



smarter chemistry | smarter decisions™

# Analysing selectivity through multi-dimensional activity cliff analysis

Tim Cheeseright

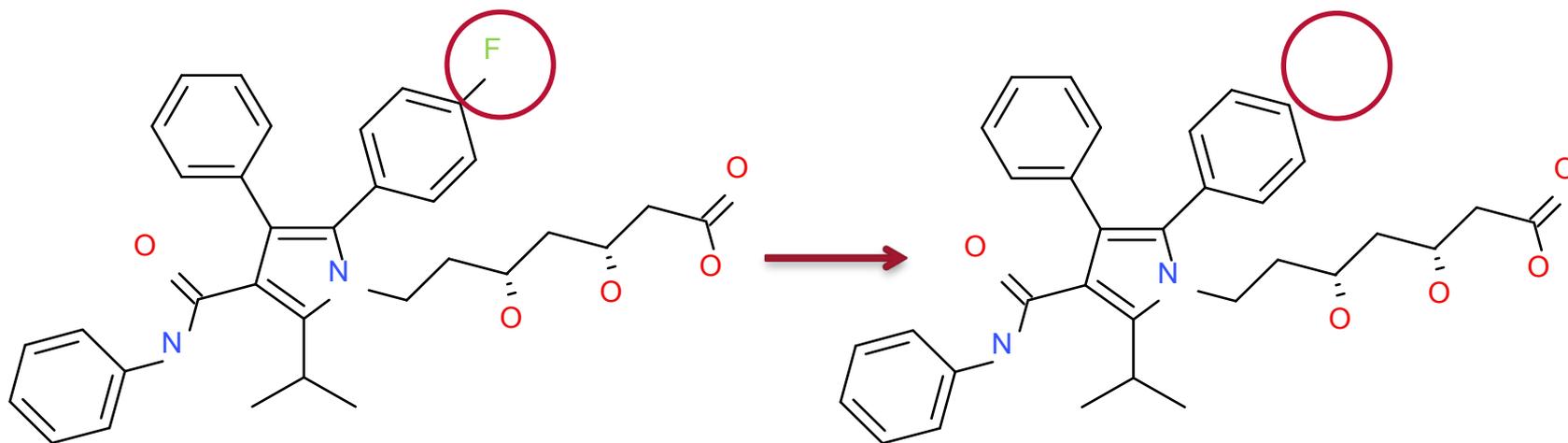
# Cresset summary

---

- > **Growing and profitable company**
  - > 20% year on year growth since 2009
  - > 20 People, 12 with PhDs
- > **Primary market pharmaceutical and biotech R&D**
  - > **Software:**
    - > 14 of the top 20 pharmaceutical companies use Cresset's technology in their research programmes
  - > **Consultancy Services:**
    - > ~200 collaborative projects delivered to global clients
- > **Secondary markets: agrochemicals, flavours and fragrances, consumer health and fine chemicals**

# Drug discovery's similarity hypothesis

---



- > Similar molecules have similar activities
- > Small changes lead to small changes
- QSAR, virtual screening, lead optimization

# (Un)Interesting SAR

---

What about the bits where the similarity hypothesis breaks down?

Nothing happens



Something dramatic happens



# Activity Cliffs – interesting regions of SAR

---

## > Many names:

- > Disparity (Merck 1990s)
- > SALI (Guha/Drie 2008)
- > Activity Landscapes
- > Activity Cliffs

## > Definition:

- > For each pair of molecules  $K = \frac{Act_1 - Act_2}{Distance_{12}}$

## > Usually distance = 1 – similarity

- > Similarity from 2D fingerprints, tanimoto etc
- > Large K indicates an activity cliff

# Gaining understanding of Activity Cliffs

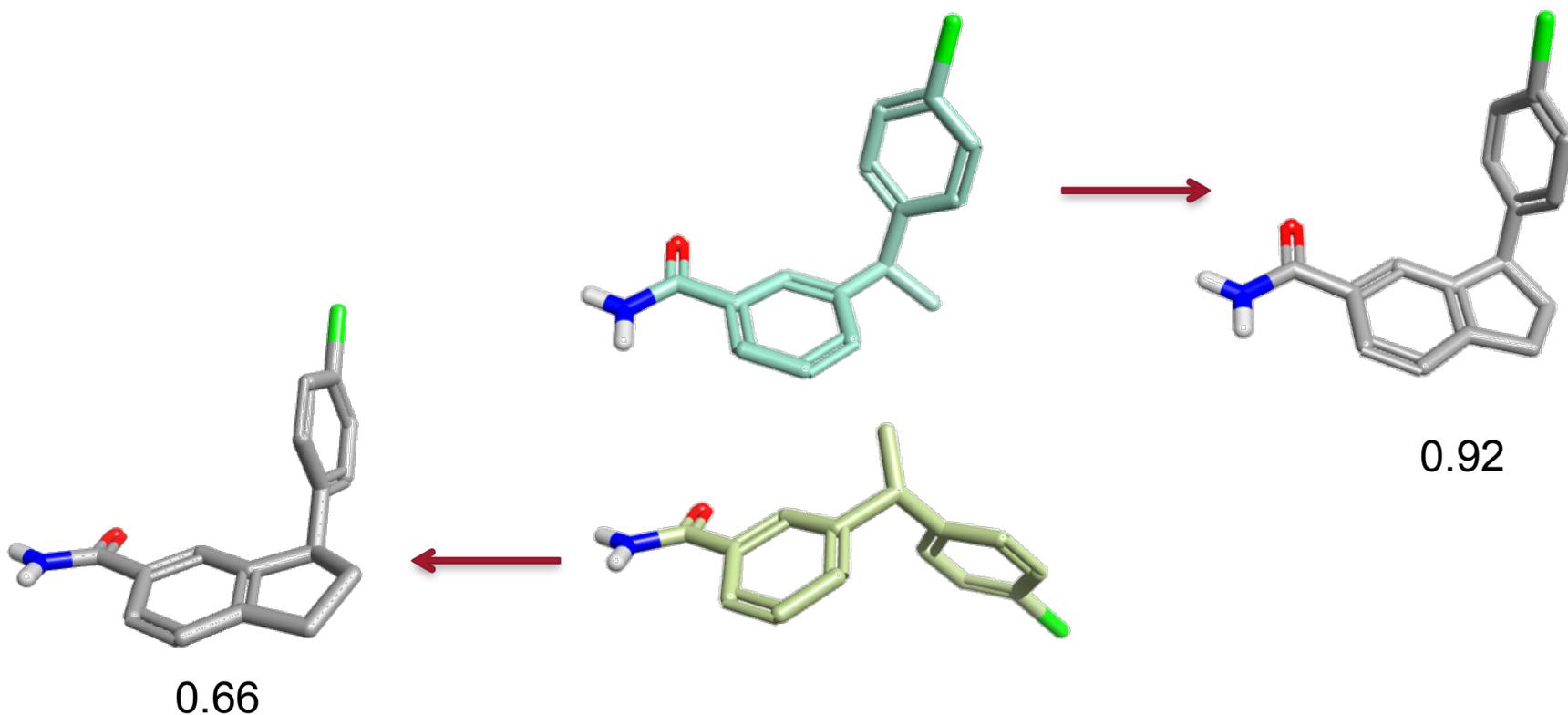
---

- > Activity cliffs from 2D similarity highly valuable
- > But no explanation for why the cliff is present
- > Without an explanation we cannot use the cliff to design new compounds with confidence
- > True understanding can come from 3D metrics
  - > Shape
  - > Electrostatics
- > What about using 3D similarity from the outset?

# Using 3D similarity

---

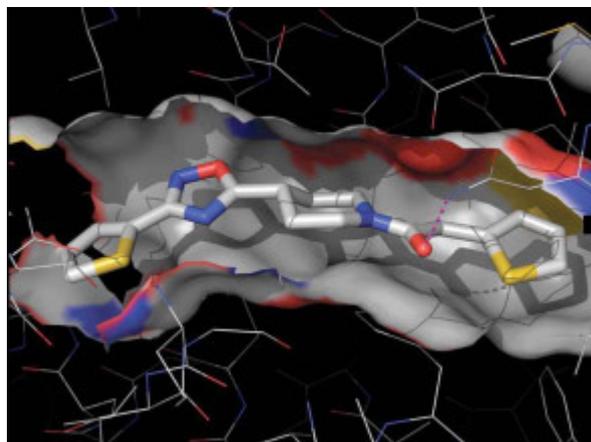
- > 2D metrics are easy: 1:1 map to topology
- > 3D is defined for **conformers**, not for **molecules**



# Context is everything

---

- > Don't need/want **generic** 3D similarity
  - > Have activity context – bound to the protein



- > Align all molecules to known bioactive reference conformer
- > Provides a conformation context to each molecule

# 3D disparity

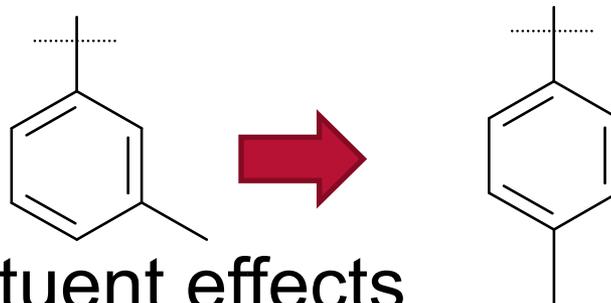
---

1. Generate conformers
2. Align to reference(s)
3. Calculate 3D similarity matrix on aligned conformations

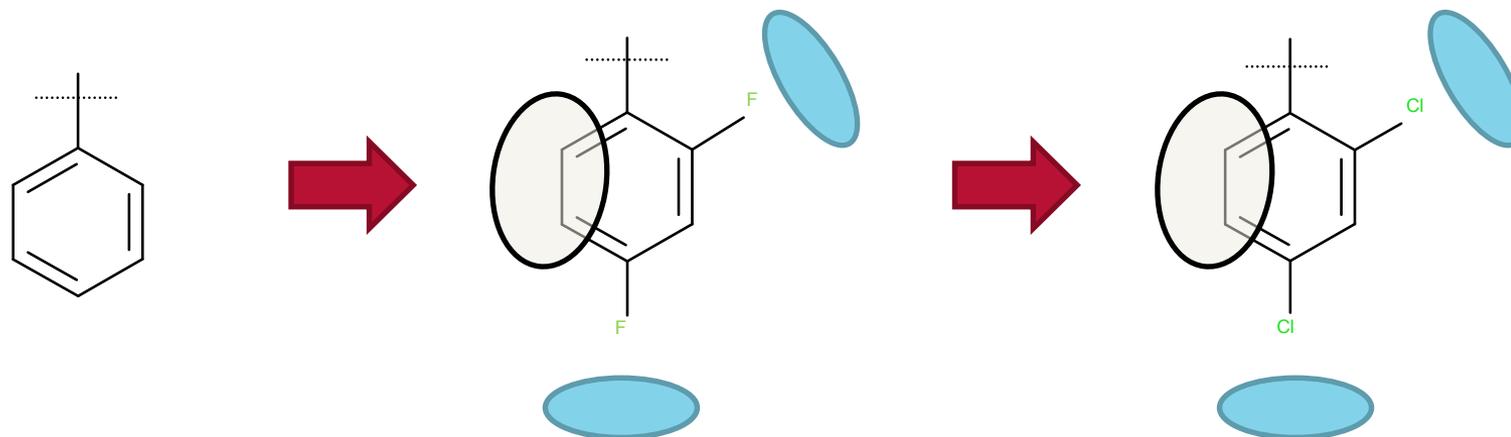
What 3D properties do we want to capture?

# Properties of a 3D similarity

> Shape / Sterics



> Electrostatics – substituent effects

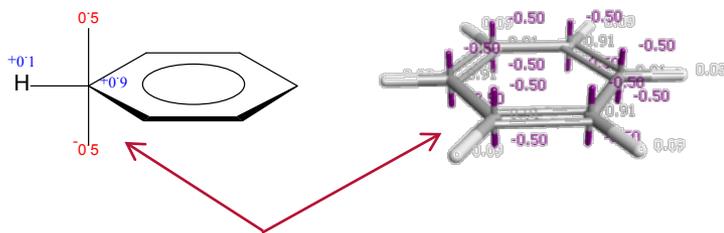


Changes to potential interactions from new atoms

Changes induced in retained portions

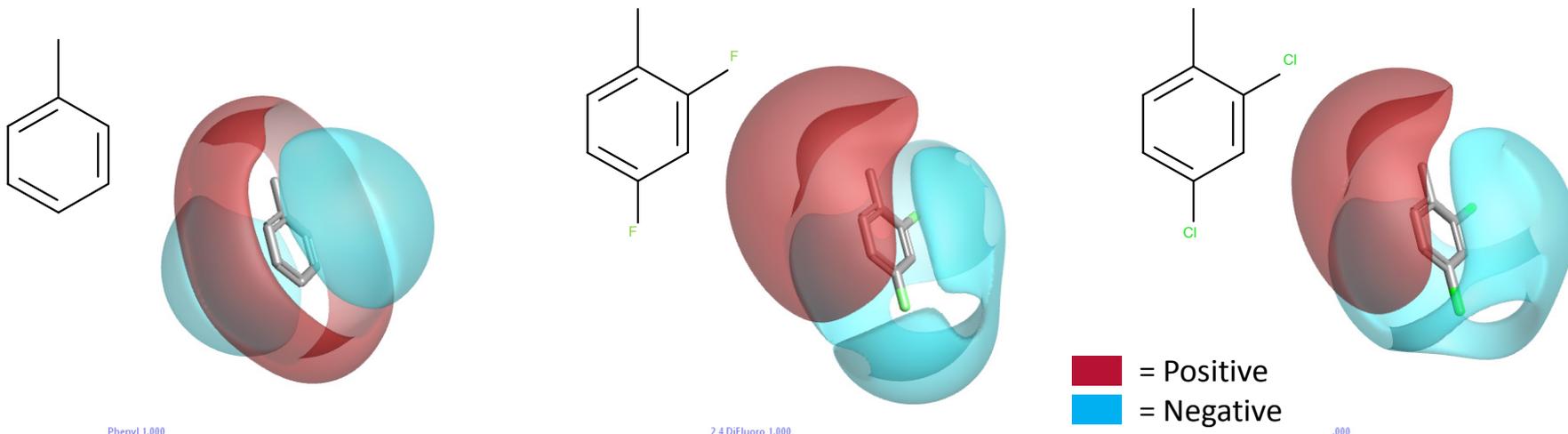
# Detailed electrostatics from XED

> eXtended Electron Distribution gives detailed electrostatic interaction patterns



Separation of  $\pi$ - and  $\sigma$ - charges enables modelling of substituent effects

XED adds p-orbitals to get detailed representation of atoms



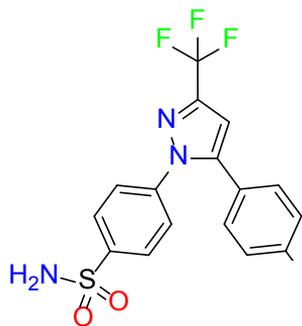
Phenyl 1.000

2,4-Difluoro 1.000

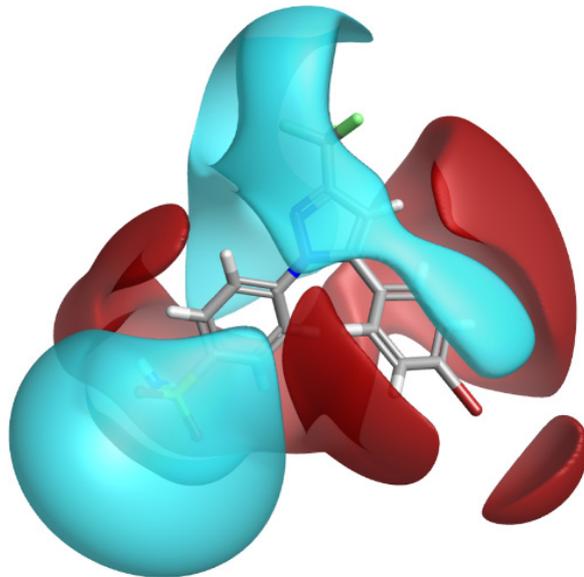
.000

# Field points

---



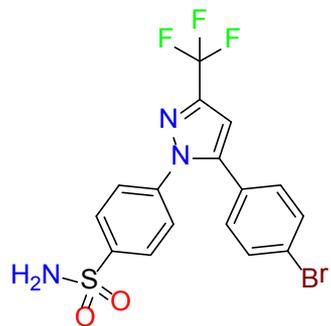
MIP contains too much information to use computationally in a reasonable time



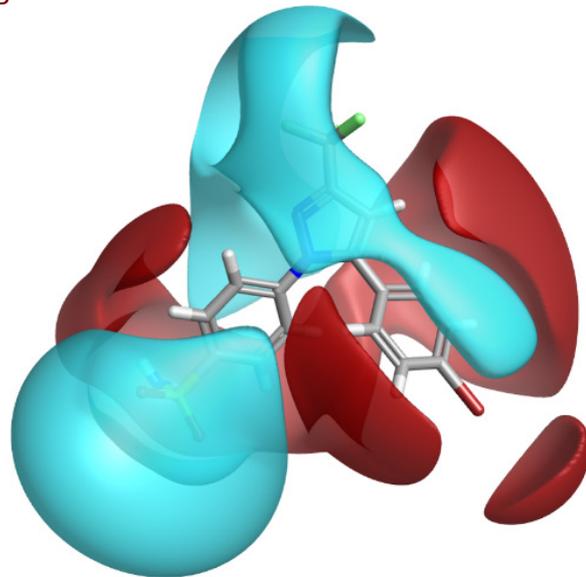
3D Molecular Electrostatic Interaction Potential (MIP)

 = Positive  
 = Negative

# Field points

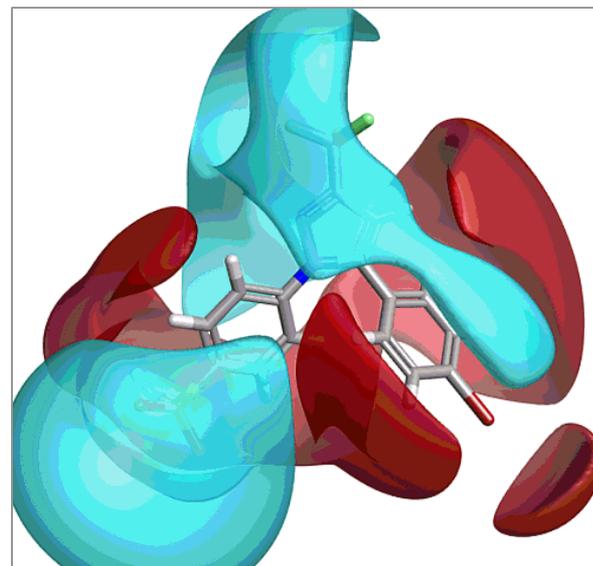


MIP contains too much information to use computationally in a reasonable time



3D Molecular Electrostatic Interaction Potential (MIP)

Field Points provide computationally tractable framework for electrostatic similarity



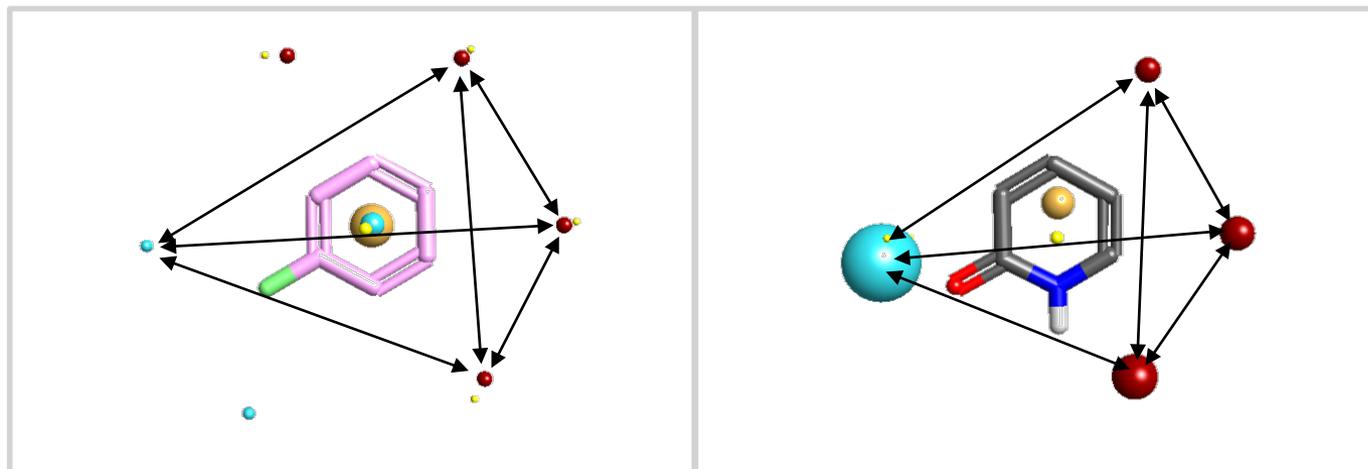
Field Points

■ = Positive  
■ = Negative

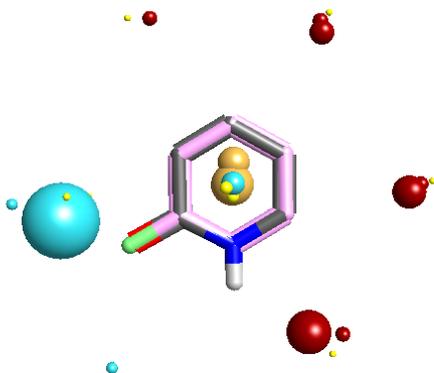
■ = Shape  
■ = Hydrophobic

# Alignment, scoring and comparisons

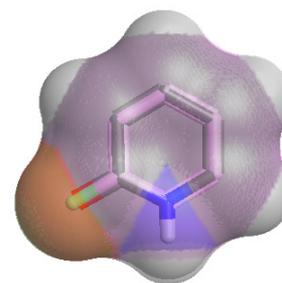
Clique based alignment



Fields  
0.66



Cheeseright et al,  
*J. Chem Inf. Mod.*, 2006, 665

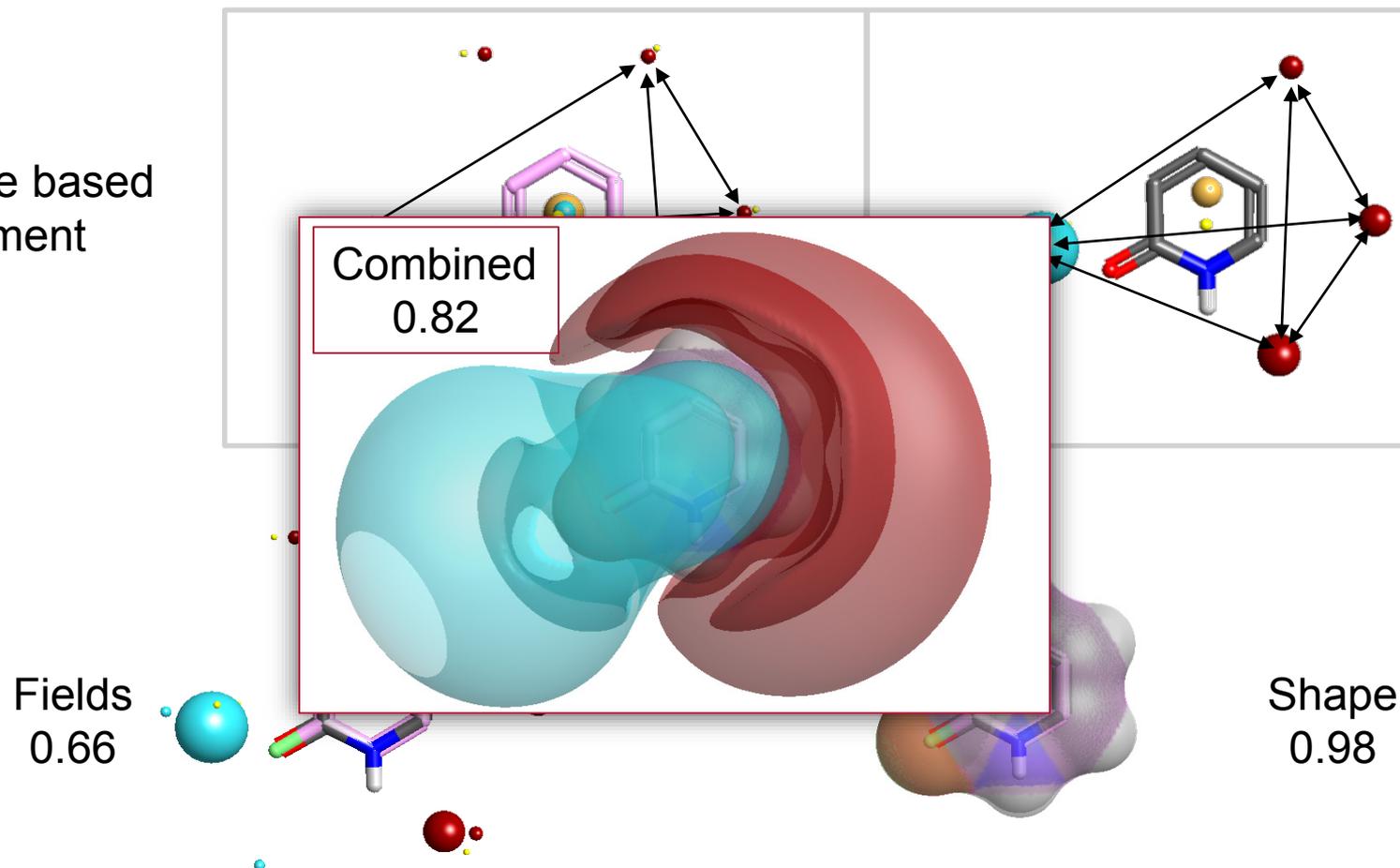


Shape  
0.98

Grant, Gallardo, Pickup,  
*J. Comp. Chem.*, 1996, 1653

# Alignment, scoring and comparisons

Clique based alignment



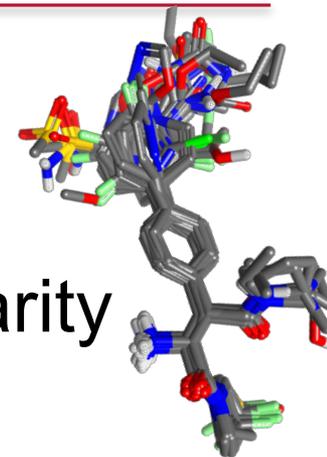
Cheeseright et al,  
*J. Chem Inf. Mod.*, 2006, 665

Grant, Gallardo, Pickup,  
*J. Comp. Chem.*, 1996, 1653

# 3D disparity workflow

---

1. Generate conformers
2. Align to reference(s)
3. Calculate 3D shape & electrostatic similarity matrix
  - > Allow small movements
4. Calculate disparity matrix from similarity numbers
  - > Similarity cutoff of 0.95 (Distance cutoff of 0.05)
5. Visualize
  - > Difficult – 100 molecules gives 4950 pairs!

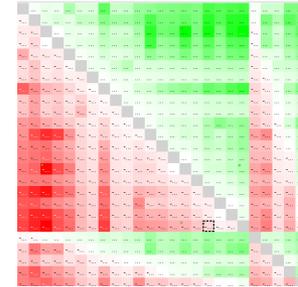


$$\frac{Act_1 - Act_2}{Distance_{12}}$$

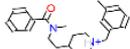
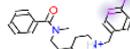
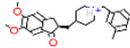
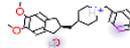
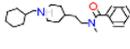
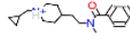
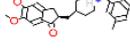
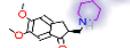
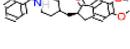
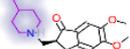
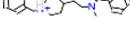
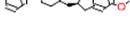
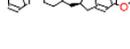
# Visualization

---

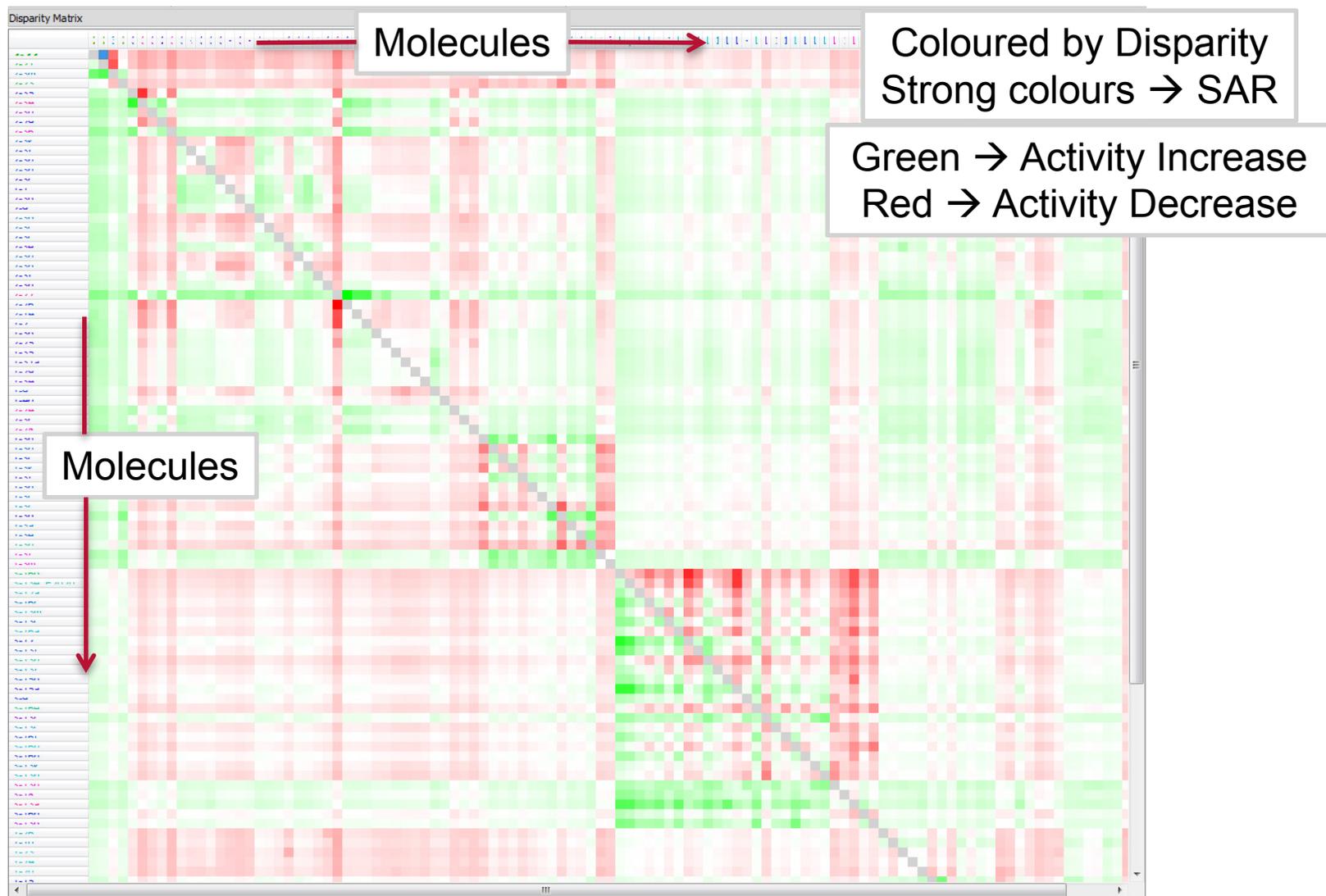
- > Existing ways to visualize
  - > Table & Matrix views



# Top pairs table

Disparity Table												
Good	Good Activity	Bad	Bad Activity	Disparity	Similarity	Fav	$\Delta$ Activity	$\Delta$ LE	$\Delta$ LLE	$\Delta$ TPSA	$\Delta$ SlogP	2D Sim
2-26 	6.84	2-27 	4.39	-49	0.951	☆	-2.45	-0.094	-0.094	0	0	0.776
3-16b 	8.7	3-17 	6.52	-41.2	0.947	☆	-2.18	-0.067	-0.07	3.2	-0.1	0.633
2-35 	6.39	2-34 	4.42	-39.4	0.959	☆	-1.97	-0.055	-0.026	0	-1.2	0.77
3-16b 	8.7	3-15a 	6.32	-38.7	0.939	☆	-2.38	-0.074	-0.057	0	-0.7	0.701
3-13e_E2020 	8.24	3-15a 	6.32	-38.4	0.952	☆	-1.92	-0.069	-0.056	0	-0.3	0.835
1-2 	6.77	2-27 	4.39	-34.5	0.931	☆	-2.38	-0.102	-0.108	0	0.3	0.791
3-13e_E2020 	8.24	3-17 	6.52	-34.4	0.963	☆	-1.72	-0.061	-0.07	3.2	0.2	0.762

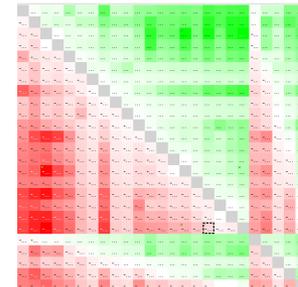
# Disparity matrix



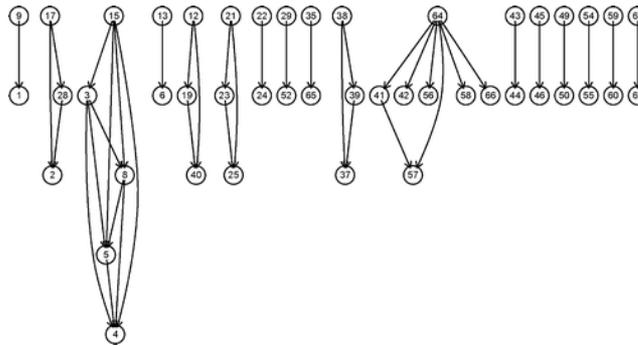
# Visualization

## > Existing ways to visualize

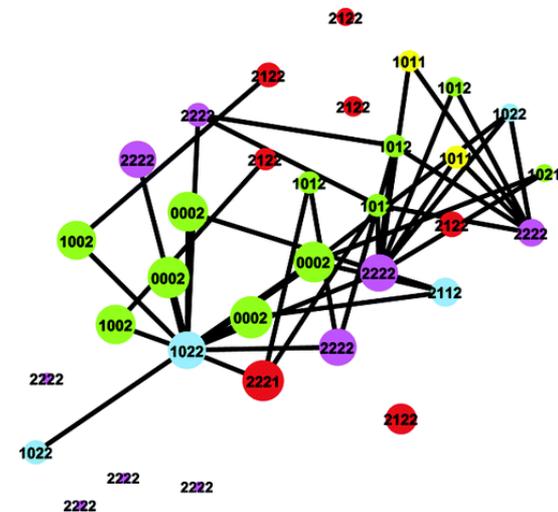
> Table & Matrix views



> Graph view (Guha/van Drie 2008)



> Activity landscapes (Bajorath)



# Activity View

Activity Miner

Matrix Settings, Matrix Color

Molecule Pair

Reset Atoms H XEDS Fields Shape, Display, Color Labels, Ospa2

Clear Mol +ve -ve vdW Hyd 2.0 Diff

Activity View

Activity: Similarity: Disparity:

0.935

Auto

Current Focus Compound

Comparator compound

Ten Nearest Compounds, height = distance

Shade = Disparity  
Strong colours = Strong SAR

Disparity Scale

-26.6 0.0 26.6

Comparison 5, #25 : activity 8.100

Prot Grid HBinds

Left mouse rotate, right mouse translate, middle mouse/wheel/Alt-left scales. Shift-mouse z-clip.

# Electrostatic comparison

The screenshot displays the Activity Miner software interface. The top toolbar includes options for Matrix Settings, Matrix Color, Molecule Pair, Reset, Atoms, H, XEDS, Fields, Shape, Display, Color, Labels, Ospa2, Clear, Mol, +ve, -ve, vdW, Hyd, 2.0, and Diff. The main window is titled 'Activity View' and shows a cluster of molecules around a central pie chart. The pie chart segments are labeled with activity, similarity, and disparity values:

- Segment 1: 20, #39; A: 9.500; S: 0.939; D: 1.6
- Segment 2: 10, #31; A: 8.800; S: 0.949; D: -1.1
- Segment 3: 16, #3; A: 9.333; S: 0.933; D: -2.0

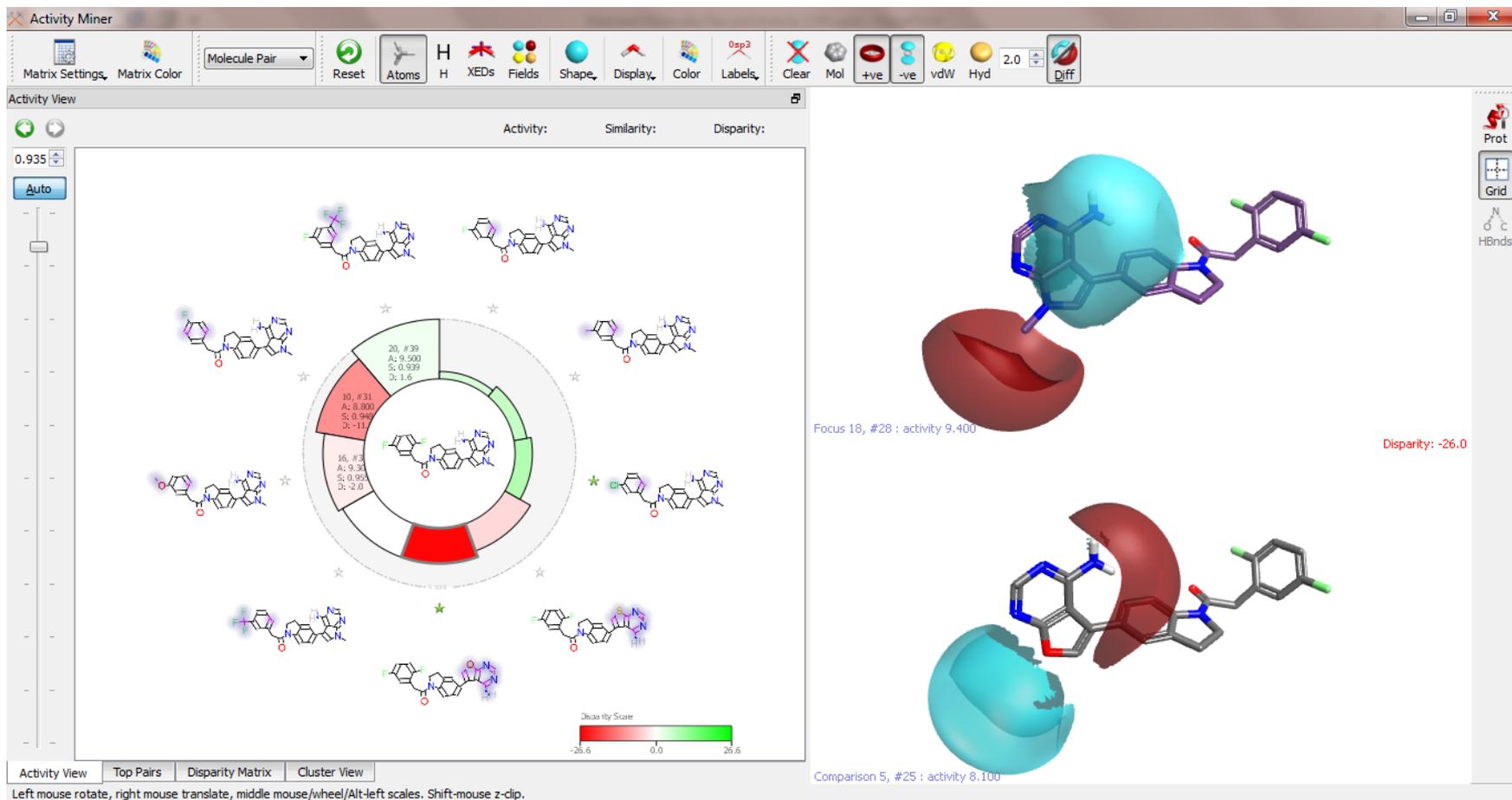
A 'Disparity Scale' at the bottom right ranges from -26.6 (red) to 26.6 (green). The right sidebar contains 'Prot', 'Grid', and 'HBinds' options. Two electrostatic comparison visualizations are shown on the right:

- Top visualization: Focus 18, #28 : activity 9.400. Disparity: -26.0
- Bottom visualization: Comparison 5, #25 : activity 8.100

At the bottom of the interface, navigation tabs include 'Activity View', 'Top Pairs', 'Disparity Matrix', and 'Cluster View'. A note at the bottom left states: 'Left mouse rotate, right mouse translate, middle mouse/wheel/Alt-left scales. Shift-mouse z-clip.'

# Electrostatic comparison

Difference plot – Regions where each molecule has stronger electrostatics



# Selectivity Cliffs

---

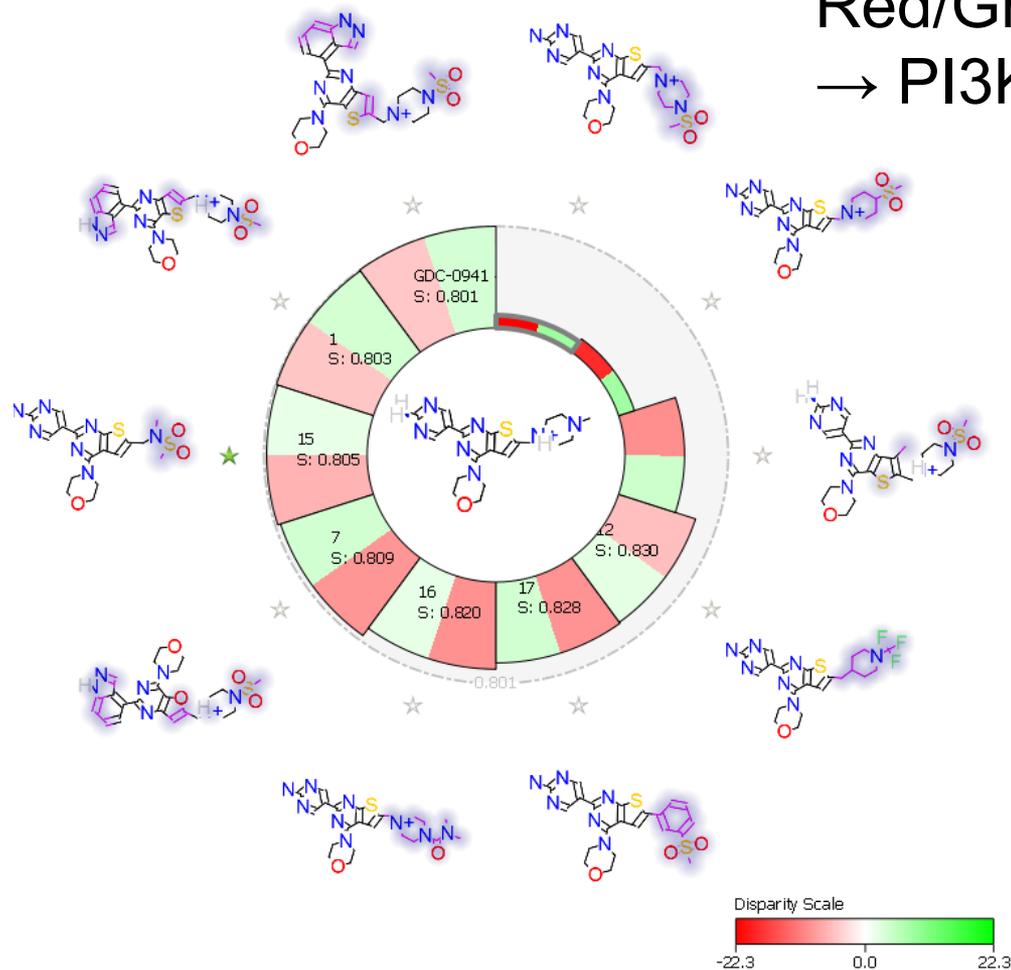
- > Selectivity often as important as potency
- > Look at what structural changes caused large changes in selectivity
- > Use Selectivity Endpoint as Activity?

$$\kappa \approx \frac{\Delta \textit{Selectivity}}{\Delta \textit{Structure}} = \frac{\left(\frac{\textit{Activity}_\beta}{\textit{Activity}_\alpha}\right)_A - \left(\frac{\textit{Activity}_\beta}{\textit{Activity}_\alpha}\right)_B}{(1 - \textit{Similarity})}$$

- > What about 3 activities?
- > How would we visualize that?

# Activity View – 2 activities

Red/Green =  $\downarrow pK_{i\beta}$ ,  $\uparrow pK_{i\alpha}$   
→ PI3K $\alpha$  selective



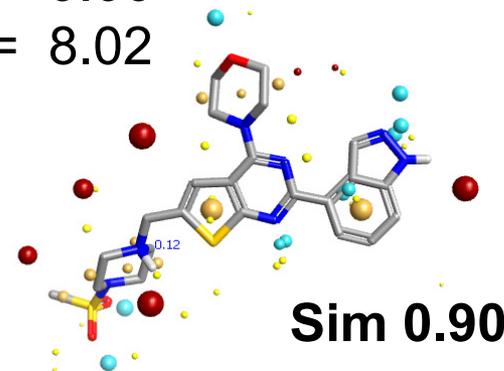
# Selectivity matrices – 2 activities



## GDC-0941

$pK_i\alpha = 9.06$

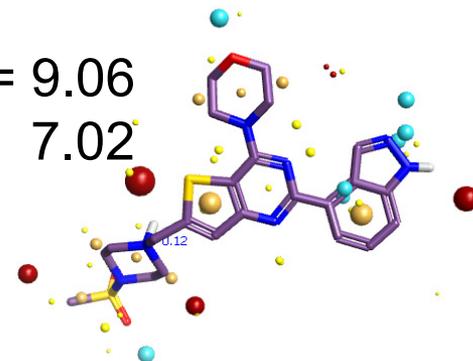
$pK_i\beta = 8.02$



## 6

$pK_i\alpha = 9.06$

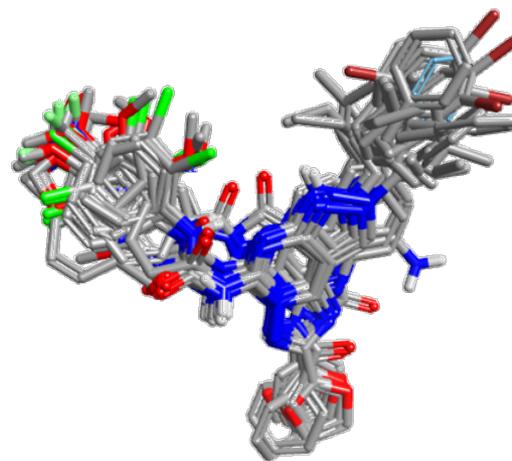
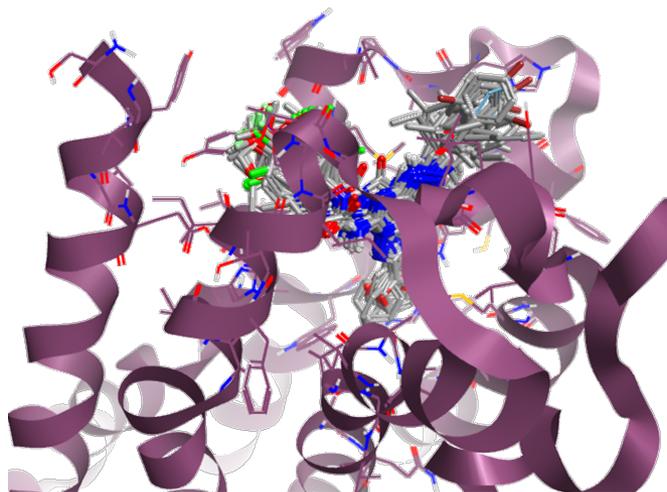
$pK_i\beta = 7.02$



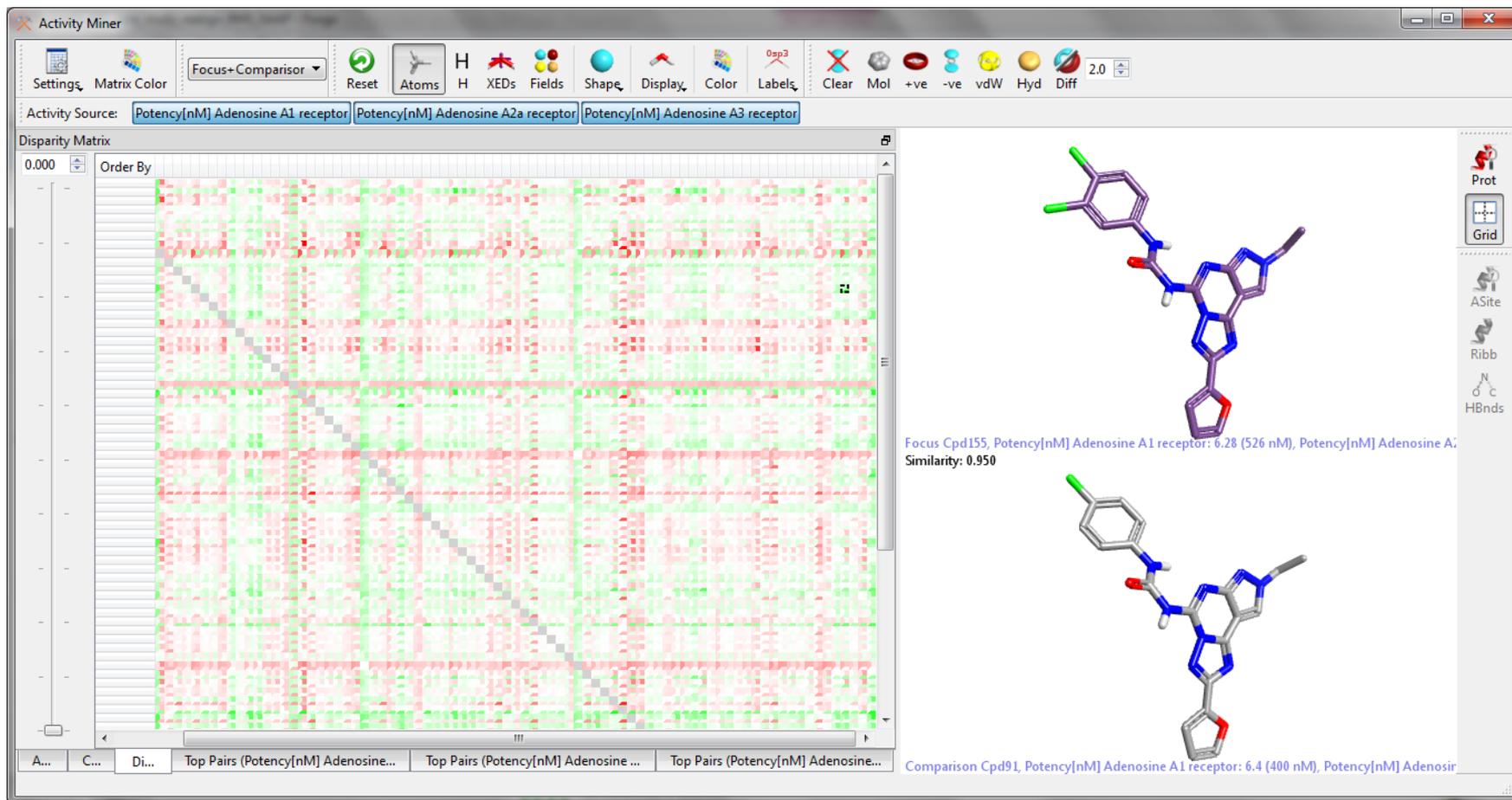
# Application to Adenosine Receptor Antagonists

---

- > Data set from Bajorath J. Chem. Inf. Model 51 258-266 2011
- > 3 Activities – A1, A2a, A3 receptors
- > Ligands aligned to x-ray structures 3PWH, 3EML
- > 89 cmpd sub-set with high 3D similarity (>0.7)



# Disparity Matrix – 11,748 data points



# Disparity Matrix – focus on highly similar pairs

The screenshot displays the Activity Miner software interface. The main window is titled "Activity Miner" and features a toolbar with various icons for settings, matrix color, focus+comparator, reset, atoms, H, XEDs, fields, shape, display, color, labels, clear, mol, +ve, -ve, vdW, hyd, and diff. The activity source is set to "Potency[nM] Adenosine A1 receptor", "Potency[nM] Adenosine A2a receptor", and "Potency[nM] Adenosine A3 receptor".

The Disparity Matrix is a heatmap showing similarity scores between compounds. The y-axis lists compound IDs from Cod100 to Cod200. The x-axis is labeled "Order By" and contains a series of "d:" and "id" labels. The matrix shows a diagonal of high similarity (0.900) and several off-diagonal blocks of high similarity (red and green squares).

Two chemical structures are shown on the right side of the interface. The top structure is labeled "Focus Cpd155, Potency[nM] Adenosine A1 receptor: 6.28 (526 nM), Potency[nM] Adenosine A2a receptor: 6.4 (400 nM), Potency[nM] Adenosine A3 receptor: 6.4 (400 nM), Similarity: 0.950". The bottom structure is labeled "Comparison Cpd91, Potency[nM] Adenosine A1 receptor: 6.4 (400 nM), Potency[nM] Adenosine A2a receptor: 6.4 (400 nM), Potency[nM] Adenosine A3 receptor: 6.4 (400 nM), Similarity: 0.950".

The interface also includes a sidebar on the right with icons for Prot, Grid, ASite, Ribb, and HBnds. At the bottom, there are tabs for "A...", "C...", "Di...", and "Top Pairs (Potency[nM] Adenosine...".

# Why?

The screenshot displays the Activity Miner software interface. The top toolbar includes various icons for settings, matrix color, focus-comparator, reset, atoms, H, XEDs, fields, shape, display, color, labels, clear, mol, +ve, -ve, vdW, hyd, and diff. The activity source is set to Potency[nM] Adenosine A1 receptor, Potency[nM] Adenosine A2a receptor, and Potency[nM] Adenosine A3 receptor.

The Disparity Matrix shows a grid of colored squares (red, green, white) representing similarity between compounds. The y-axis lists compounds from Cpd108 to Cpd197. The x-axis is labeled with 'Order By' and 'id3'.

Two 3D molecular models are shown on the right. The top model is labeled: Focus Cpd155, Potency[nM] Adenosine A1 receptor: 6.28 (526 nM), Potency[nM] Adenosine A1 receptor: 6.28 (526 nM), Similarity: 0.950. The bottom model is labeled: Comparison Cpd91, Potency[nM] Adenosine A1 receptor: 6.4 (400 nM), Potency[nM] Adenosine A1 receptor: 6.4 (400 nM), Similarity: 0.950.

At the bottom, there are tabs for 'A...', 'C...', 'Di...', and 'Top Pairs (Potency[nM] Adenosine...)'. A legend at the bottom left states: Left mouse rotate, right mouse translate, middle mouse/wheel/Alt-left scales. Shift-mouse z-clip.

# Limitations

---

- > 2 Activities work well
- > 3 is OK
  
- > 7 is too many!

# Limitations

The screenshot displays the Activity Miner software interface. The top toolbar includes various icons for settings, matrix color, focus+comparator, reset, atoms, H, XEDs, fields, shape, display, color, labels, clear, mol, +ve, -ve, vdW, hyd, and diff. Below the toolbar, the 'Activity Source' section lists several targets: PIM1, GSK3B, CDK2, STK3, CDK7, CAMK2D, CSNK1A1, and CDK5. The 'Disparity Matrix' on the left shows a heatmap of similarity values between different molecule configurations. The 'Activity View' on the right displays a central molecule structure surrounded by other structures, with a circular chart showing similarity and activity values for each configuration. A 'Disparity Scale' legend is located at the bottom right of the Activity View, ranging from -7.5 to 7.5.

Activity Miner

Settings Matrix Color Focus+Comparator Reset Atoms H XEDs Fields Shape Display Color Labels Clear Mol +ve -ve vdW Hyd Diff 2.0

Activity Source: PIM1 GSK3B CDK2 STK3 CDK7 CAMK2D CSNK1A1 CDK5

Disparity Matrix

Order By 30\_conf 31\_conf 32\_conf 33\_conf 34\_conf 35\_conf 36\_conf 37\_conf 38\_conf

0002\_conf:1  
0011\_conf:1  
0012\_conf:1  
0029\_conf:1  
0030\_conf:1  
0031\_conf:1  
0032\_conf:1  
0033\_conf:1  
0034\_conf:1  
0035\_conf:1

Activity View

0002\_conf:1 Similarity: 0.720 Source: CAMK2D Activity: 6.600 Disparity: 0.0

Order By 0.615 Auto

Disparity Scale -7.5 0.0 7.5

Focus 0029\_conf:1, PIMI: 6.9 (125.89 nM), GSK3B: 5.5 (3162.28 n Comparison 0037\_conf:1, PIMI: 5.2 (6309.57 nM), GSK3B: 7.1 (7.1

Left mouse sets '0002\_conf:1' as comparison molecule, double-click sets '0002\_conf:1' as focus molecule

# Conclusions

---

- > **Activity Cliff/Disparity analysis provides quick insights into SAR**
  - > Focus on understanding the reason for a cliff
  - > Drive design decisions
- > **Multiple ways to navigate the data**
  - > Compound focus
  - > Most significant changes
  - > Global overview
  - > Cluster analysis
- > **2D and 3D both useful**
  - > 2D provides insights into conformational changes
  - > 3D provides insights into electrostatic effects
- > **Visualizing multiple activities simultaneously allows selectivity analysis**
  - > Large amounts of data difficult to visualize

# Acknowledgements

---

- > Mark Mackey
- > Nigel Palmer
- > Rae Lawrence
- > Susana Tomasio
- > Giovanna Tedesco

# Thank you!

---

Questions Welcomed



Follow Cresset



[tim@cresset-group.com](mailto:tim@cresset-group.com)