium.com/stardrop](http://www.optibrium.com/stardrop)  
[2] J. Chem. Inf. Model. (2011) 51(11) pp. 2967-2976