



Guiding Compound Design in 2, 3... N Dimensions

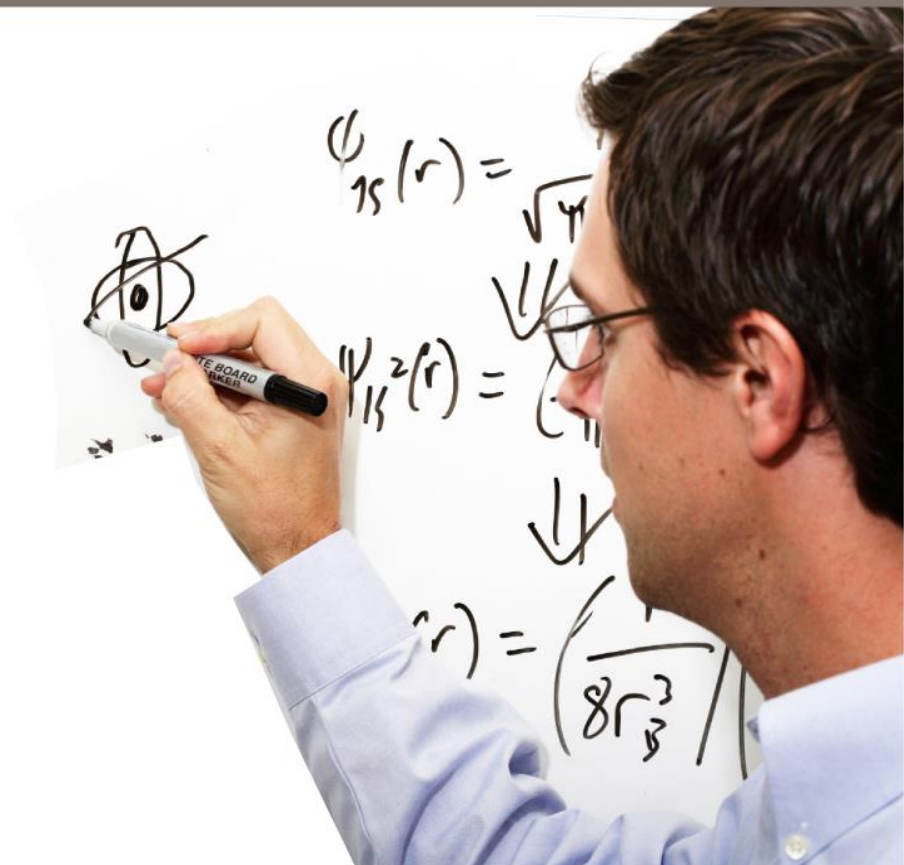
Streamlining Drug Discovery and Development – Boston, April 11th 2016

Matthew Segall - matt.segall@optibrium.com

Overview

- 2-Dimensions
 - Qualitative SAR: Activity cliffs, matched molecular pair analysis...
 - Quantitative SAR: Predictive models
- 3-Dimensions
 - Structure-based design
 - Scoring/affinity prediction
 - Understanding 3D SAR
- N-Dimensions
 - Multi-parameter optimisation
- Linking 2D, 3D,... 'N'D SAR to guide design
- Conclusions

2-Dimensions



Qualitative 2D SAR

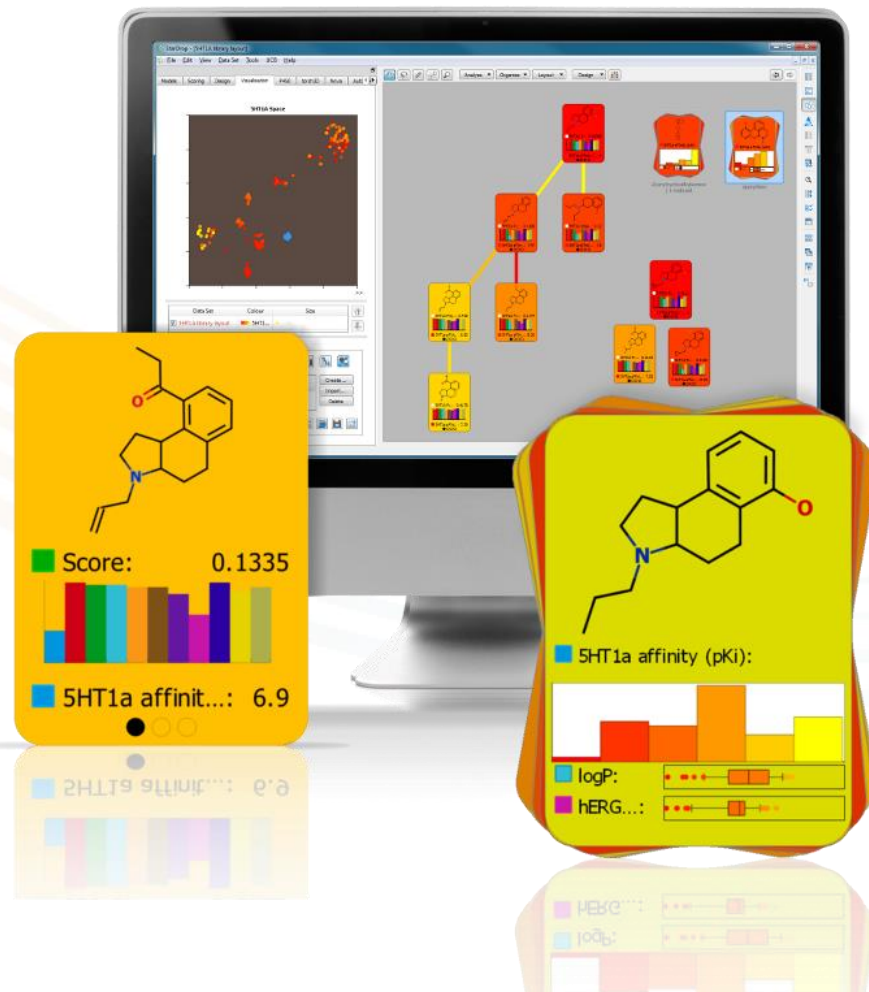
Many methods are routinely used for analysis of data to reveal patterns and trends to guide compound optimisation, e.g.

- Clustering
 - Group 'similar' compounds to identify series with interesting SAR
- Activity cliff detection
 - Small changes in structure that cause a large change in activity
- Matched molecular pair analysis
 - Pairs of compounds that are identical except for one small change at the same position

Visualising 2D SAR

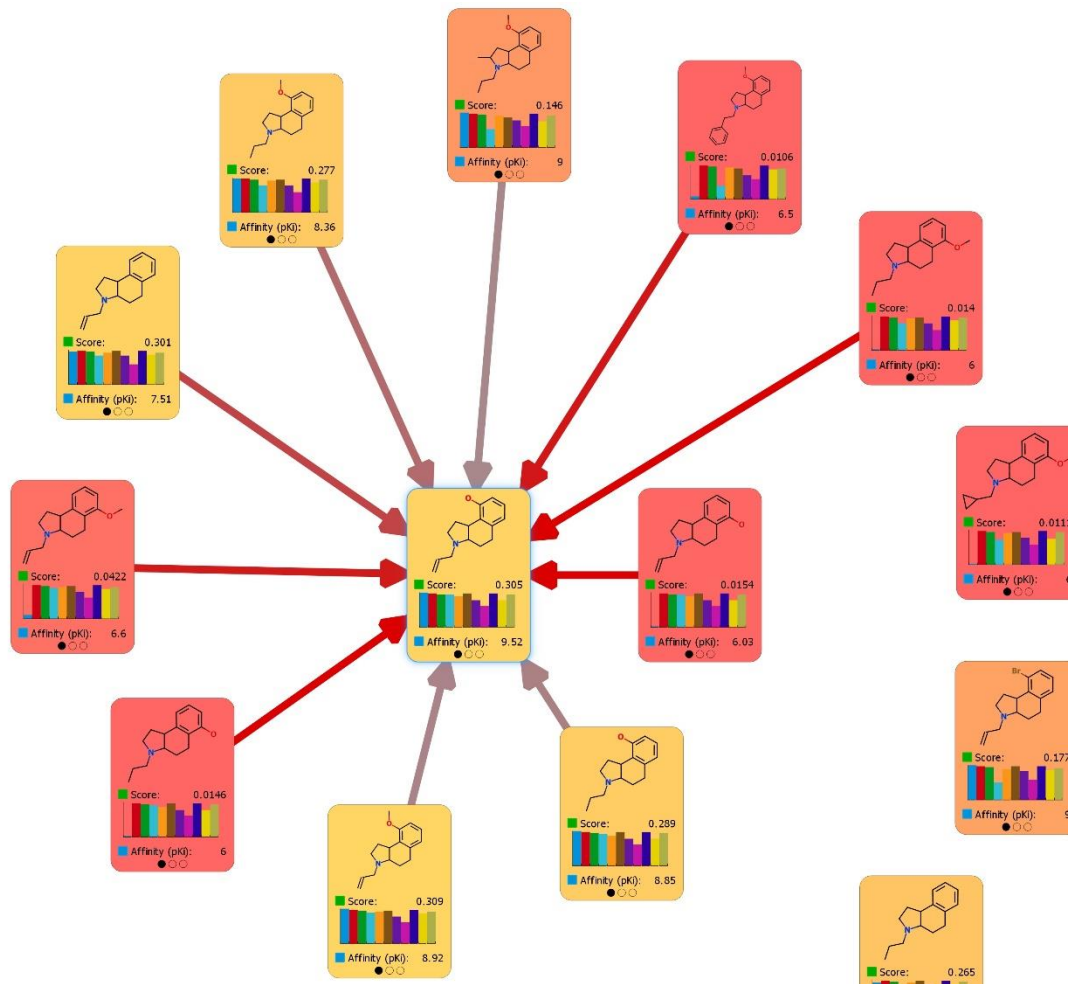
Card View™

- Freedom from the constraints of ‘chemical spreadsheets’
 - Represent compound relationships
- Work the way you think
 - Cards: Display key compound data
 - Stacks: Summarise and compare data for groups of compounds
 - Links: Highlight compound relationships
- Intuitive visualisation of analyses
 - Clustering, activity cliffs, matched molecular pairs...
- Quickly identify optimisation strategies

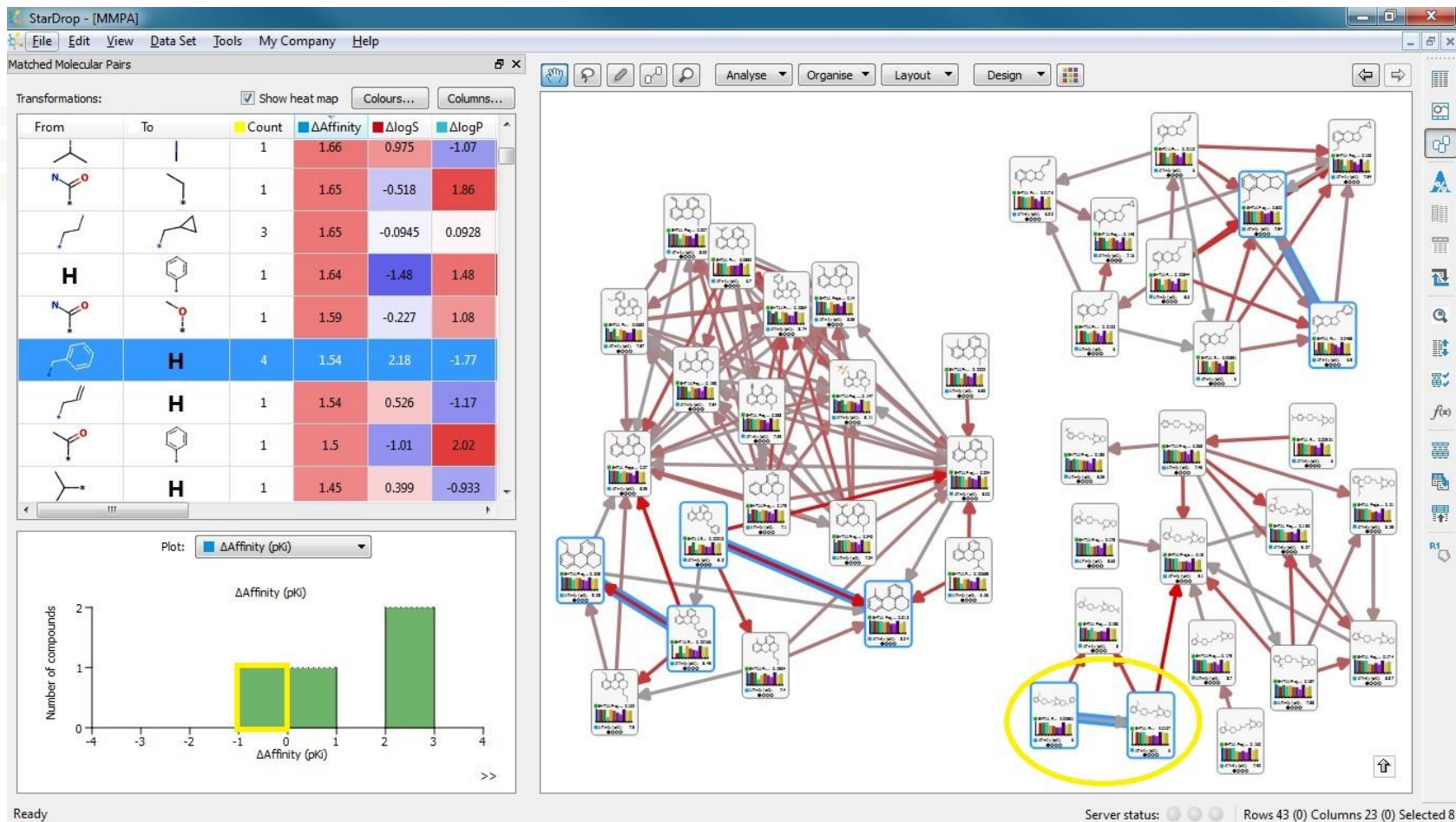


Activity Neighbourhood

Activity Cliff Visualisation



Matched Molecular Pair Analysis



Quantitative Structure-Activity Relationships

Principles

$$y = f(x_1, x_2, x_3, \dots) \pm \varepsilon$$

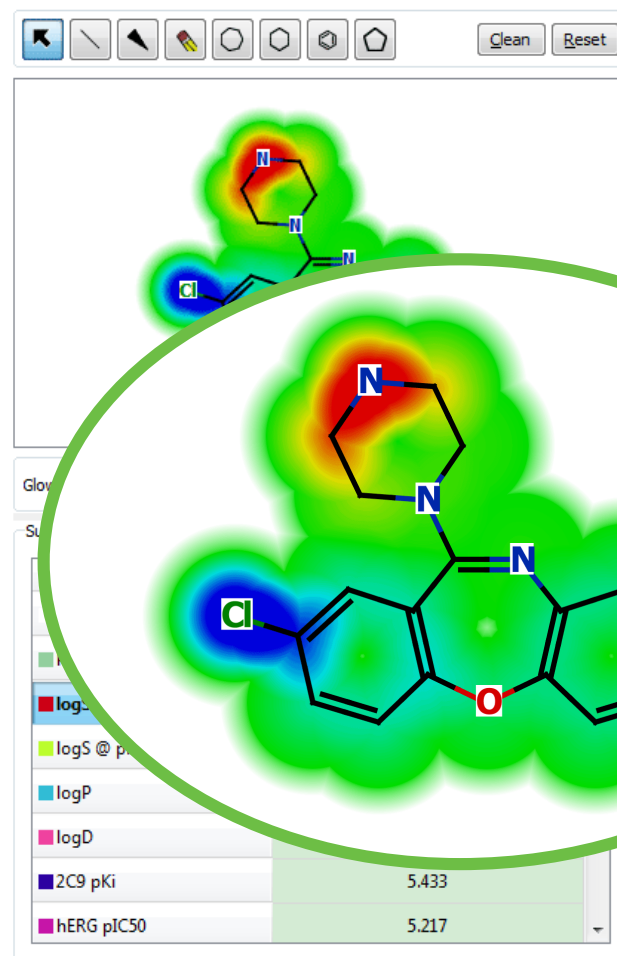
Statistical
uncertainty



- Data
 - Quality data is essential
 - Public data needs very careful curation (and may not be good enough)
- Descriptors, e.g.
 - Whole molecule properties, e.g. logP, MW, PSA...
 - Structural descriptors, SMARTS, fingerprints...
- Statistical fitting or machine learning method, e.g.
 - Partial least squares, artificial neural networks, support vector machines, random forest, Gaussian processes...
- Widely applied to prediction of ADME and physicochemical properties

Interactive Redesign

- QSAR models provide estimates of compounds' properties
 - Explore strategies for redesign
- Instant feedback on how properties are likely to change
 - “Why is a property value predicted?”
 - “Where can I change this property?”
- Glowing Molecule™:
 - Visual indication of structural influences on predicted properties

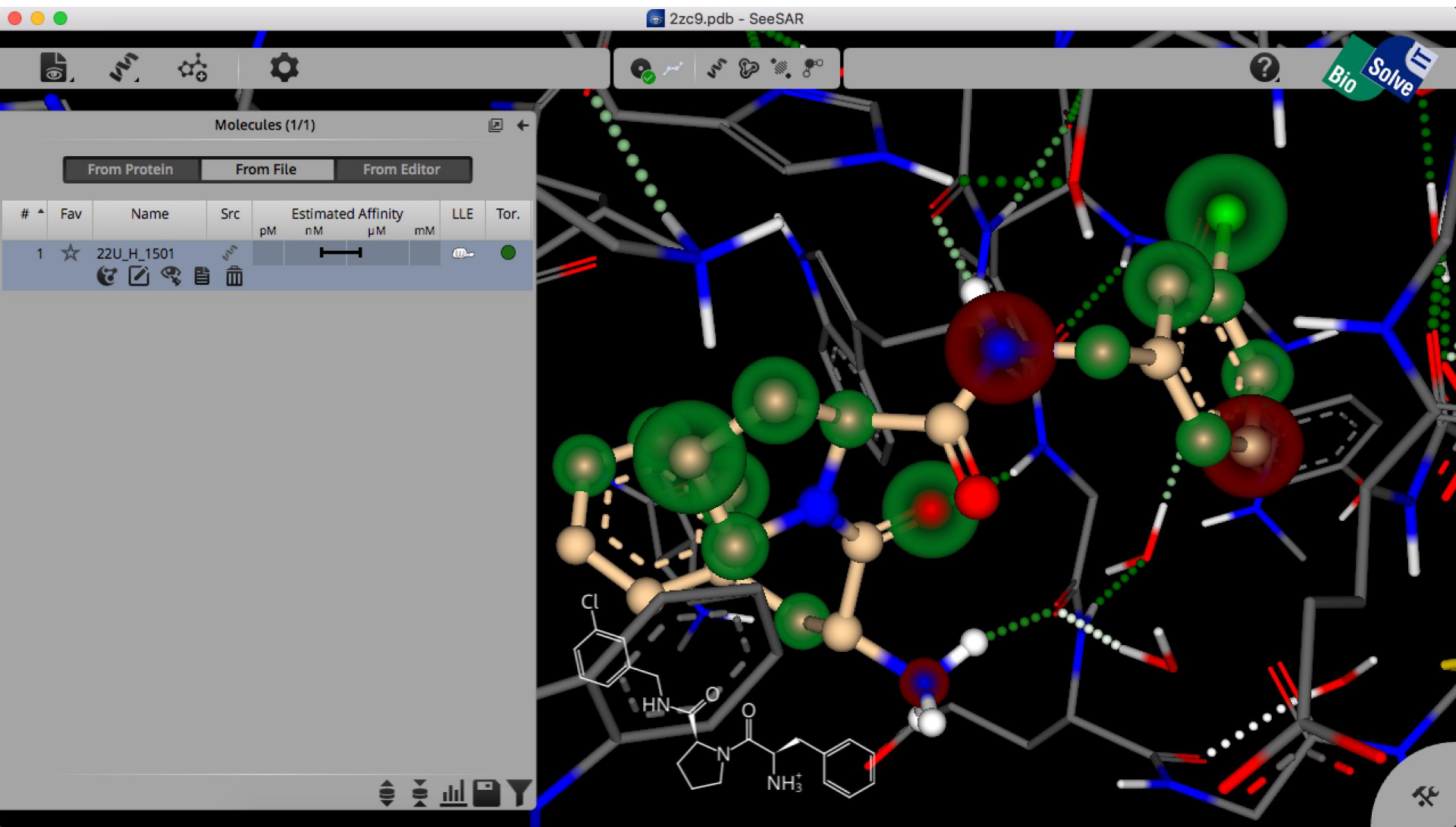




3-Dimensions



Visual Understanding of 3D Affinity Data

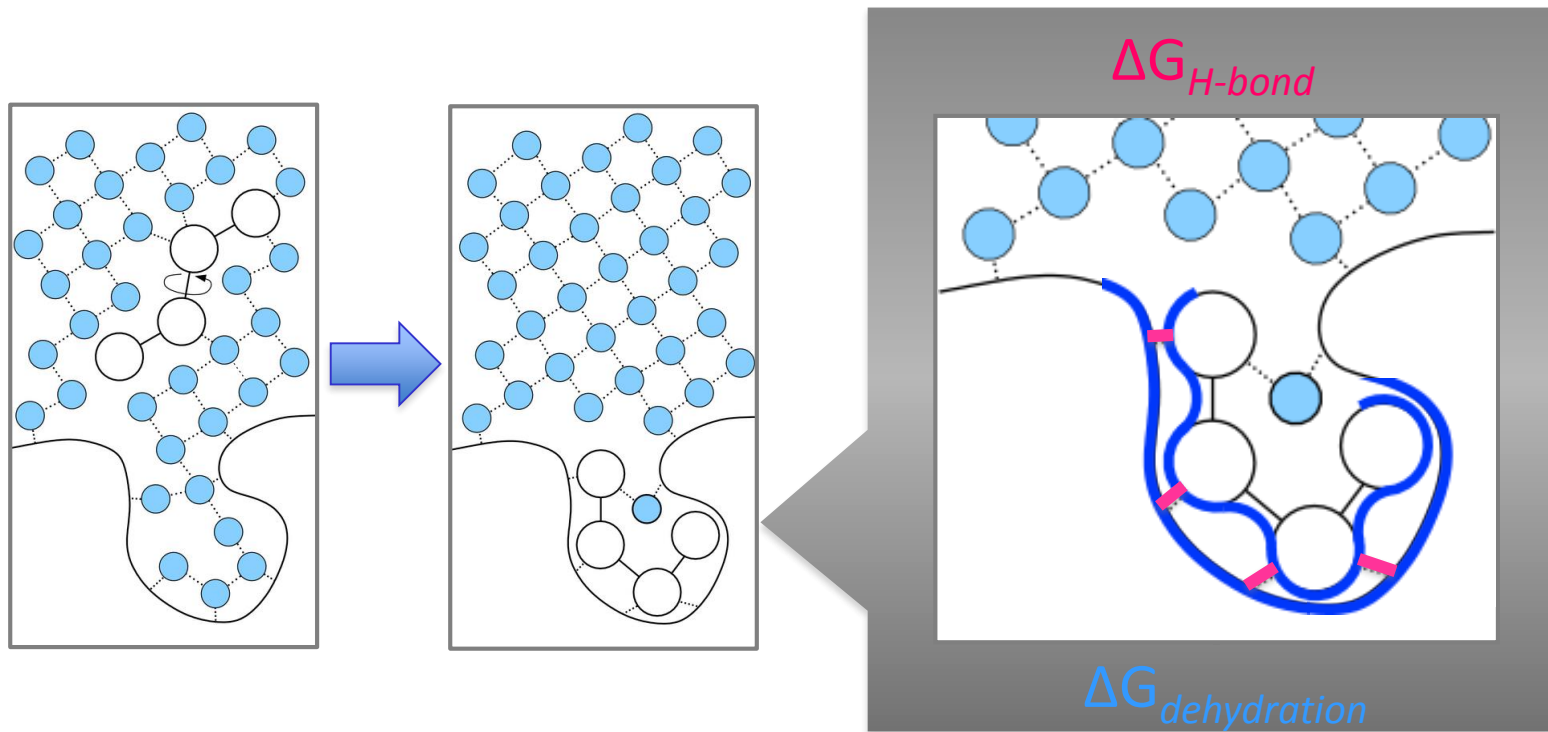


HYDE: A Different View @ Energetics*

- ★ Physics only
- ★ No calibration to complexes
- ★ Relates to real Free Energies

Please see Carsten Detering's talk for more details

HYDE Scoring Function – Concept*

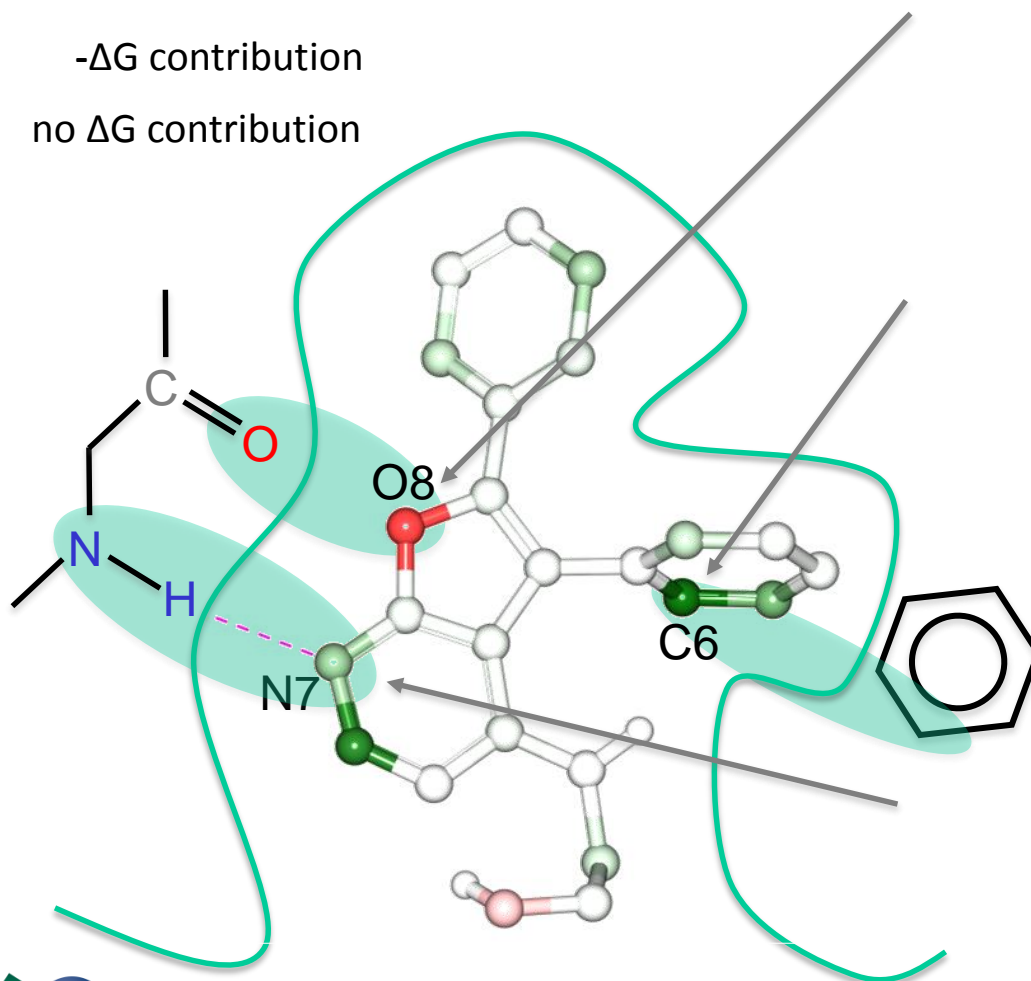


$$\Delta G_{HYDE}^i = \sum_{atom\ i} \Delta G_{dehydration}^i + \Delta G_{H-bond}^i$$

Hyde - Visual Affinities

HYDE color code:

- + ΔG contribution
- - ΔG contribution
- no ΔG contribution

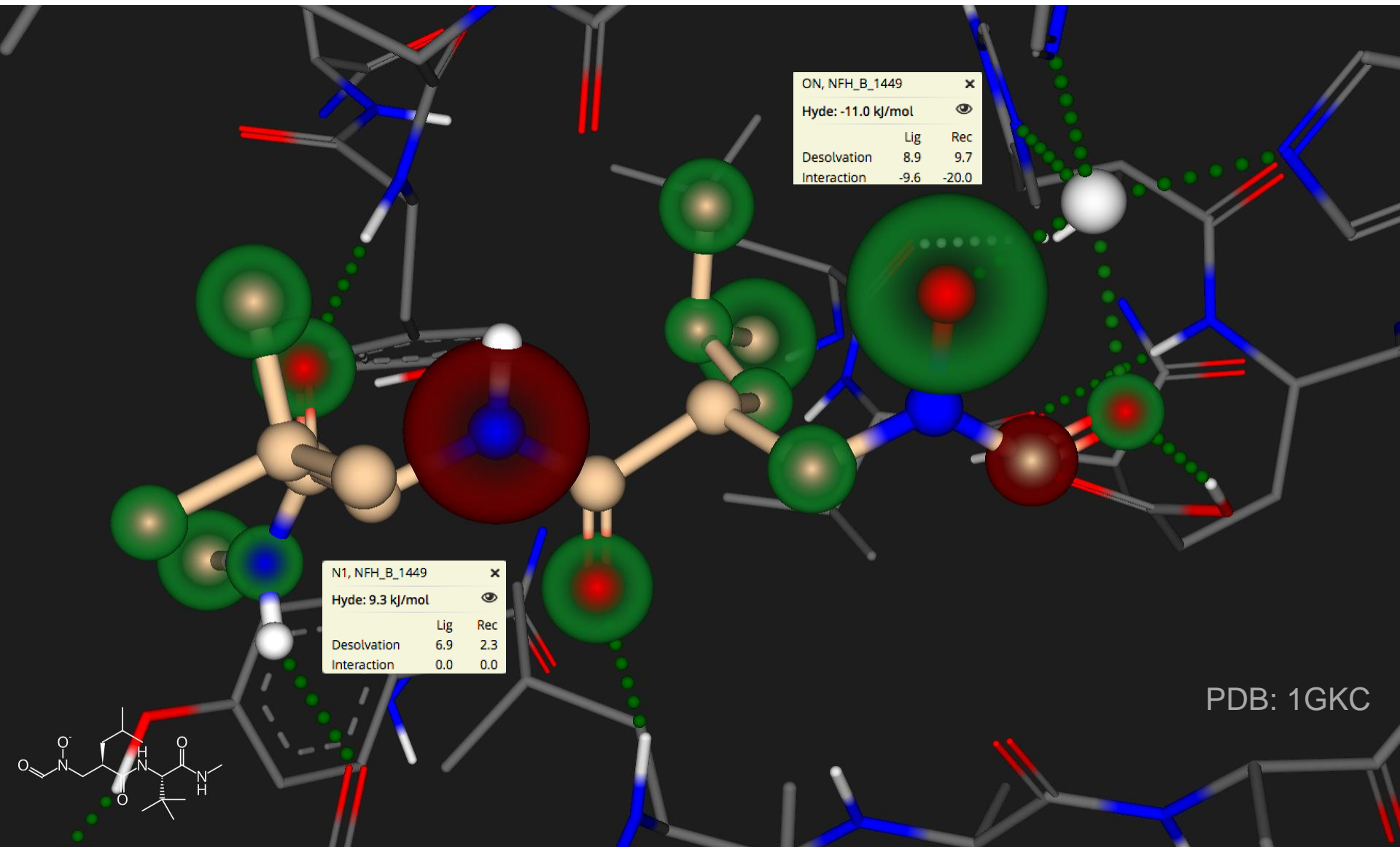


receptor carbonyl oxygen	8.2 kJ/mol
ligand aromatic oxygen	2.4 kJ/mol
total desolvation cost	<u>10.6</u> kJ/mol

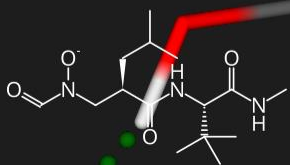
receptor aromatic carbons	-5.2 kJ/mol
ligand aromatic carbon	-2.0 kJ/mol
total desolvation gain	<u>-7.2</u> kJ/mol

receptor amide N dehydrat	6.3 kJ/mol
interaction energy	-7.4 kJ/mol
ligand aromatic N dehydrat	6.4 kJ/mol
interaction energy	-7.5 kJ/mol
total H-bond energy	<u>-2.2</u> kJ/mol

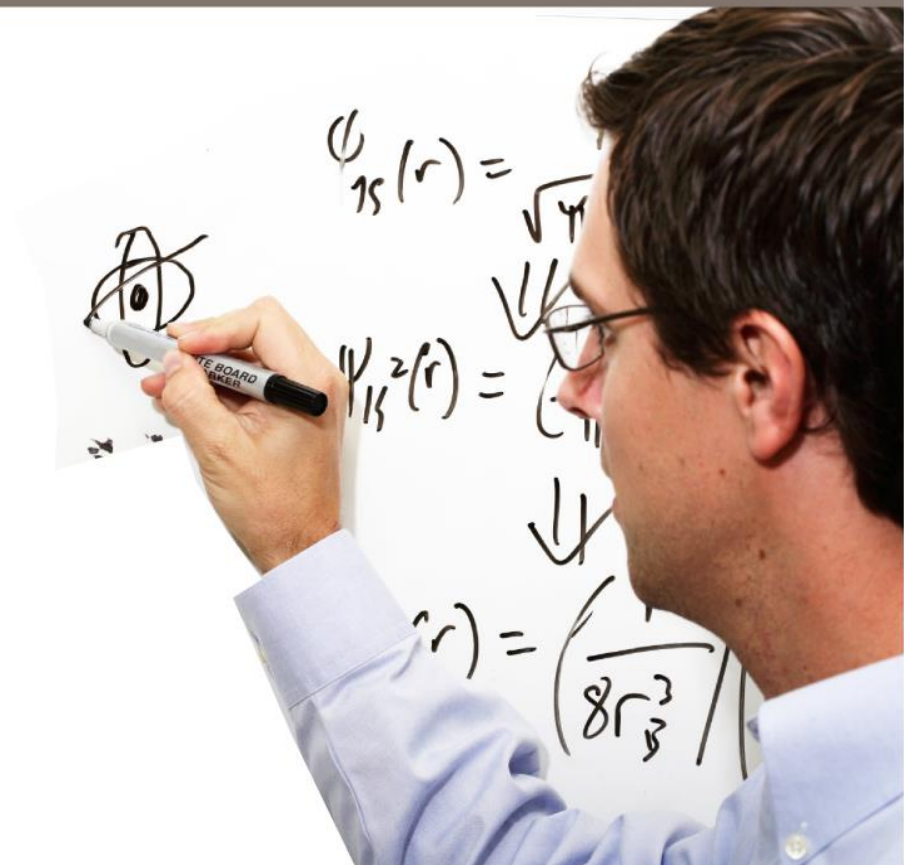
Visual Understanding of 3D Affinity Data



PDB: 1GKC



N-Dimensions



The Objectives of Drug Discovery

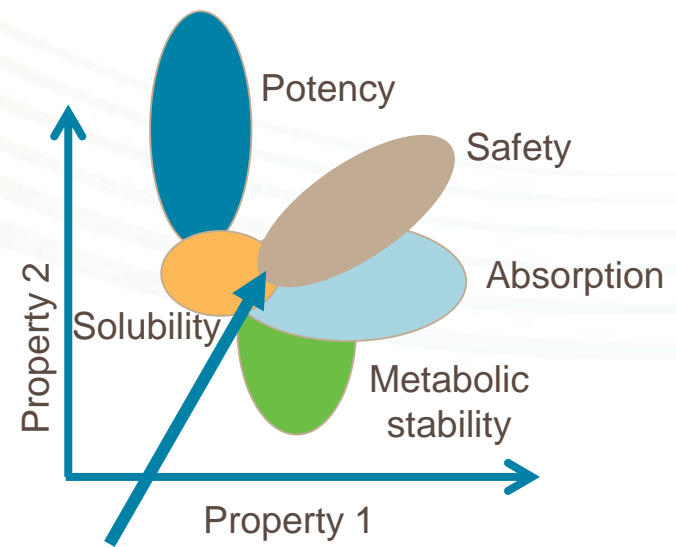
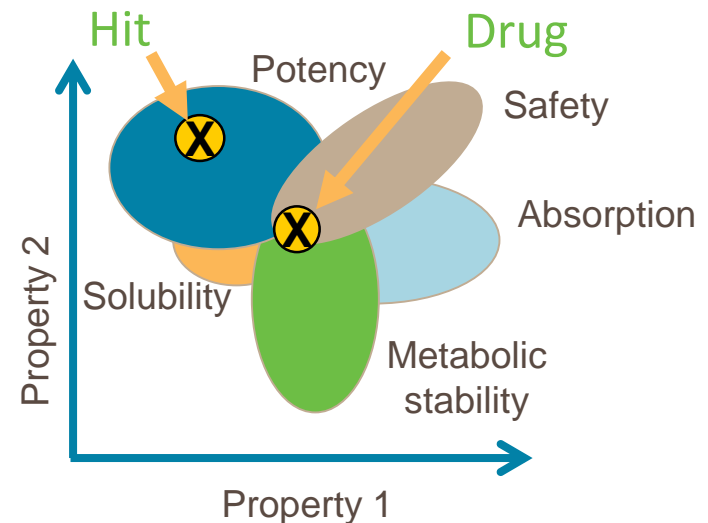
Multi-parameter optimisation

- Identify chemistries with an optimal **balance** of properties

- Quickly identify situations when such a balance is not possible

 - Fail fast, fail cheap

 - Only when **confident**



No good drug

Multi-Parameter Optimisation

Probabilistic Scoring

Integrated assessment of data against project criteria

Accounts for the uncertainties in all compound-related data
(experimental or calculated)

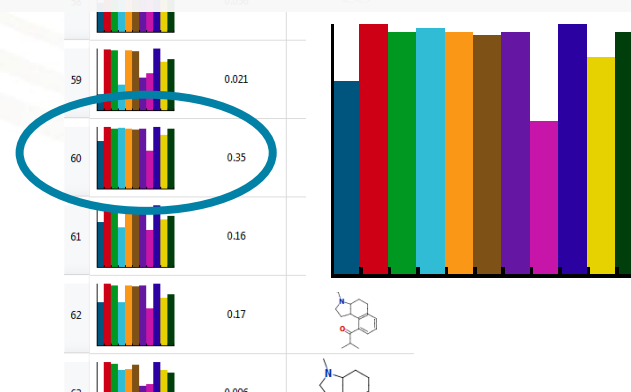
Project specific scoring profile

Profile: SHT1a Scoring Profile *

Property	Desired Value	Importance
pKi SHT1a affinity	8.00 -> inf	High
logS	> 1	High
HIA category	+	High
logP	0.0 -> 3.5	Medium
BBB category	+	Medium
BBB log([brain]:[blood])	-0.20 -> 1.00	Medium
P-gp category	no	Medium
hERG pIC50	≤ 5	Medium
2C9 pKi	≤ 6	Medium
2D6 affinity category	low medium	Low
PPB category	low	Low

Compounds ranked by
likelihood of success

Histograms for quick visual guide to
compound properties



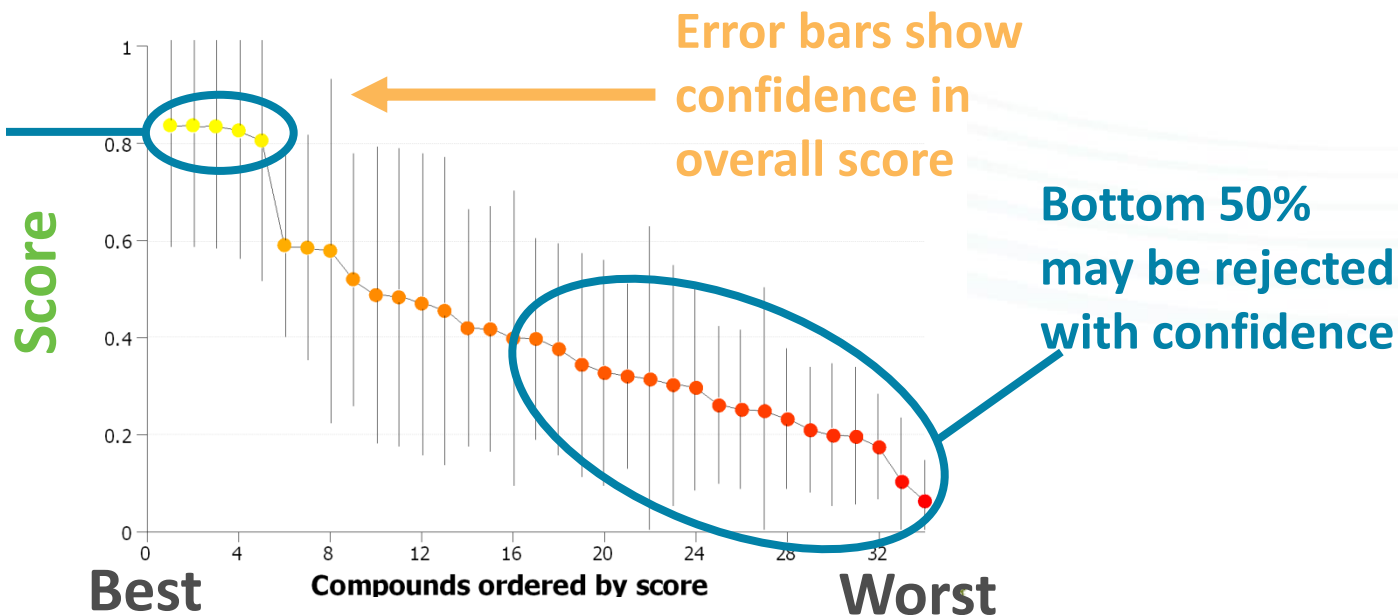
Multi-Parameter Optimisation

Probabilistic Scoring

- **Property data**
 - Experimental or predicted
- **Criteria for success**
 - Relative importance
- **Uncertainties in data**
 - Experimental or statistical

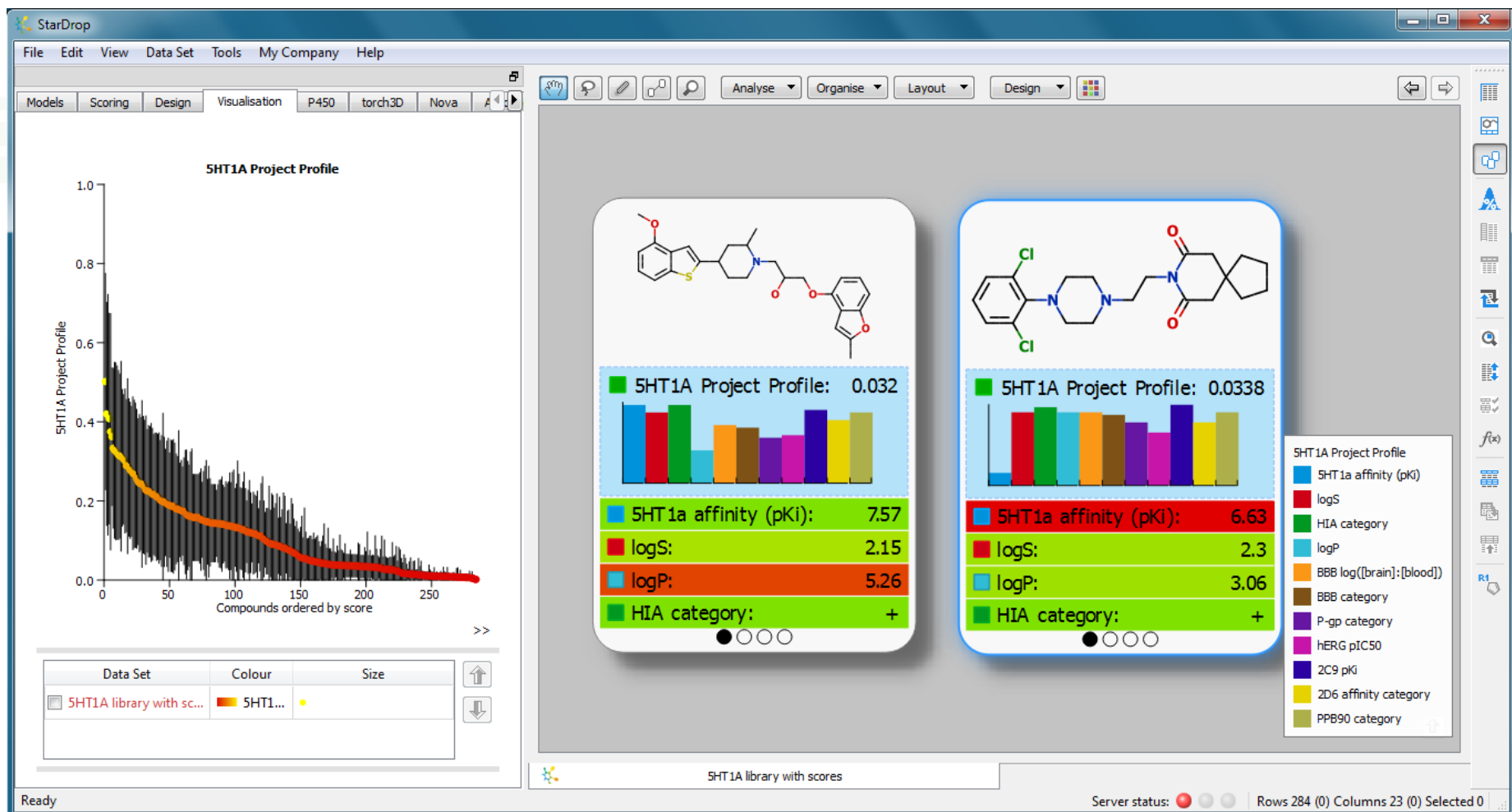
- **Score (Likelihood of Success)**
- **Confidence in score**

Data do not separate these as error bars overlap

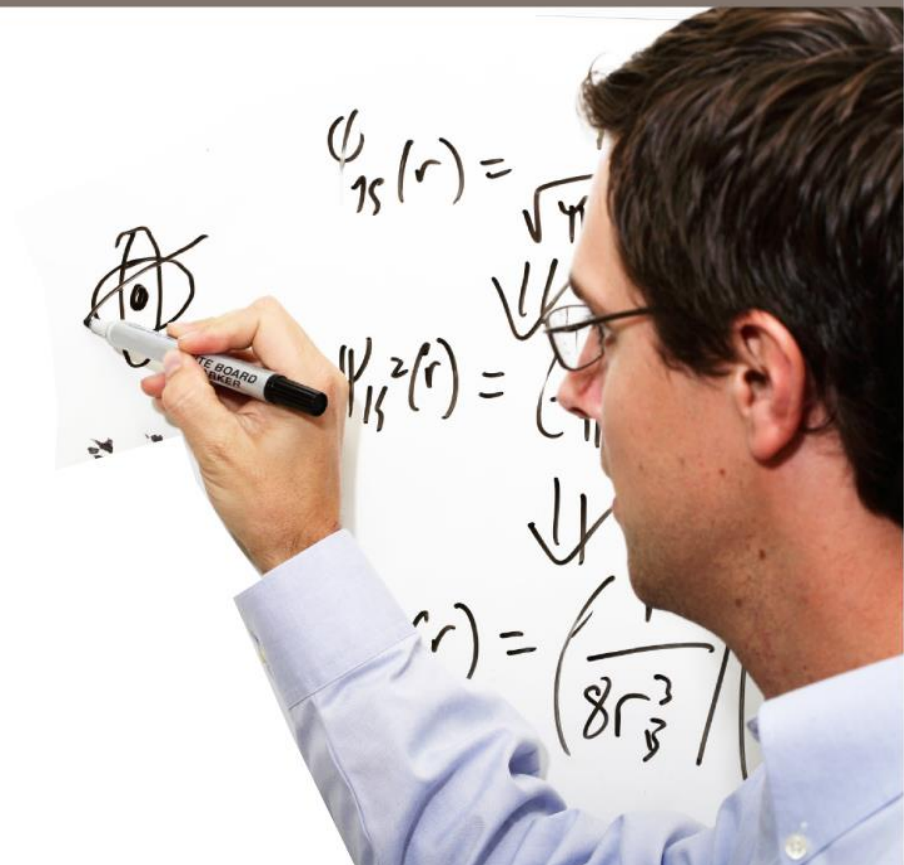


Probabilistic Scoring

Guide redesign to improve chance of success



Linking 2D and 3D SAR to Guide Design



Understanding Activity Cliffs in 3D

PPAR_γ PDB 4EMA

The image displays the StarDrop software interface for a PPAR_γ activity cliff analysis. The left panel shows a 3D molecular model of the protein-ligand complex, with the protein in grey and the ligand in orange. The right panel shows a network diagram where nodes represent ligands and edges represent activity cliffs. The central node is highlighted with a blue border and has a pIC50 of 7.244. Other nodes are connected to it by red arrows, indicating activity cliffs. The pIC50 values for these nodes are: 6.337, 6.319, 6.886, 5.585, 6.886, 6.745, 6.481, 7.119, 6.71, and 6. The bottom status bar shows 'Server status: [green circles] Rows 27 (0) Columns 3 (0) Selected 1'.

StarDrop - PPAR Example

File Edit View Data Set Tools Query Tool Help

Models Scoring Design Visualisation SeeSAR3D P450 torch3D Nova

Protein Display Show Binding Powered by BioSolve

Structure Pose ID

Ready

PPAR_SeeSAR_example

Server status: [green circles] Rows 27 (0) Columns 3 (0) Selected 1

Understanding Activity Cliffs in 3D

PPAR_γ PDB 4EMA

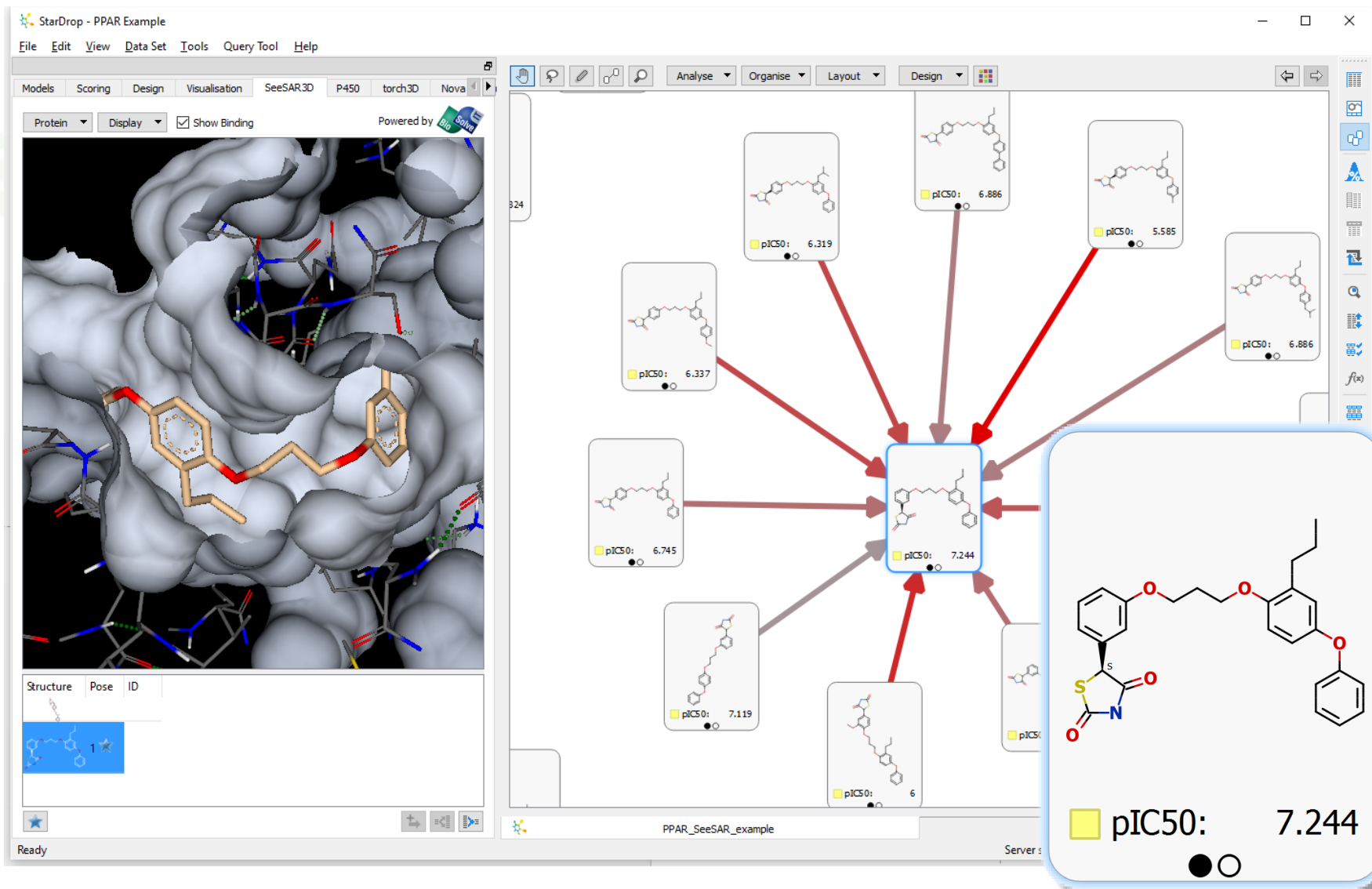
The image displays the StarDrop software interface for a PPAR_γ example. On the left, a 3D molecular model shows a protein surface (grey) with a ligand (orange and red) docked in a binding pocket. The interface includes a menu bar (File, Edit, View, Data Set, Tools, Query Tool, Help) and a toolbar with various visualization and analysis tools. Below the 3D model is a table with columns for Structure, Pose, and ID, showing a single entry with a star icon. On the right, a network diagram visualizes activity cliffs. The central node is highlighted with a blue border and contains a chemical structure with a pIC50 value of 7.244. Eight surrounding nodes, each containing a chemical structure and a pIC50 value, are connected to the central node by red arrows pointing towards it. The pIC50 values for these surrounding nodes are: 6.337, 6.319, 6.886, 5.585, 6.745, 6.481, 7.119, and 6.71. The StarDrop logo and the text 'PPAR_SeeSAR_example' are visible at the bottom of the interface. The status bar at the bottom right shows 'Server status: [green circles] Rows 27 (0) Columns 3 (0) Selected 1'.

Structure	Pose	ID
	1	★

Chemical Structure	pIC50
	6.337
	6.319
	6.886
	5.585
	6.745
	6.481
	7.119
	6.71
	7.244

Understanding Activity Cliffs in 3D

PPAR_γ PDB 4EMA



Understanding Activity Cliffs in 3D

PPAR_γ PDB 4EMA

The image displays the StarDrop software interface for a PPAR_γ docking study. The left panel shows a 3D molecular model of the ligand (highlighted in orange) bound to the protein (grey surface). The right panel shows a 2D activity cliff visualization of the ligand, with various chemical features highlighted in yellow boxes and labeled with their pIC50 values. The central feature is highlighted with a blue border and labeled with a pIC50 of 5.585.

StarDrop - PPAR Example

File Edit View Data Set Tools Query Tool Help

Models Scoring Design Visualisation SeeSAR3D P450 torch3D Nova

Protein Display Show Binding Powered by BioSolve

Structure Pose ID

Ready

PPAR_SeeSAR_example

Server status: ●●●● Rows 27 (0) Columns 3 (0) Selected 1

Chemical Feature	pIC50 Value
Central Feature (Blue Border)	5.585
Top-Right Feature	5.585
Top-Middle Feature	6.886
Top-Left Feature	6.319
Middle-Left Feature	6.337
Bottom-Left Feature	6.745
Bottom-Middle Feature	6.481
Bottom-Right Feature	6.886
Bottom-Far-Right Feature	6.71
Bottom-Far-Left Feature	7.1
Bottom-Far-Middle Feature	6

Understanding Activity Cliffs in 3D

PPAR_γ PDB 4EMA

The screenshot displays the StarDrop software interface for a PPAR_γ example. The left panel shows a 3D molecular model of the protein (grey surface) with a ligand (orange sticks) bound in the pocket. The right panel shows a 2D activity cliff visualization of the ligand space, with the most active molecule (pIC50: 6.319) highlighted in a blue box. The activity cliff shows several other molecules with pIC50 values ranging from 5.585 to 7.119. The interface includes a menu bar (File, Edit, View, Data Set, Tools, Query Tool, Help) and a toolbar with various visualization and analysis tools. The status bar at the bottom indicates 'Ready' and 'Server status: [green circles] Rows 27 (0) Columns 3 (0) Selected 1'.

StarDrop - PPAR Example

File Edit View Data Set Tools Query Tool Help

Models Scoring Design Visualisation SeeSAR3D P450 torch3D Nova

Protein Display Show Binding Powered by BioSolve

Structure Pose ID

Ready

PPAR_SeeSAR_example

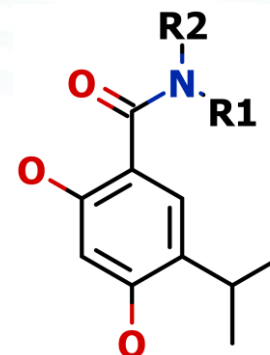
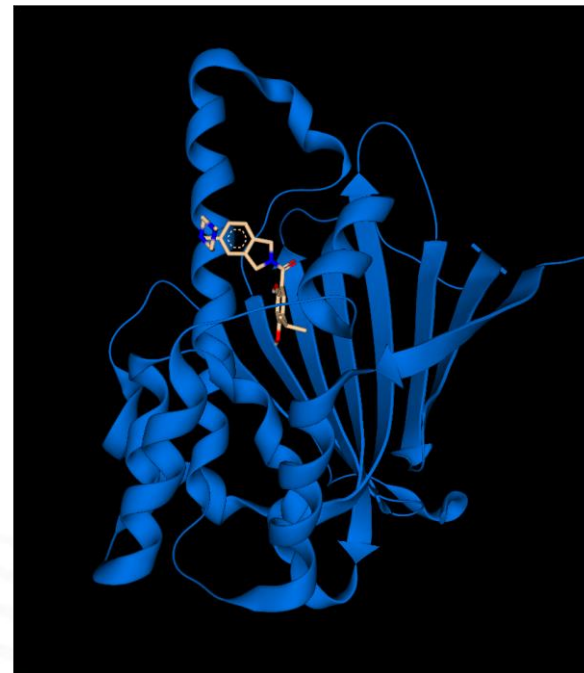
Server status: [green circles] Rows 27 (0) Columns 3 (0) Selected 1

Molecule	pIC50
Highlighted (Blue Box)	6.319
Top Left	6.337
Top Middle	6.319
Top Right	5.585
Right	6.886
Bottom Left	6.745
Bottom Middle	7.119
Bottom Right	6.71
Bottom Far Right	6.886

Exploration of Virtual Screening Results

HSP90

- Crystal structure PDB ref. 2XJX
- Virtual library generated using STORM workflow in KNIME
 - Amide substitution using Schotten Baumann reaction on beta resorcylic core
 - Building blocks from vendor catalogues
 - ‘Tail’ of molecule not contributing to affinity
- Resulting library docked with FlexX
- Scored using SeeSAR and HYDE to estimate pK_i





Matched Molecular Pair Analysis

StarDrop - HSP90 virtual library




File Edit View Data Set Tools Query Tool Help

Models Scoring Design Visualisation SeeSAR3D P450 torch3D Nova

Protein Display Show Binding Powered by 

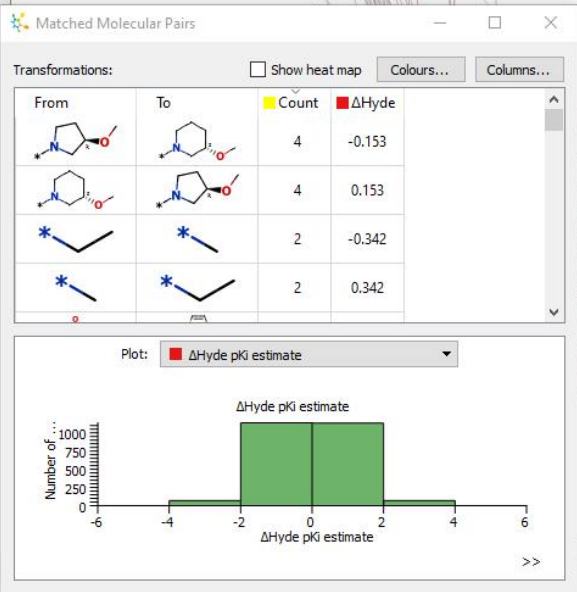
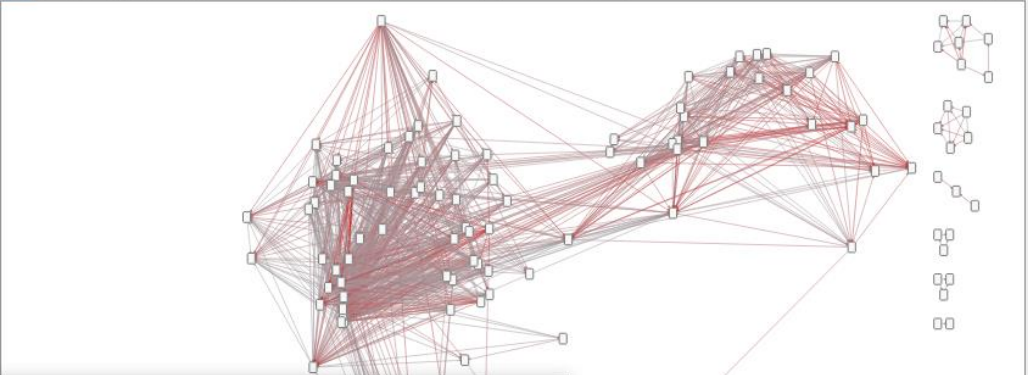


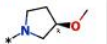
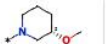
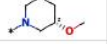
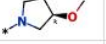




Structure Pose ID

Structure	Pose	ID
		

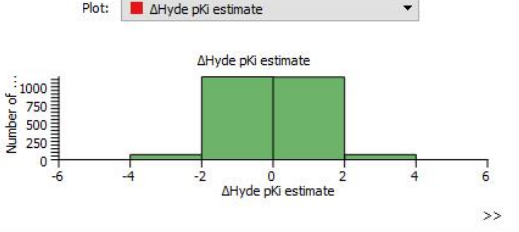
Ready

Analyse Organise Layout Design



From	To	Count	ΔHyde
		4	-0.153
		4	0.153
		2	-0.342
		2	0.342

Plot: ■ ΔHyde pKi estimate



Server status: ● ● ● ● Rows 124 (0) Columns 22 (2) Selected 0

Matched Molecular Pair Analysis

StarDrop - HSP90 virtual library

File Edit View Data Set Tools Query Tool Help

Models Scoring Design Visualisation SeeSAR3D P450 torch3D Nova

Protein Display Show Binding Powered by BioSolve

Row731

CC1CN(C1)C(=O)c2cc(O)c(O)c2

Hyde pKi e...: 8.142

Row208

CC1CN(C1)C(=O)c2cc(O)c(O)c2

Hyde pKi e...: 4.138

Matched Molecular Pairs

Transformations: Show heat map Colours... Columns...

From	To	Count	ΔHyde
		1	4
		1	3.73
		1	3.72
		1	3.67

Plot: ΔHyde pKi estimate

Number of C... vs ΔHyde pKi estimate

MiniMap

Server status: Rows 124 (0) Columns 22 (2) Selected 1

Matched Molecular Pair Analysis

StarDrop - HSP90 virtual library

File Edit View Data Set Tools Query Tool Help

Models Scoring Design Visualisation SeeSAR3D P450 torch3D Nova

Protein Display Show Binding Powered by BioSolve

Row 731

Hyde pKi e...: 8.142

Row 208

Hyde pKi e...: 4.138

Matched Molecular Pairs

Transformations: Show heat map Colours... Columns...

From	To	Count	ΔHyde
		1	4
		1	3.73
		1	3.72
		1	3.67

Plot: ΔHyde pKi estimate

Number of C... vs ΔHyde pKi estimate

MiniMap

Server status: Rows 124 (0) Columns 22 (2) Selected 1

Matched Molecular Pair Analysis

The screenshot displays the StarDrop - HSP90 virtual library interface. The main window shows a 3D molecular model of a protein binding site with a ligand. The interface includes a menu bar (File, Edit, View, Data Set, Tools, Query Tool, Help) and a toolbar with options like Analyse, Organise, Layout, and Design. The left panel shows the protein structure and a list of molecules, with Row 731 selected. The right panel shows the chemical structure of Row 208 and its properties: Hyde pKi e...: 4.138. The Matched Molecular Pairs window is open, showing a table of transformations and a plot of the number of molecules versus the change in Hyde pKi estimate.

Row 731
Hyde pKi e...: 8.142

Row 208
Hyde pKi e...: 4.138

Matched Molecular Pairs

From	To	Count	Δ Hyde
		1	4
		1	3.73
		1	3.72
		1	3.67

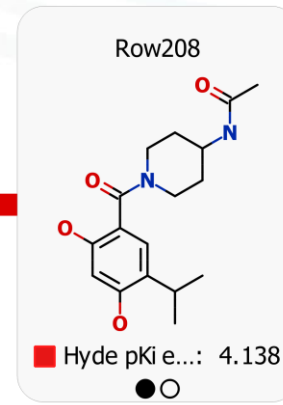
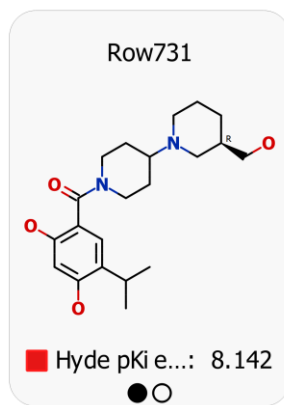
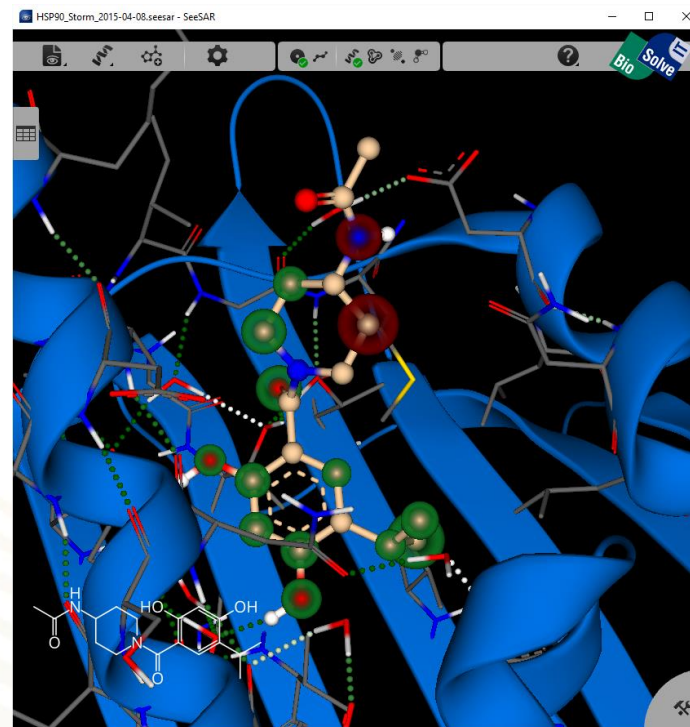
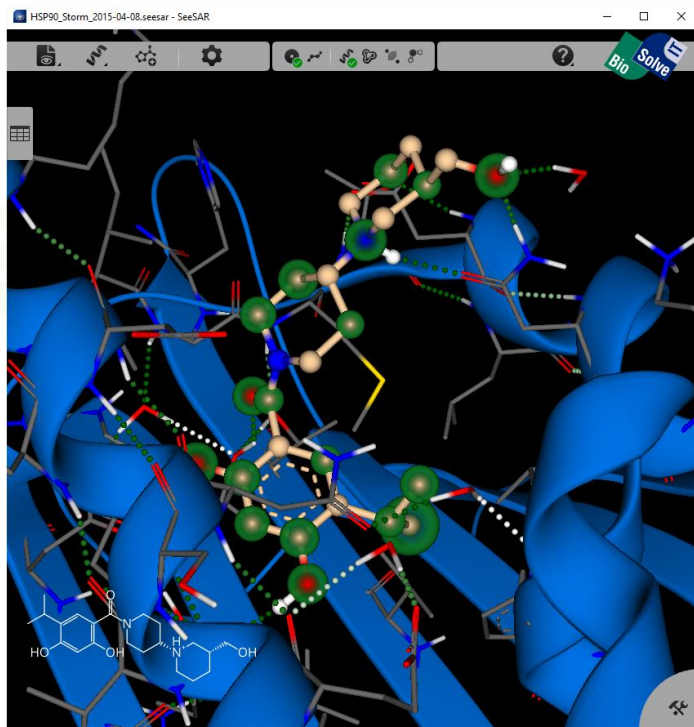
Plot: Δ Hyde pKi estimate

Number of C... vs Δ Hyde pKi estimate

MiniMap

Server status: Rows 124 (0) Columns 22 (2) Selected 1

HYDE Analysis in SeeSAR



Combine with 2D QSAR Predictions

Multi-parameter optimisation

StarDrop - HSP90 virtual library

File Edit View Data Set Tools Query Tool Help

Models Scoring Design Visualisation SeeSAR3D P450 torch3D

Profile: Hyde pKi + Oral Non CNS Scoring Profile

Property	Desired Value	Importance
Hyde pKi estimate	> 7	
logS	> 1	
HIA category	+	
logP	0 -> 3.5	
hERG pIC50	≤ 5	
2D6 affinity category	low medium	
2C9 pKi	≤ 6	
P-gp category	no	
PPB90 category	low	
BBB category	-	
BBB log([brain]:[blood])	≤ -0.5	

Available Properties Criteria Importance

- hERG pIC50
- HIA category
- BBB log([bra...]
- logD
- logP
- logS
- logS @ pH7.4
- MW

Scoring Profiles Location

Oral Non CNS Scoring Profile	File
Oral CNS Scoring Profile	File
Lipinski Rule of Five	File
Intravenous Non CNS Scoring Profile	File
Intravenous CNS Scoring Profile	File
Hyde pKi + Oral Non CNS Scoring ...	Project

MPO Explorer: Build profile... Analyse... Sensitivity...

Ready

	Hyde pKi + Oral Non CN	Structure	ID	Hyde pKi	logS	logP	hERG pIC50	BBB log([brain]:[blood])	2C9 pKi
1	0.4134		Row36	7.52	4.185	1.53	4.55	-0.7597	4.301
2	0.4004		Row316	7.217	3.453	1.177	4.478	-0.6378	4.627
3	0.3941		Row842	7.228	4.528	0.9073	4.524	-0.6858	4.424
4	0.3913		Row76	7.94	4.194	0.956	5.271	-0.6116	4.525
5	0.3868		Row372	7.097	4.163	1.517	4.524	-0.2485	4.503
6	0.3726		Row583	7.041	4.153	1.382	4.376	-0.4906	4.397
7	0.367		Row251	7.374	3.924	1.79	5.058	-0.4092	4.37
8	0.3647		Row67	7.918	4.872	1.655	5.575	-0.4405	4.418
9	0.359		Row421	7.094	4.412	1.537	4.857	-0.5003	4.146
10	0.3527		Row264	7.155	3.935	1.955	4.683	-0.4545	4.33

HSP90 ligands

Server status: Rows 124 (0) Columns 22 (2) Selected 0

3D View... Optimisation opportunities

The screenshot displays the StarDrop - HSP90 virtual library interface. On the left, a 3D molecular model shows a protein surface (grey) with a ligand (orange and blue) docked in a binding pocket. The interface includes a menu bar (File, Edit, View, Data Set, Tools, Query Tool, Help) and a toolbar with various interaction tools. Below the 3D view, a table lists the structure, pose, and ID of the ligands.

Structure	Pose	ID
		Row3

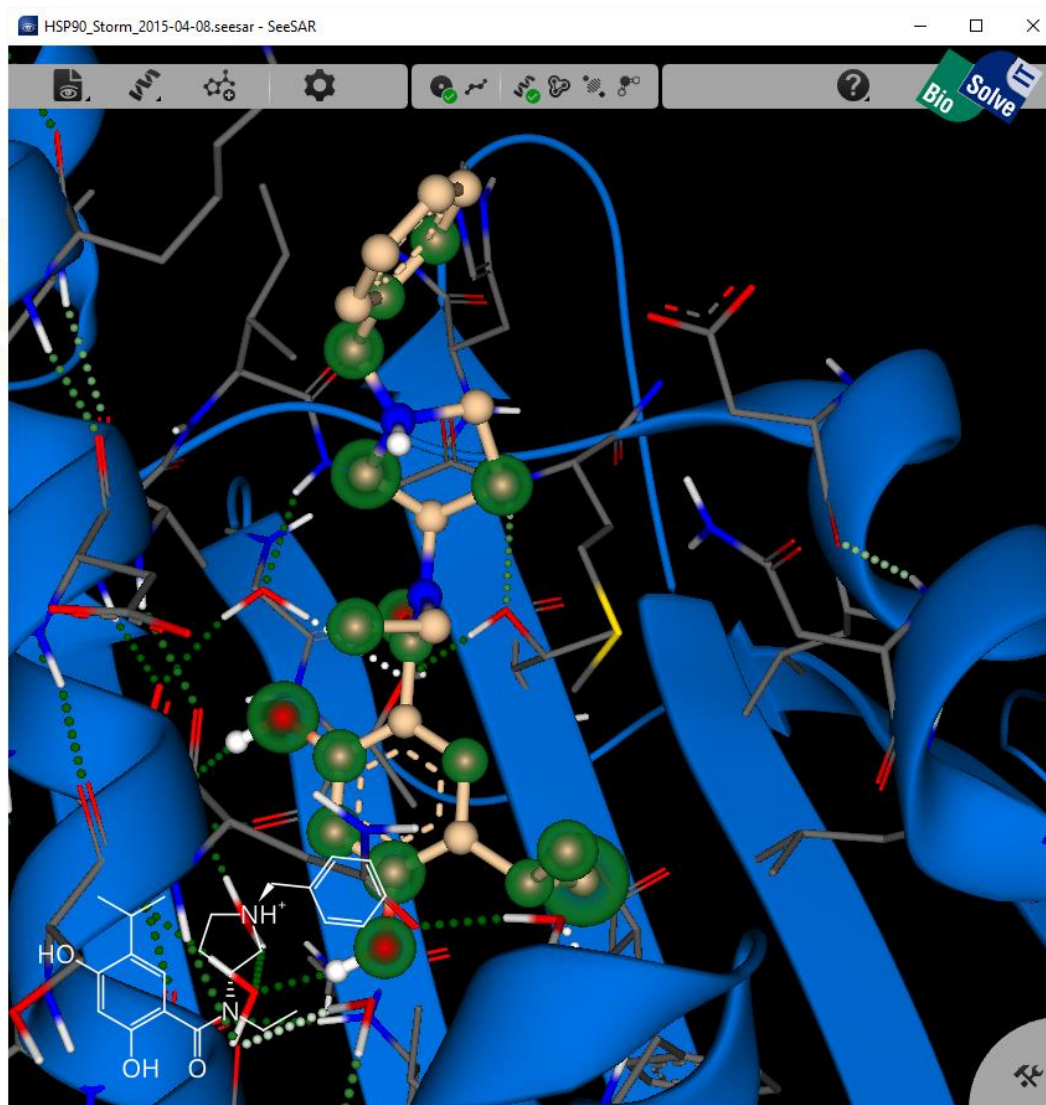
The right panel provides a detailed view of the selected ligand, Row3. It features a 3D molecular model with a color-coded electrostatic potential map (red for negative, blue for positive, green for neutral). Below the model, a bar chart compares the ligand's properties to a reference set. The properties are:

- Hyde pKi + Oral Non C...: 0.139
- Hyde pKi estimate: 8.011
- hER...: 5.862
- logP: 3.705

The bar chart shows the relative values of these properties for the selected ligand (purple bar) compared to other ligands in the set. The Hyde pKi estimate (8.011) is significantly higher than the reference value (0.139), while the hER... (5.862) and logP (3.705) are also higher than their respective reference values.

StarDrop - HSP90 virtual library
File Edit View Data Set Tools Query Tool Help
Models Scoring Design Visualisation SeeSAR3D P450 torch3D
Protein Display Show Binding Powered by BioSolve
Ready
HSP90 ligands
Server status: [green] [green] [green] [grey] Rows 124 (0) Columns 22 (2) Selected 1

HYDE Analysis in SeeSAR



Optimisation Idea

Modify pKa of Nitrogen by amide substitution

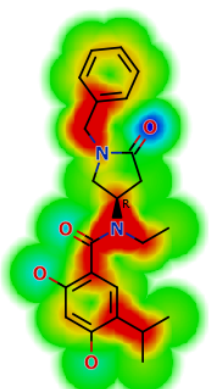
StarDrop - HSP90 virtual library

File Edit View Data Set Tools Query Tool Help

Models Scoring Design Visualisation SeeSAR3D P450 torch3D

Hand Rotate Translate Zoom Analyse Organise Layout Design

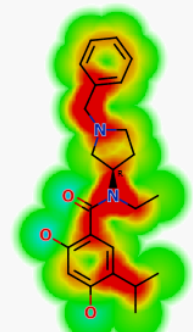
Clean Reset



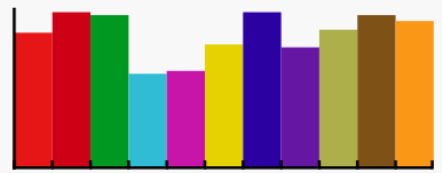
Glowing Molecule: - +

	Results
Hyde pKi + Oral Non CNS Scoring Profile	?
logS @ pH7.4	2.46
logD	3.324
logS	2.46
logP	3.324
2C9 pKi	5.531
hERG pIC50	4.971
BBB log([brain]:[blood])	-0.8076
BBB category	-

Row3



Hyde pKi + Oral Non C...: 0.139



Hyde pKi estimate: 8.011

hERG...: 5.862 logP: 3.705

HSP90 ligands

Server status: Rows 124 (0) Columns 22 (2) Selected 1

Optimisation Idea

Add polar group to phenyl ring

StarDrop - HSP90 virtual library

File Edit View Data Set Tools Query Tool Help

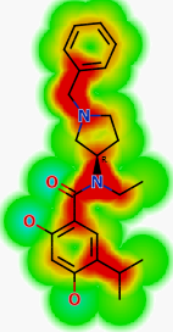
Models Scoring Design Visualisation SeeSAR3D P450 torch3D

Clean Reset

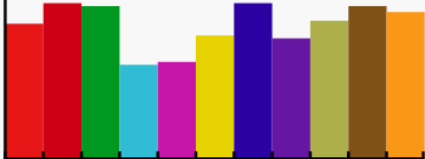
Glowing Molecule: - +

	Results
Hyde pKi + Oral Non CNS Scoring Profile	?
logS @ pH7.4	1.852
logD	-0.1161
logS	2.864
logP	3.263
2C9 pKi	5.321
hERG pIC50	4.601
BBB log([brain]:[blood])	-0.9159
BBB category	-

Row3



Hyde pKi + Oral Non C...: 0.139



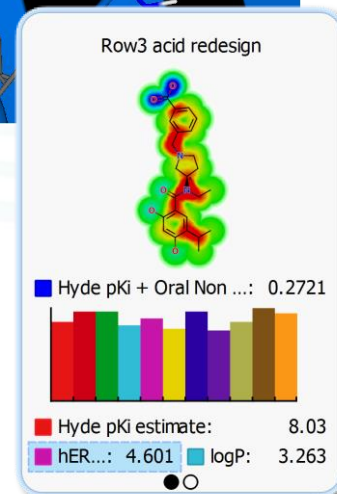
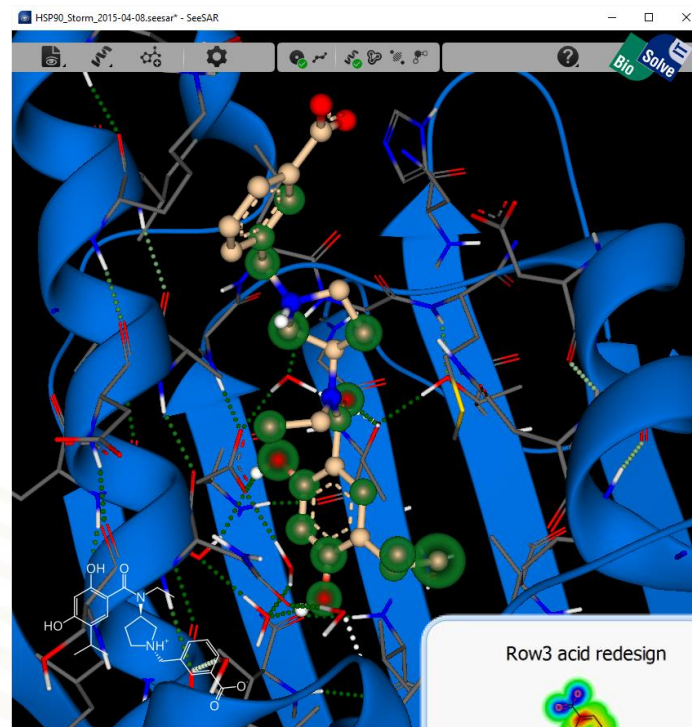
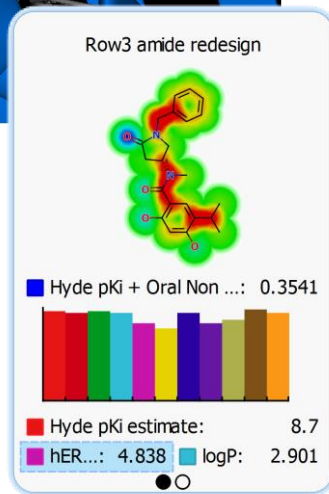
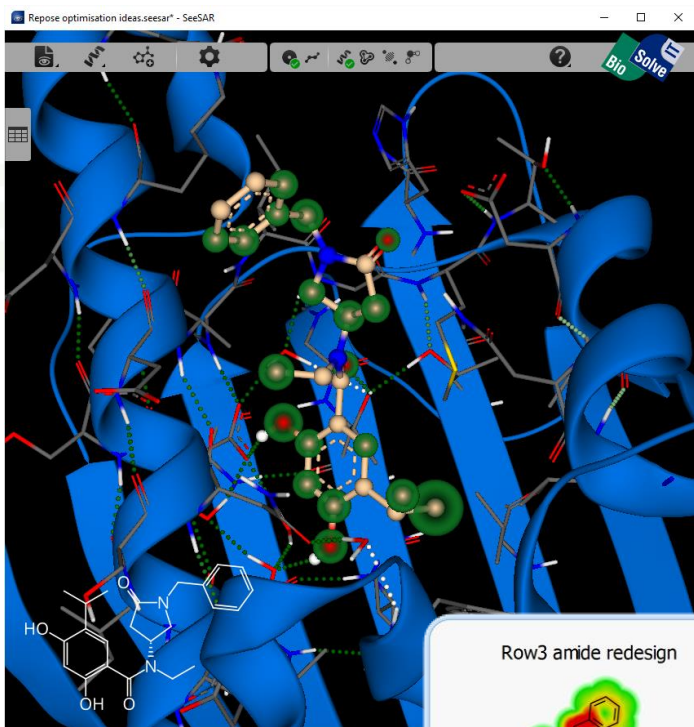
Hyde pKi estimate: 8.011

hERG...: 5.862 logP: 3.705

HSP90 ligands

Server status: [green] [green] [green] [green] Rows 124 (0) Columns 22 (2) Selected 0

Repose in SeeSAR



Conclusions

- Both 2D and 3D information are important to interpret SAR and guide design
- A seamless combination between these two views of the chemical world maximises the benefits that they bring
 - Understanding SAR from experimental data
 - Analysis of virtual screening/docking results
- Multi-parameter optimisation of potency, physicochemical and ADMET properties
 - Quickly target high quality compounds



Acknowledgements

- BioSolve IT

- Christian Lemmen
- Markus Gastreich
- Carsten Detering

- Optibrium

- Peter Hunt
- James Chisholm
- Ed Champness
- Chris Leeding
- Nick Foster
- Tamsin Mansley
- Alex Elliot
- Sam Dowling
- Fayzan Ahmed
- Many more....

