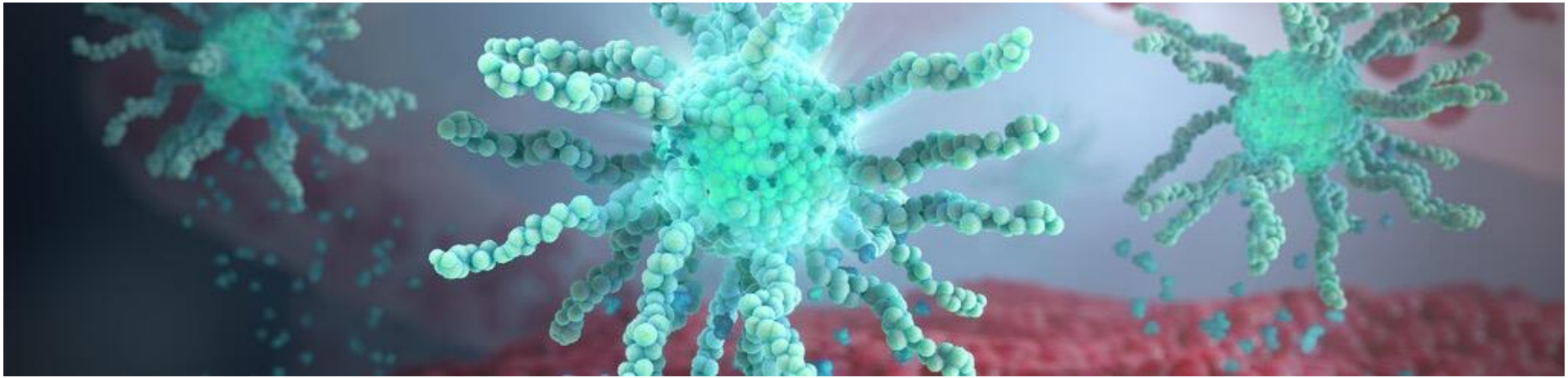
AZ workflows involving D360

William McCoull (william.mccoull@astrazeneca.com)
Nick Tomkinson (nick.tomkinson@astrazeneca.com)

D360 UGM, Boston

8th October 2019



About us...

Bill McCoull

- AstraZeneca D360 business lead
- Do data analysis every day in projects
- 15+ year involvement with AZ analysis tools
- End user - “I care what the tool does and how to use it”

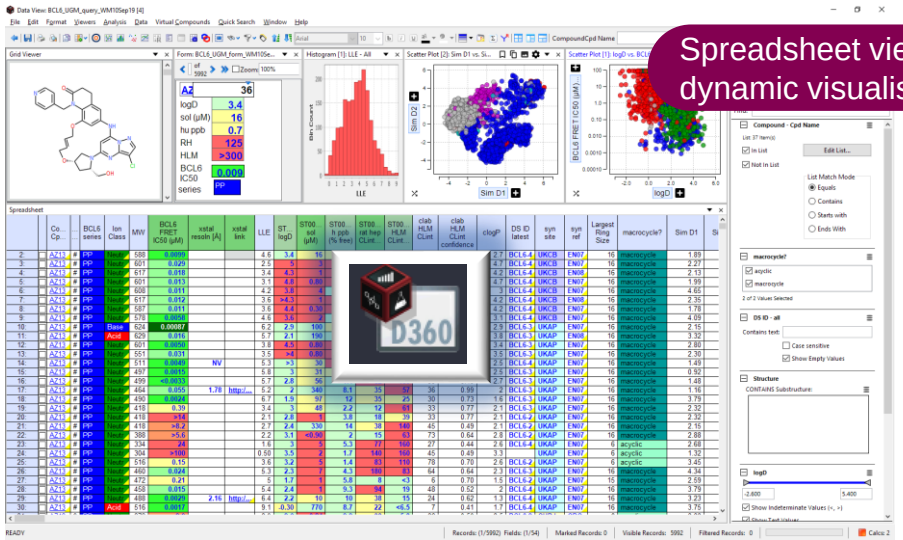
Nick Tomkinson

- AstraZeneca D360 tools workstream lead
- Informatics analyst
- 20+ year involvement with AZ analysis tools
- Work with users to provide analysis & visualisation

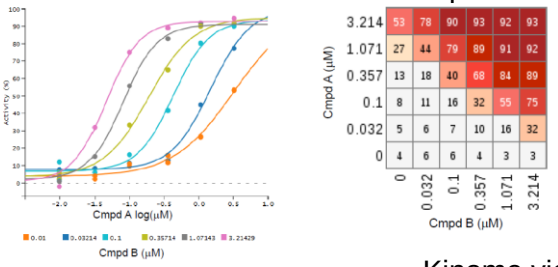
Outline

1. Sample availability – switching a data category folder
2. Design workflow – calculation cache & virtual/real compounds
3. 3D visualisation requirements
4. R-group stripping
5. Medchemica Matched Molecular Pairs
6. Summary

D360 status at AZ @ last UGM

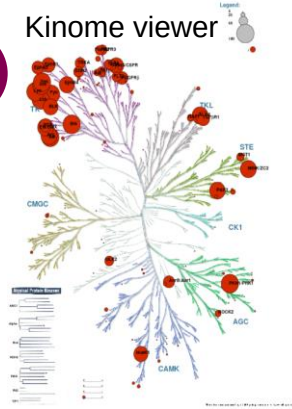


Combination curves / heatmaps

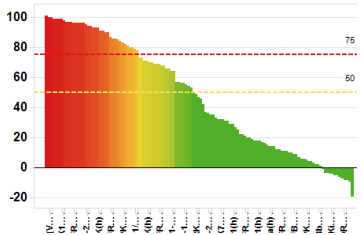


Extendable plug-ins to support new data analysis

Kinome viewer

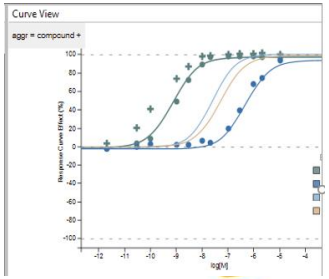


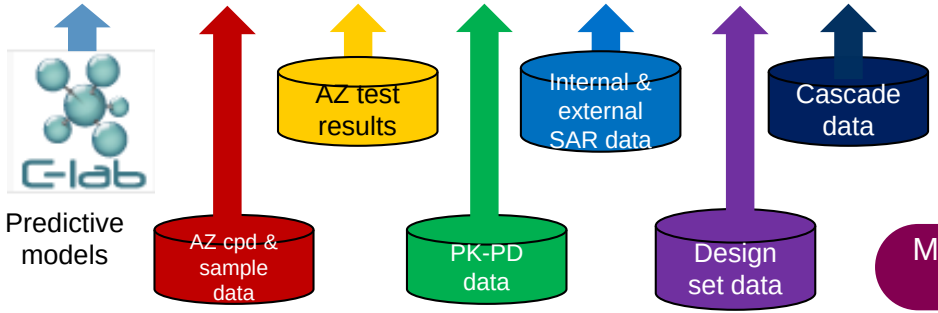
Kinome waterfall



Widgetised integrated spotfire queries

Curve viewer

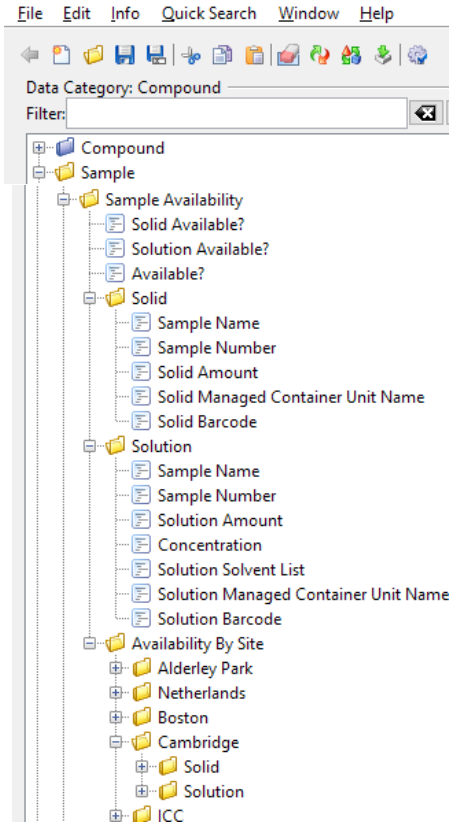




Multiple data sources

- Since launch in 2015, D360 has become a key platform at AZ

GSM (Global Sample Management) availability



- Previously used an ETL process to populate a static hierarchy of folders in D360 data category
  - Prone to errors lacked full reliability

- Moved to a real-time query of GSM availability
  - Kept the same folders (avoid user confusion)
  - Can add fields but not take away
  - Existing queries did not break

- Pro's
  - Improved reliability & accuracy
  - D360 query functionality allows easier & more powerful querying of availability than existing GSM tool
- Cons
  - Slow if large #cpds & complex query
  (not an issue for vast majority of use)

GSM (Global Sample Management) availability - output

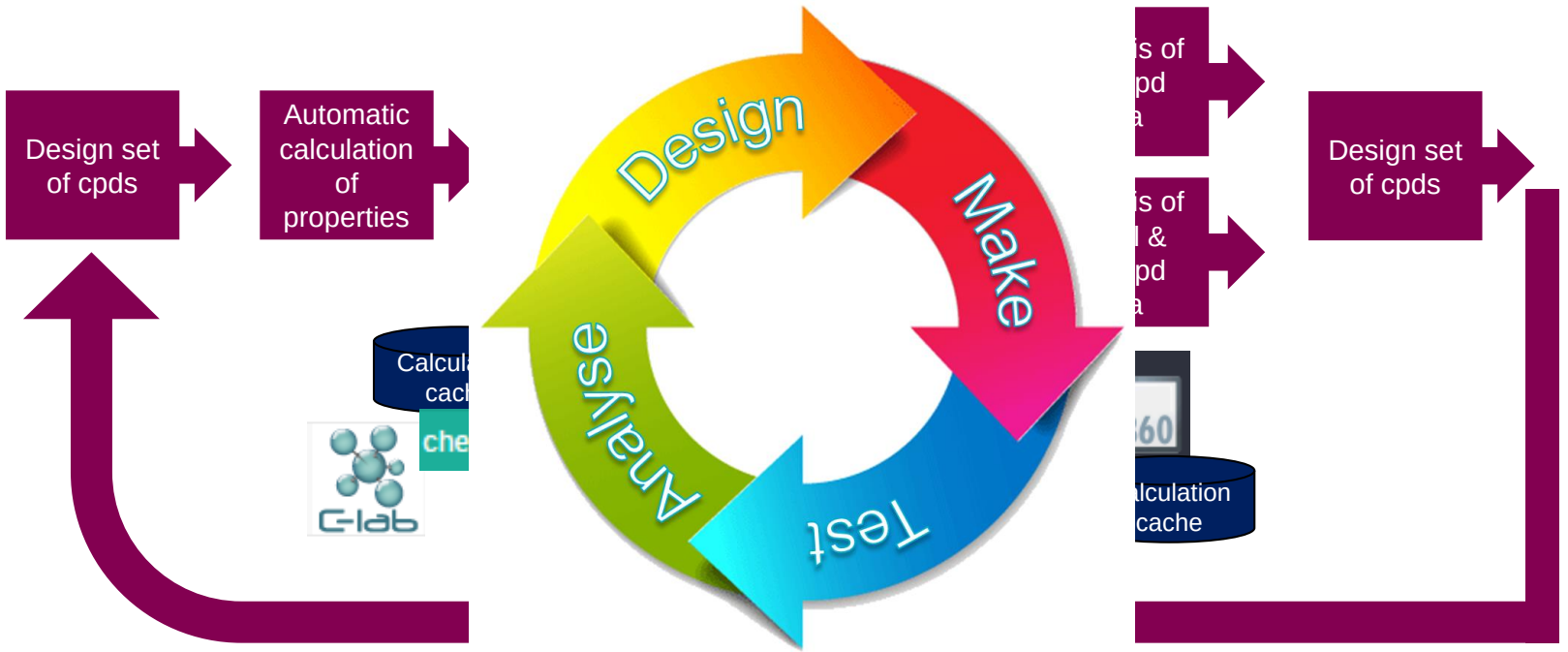
- Easy to display all samples for a specified cpd

		Compound Name	CSS SName	CSS SN	CSS amt (mg)	CSS site	CSS Barcode	PLS SName	PLS SN	PLS amt (ml)	PLS concn...	PLS site	PLS Barcode
1:	☐	Cpd 11	Cpd 11 -004	SN105227[illegible]	17.4	Specs Netherlands	5006004946	Cpd 11 -004 Cpd 11 -004 Cpd 11 -004	SN105227 SN105227 SN105227	0.163 0.726 0.556	10 10 10	Boston Cambridge Alderley Park	30797712 30954946 31518156

- Or aggregated data e.g. largest solid sample or largest solution sample at specified concentration/solvent available

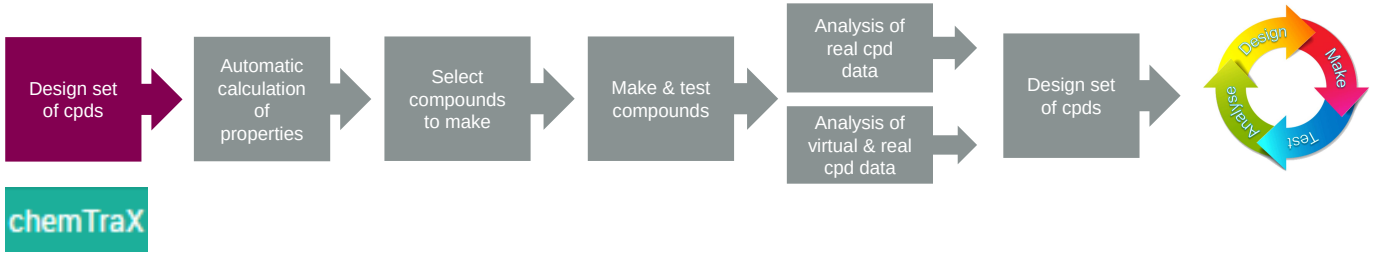
		cpd no.	NLSP ... max (mg)	GHP CSS max (mg)	Mo CSS max (mg)	Cam CSS max (mg)	AP PLS max 10 mM (ml)	GHP PLS max 10 mM (ml)	Mo PLS max 10 mM (ml)	Cam PLS max 10 mM (ml)
1:	☑	13	14.7				0.31		0.14	
2:	☑	11	17.4				0.56	0.16		0.73
3:	☑	5	78.1			2.8	0.08	0.05		

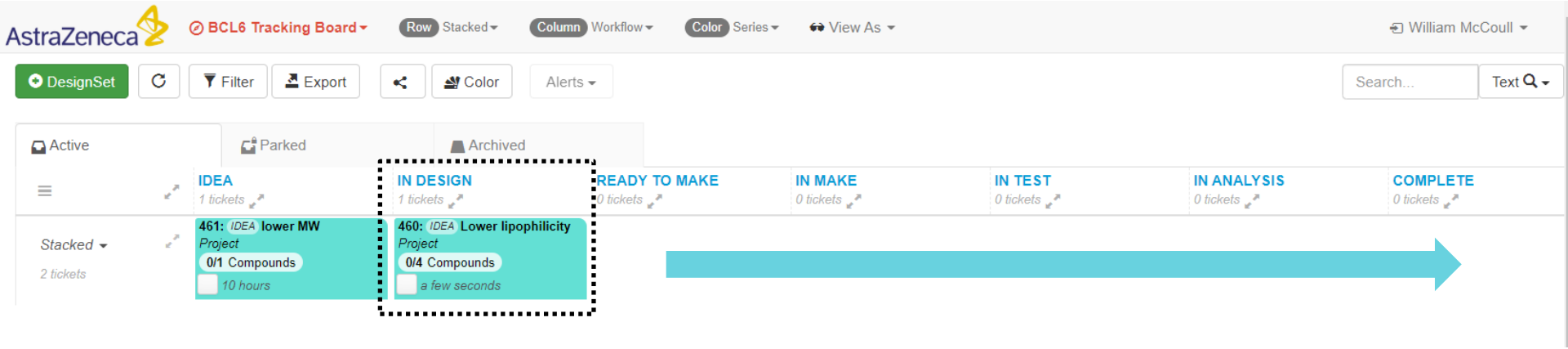
Design workflow



- D360 facilitates the analysis step of the design-make-test-cycle
- D360 is a project query, visualisation and analysis tool

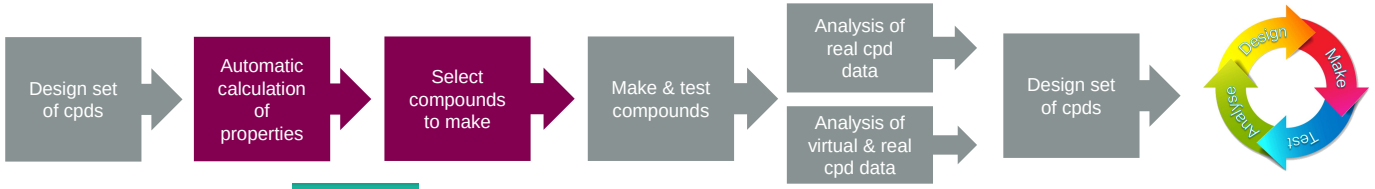
Design workflow

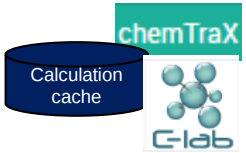


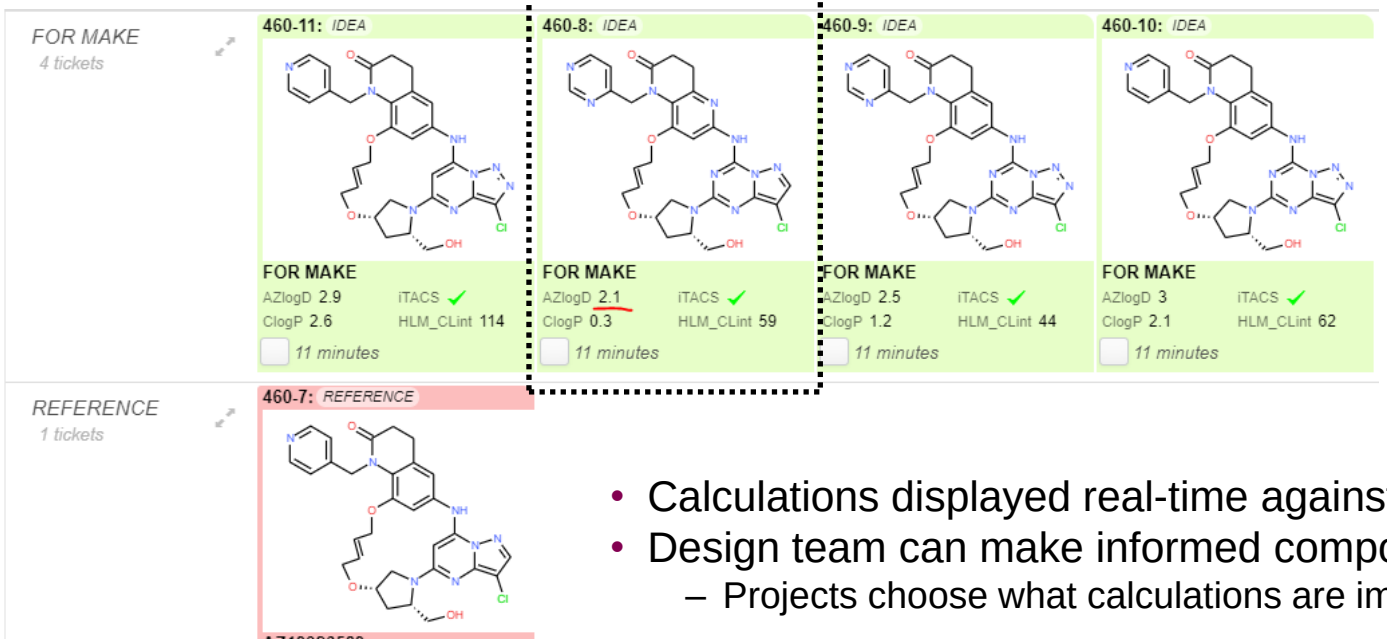


- ChemTrax used to manage progress of design sets through DMTA process
- Design sets of compounds created against design hypotheses
- Swim lanes used for each step in the process

Design workflow

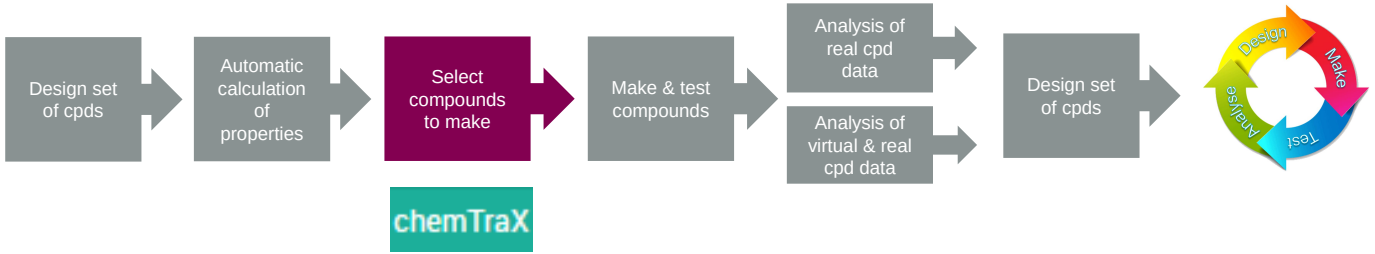


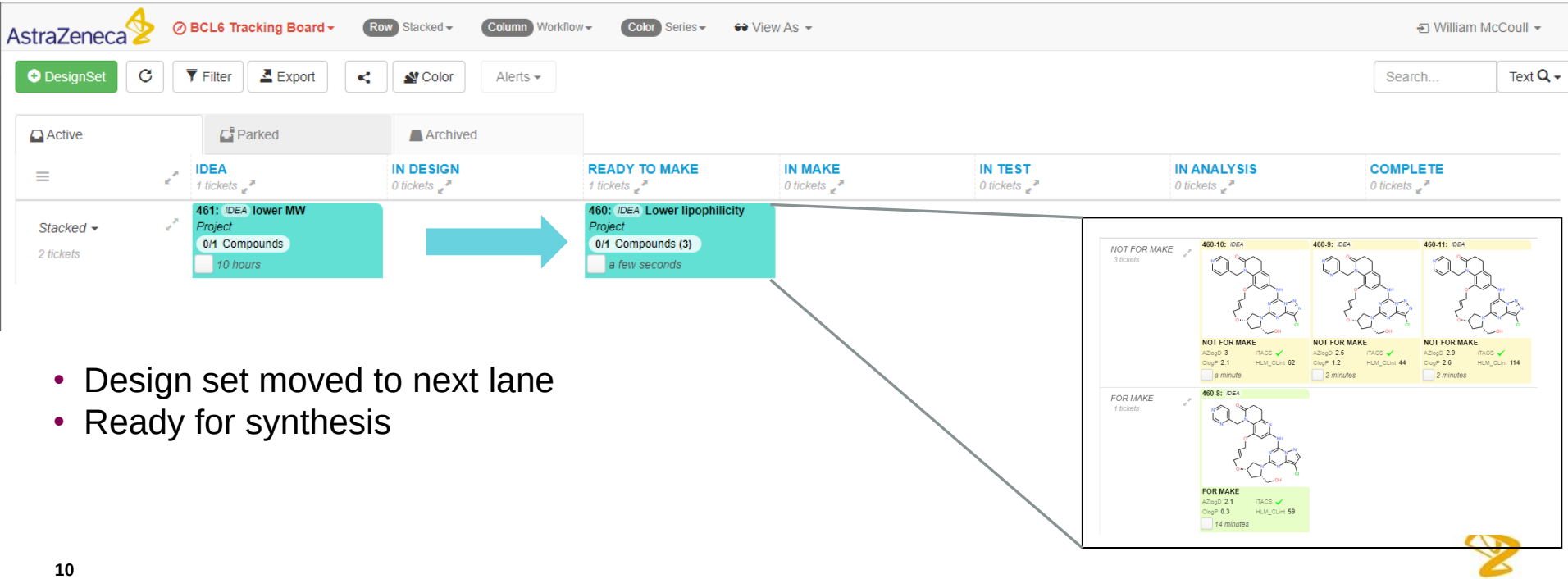




- Calculations displayed real-time against each compound
- Design team can make informed compound selection
  - Projects choose what calculations are important to them

Design workflow

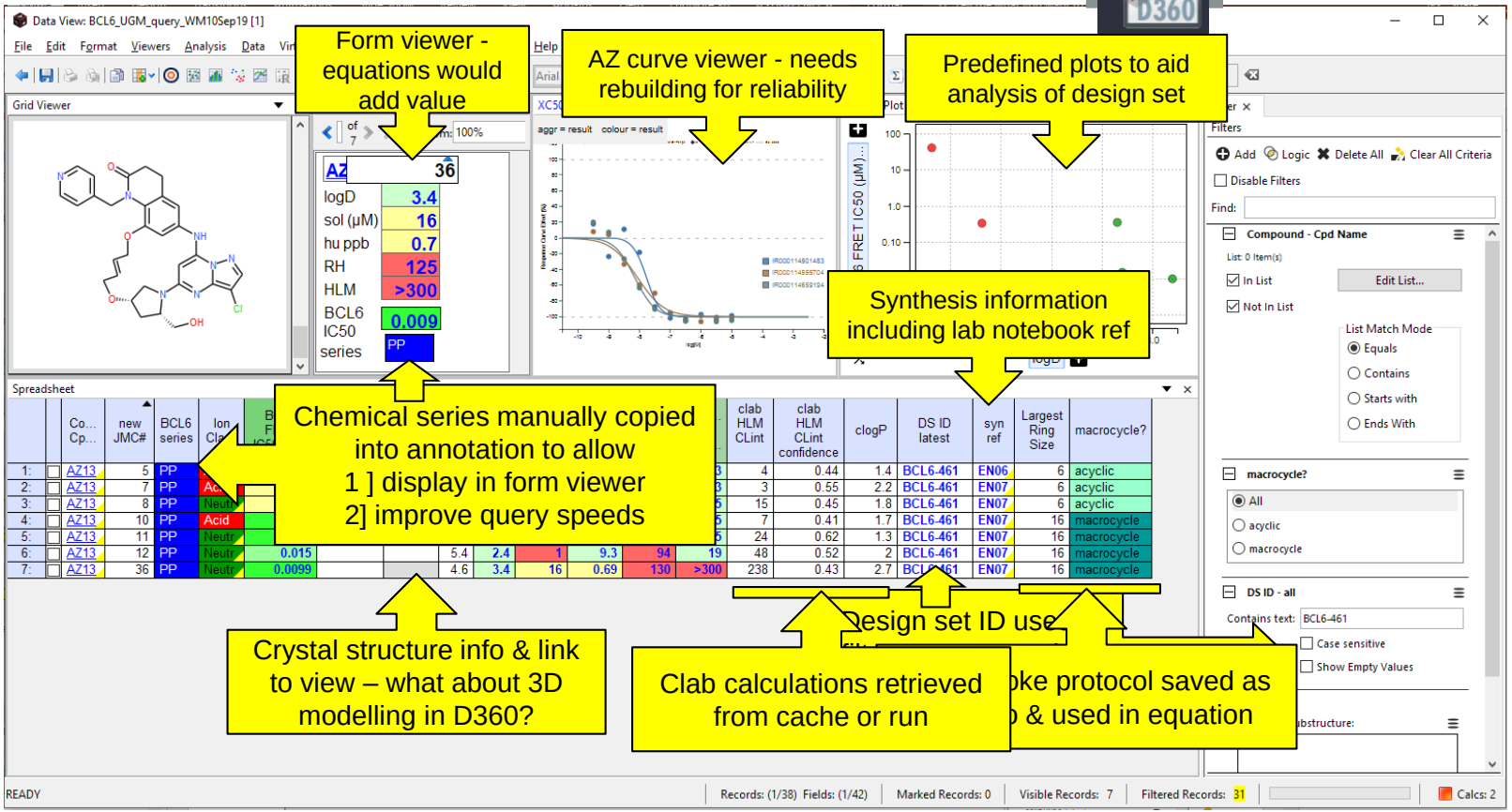




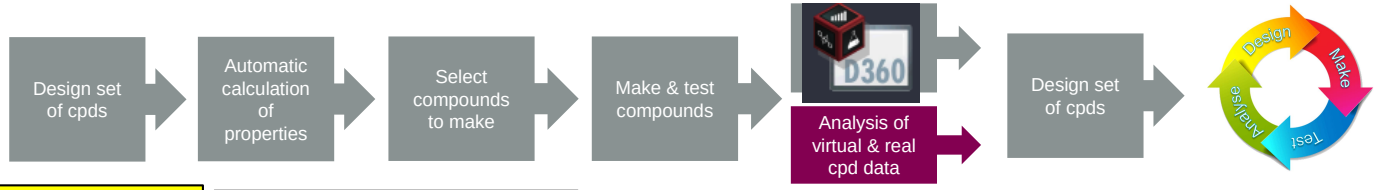
- Design set moved to next lane
- Ready for synthesis

Design workflow: R

Design set of cpds → Automatic calculation of properties → Select compounds to make → Make & test compounds → Analysis of real cpd data → Design set of cpds



Design workflow: V&R



Curve viewer & plots import
Form viewer needs recreating – calculations would be useful to display

Can assess if V cpds not made from original design are worth revisiting

Can assess how well calculation predict R cpd experimental values

V cpd with pred low logD & CLint (high confidence) thus worth making

Can toggle off/on virtual or real cpds

	Cpd… Cpd…	Cpd Union SMILES	Ch… ISAC Mat…	Virtual cpd ID	BCL6 series	BCL6 FRET IC50 (µM)	xst link	resolm [Å]	JMC#	logD	sol (µM)	fr ppb (% free)	ST00… rat hep CLint…	ST00… HLM CLint…	DS ID latest	syn ref	LLE	clab HLM CLint	clab HLM CLint confidence	Largest Ring Size	macrocycle?	V&R AZlogD	V&R Ion class
1:	AZ13	c1c(nc2c(	AZ13	BCL6-461-1	PP	45		1.59	5	-1.4	900	19	<1	<3	BCL6-461	EN06	5.7	3.6	0.44	6	acyclic	-0.38	Acid
2:	AZ13	c1cc2c(cc	AZ13	BCL6-461-2	PP	0.35	http…	1.98	7	-0.40	>1000	2.3	5.8	<3	BCL6-461	EN07	6.9	3.1	0.55	6	acyclic	-0.077	Acid
3:	AZ13	c1cc2c(cc	AZ13	BCL6-461-3	PP	0.37			8	2.3	3	3.5	3.6	5.5	BCL6-461	EN07	4.1	15	0.45	6	acyclic	2.1	Neutr
4:	AZ13	c1c2cc(c3	AZ13	BCL6-461-4	PP	0.0017			10	-0.30	770	8.7	22	<6.5	BCL6-461	EN07	9.1	6.7	0.41	16	macrocycle	0.31	Acid
5:	AZ13	c1c2cc(c3	AZ13	BCL6-461-5	PP	0.0029	http…	2.16	11	2.2	10	10	38	15	BCL6-461	EN07	6.4	24	0.62	16	macrocycle	2.1	Neutr
6:	AZ13	c1c2cc(c3	AZ13	BCL6-461-6	PP	0.015			12	2.4	1	9.3	94	19	BCL6-461	EN07	5.4	48	0.52	16	macrocycle	2.6	Neutr
7:	AZ13	c1cnccc1	AZ13	BCL6-461-7	PP	0.0099			36	3.4	16	0.69	130	>300	BCL6-461	EN07	4.6	240	0.43	16	macrocycle	3.6	Neutr
8:	DC00	1c2cc(c3		BCL6-461-8											BCL6-461			15	0.67	16	macrocycle	2.2	Neutr

Cpd Union has mix of virtual (DC..) & real (AZ..) ID with unique SMILES ID

Real data imported from original query

Chemtrax data imported from original query

Calculations and macros need to be recreated to run V&R

3D Visualisation - Requirements

- Dynamic selection & editing of ligand (2-way)
- Save and share visualisations
- Read target and electron density/grid information
- Access to computational methods (docking/minimisation/FEP)
- Customizable interface
- Easy ways to focus in on binding site
- High quality rendering & antialiasing, display interactions, grids & surfaces

Data View: GVK Patent Structure 15-Sep-2019 20:07 [1]

File Edit Format Viewers Analysis Data Quick Search Window Help

Mock-up

2-way link between 2D-plots, 3D ligand visualisation and D360 dataset rows

Spreadsheet

		GVK Id	GVK Structure	Concat Chem Connect ID	Concat AZ Number	Concat Journal Patent N...	Concat Normalized Jou...	Concat EntrezGene ID	Concat Official Name	Concat Common Name	Concat Standard Name	GVK Compound ...	WO2019084026 GMean IC50 (uM) Protein: VHL-EL...	WO2019084026 GMean IC50 (uM) Protein: VHL-EL...	WO2019084026 GMean KD (uM) Protein: VHL-EL... Assay Type: B Assay Id: 119416...
1:		124222798		407										100.000	83.926
2:		124222802		407										45.400	12.800
3:		124222804		407										100.000	150.000
4:		124222805		407										73.200	16.400
5:		124222806		407											150.000
6:		124222808		407										16.191	6.538

GVK_Patent_Structure_15-Sep-2019_12_21 - TIBCO Spotfire

File Edit Data Visualizations View Tools Help User Editing

load structure
load pdb id
full structure
select ligand
select chain
select residue
pocket near clipping
pocket radius clipping
pocket opacity
cartoon
backbone
hydrogen
sidechainAttached
hydrophobic
hbond
weak hbond
water-water hbond
backbone-backbone hbond
halogen bond
metal interaction
salt bridge
cation-pi
pi-stacking

GVK_Patent_Structure_15-Sep-2019_12_21

Marking: Marking
Marker by: (None)
Color by: (None)
All values
Shape by: (None)
All values
Size by: (None)

WO2019084026 GMean IC50 (uM) Protein: VHL-ELONGINC COMPLEX Assay Type: B ...

WO2019084026 GMean IC50 (uM) Protein: VHL-ELON...

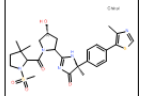
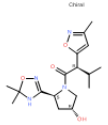
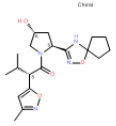
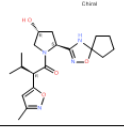
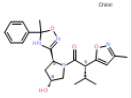
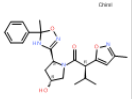
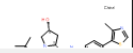
GVK_Patent_Structure_15-Sep-2019_12_21 | 61 of 61 rows | 1 marked | 32 columns

Data View: GVK Patent Structure 15-Sep-2019 20:07 [1]

File Edit Format Viewers Analysis Data Quick Search Window Help

Mock-up

Spreadsheet

		GVK Id	GVK Structure	Concat Chem Connect ID	Concat AZ Number	Concat Journal Patent N...	Concat Normalized Journ...	Concat EntrezGene ID	Concat Official Name
1:	☐	124222798		407591518		WO 2019/084026 A1	WO2019084026	6921, 6923, 6923, 6923, 6923, 6923, 6923, 6923, 6923, 7428, 7428,	
2:	☐	124222802		407597369		WO 2019/084026 A1	WO2019084026	6921, 6923, 6923, 6923, 7428, 7428, 7428, 8453, 8453, 8453	
3:	☐	124222804		407588402		WO 2019/084026 A1	WO2019084026	6921, 6923, 6923, 6923,	
4:	☐	124222805		407589535		WO 2019/084026 A1	WO2019084026	6923, 6923, 7428, 7428, 7428, 8453, 8453, 8453	
5:	☐	124222806		407599052		WO 2019/084026 A1	WO2019084026	6921, 6923, 6923, 7428, 7428, 8453, 8453	
6:	☐	124222808		407597371		WO 2019/084026 A1	WO2019084026	6921, 6923, 6923, 6923, 6923, 7428, 7428, 7428, 7428, 8453, 8453, 8453,	
7:	☐	124222810		407599051		WO 2019/084026	WO2019084026	6921, 6923, 6923, 6923, 6923, 7428,	

2-way link

3D Structure Viewer

load structure

load pdb id

full structure

select ligand

select chain

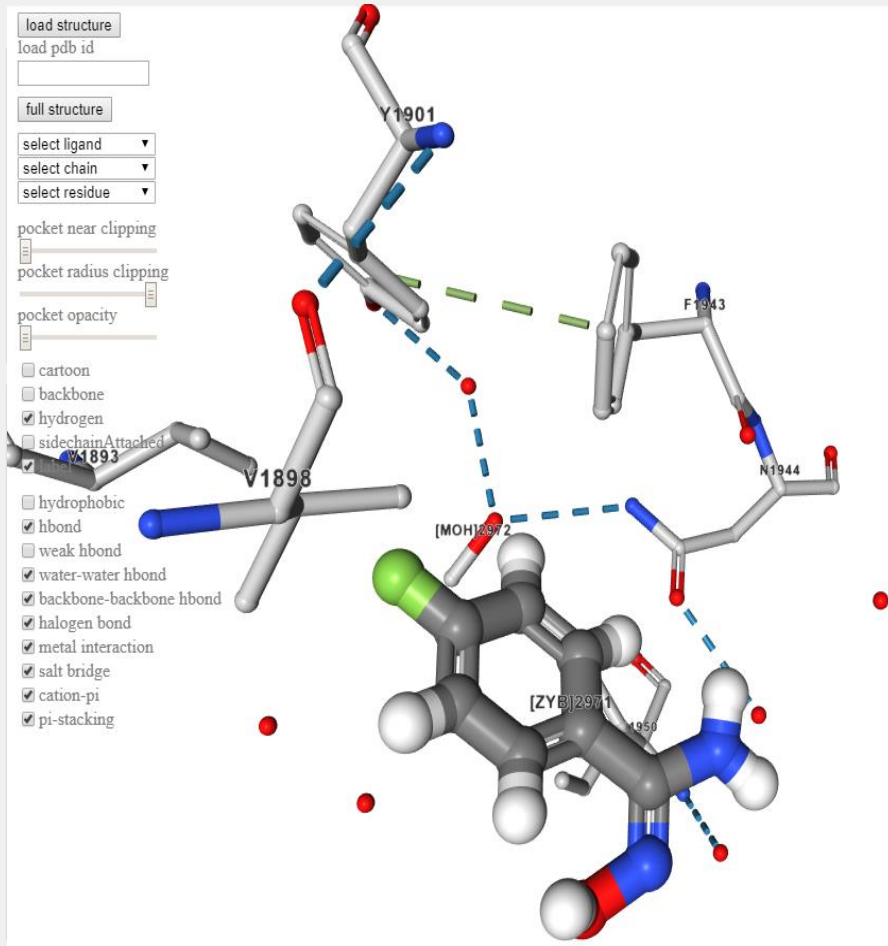
select residue

pocket near clipping

pocket radius clipping

pocket opacity

- ☐ cartoon
- ☐ backbone
- ☑ hydrogen
- ☐ sidechainAttached
- ☑ other
- ☐ hydrophobic
- ☑ hbond
- ☐ weak hbond
- ☑ water-water hbond
- ☑ backbone-backbone hbond
- ☑ halogen bond
- ☑ metal interaction
- ☑ salt bridge
- ☑ cation-pi
- ☑ pi-stacking

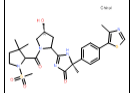
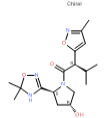
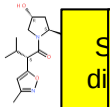


Data View: GVK Patent Structure 15-Sep-2019 20:07 [1]

File Edit Format Viewers Analysis Data Quick Search Window Help

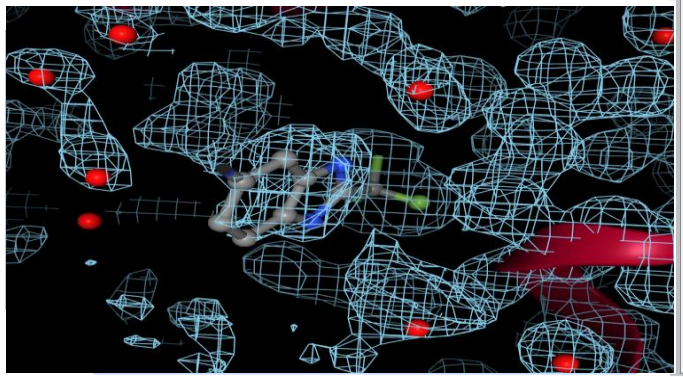
Mock-up

Spreadsheet

		GVK Id	GVK Structure	Concat Chem Connect ID	Concat AZ Number	Concat Journal Patent N...	Co Normaliz
1:	☐	124222798		407591518		WO 2019/084026 A1	WO2019
2:	☐	124222802		407597369		WO 2019/084026 A1	WO2019
3:	☐	124222804				WO 2019/084026 A1	WO2019

3D vis linked to all plots

3D Structure Viewer



3D Structure Viewer

load structure

load pdb id

full structure

select ligand

select chain

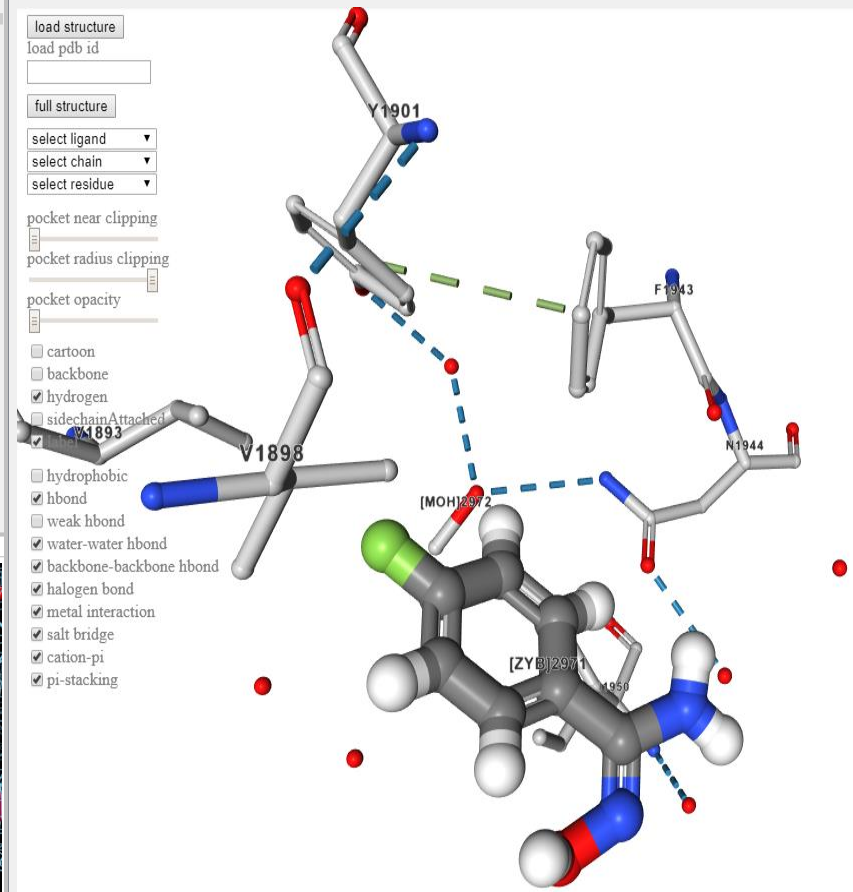
select residue

pocket near clipping

pocket radius clipping

pocket opacity

- ☐ cartoon
- ☐ backbone
- ☑ hydrogen
- ☐ sidechainAttached
- ☐ hydrophobic
- ☑ hbond
- ☐ weak hbond
- ☑ water-water hbond
- ☑ backbone-backbone hbond
- ☑ halogen bond
- ☑ metal interaction
- ☑ salt bridge
- ☑ cation-pi
- ☑ pi-stacking



READY

Records: (0/61) Fields: (0/21) Marked Records: 0 Visible Records: 61 Filtered Records: 0

R-Group Analysis

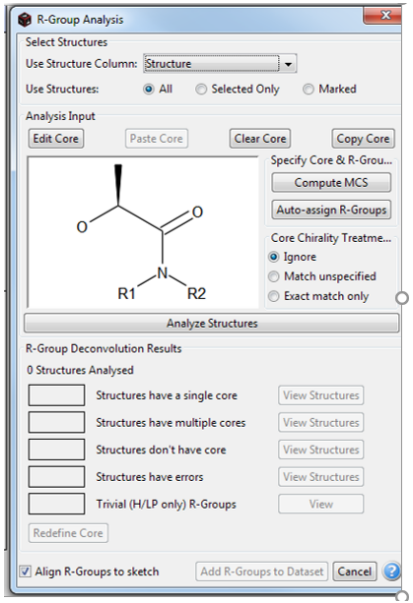
- fRGS
- Pipeline Pilot/Spotfire
- SARPlatform
- D360
- Moesaic

Related applications

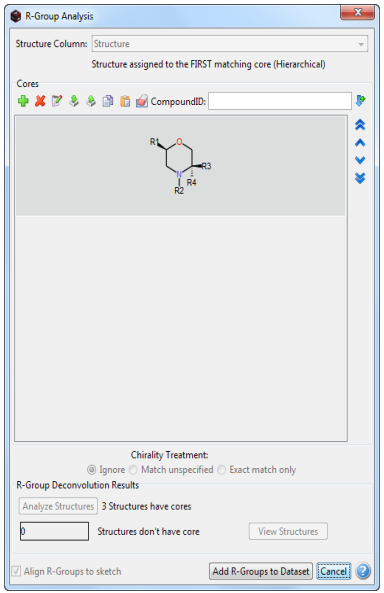
- Matched pairs
- Spark

D360 R-Group Analysis

v 9



v 10



Recent changes have removed Auto generation of MCS and R-Groups

fRGS Workflow(s)

Import Smiles

SDFile Patents, ibex etc

Paste IDs

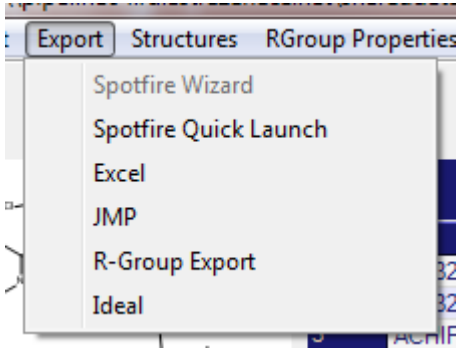
Import text IDs

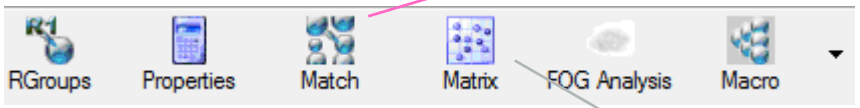
Define, Store, Share & Manage Cores

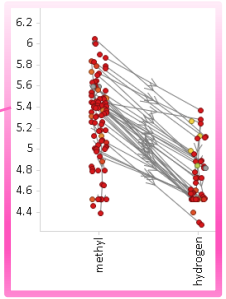
Strip

JMP, FW

Predictions







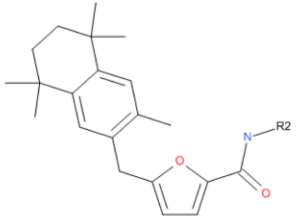
- Chemminer – hierarchical environment visualisation
- 2D pharmacophore and plotting
- Rgroup fingerprint – also says where pharmacophore features are missing (known unknowns)

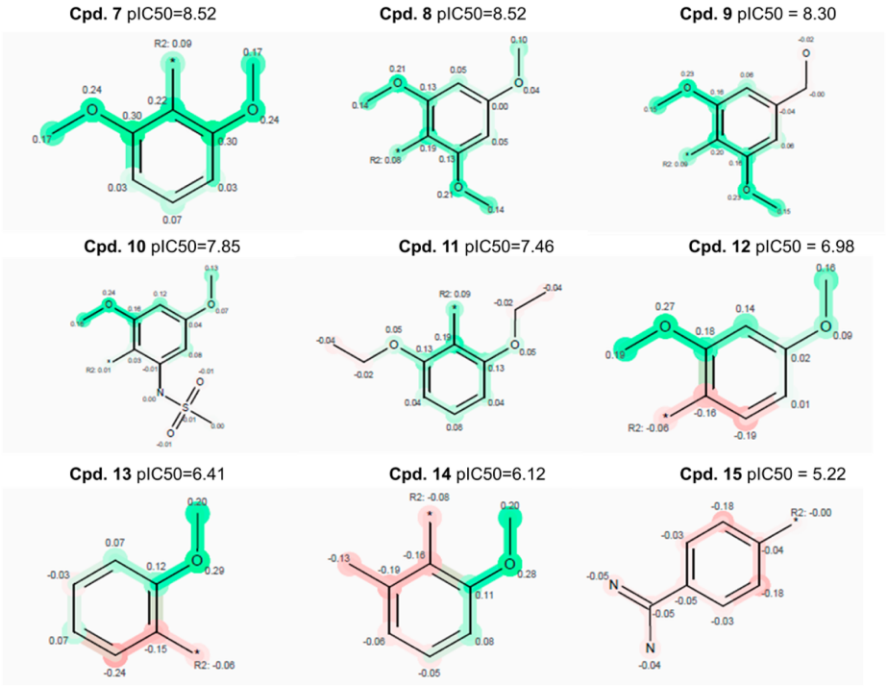
Hole-filling and creates smiles for Chemtrax

- Some of this functionality is available in D360 but still significant gaps

Machine-Learning Atom Contributions

Model Interpretation





Useful to see what data has contributed to the visualisation

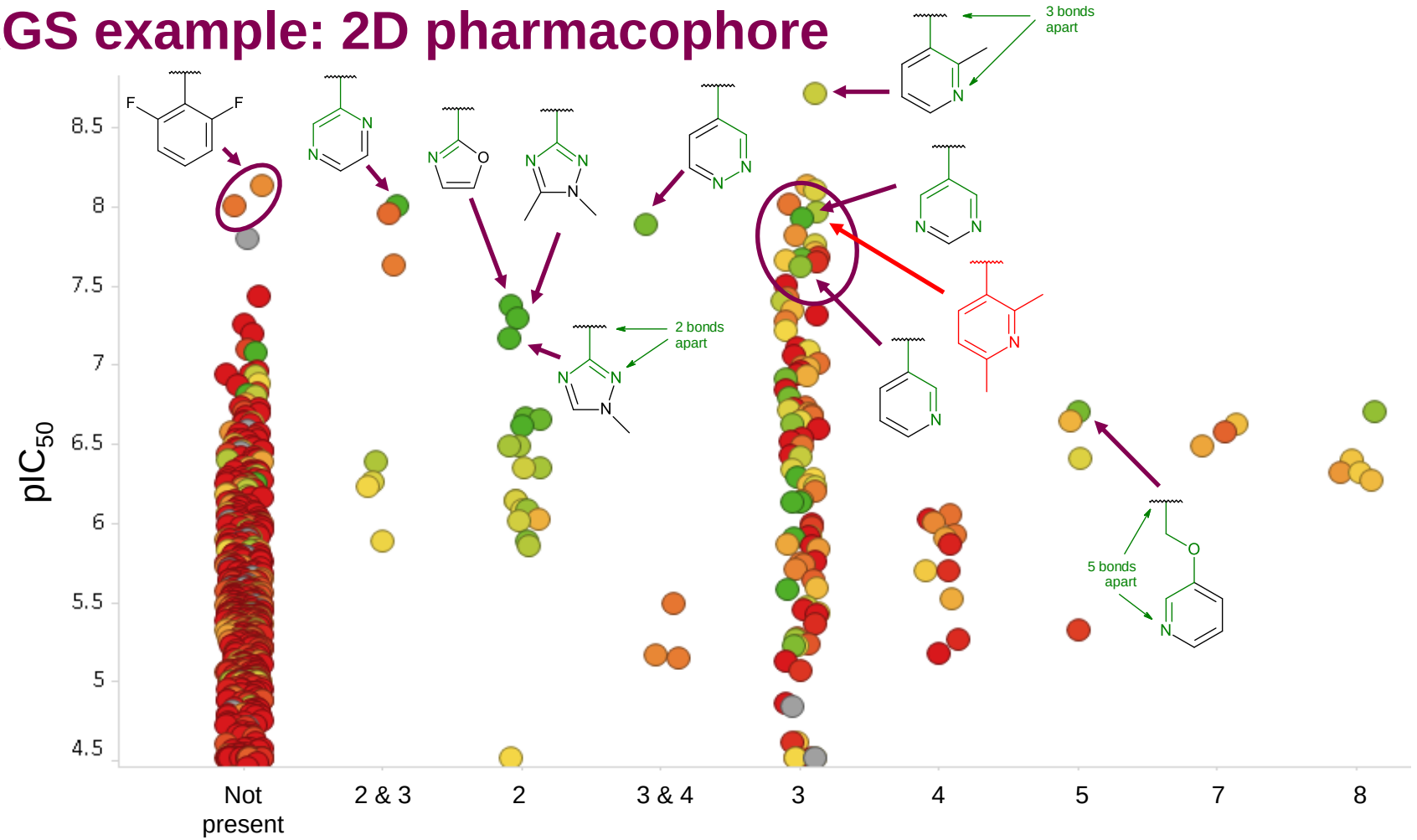
Mats Eriksson, Hongming Chen, Lars Carlsson, J. Willem M. Nissink, John G. Cumming, and Ingemar Nilsson,Ulf Norinder, Peter Varkonyi

Journal of Chemical Information and Modeling (2014), 54(4), 1117-1128,
Journal of Chemical Information and Modeling (2013), 53(6), 1324-1336

Comparative performance

		fRGS	D360	SAR Platform
1	Multiple mappings	Yes	Yes	Yes
2	Chiral R groups	Yes	Yes	Yes
3	Chirality in the core	No	Yes	Yes
4	Chirality at attachment point	No	No – scrambles core	Yes
5	Symmetry	Yes but needs improving	?	Yes
6	Double bond geometry in core	?	?	?
7	R groups as linkers	Yes except if linker is single atom	Possibly – seems odd	No – not in Auto RGroup detection.
8	R groups that cyclise onto core	Yes	Yes -but odd	No
9	“C vs N in aromatic ring”	No – need to draw 2 cores	OK for aliphatic C/O not checked for c vs n	Yes – input as SMARTS
10	Automated core perception	No	MCS removed from v10	Yes
11	#compounds matching MCS	No	No	Yes
12	Alternative tautomers in core or R groups	?	?	?
13	Double/triple bond to R groups	Yes	Yes	?

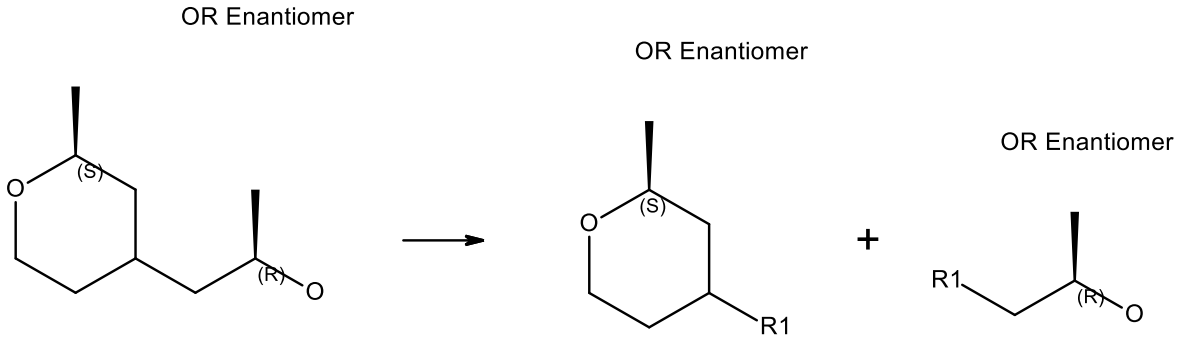
fRGS example: 2D pharmacophore



Number of bonds between linker and the 'aromatic nitrogen' pharmacophore

R-Group Requirements:

- Core entry, management and perception:
  - Autodetect core
  - Edit core or supply from scratch. Manage (save, share) cores.
- AZIDs, virtual compounds (mol/smiles), external compounds etc. from Database, SD, text and clipboard. Ideally from within D360.
- Handle poorly defined compounds – chiral mixtures etc. and any enhanced features.
- Provide user-defined ability to handle equivalencing of structures, cores and R-Groups. For example, R and Racemate in R-Group.
- Also need to retain enhanced stereo in R-Group and core. (What to do about relative stereo between core and R-Group? The following example loses the relative relationship between the 2 stereo centres.) This is NOT currently handled by any approach.



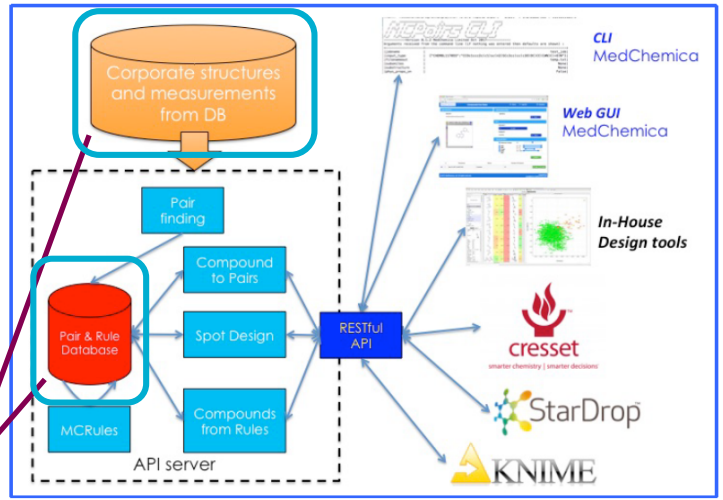
MCPAIRS
ENTERPRISE

MedChemica

CREATING A STEP CHANGE IN MEDICINAL CHEMISTRY

MCPairs Enterprise provides access to MedChemica's leading Artificial Intelligence and exploitation platform. For organisations with compounds and measurements, MCPairs will extract knowledge and build a unique database of corporate knowledge. MCPairs uses the method most trusted by medicinal chemists: Matched Molecular Pair Analysis and puts it to work helping their chemists design and progress better compounds faster.

Database, API and GUI's. MCPairs can be configured to automatically update from internal databases; constantly keeping the knowledge up to date. This enables all users to instantly look up matched pairs and perform SAR analysis and free up computational chemists to concentrate on high value activities.



Matched-Pairs

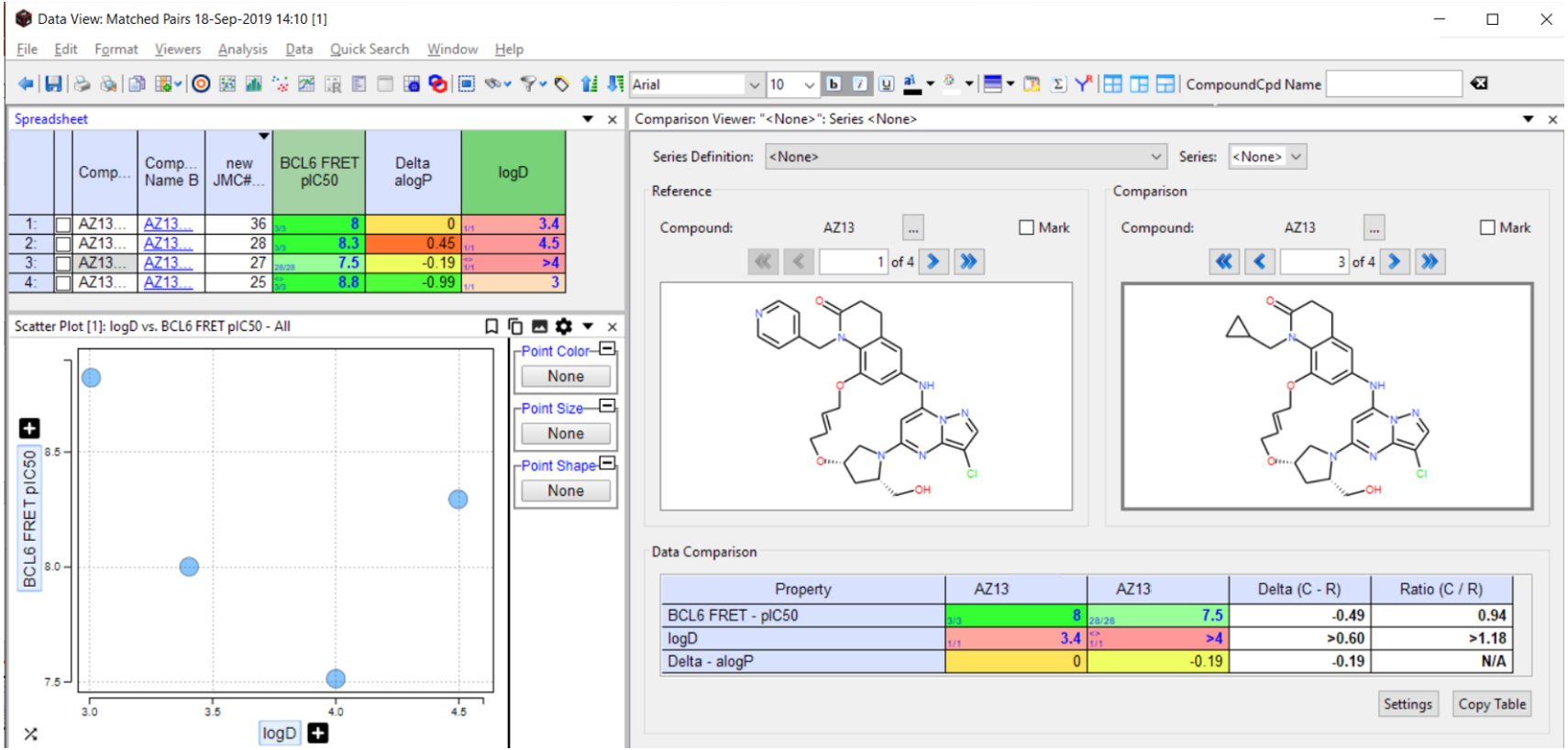
“AZ have especially pair-dense data.

Prediction is 2000 CPUs for 1 month. With current 140 CPUs on HPC this will take all year (840K done so far and expected to slow down).

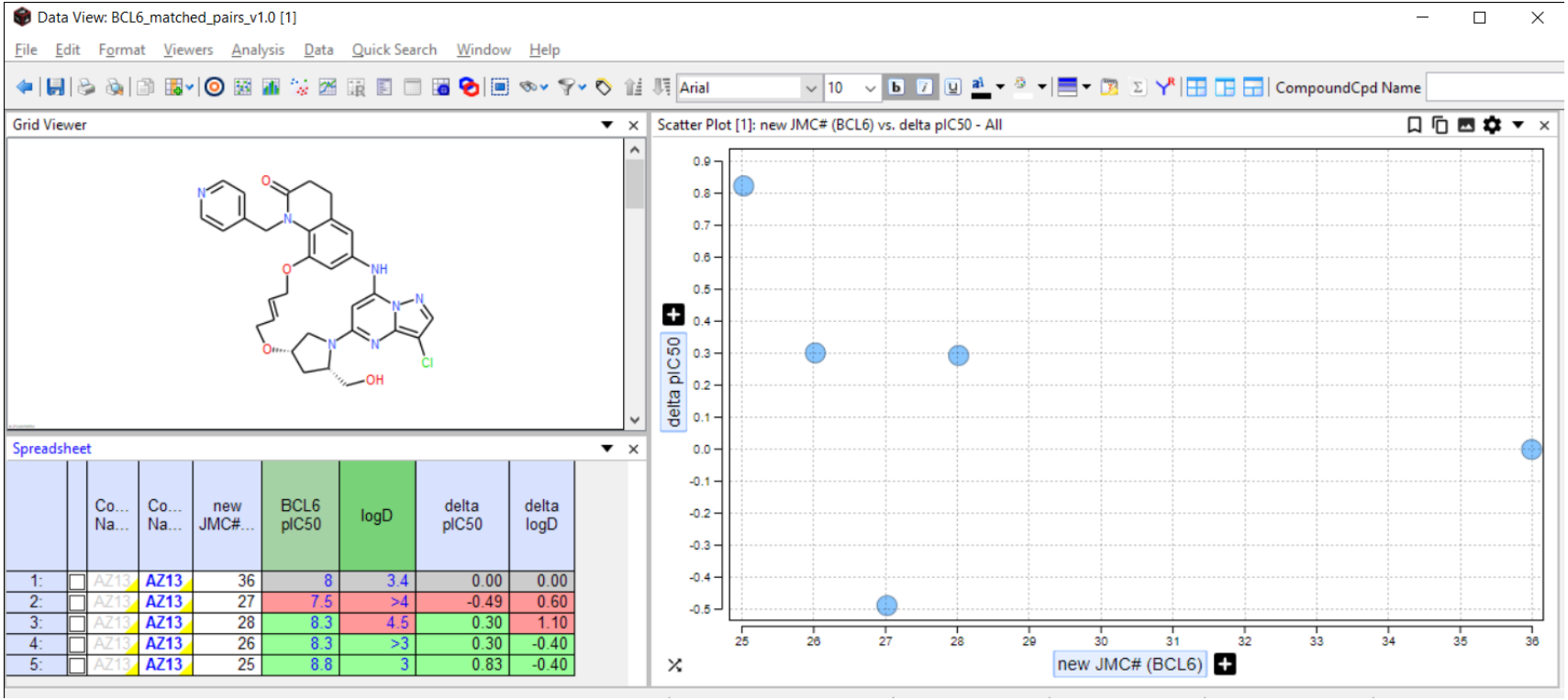
Running pair-finding interferes with update of existing information for current projects. (Data currently available through Ideal.)”

- Data availability through D360
- Performance of federated teeid view unacceptable due to Union SQL statement
- ETL daily build into datamart using Pipeline PIlot

Matched-Pairs: D360 Comparison viewer



Matched-Pairs: D360 formulas



What next for AZ & D360?

- v19 rollout in Oct19
- HTS data integration
- More reliable curve viewer
- Integration of AI/ADD tools, 3D vis tools
- Future infrastructure

Summary

- Swapping/modifying data sources (GSM) requires significant planning
- D360 useful in design meetings for visualising data on real cpds
- Virtual & Real (Cpd Union) data category aids design/analysis – still WIP
- 3D visualisation is important to AZ going forward
- R-group analysis and MMPA still WIP

Acknowledgements

- William McCoull
- Nick Tomkinson
- Andrew Poirrette
- Emma Evertsson
- Scott Throner
- Wolfgang Klute
- Jon Winter-Holt
- Hongming Chen
- Frank Kilty
- Kalyan Ponamalli
- Science & Enabling Units IT
- Ian Ingram
- Dennis Powell
- Dave Lowis
- Certara D360 team

Confidentiality Notice