



Closing the loop between synthesis and design: Helping chemists to use all the information in compound optimization

COMP 24: ACS National Meeting, San Diego, CA - 13th March 2016

Tamsin Mansley, Edmund Champness, Peter Hunt, James Chisholm,
Chris Leeding, Alex Elliot, Sam Dowling, Fayzan Ahmed & Matthew Segall

Overview

- Visualising and understanding SAR
 - Card View™
 - Linking 2D and 3D SAR
 - o Matched molecular pairs
 - o Activity neighbourhood
- Design of new compound ideas with an improved balance of properties
 - Multi-Parameter Optimisation (MPO)
 - Glowing Molecule™
 - Medicinal chemistry idea generation
- Conclusions

The Challenges

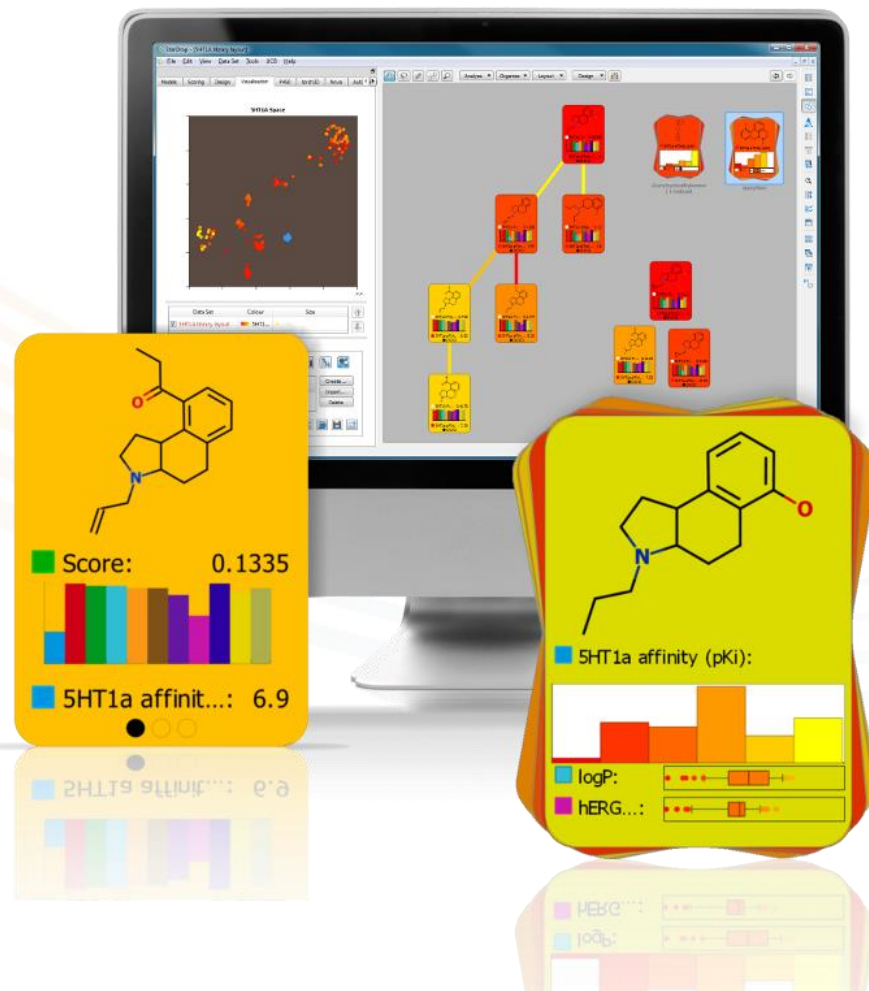
- More compounds
 - Synthesised in-house or at CRO
 - Purchased
 - *In silico* libraries
- More data and data types to understand
- Shorter project timelines
- Need to make intelligent decisions that guide selection and design of compounds with a balance of properties
 - Multi-Parameter Optimisation (MPO)

Visualising and Understanding SAR



Card View Concepts

- Freedom from the constraints of 'chemical spreadsheets'
 - Represent compound relationships
- Cards
 - Show relevant compound data
 - Complete freedom to move
- Stacks
 - Group compounds
 - Summarise and compare data
- Links
 - Highlight compound relationships
- Intuitive visualisation of SAR
 - Clustering, activity landscapes, matched molecular pairs...



New SeeSAR 'Viewer' Module

Visualise 3D Structure Information

Import docking results or crystal structures, from SD/PDB files

Include docking scores or affinity predictions in MPO

Select between multiple poses

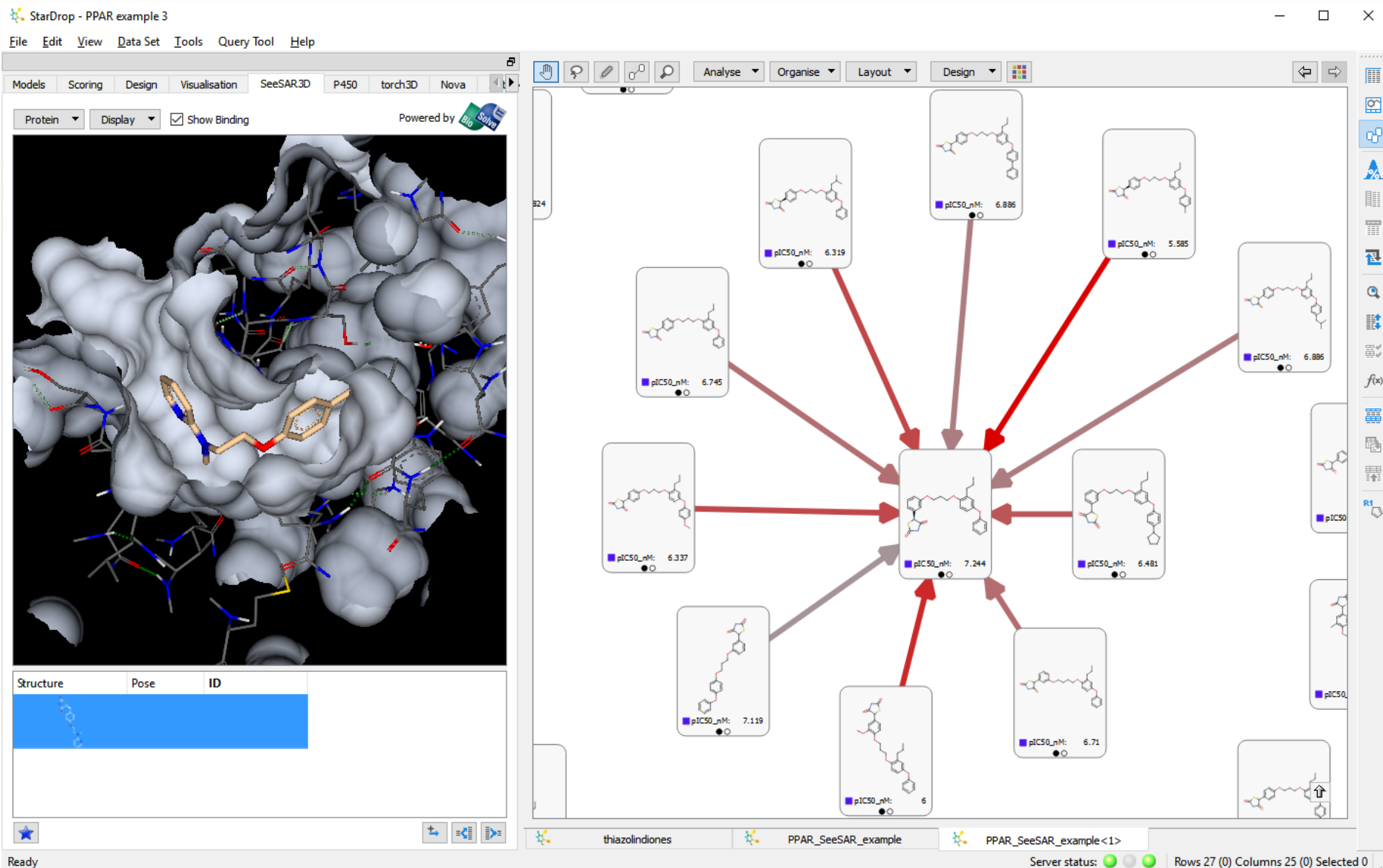
The screenshot displays the SeeSAR Viewer interface. On the left, a 3D molecular structure is shown in a black box. Below it, a 'Structure' tab is selected, and a list of poses is visible. On the right, a table displays docking results for 10 rows. The table has columns for 'Hyde pKi + Oral Non CN', 'Structure', 'ID', 'Hyde pKi estimate', 'logP', and 'logS'. The 'logP' column is highlighted in orange for Row 2. The 'logS' column is highlighted in green for Row 2. The 'logP' column is highlighted in green for Row 2. The 'logS' column is highlighted in green for Row 2.

	Hyde pKi + Oral Non CN	Structure	ID	Hyde pKi estimate	logP	logS
1	0.2905		Row731	8.142	3.837	2.126
2	0.139		Row3	8.011	3.078	3.705
3	0.3913		Row76	7.94	4.194	0.956
4	0.2637		Row6	7.928	4.501	1.976
5	0.3647		Row67	7.918	4.872	1.655
6	0.2537		Row137	7.676	3.286	2.559
7	0.1824		Row462	7.673	3.336	3.28
8	0.2824		Row139	7.671	2.141	1.959
9	0.2337		Row45	7.536	3.207	2.652
10	0.2337		Row46	7.536	3.207	2.652

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
Ready
HSP90 ligands
Server status: | Rows 124 (0) Columns 22 (2) Selected 1

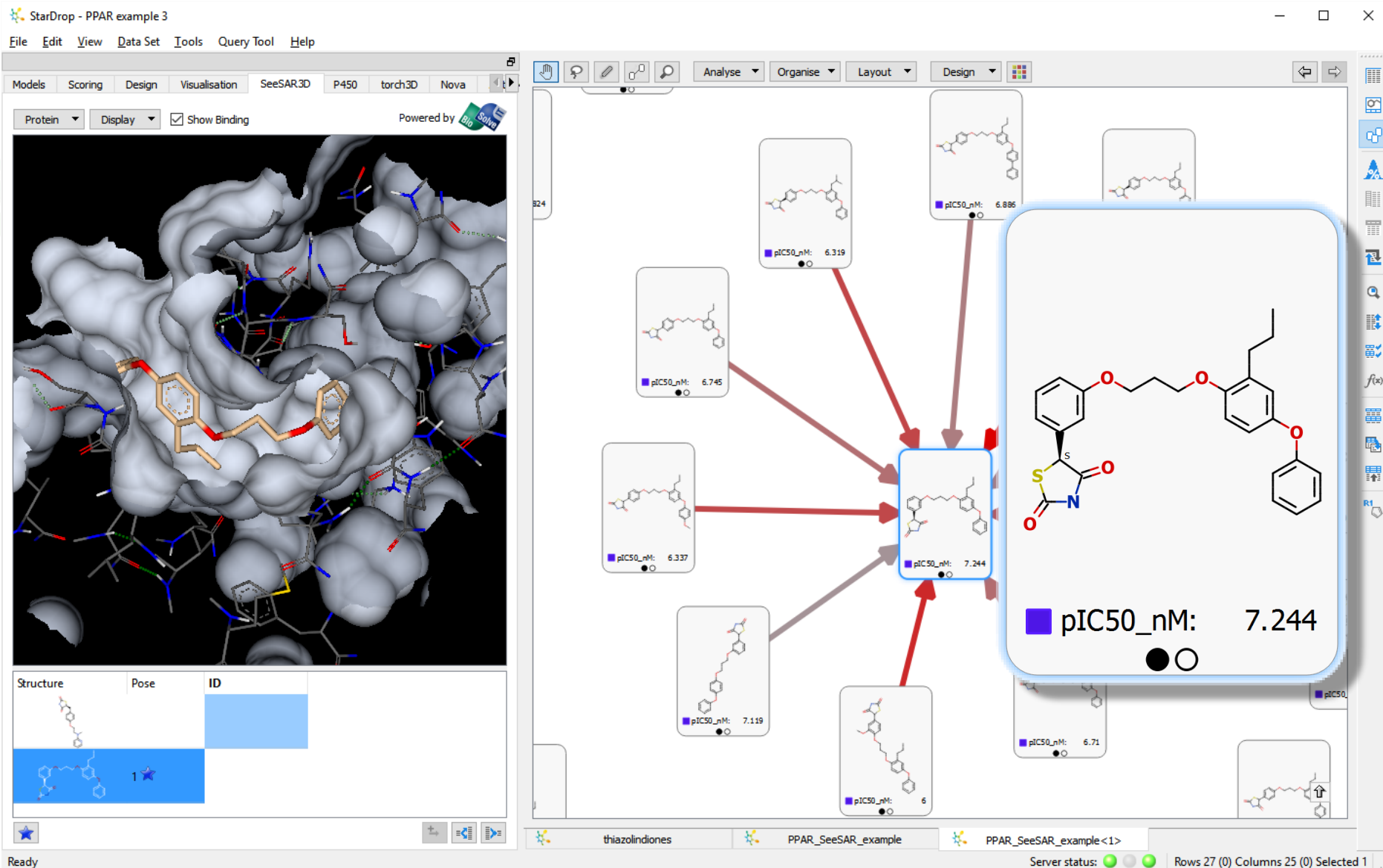
Linking 2D and 3D SAR

Understanding Activity Cliffs in 3D



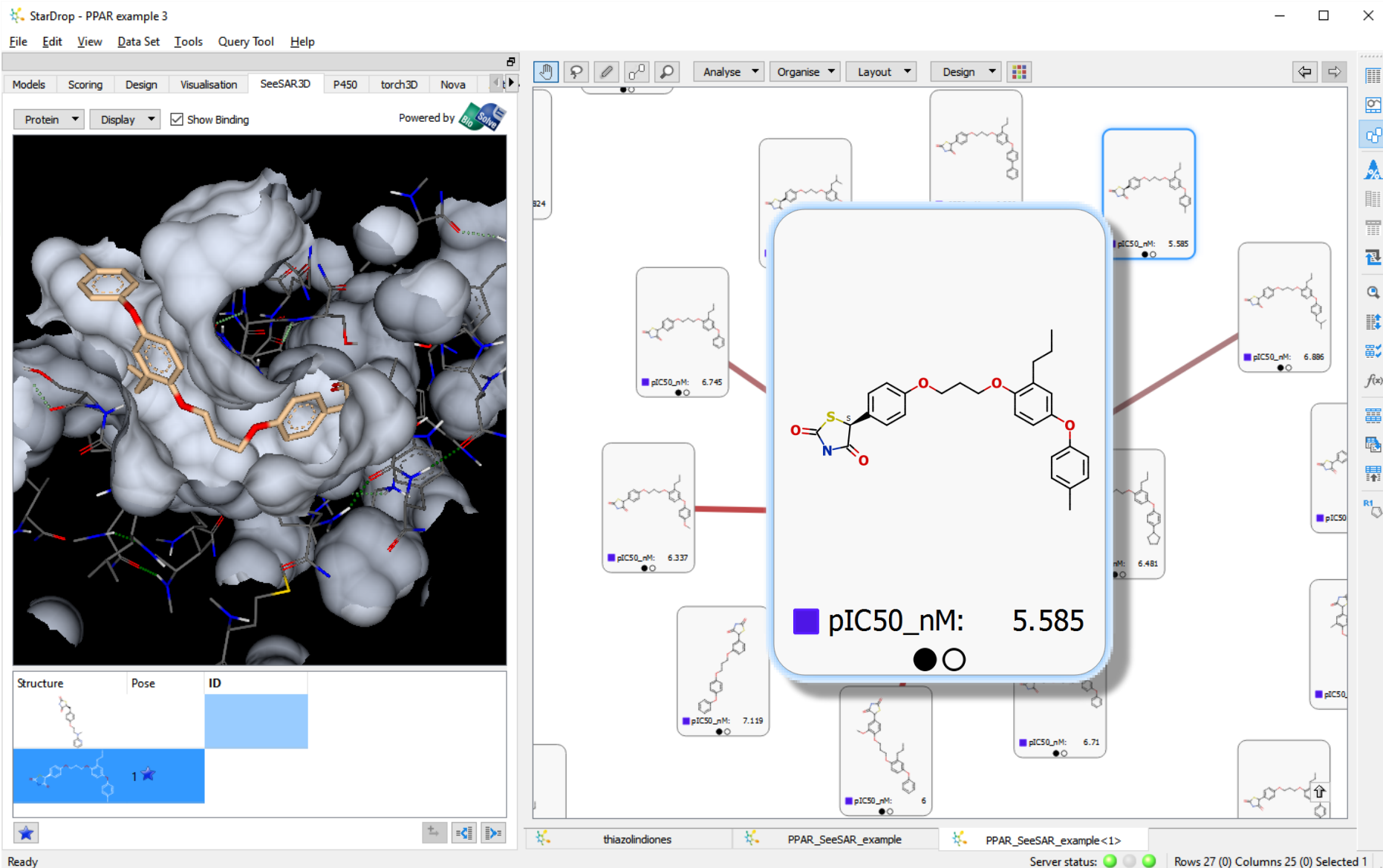
Linking 2D and 3D SAR

Understanding Activity Cliffs in 3D



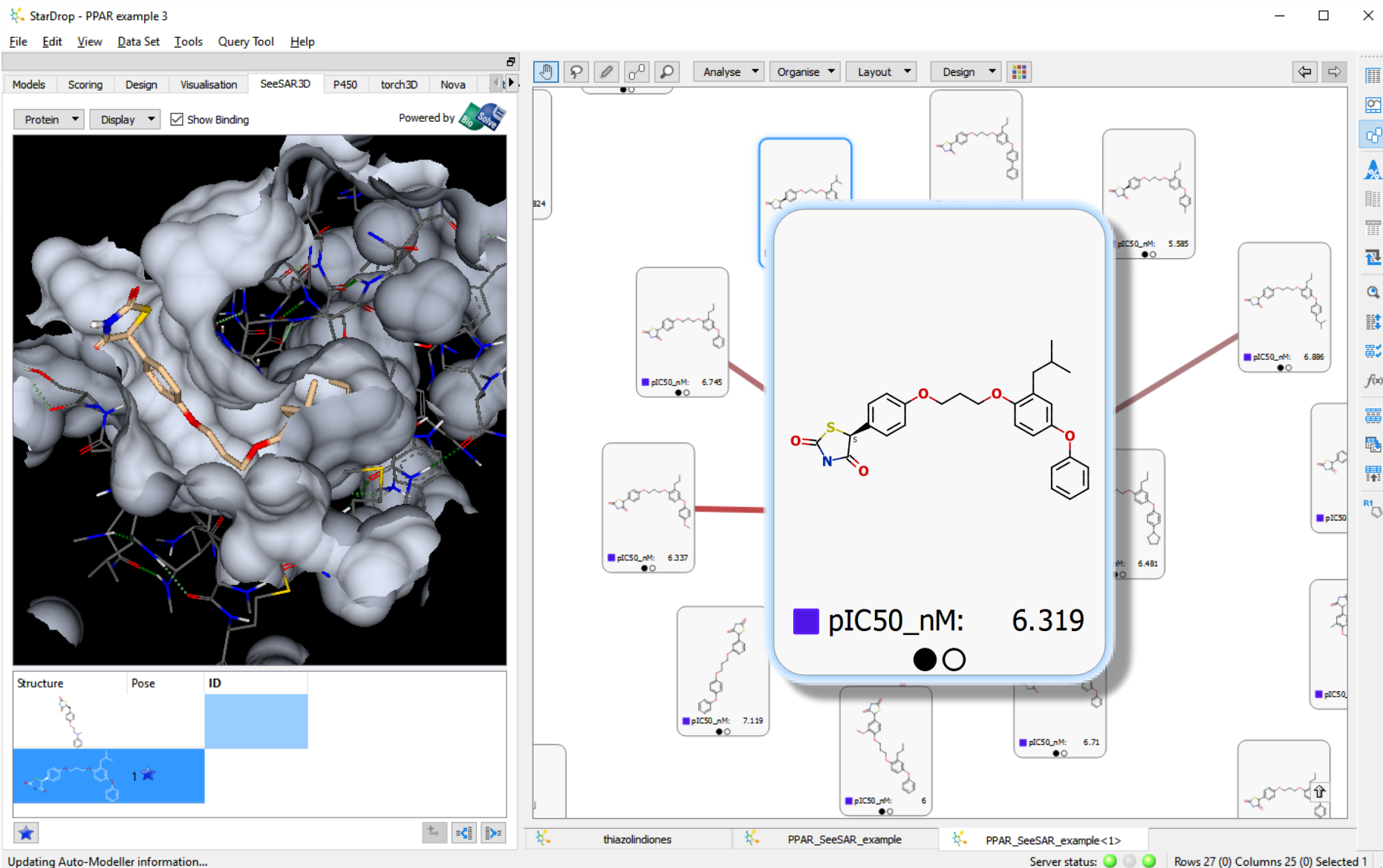
Linking 2D and 3D SAR

Understanding Activity Cliffs in 3D



Linking 2D and 3D SAR

Understanding Activity Cliffs in 3D



But What About Designing New Compounds?

With an Improved Balance of Properties



The Objectives

Multi-Parameter Optimisation

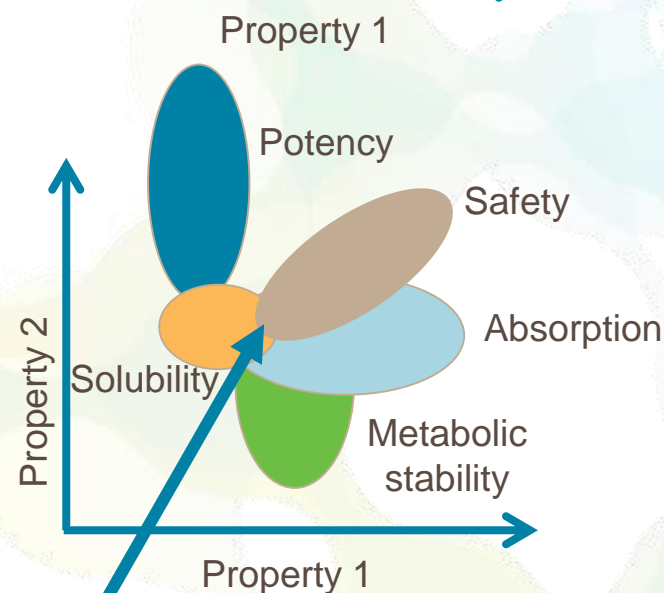
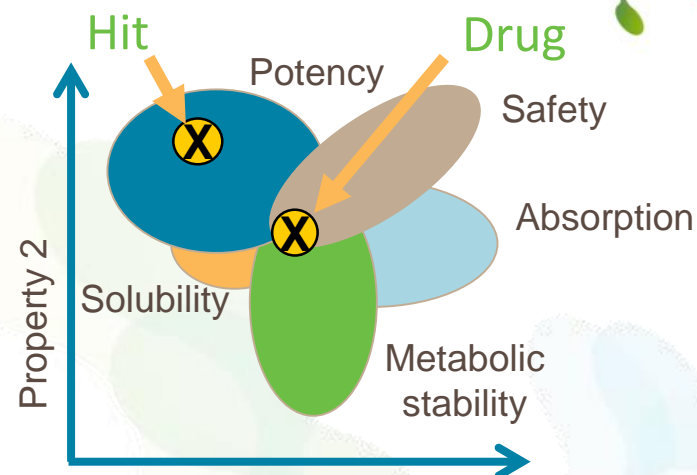


- Identify chemistries with an optimal **balance** of properties

- Quickly identify situations when such a balance is not possible

- Fail fast, fail cheap

- Only when **confident**



No good drug

StarDrop Prioritisation

Probabilistic Scoring



Integrated assessment of data against project criteria

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

User-defined scoring profile

Profile: SHT1a Scoring Profile *

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

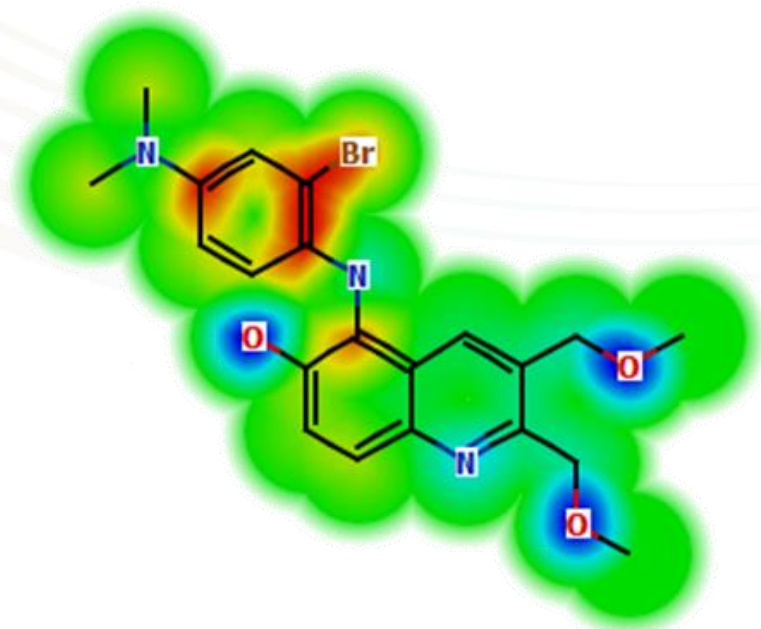
Compounds ranked by likelihood of success

Histograms for quick visual guide to compound properties



The Glowing Molecule

- Models provide estimates of compound's properties (if within the chemical space)
- However, StarDrop models also gives a visual indication of structural influences on predicted properties
 - “Why is a property value predicted?”
 - “Where can I change this property?”
- Glowing Molecule:
 - Can be applied to:
 - StarDrop models
 - Scores
 - Auto-Modeller models



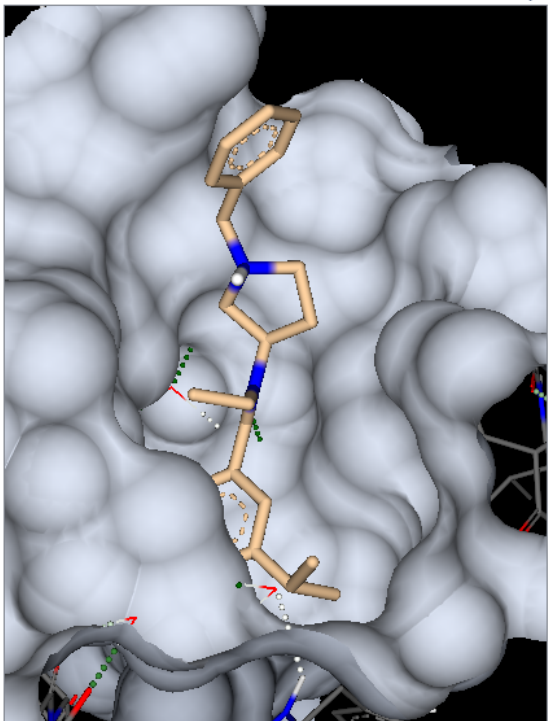
3D View... Optimisation Opportunities

StarDrop - HSP90 project 2-3-16

File Edit View Data Set Tools Query Tool Help

Models Scoring Design Visualisation SeeSAR3D P450 torch

Protein Display ☒ Show Binding Powered by BioSolve



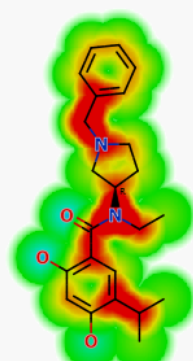
Structure Pose ID

1 ★ Row3

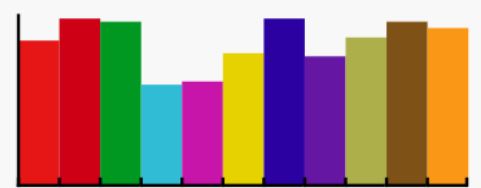
Ready

Analyse Organise Layout Design

Row3



☐ Hyde pKi + Oral Non C...: 0.139



Property	Value
Hyde pKi estimate	8.011
hER...	5.862
logP	3.705

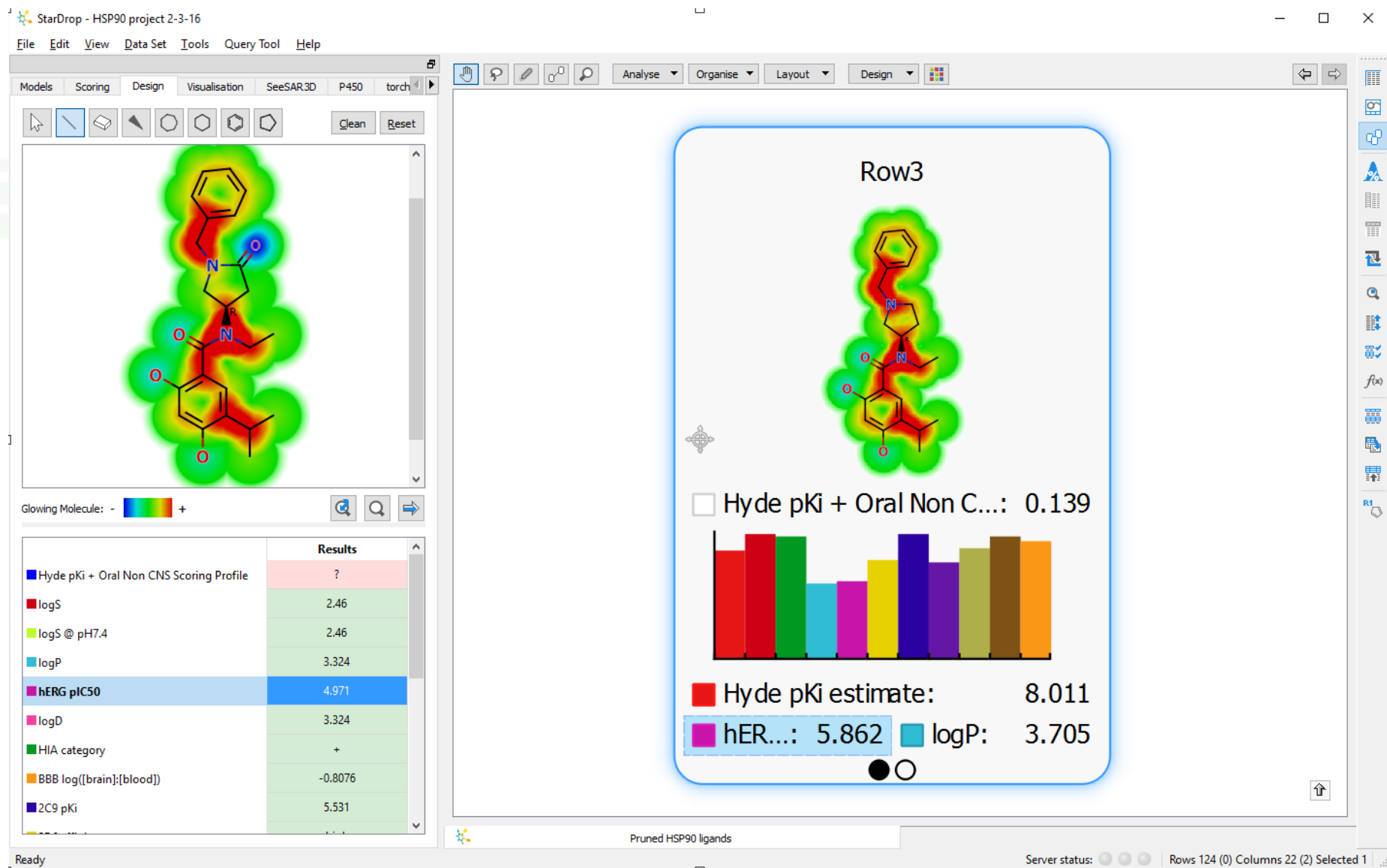
● ○

Pruned HSP90 ligands

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

Optimisation Idea

Modify pKa of Nitrogen by Amide Substitution



Optimisation Idea

Add Polar Group to Phenyl Ring

StarDrop - HSP90 project 2-3-16

File Edit View Data Set Tools Query Tool Help

Models Scoring Design Visualisation SeeSAR3D P450 torch

Glowing Molecule: - +

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

Row3

Hyde pKi + Oral Non C...: 0.139

Hyde pKi estimate: 8.011

hERG...: 5.862 logP: 3.705

Pruned HSP90 ligands

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

Automatic Generation of Compound Ideas

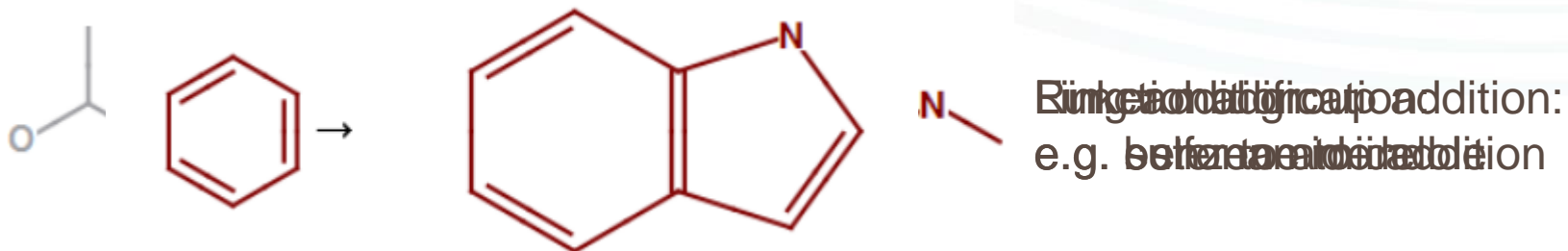
Objectives

- Traditionally have a scarcity of data
 - Time consuming and expensive to generate data
 - Easy to think of more ideas than can be synthesised and tested
 - Manually analyse all of the available data to select compounds
- With predictive models and MPO, it is easy to evaluate very large numbers of ideas
 - More ideas than one scientist can design and draw
- Generate new ideas to stimulate exploration of chemistry
 - *In silico* analysis helps to prioritise ideas for detailed consideration
- But... compound ideas must be **relevant and accessible**
 - First generation of *de novo* design methods tended to generate poor quality, non-synthesizable structures

Generating Compound Ideas

Applying Med. Chem. 'Transformation Rules'

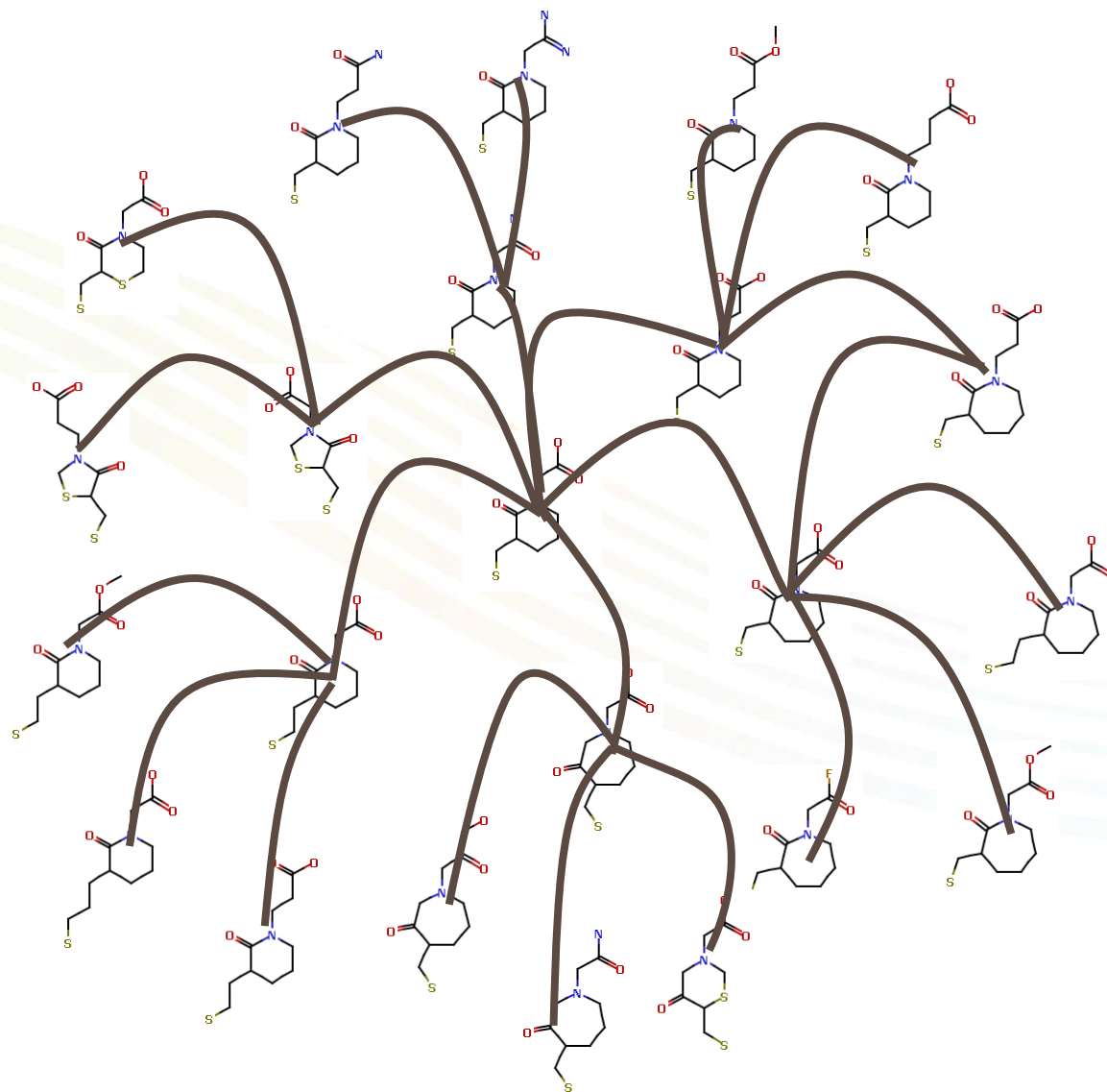
- Compounds generated must 'make sense' from a medicinal chemistry perspective
- Apply 'transformation rules', derived from medicinal chemistry experience, to initial compound(s)*
 - Library of >200 transformations, generate ~180 new compounds per input
 - >94% of structures generated acceptable to med. chemists
 - Not only functional group replacement but also framework transformations



*Stewart *et al.* Bioorg. Med. Chem. (2006) **14** p. 7011

*Segall *et al.* J. Chem. Inf. Model. (2011) **51** pp. 2967-2976

Exponential Growth!



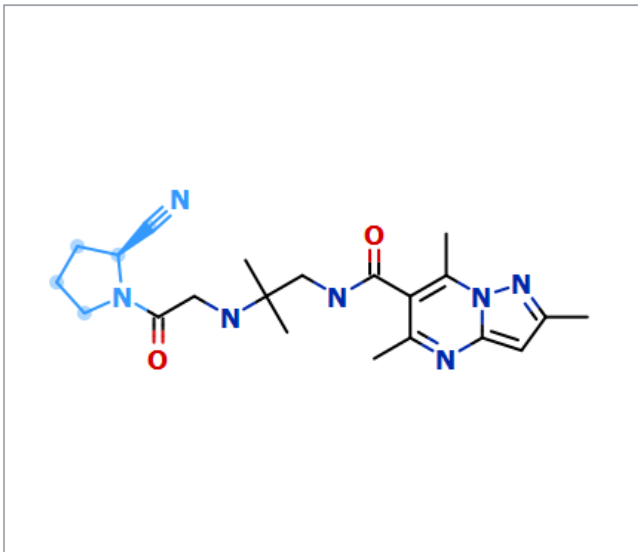
Controlling the Process

Nova Setup Wizard

Specify Input Structure

Lasso a portion of the molecule to mask it from any transformations

☐ Strict masking



< Back Next > Finish Cancel

Nova Setup Wizard

Control Output

Generations

☒ Select compounds at each generation

Method

☒ Biased Diverse Value

☐ Random

Select compounds with

Selection Criteria

☒ The best compounds

☐ The best % of compounds

☐ Compounds with values higher than

☒ Attempt all transformations after generation 1

☒ Limit atom count change Maximum:

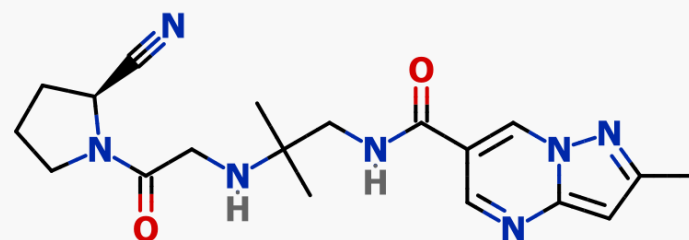
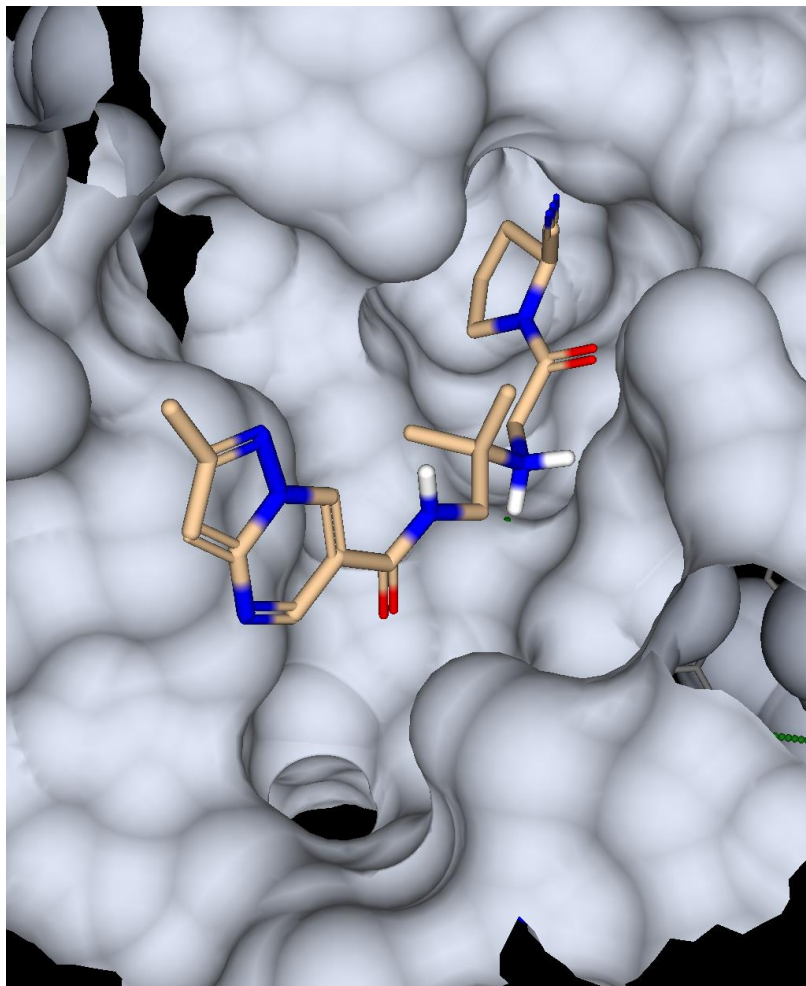
☒ Show results in Card View

< Back Next > Finish Cancel

- Masked substructure is retained unchanged
- Apply multiple generations of transformations
- Bias selection in favour of property, score or diversity

3WQH

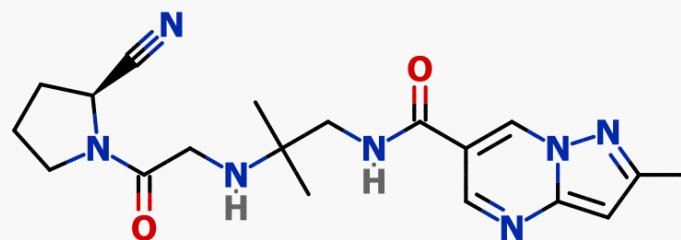
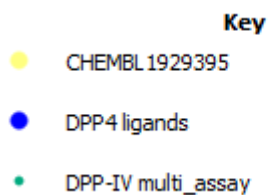
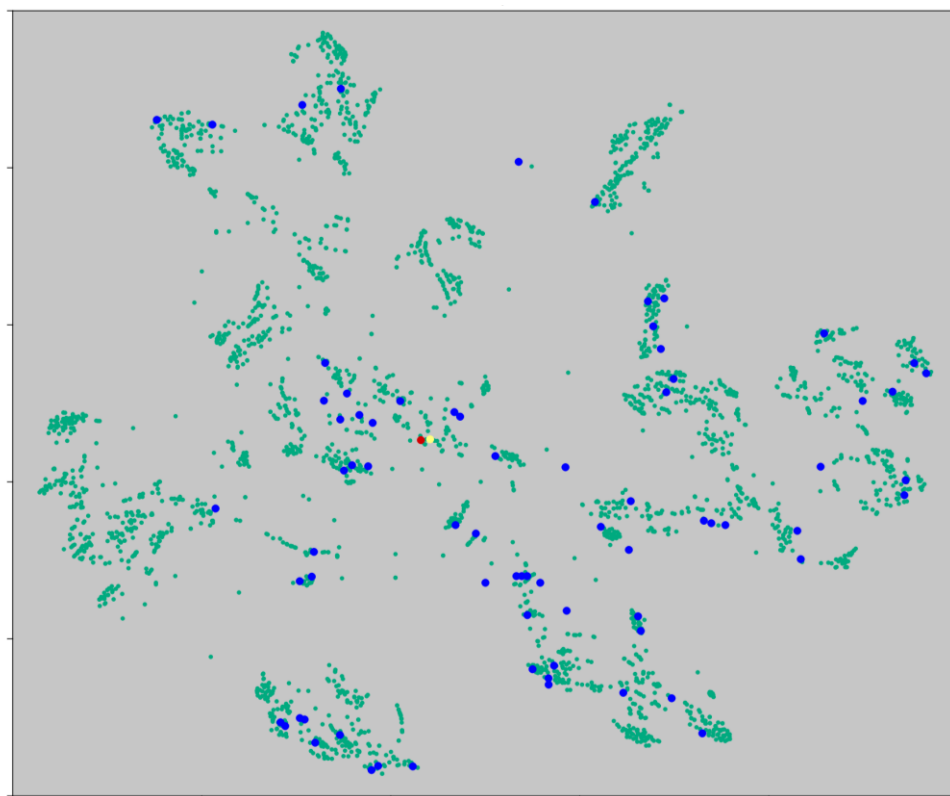
Crystal Structure of Human DPP-4 in Complex with Anagliptin



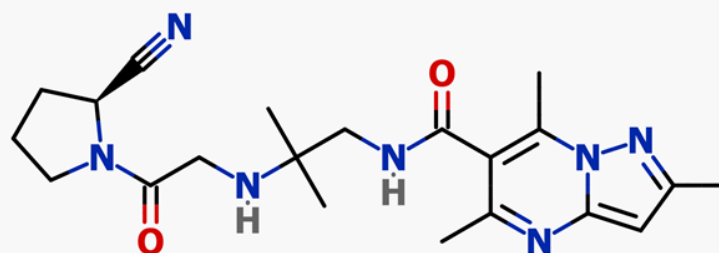
■ Name: Anagliptin

Watanabe, Y.S. *et al.*
J. Enzyme Inhib. Med. Chem. (2015) pp 1-8

DPP-4 Chemical Space



Name: Anagliptin



CHEMBL1929395

Selectivity_DPP:

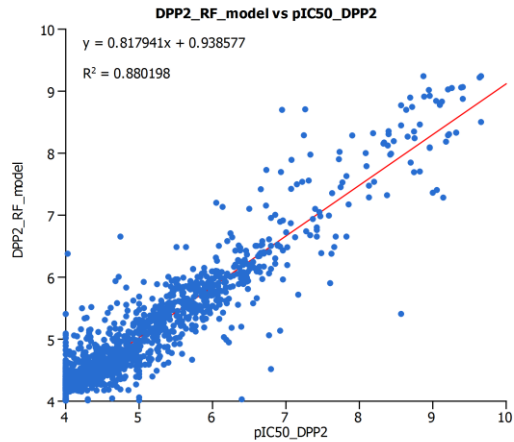
0.141

DPP4_pIC50:	6.96
DPP2_pIC50:	
DPP8_pIC50:	7.35
DPP9_pIC50:	7.55

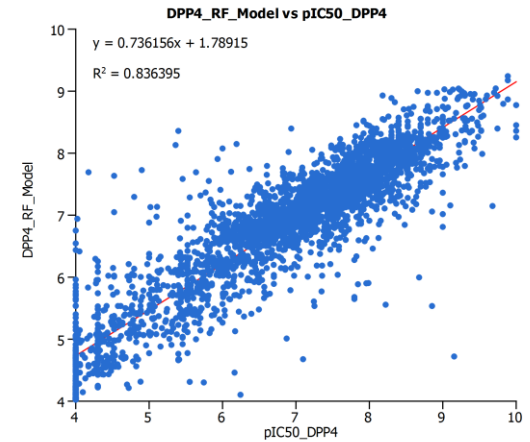


DPP Models

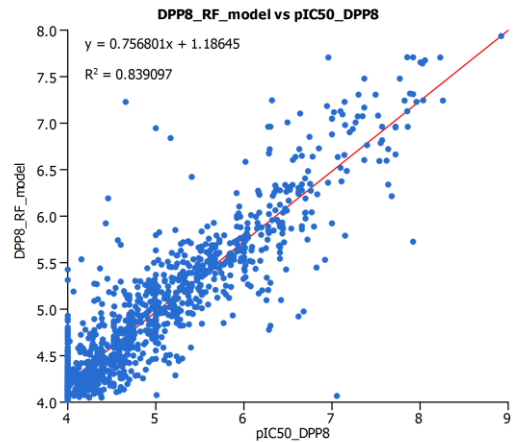
• DPP-2



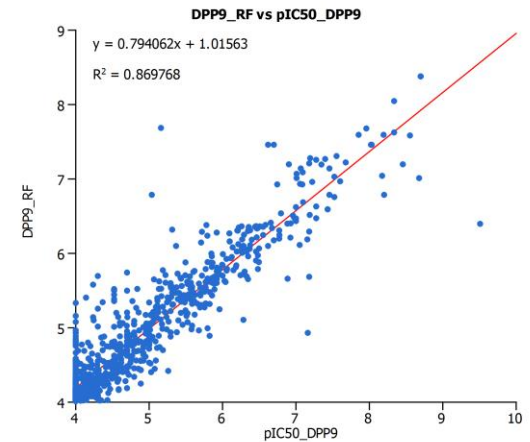
• DPP-4



• DPP-8

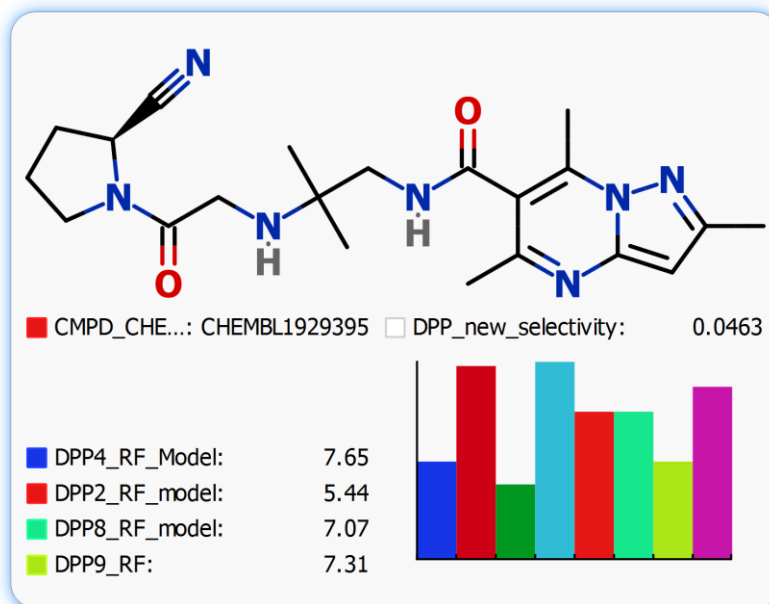


• DPP-9



Probabilistic Scoring

Designing to Achieve a Balance of Properties



Profile: DPP_new_selectivity

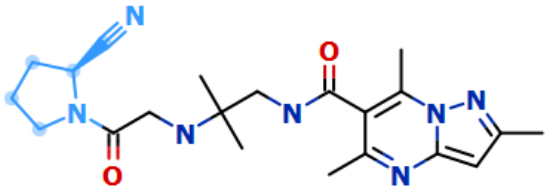
Property	Desired Value	Importance
■ DPP4_RF_Model	6 -> inf	
■ logS	> 1	
■ HIA category	+	
■ logP	0 -> 3.5	
■ DPP2_RF_model	≤ 6	
■ DPP8_RF_model	≤ 6	
■ DPP9_RF	≤ 6	
■ hERG pIC50	≤ 5	

Nova Set-up

Nova Setup Wizard

Specify Input Structure

Lasso a portion of the molecule to mask it from any transformations



Strict masking ☐

< Back Next > Finish Cancel

Nova Setup Wizard

Control Output

Generations

☒ Select compounds at each generation

Method

☒ Biased Diverse Value

☐ Random

Select compounds with

Selection Criteria

☒ The best compounds

☐ The best % of compounds

☐ Compounds with values higher than

☒ Attempt all transformations after generation 1

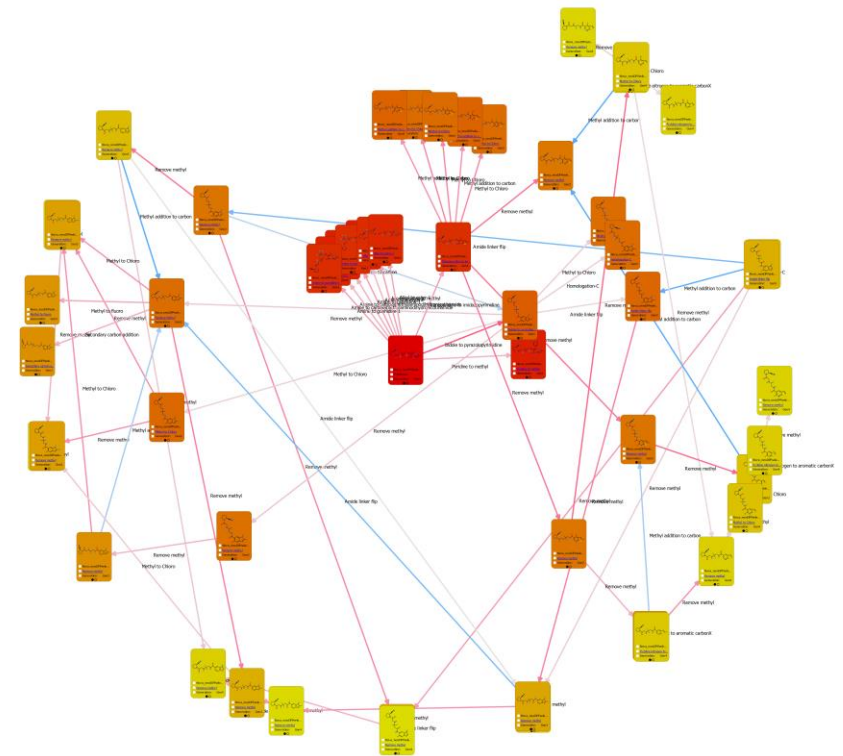
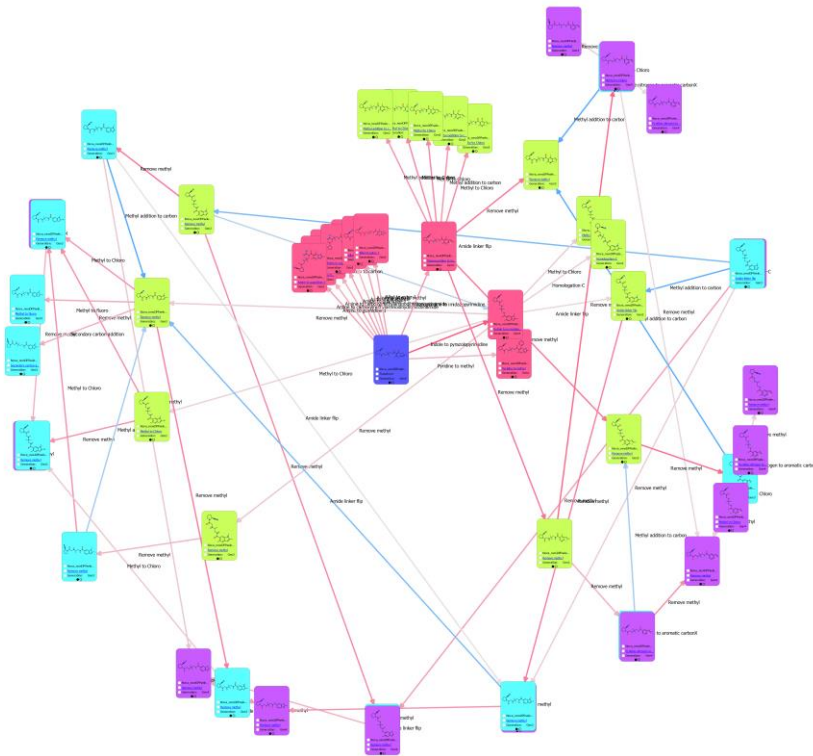
☒ Limit atom count change Maximum:

☒ Show results in Card View

< Back Next > Finish Cancel

Card View

Visualising Nova Results



Colour by: ☐ Generation

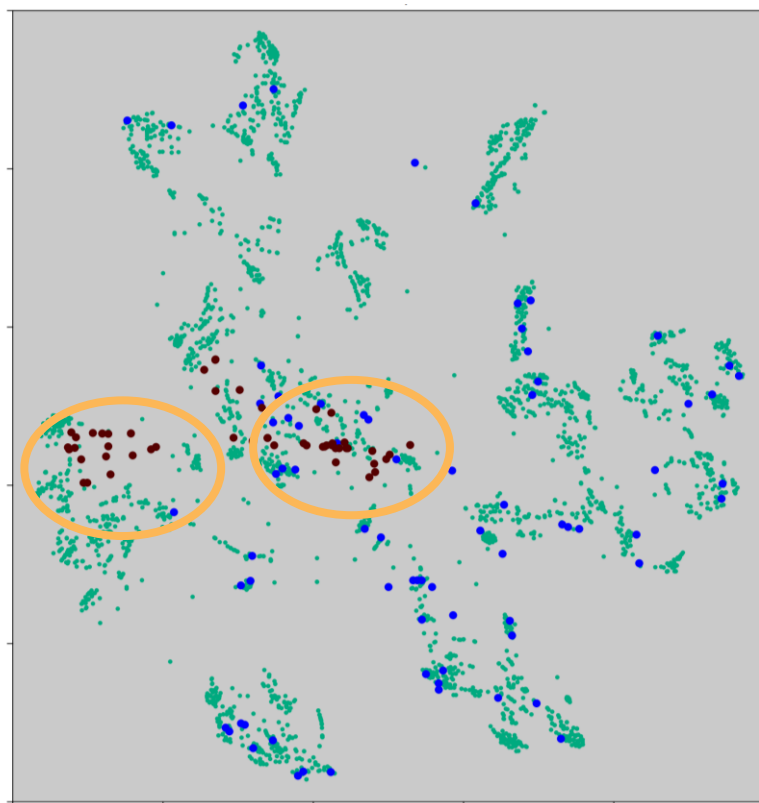
	Name	Colour
1	Gen0	■
2	Gen1	■
3	Gen2	■
4	Gen3	■
5	Gen4	■

Colour by: ☐ DPP_new_selectivity

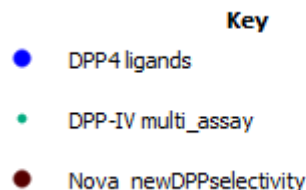
	Name	Colour
1	High	■
2	Low	■

New Selective DPP-4 Ideas

Explore Novel Chemistry Space



- New compound ideas sample multiple areas of chemistry space
 - Close to the known DPP-4 ligands
 - Areas of greater diversity



Selective DPP-4 Ideas

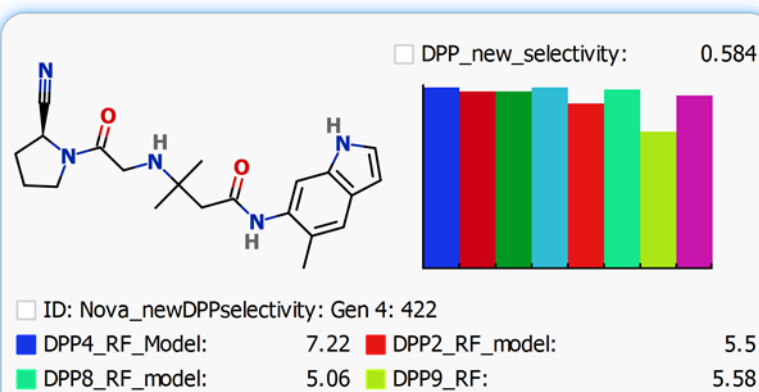
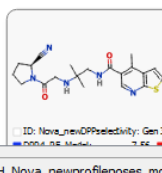
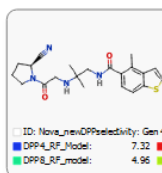
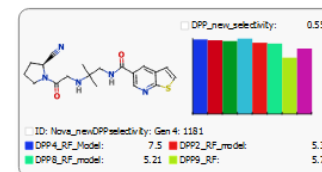
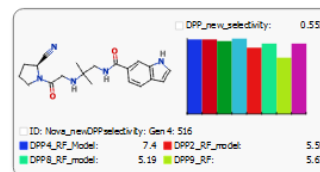
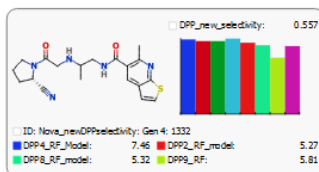
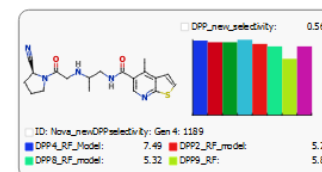
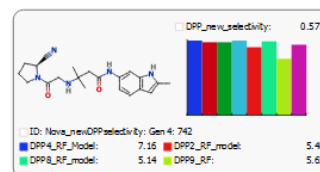
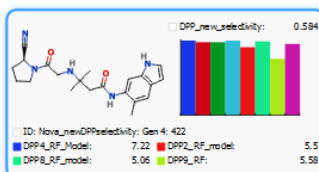
StarDrop - DPP-IV chembl IC50_workshopNovaonly

File Edit View Data Set Tools Help

Models	Scoring	Design	Visualisation	SeeSAR3D	P450	torch3D	Nova	Auto-Modeller
--------	---------	--------	---------------	----------	------	---------	------	---------------

Protein ▼ Display ▼ ☒ Show Binding

Powered by Biosolve



Columns 13 (0) Selected 1

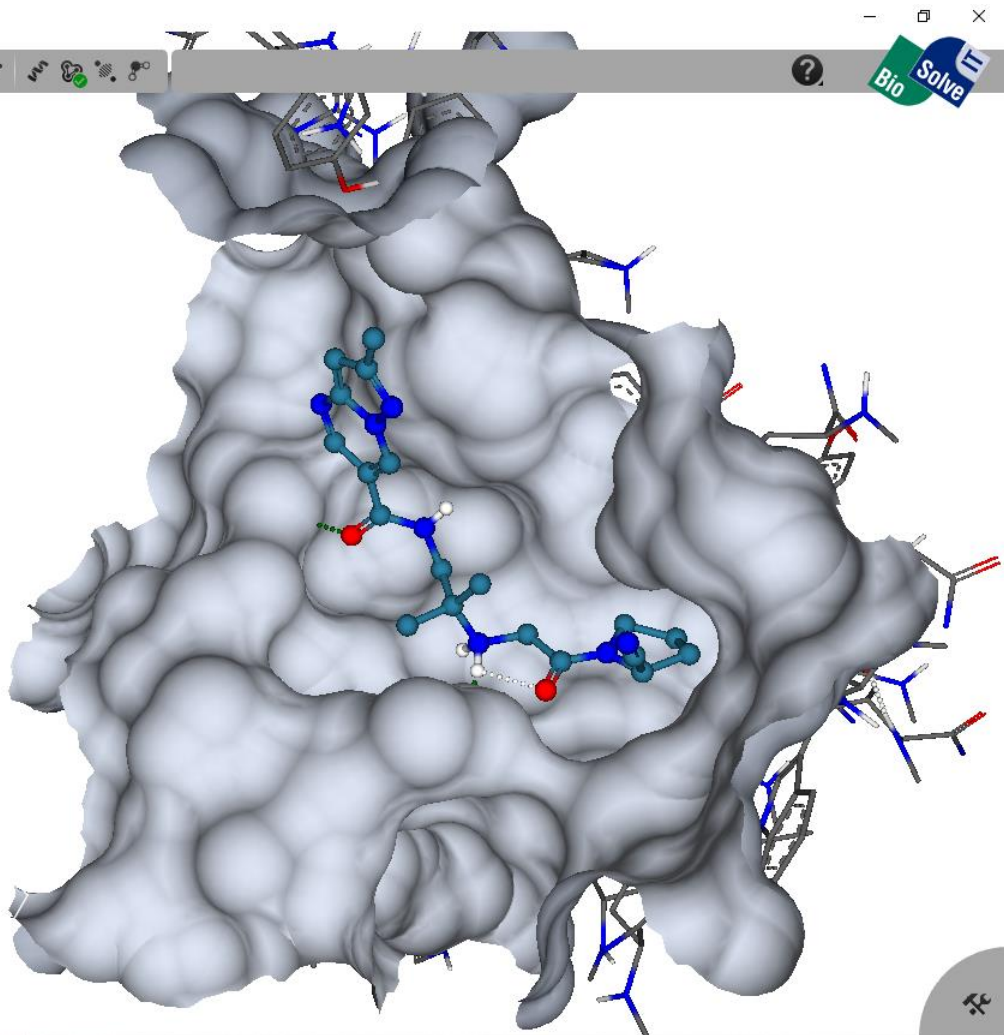
Hyde Scores in SeeSAR

DPP4_3WQH_Nova_newprofileposes_viaEditing.seesar - SeeSAR

Molecules (42/42)

From Protein From File From Editor

#	Fav	Name	Src	Estimated Affinity	LLE	Tor
				pM nM μ M mM		
1	☆	NAG_A_807	Protein	✗ optimization moved pose too far off		
2	☆	SKK_A_808	Protein		T	●
3	☆	SKK_B_807	Protein		T	●
4	☆	SKK_B_807	Protein		T	●
8	☆	SKK_B_807_1_3	Protein		T	●
10	☆	SKK_B_807_1_4	Protein		T	●
12	☆	SKK_B_807_5	Protein		T	●
13	☆	SKK_B_807_6	Protein	✗ optimization moved pose too far off		
14	☆	SKK_B_807_7	Protein		T	●
15	☆	SKK_B_807_8	Protein		T	●
16	☆	SKK_B_807_9	Protein		T	●
17	☆	SKK_B_807_10	Protein	✗ optimization moved pose too far off		
18	☆	SKK_B_807_11	Protein		T	●
19	☆	SKK_B_807_12	Protein		T	●
27	☆	SKK_B_...1_3_8	Protein		T	●
29	☆	SKK_B_...1_3_10	Protein		T	●
35	☆	SKK_B_...1_4_6	Protein		T	●
37	☆	SKK_B_...1_4_8	Protein		T	●
41	☆	SKK_B_807_5_2	Protein		T	●
59	☆	SKK_B_807_6_10	Protein		T	●
60	☆	SKK_B_807_7_1	Protein		T	●
66	☆	SKK_B_807_7_7	Protein		T	●
67	☆	SKK_B_807_7_8	Protein		T	●
68	☆	SKK_B_807_7_9	Protein		T	●
69	☆	SKK_B_807_7_10	Protein		T	●
77	☆	SKK_B_807_8_8	Protein		T	●




Selective DPP-4 Ideas

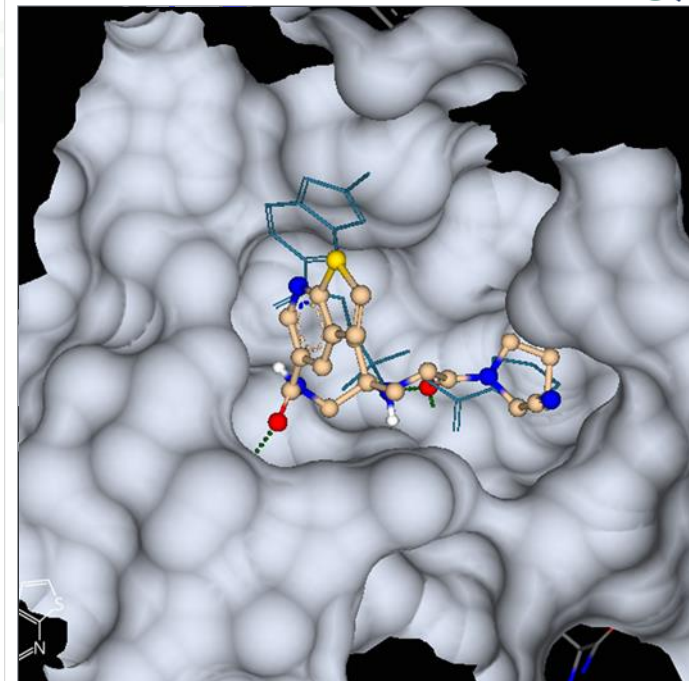
StarDrop - DPP-IV chembl IC50_workshopNovaonly



File Edit View Data Set Tools Help

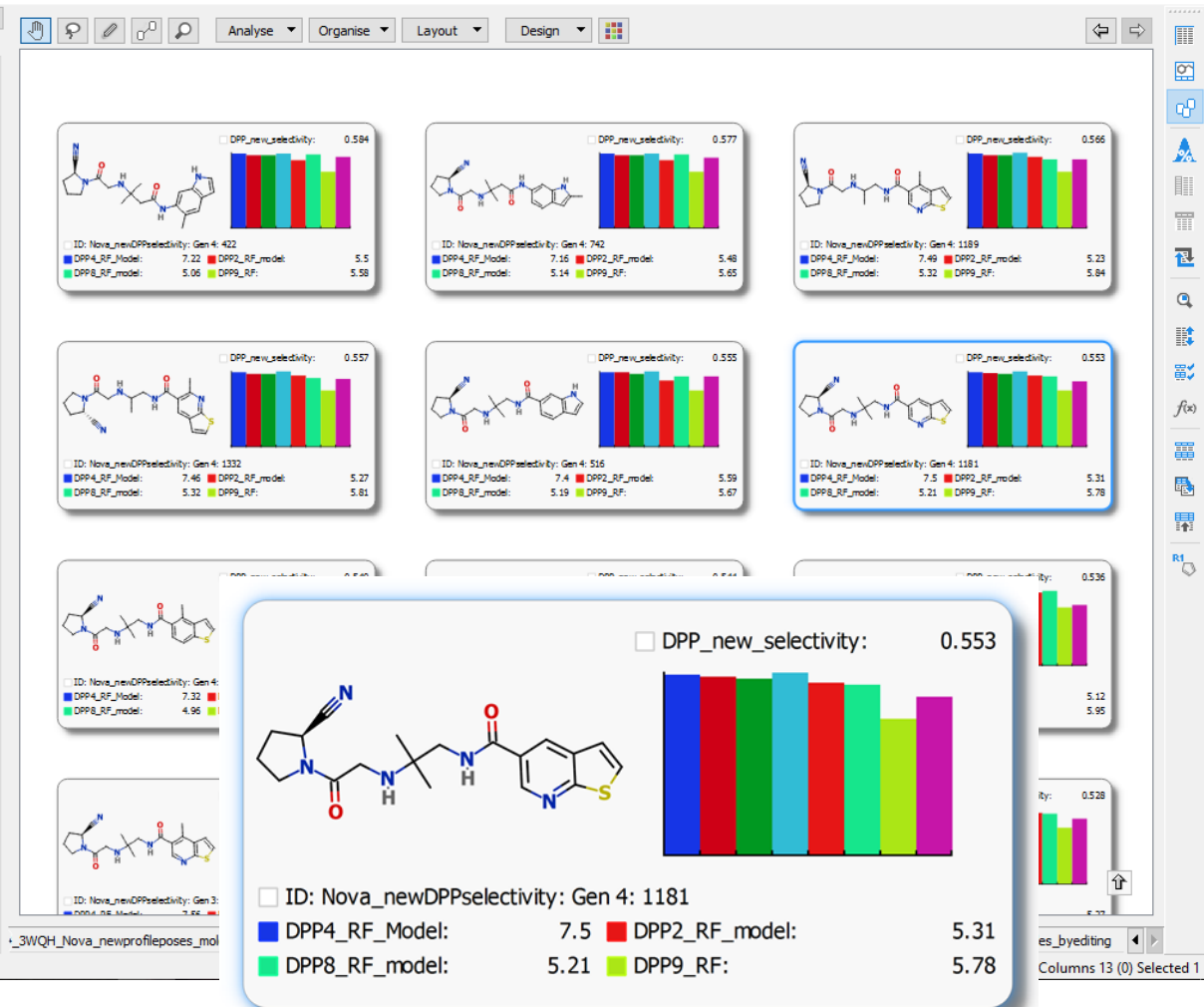
Models Scoring Design Visualisation SeeSAR3D P450 torch3D Nova Auto-Modeller

Protein Display ☒ Show Binding

Powered by 



Structure	Pose	ID
		
		



Conclusions

- Seamless integration of 2D and 3D information aids interpretation of SAR
 - Intuitive visualisations identify patterns
 - Understanding of interactions (3D) and relationships (2D)
- Using all the available information guides selection and design of compounds
 - Understand and rationalise structural modifications
 - Improved balance of properties
 - Avoid missing opportunities
- For more information:
 - www.optibrium.com/stardrop and www.biosolveit.com/SeeSAR
 - Optibrium: Booth 1227 or outside of room 6E (MEDI)

Acknowledgements



Edmund Champness

Peter Hunt

James Chisholm

Chris Leeding

Alex Elliott

Samuel Dowling

Fayzan Ahmed

Matthew Segall



Christian Lemmen

Marcus Gastreich

Carsten Detering