

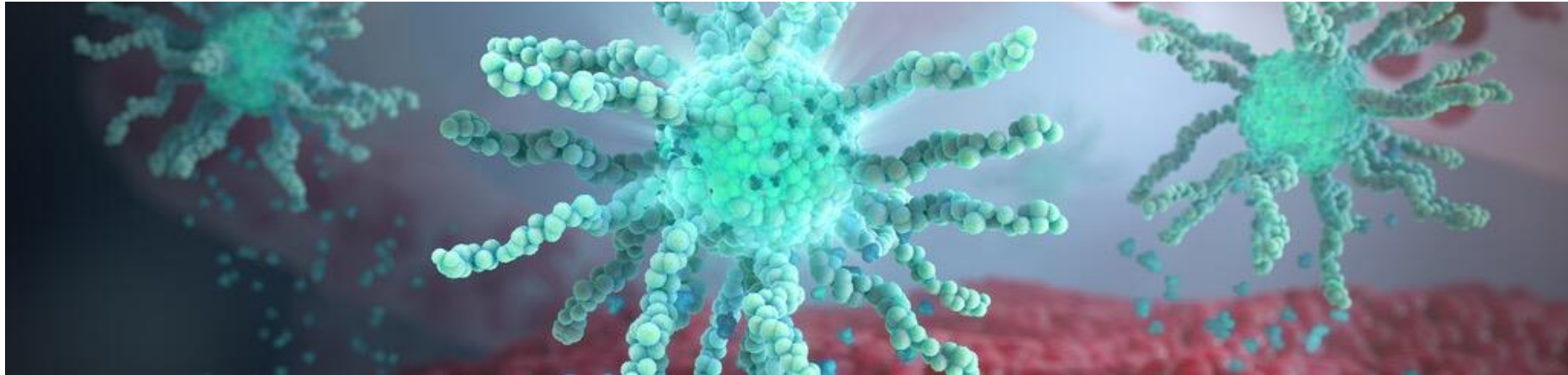
AZ workflows involving D360

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

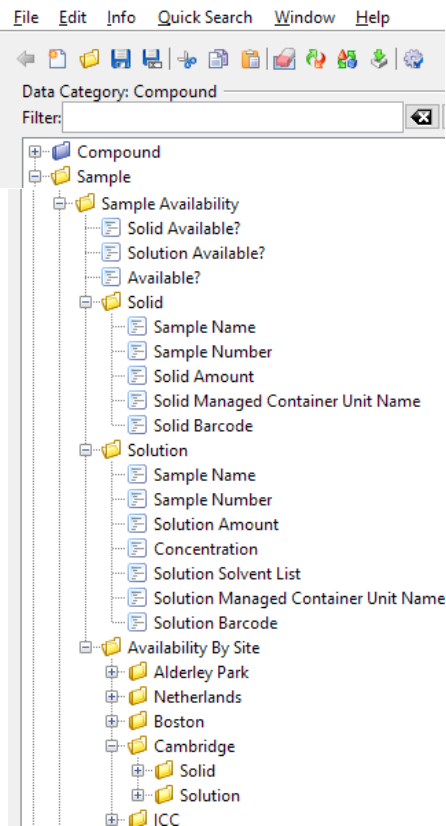
Medchemica Matched Molecular Pairs

6

Summary



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

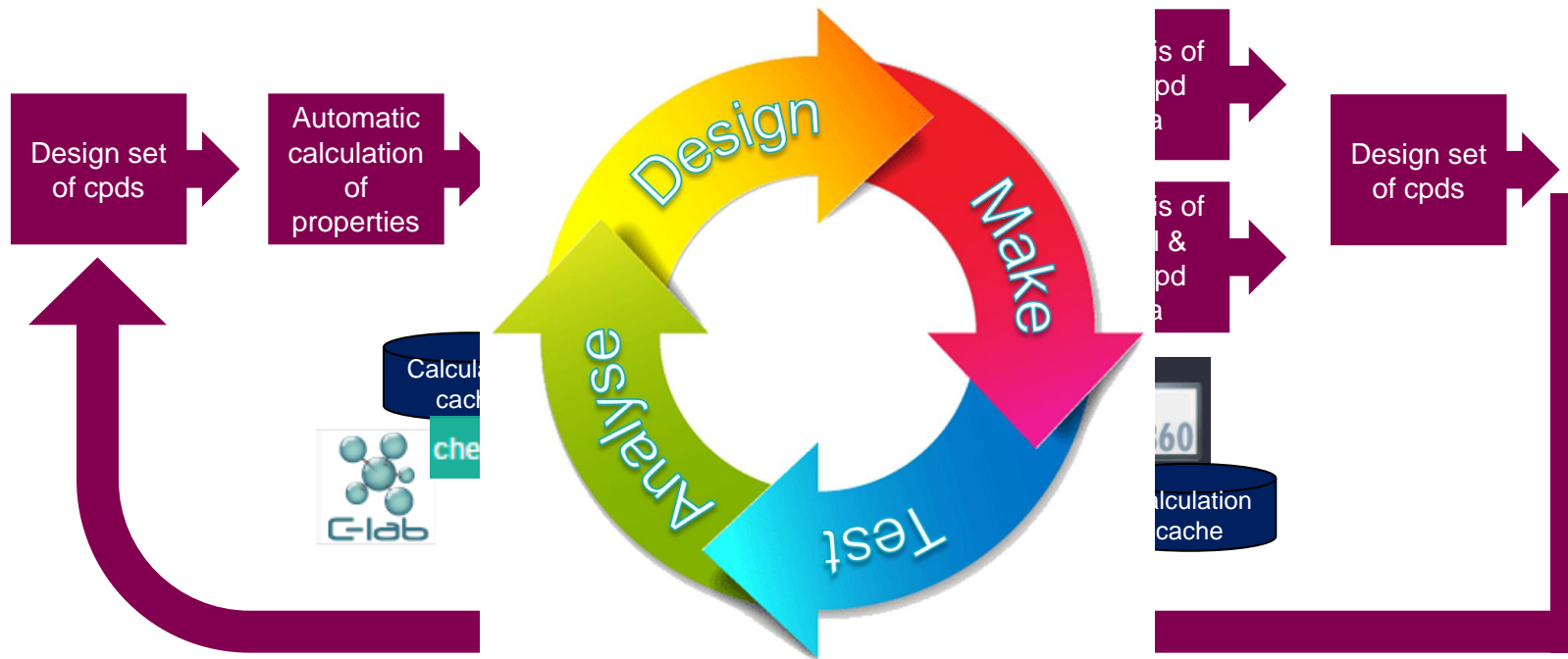
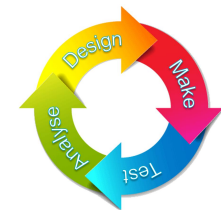
	Compound Name	CSS SName	CSS SN	CSS amt (mg)	CSS site	CSS Barcode	PLS SName	PLS SN	PLS amt (ml)	PLS concn...	PLS site	PLS Barcode
1:	<input type="checkbox"/> Cpd 11	Cpd 11 -004	SN1052278	17.4	Specs Netherlands	5006004946	Cpd 11 -004 Cpd 11 -004 Cpd 11 -004	SN105227 SN105227 SN105227	0.163 0.726 0.556	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:	<input checked="" type="checkbox"/>	13	14.7			0.31		0.14	
2:	<input checked="" type="checkbox"/>	11	17.4			0.56	0.16		0.73
3:	<input checked="" type="checkbox"/>	5	78.1		2.8	0.08	0.05		



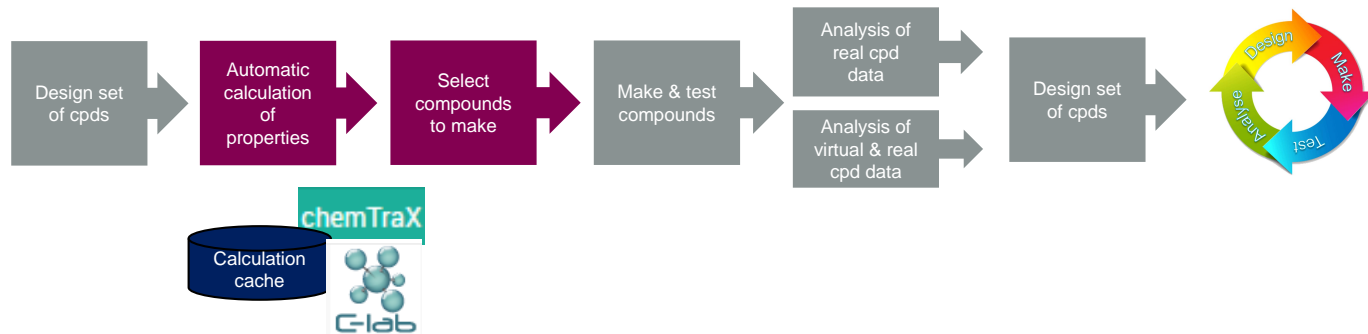
Design workflow

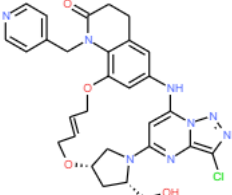
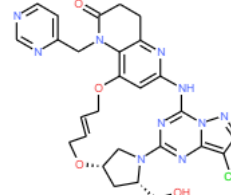
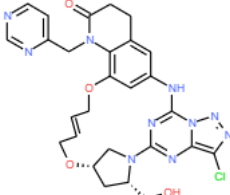
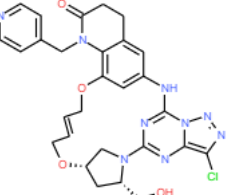
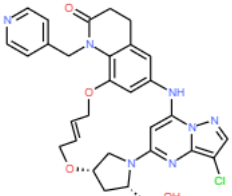


- D360 facilitates the analysis step of the design-make-test-cycle
- D360 is a project query, visualisation and analysis tool



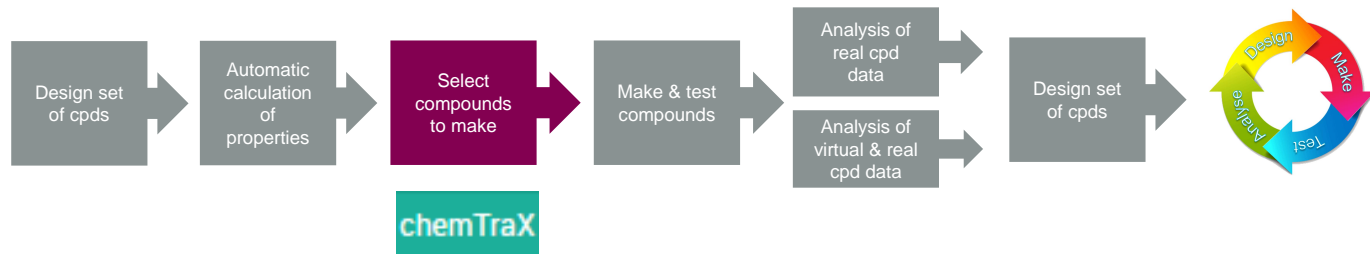
Design workflow



FOR MAKE 4 tickets	<p>460-11: IDEA</p>  <p>FOR MAKE AZlogD 2.9 ITACS ✓ ClogP 2.6 HLM_CLint 114 11 minutes</p>	<p>460-8: IDEA</p>  <p>FOR MAKE AZlogD 2.1 ITACS ✓ ClogP 0.3 HLM_CLint 59 11 minutes</p>	<p>460-9: IDEA</p>  <p>FOR MAKE AZlogD 2.5 ITACS ✓ ClogP 1.2 HLM_CLint 44 11 minutes</p>	<p>460-10: IDEA</p>  <p>FOR MAKE AZlogD 3 ITACS ✓ ClogP 2.1 HLM_CLint 62 11 minutes</p>	
	<p>460-7: REFERENCE</p> 				
	REFERENCE 1 tickets				
	<ul style="list-style-type: none"> • Calculations displayed real-time against each compound • Design team can make informed compound selection <ul style="list-style-type: none"> – Projects choose what calculations are important to them 				



Design workflow



AstraZeneca BCL6 Tracking Board

Row: Stacked Column: Workflow Color: Series View As: William McCoull

DesignSet Filter Export Color Alerts Search... Text Q

Active	Parked	Archived	READY TO MAKE	IN MAKE	IN TEST	IN ANALYSIS	COMPLETE
			1 tickets	0 tickets	0 tickets	0 tickets	0 tickets
Stacked 2 tickets	461: IDEA lower MW Project 0/1 Compounds 10 hours		460: IDEA Lower lipophilicity Project 0/1 Compounds (3) a few seconds				

NOT FOR MAKE 3 tickets

460-10: IDE4 A2logD 3 ITACS ✓ CllogP 2.1 HLM_Clit 62 a minute	460-9: IDE4 A2logD 2.5 ITACS ✓ CllogP 1.2 HLM_Clit 44 2 minutes	460-11: IDE4 A2logD 2.9 ITACS ✓ CllogP 2.8 HLM_Clit 114 2 minutes
---	---	---

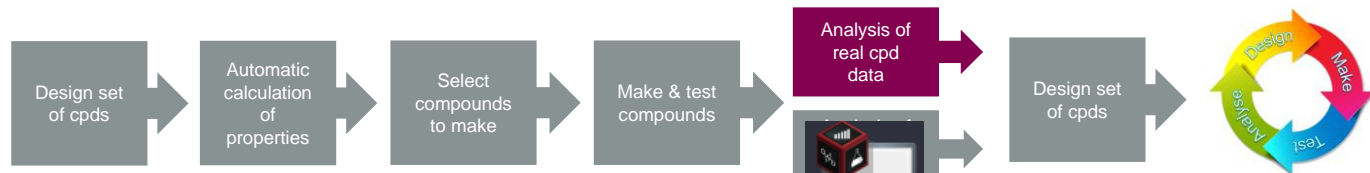
FOR MAKE 1 tickets

460-8: IDE4 A2logD 2.1 ITACS ✓ CllogP 0.3 HLM_Clit 59 14 minutes
--

- Design set moved to next lane
- Ready for synthesis



Design workflow: R



Data View: BCL6_UGM_query_WM10Sep19 [1]

Form viewer - equations would add value

AZ curve viewer - needs rebuilding for reliability

Predefined plots to aid analysis of design set

Synthesis information including lab notebook ref

Chemical series manually copied into annotation to allow
 1] display in form viewer
 2] improve query speeds

Crystal structure info & link to view – what about 3D modelling in D360?

Design set ID use

Clab calculations retrieved from cache or run

Make protocol saved as & used in equation

Grid Viewer

Spreadsheet

	Co... Cp...	new JMC#	BCL6 series	Ion Cl...	CF	CFID	clab HLM CLint	clab HLM CLint confidence	clogP	DS ID latest	syn ref	Largest Ring Size	macrocycle?
1:	AZ13	5	PP	Acid			4	0.44	1.4	BCL6-461	EN06	6	acyclic
2:	AZ13	7	PP	Neutr			3	0.55	2.2	BCL6-461	EN07	6	acyclic
3:	AZ13	8	PP	Neutr			15	0.45	1.8	BCL6-461	EN07	6	acyclic
4:	AZ13	10	PP	Acid			7	0.41	1.7	BCL6-461	EN07	16	macrocycle
5:	AZ13	11	PP	Neutr			24	0.62	1.3	BCL6-461	EN07	16	macrocycle
6:	AZ13	12	PP	Neutr	0.015		48	0.52	2	BCL6-461	EN07	16	macrocycle
7:	AZ13	36	PP	Neutr	0.0099		238	0.43	2.7	BCL6-461	EN07	16	macrocycle

Form viewer details:

AZ: 36

logD: 3.4

sol (µM): 16

hu ppb: 0.7

RH: 125

HLM: >300

BCL6 IC50: 0.009

series: PP

Spreadsheet filters:

Compound - Cpd Name

List: 0 Item(s)

In List

Not In List

List Match Mode

Equals

Contains

Starts with

Ends With

macrocycle?

All

acyclic

macrocycle

DS ID - all

Contains text: BCL6-461

Case sensitive

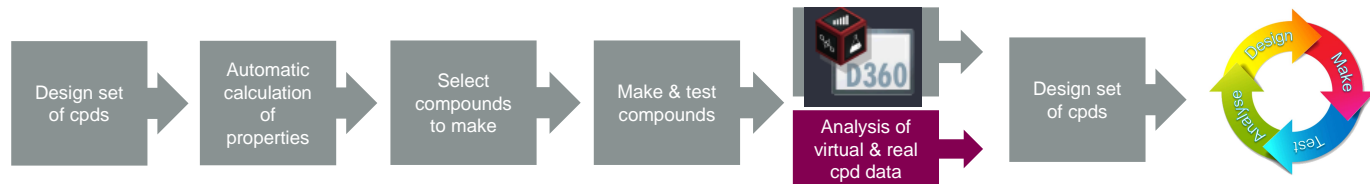
Show Empty Values

Structure:

Records: (1/38) Fields: (1/42) Marked Records: 0 Visible Records: 7 Filtered Records: 31 Calcs: 2



Design workflow: V&R



Data View: Compound Union_POC...
File Edit Format Viewers Anal

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

Grid Viewer

Form: B [Cpd Union] [Cpd Name] [36]
logD [3.4]
BCL6 [0.009]
IC50

Scatter Plot [1]: V&R AZlogD vs. clab HLM CLint (Log) - All

Scatter Plot [2]: ST00025 HLM CLint vs. clab HLM CLint (Log)

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

Spreadsheet

	Cpd...	Cpd Union	Ch...	Virtual	BCL6	BCL6	xst	resul	JMC#	logD	sol	n ppo	ST00...	ST00...	DS ID	syn	LLE	clab	clab	Largest	macrocycle?	V&R	V&R	
	Cpd...	SMILES	ISAC	cpd ID	series	FRET	link	[A]			(μM)	(% free)	rat hep	HLM	latest	ref		HLM	HLM	Ring		AZlogD	ion	
			Mat...			IC50 (μM)							CLint...	CLint...				CLint	CLint	Size			class	
1:	<input type="checkbox"/>	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:	<input type="checkbox"/>	AZ13	c1cc2c(cc...	AZ13	BCL6-461.2	PP	0.35	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:	<input type="checkbox"/>	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:	<input type="checkbox"/>	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:	<input type="checkbox"/>	AZ13	c1c2cc(c3...	AZ13	BCL6-461.5	PP	0.0029	1.98	11	2.2	10	10	38	15	BCL6-461	EN07	6.4	24	0.62	16	macrocycle	2.6	Neutr	
6:	<input type="checkbox"/>	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:	<input type="checkbox"/>	AZ13	c1nc2cc1...	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:	<input type="checkbox"/>	DC00	c1c2cc(c3...		BCL6-461.8										BCL6-461				15	0.67	16	macrocycle	2.2	Neutr

Can toggle off/on
virtual or real cpds

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

Cpd Union - Design - cpd Y/N
 N
 Y
2 of 2 Values Selected

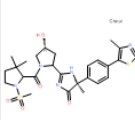
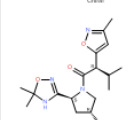
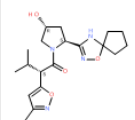
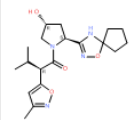
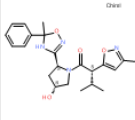
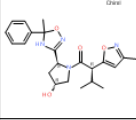
3D Visualisation - Requirements

- 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

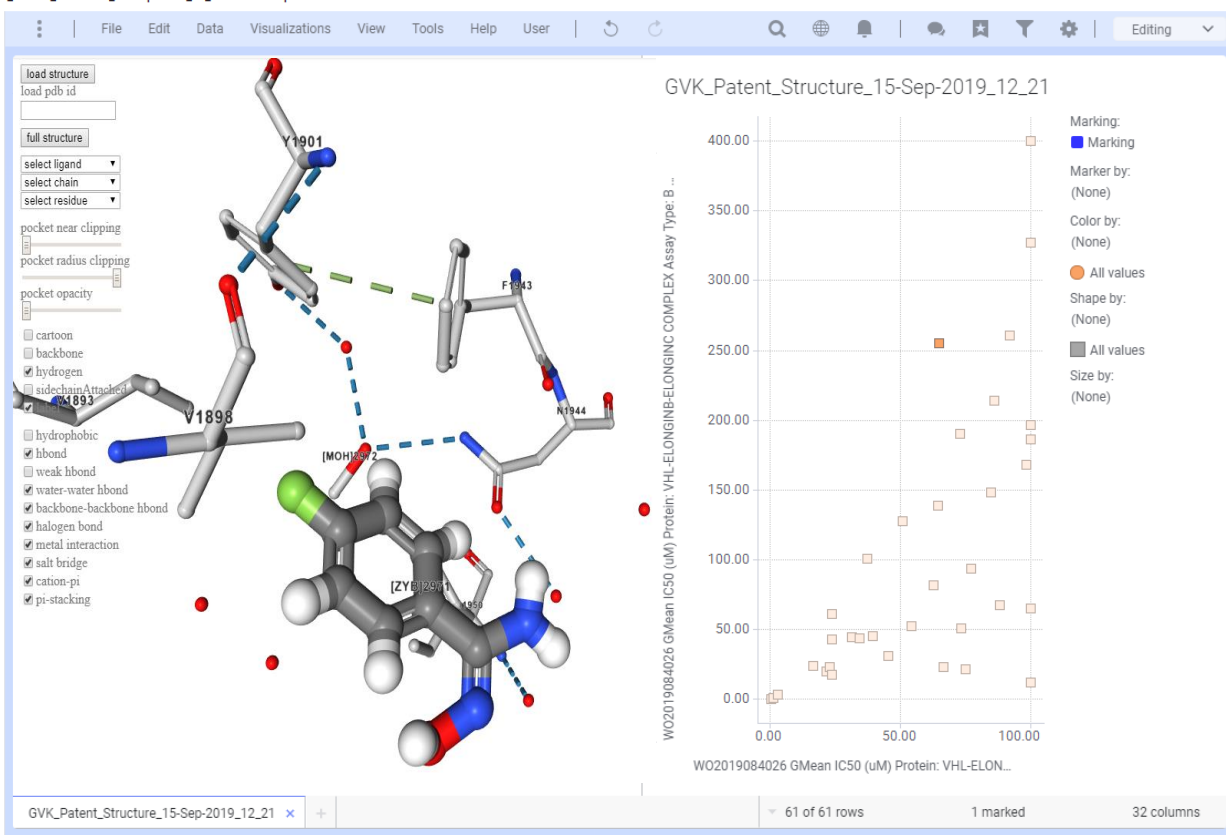


2-way link between 2D-plots, 3D ligand
visualisation and D360 dataset rows

Spreadsheet

	GVK Id	GVK Structure	Concat Chem Connect ID	Concat AZ Number	Concat Journal Patent N...	Concat Normalized Jov...	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...	Q...
1:	124222798		407									100.000	83.926		
2:	124222802		407									4/4	45.400	12.800	
3:	124222804		407									1/1	100.000	150.000	
4:	124222805		407									1/1	73.200	16.400	
5:	124222806		407									2/2		150.000	
6:	124222808		407									2/2	16.191	6.538	

GVK_Patent_Structure_15-Sep-2019_12_21 - TIBCO Spotfire



	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, 6923, 7428, 7428, 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, 6923, 6923, 7428, 7428, 7428, 7428, 7428, 7428, 7428, 8453, 8453,	
7:	124222810		407599051		WO 2019/084026	WO2019084026	6921, 6923, 6923, 6923, 6923, 6923, 6923, 7428,	

2-way link

3D Structure Viewer

load structure

load pdb id

full structure

select ligand

select chain

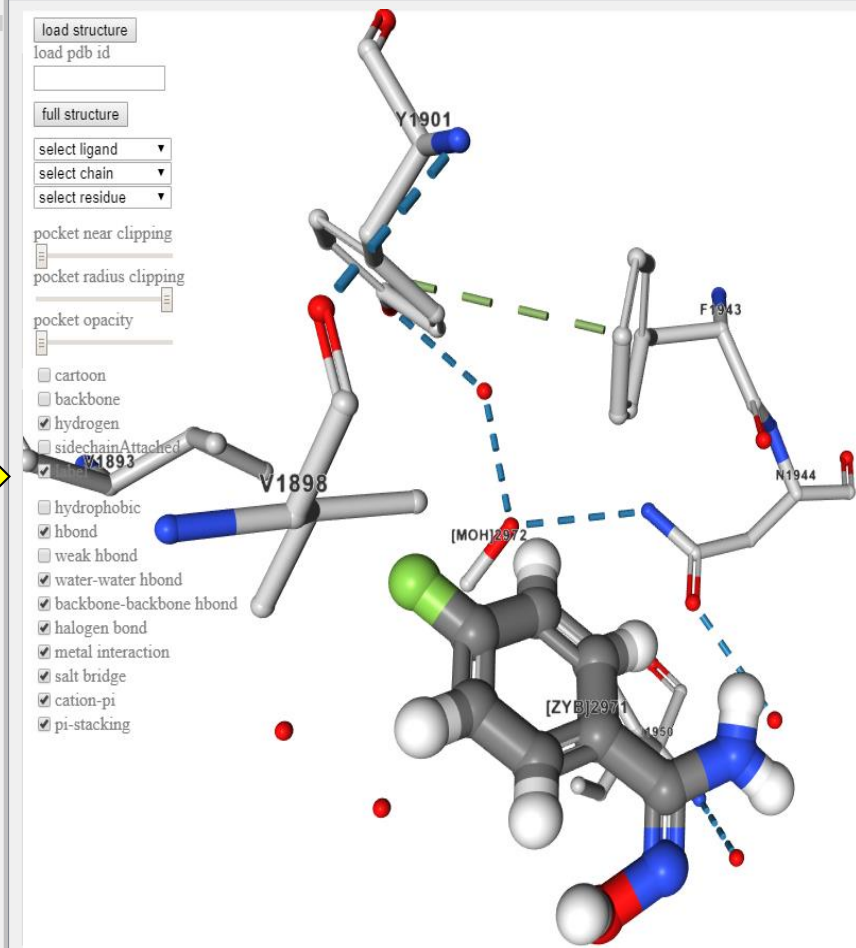
select residue

pocket near clipping

pocket radius clipping

pocket opacity

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

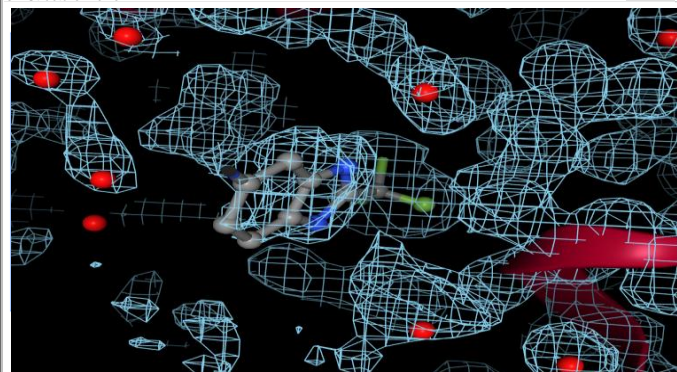


	GVK Id	GVK Structure	Concat Chem Connect ID	Concat AZ Number	Concat Journal Patent N...	Co Normaliz
1:	124222798		407591518		WO 2019/084026 A1	WO2019
2:	124222802		407597369		WO 2019/084026 A1	WO2019
3:	124222804				WO 2019/084026 A1	WO2019

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

R-Group Analysis

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

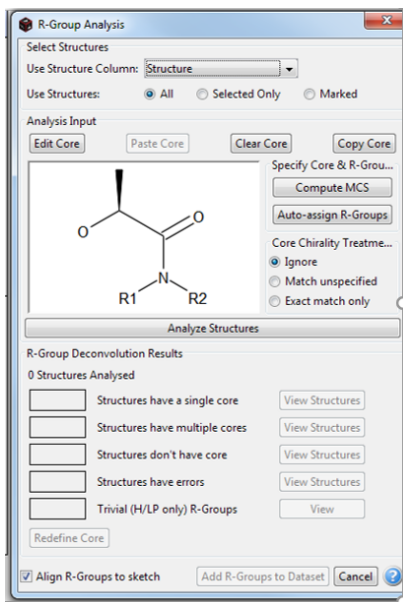
Related applications

- Matched pairs
- Spark



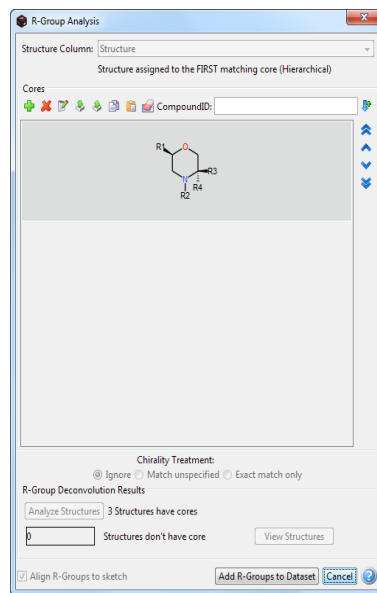
D360 R-Group Analysis

v 9



The screenshot shows the 'R-Group Analysis' dialog box in version 9. It features a 'Select Structures' section with a dropdown menu set to 'Structure' and radio buttons for 'All', 'Selected Only', and 'Marked'. Below this is an 'Analysis Input' section with buttons for 'Edit Core', 'Paste Core', 'Clear Core', and 'Copy Core'. A chemical structure of a chiral amide is displayed in the center, with R1 and R2 labels. To the right, there is a 'Specify Core & R-Group...' section with a 'Compute MCS' button and an 'Auto-assign R-Groups' button. Below that, 'Core Chirality Treatment' options are shown: 'Ignore' (selected), 'Match unspecified', and 'Exact match only'. At the bottom, there is an 'Analyze Structures' button and a section for 'R-Group Deconvolution Results' with five empty checkboxes and corresponding 'View Structures' buttons. The footer includes 'Align R-Groups to sketch', 'Add R-Groups to Dataset', and 'Cancel' buttons.

v 10

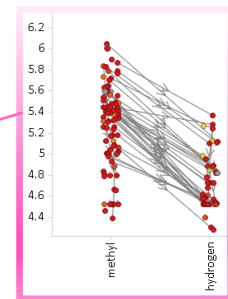
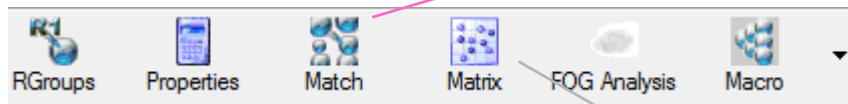
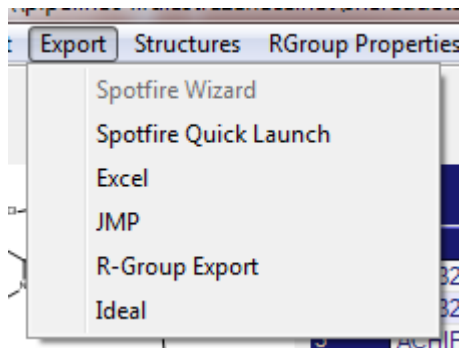
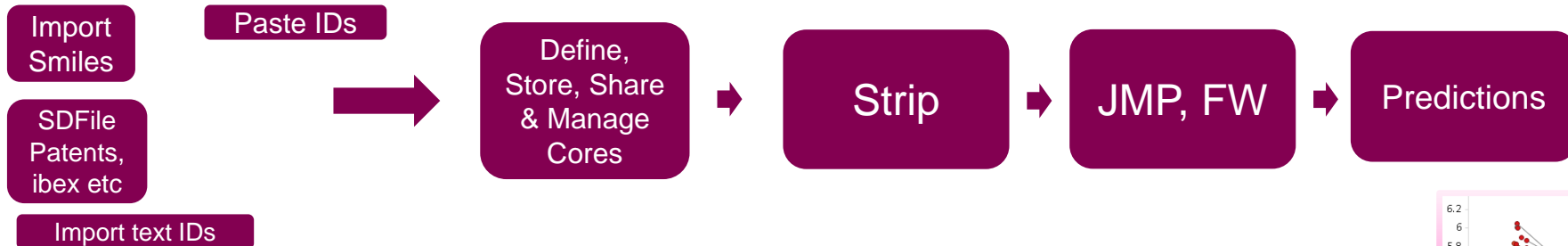


The screenshot shows the 'R-Group Analysis' dialog box in version 10. The 'Structure Column' is set to 'Structure' and the text below reads 'Structure assigned to the FIRST matching core (Hierarchical)'. The 'Cores' section has a 'CompoundID:' field. A chemical structure of a chiral amide is shown in the center, with R1, R2, R3, and R4 labels. Below the structure, 'Chirality Treatment' options are shown: 'Ignore' (selected), 'Match unspecified', and 'Exact match only'. The 'R-Group Deconvolution Results' section shows 'Analyze Structures' with '3 Structures have cores' and '0 Structures don't have core', with a 'View Structures' button. The footer includes 'Align R-Groups to sketch', 'Add R-Groups to Dataset', and 'Cancel' buttons.

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



fRGS Workflow(s)



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

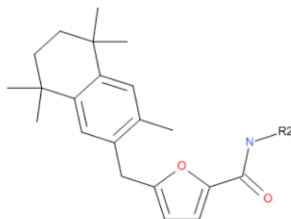
Hole-filling and creates smiles for Chemtrax

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

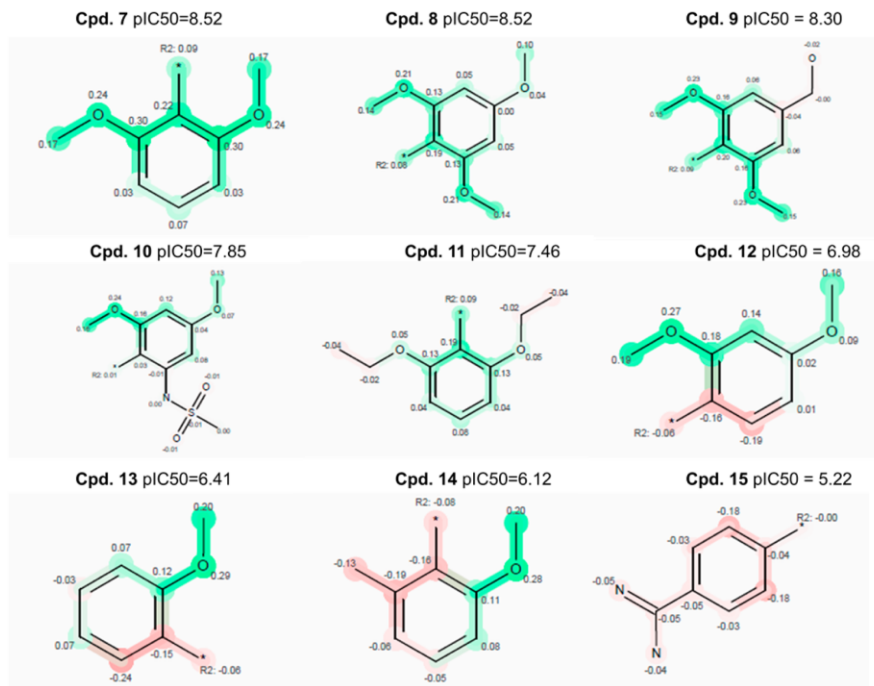


Machine-Learning Atom Contributions

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

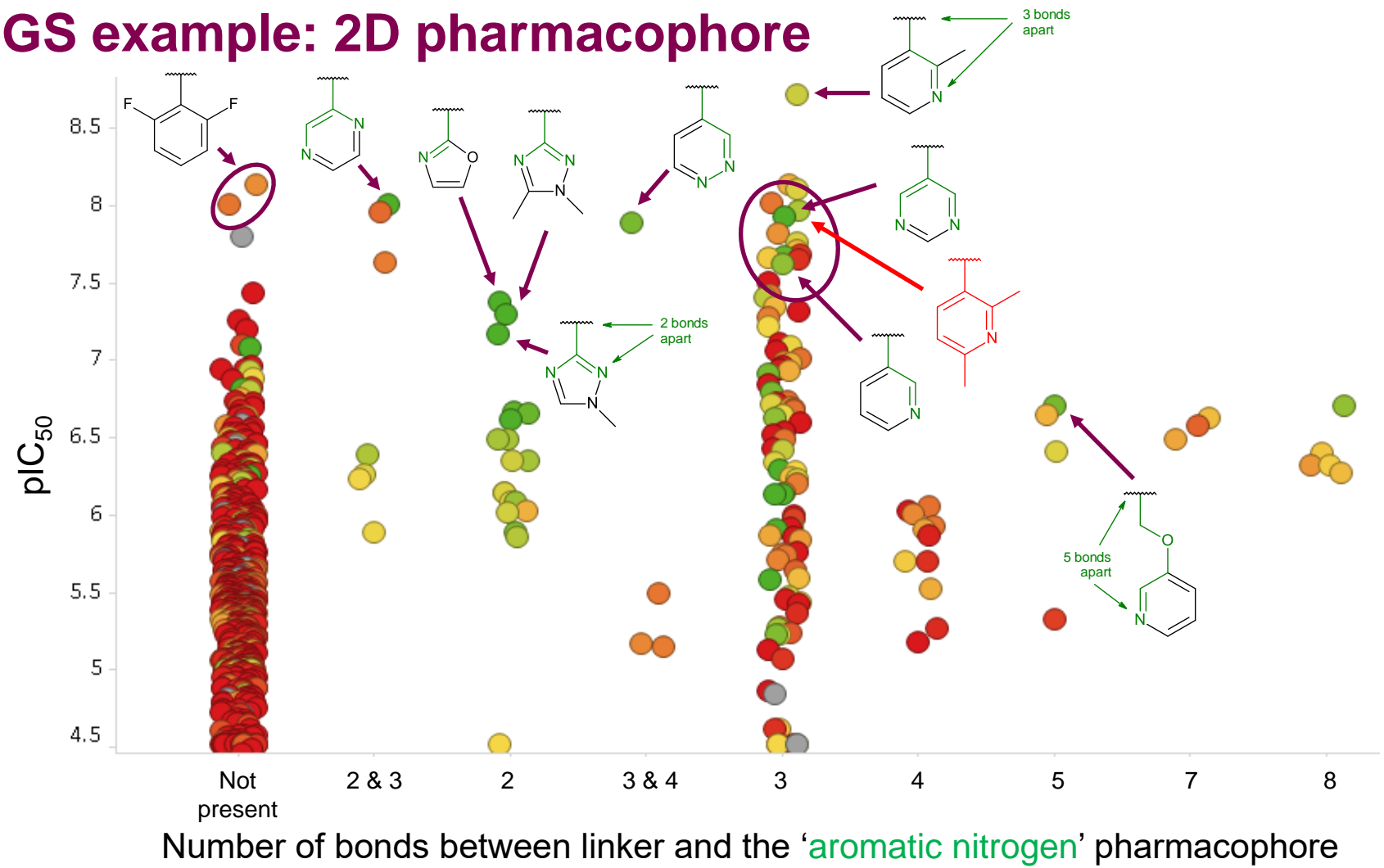


Comparative performance

		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	?

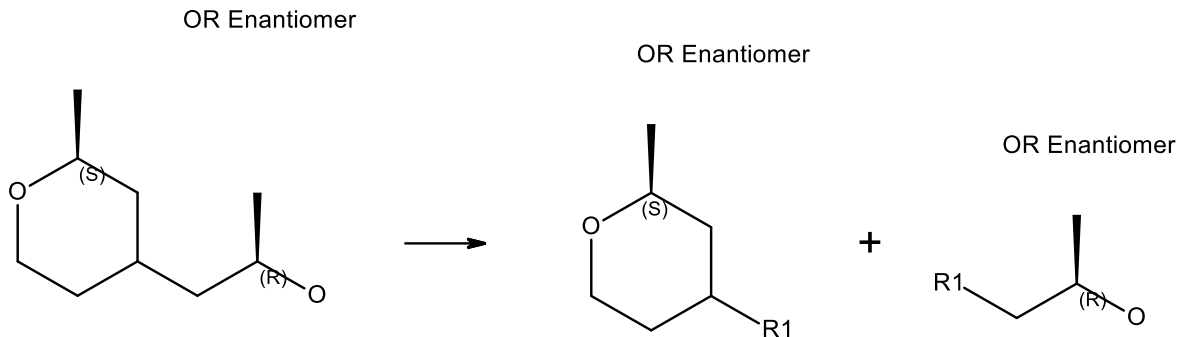


fRGS example: 2D pharmacophore



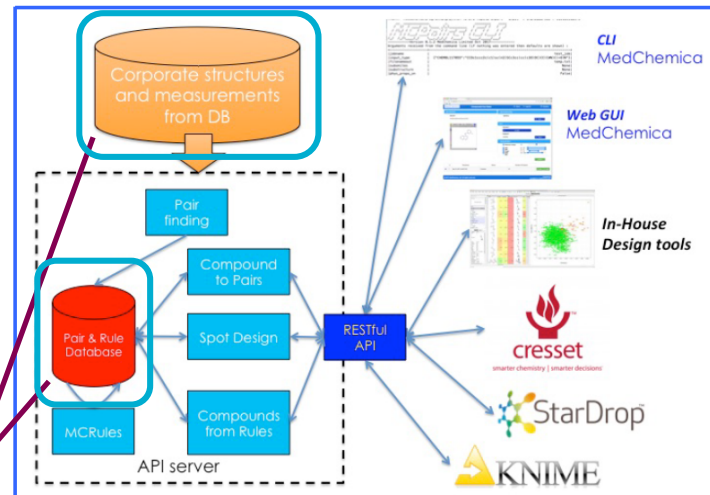
R-Group Requirements:

- 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 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

Data View: Matched Pairs 18-Sep-2019 14:10 [1]

File Edit Format Viewers Analysis Data Quick Search Window Help

Spreadsheet

	Comp...	Comp... Name B	new JMC#...	BCL6 FRET pIC50	Delta alogP	logD	
1:	<input type="checkbox"/>	AZ13...	AZ13...	36	8	0	3.4
2:	<input type="checkbox"/>	AZ13...	AZ13...	28	8.3	0.45	4.5
3:	<input type="checkbox"/>	AZ13...	AZ13...	27	7.5	-0.19	>4
4:	<input type="checkbox"/>	AZ13...	AZ13...	25	8.8	-0.99	3

Comparison Viewer: "<None>": Series <None>

Series Definition: <None> Series: <None>

Reference

Compound: AZ13 Mark

1 of 4

Comparison

Compound: AZ13 Mark

3 of 4

Scatter Plot [1]: logD vs. BCL6 FRET pIC50 - All

Point Color: None

Point Size: None

Point Shape: None

Chemical Structure 1: CC1=CC=C(C=C1)CN(C1=CC=C2C(=C1)C(=O)N(C2)C1=CC=C(C=C1)OC1=CC=C(C=C1)N1C=NC=C1Cl)C1=CC=C(C=C1)O

Chemical Structure 2: CC1=CC=C(C=C1)CN(C1=CC=C2C(=C1)C(=O)N(C2)C1=CC=C(C=C1)OC1=CC=C(C=C1)N1C=NC=C1Cl)C1=CC=C(C=C1)O

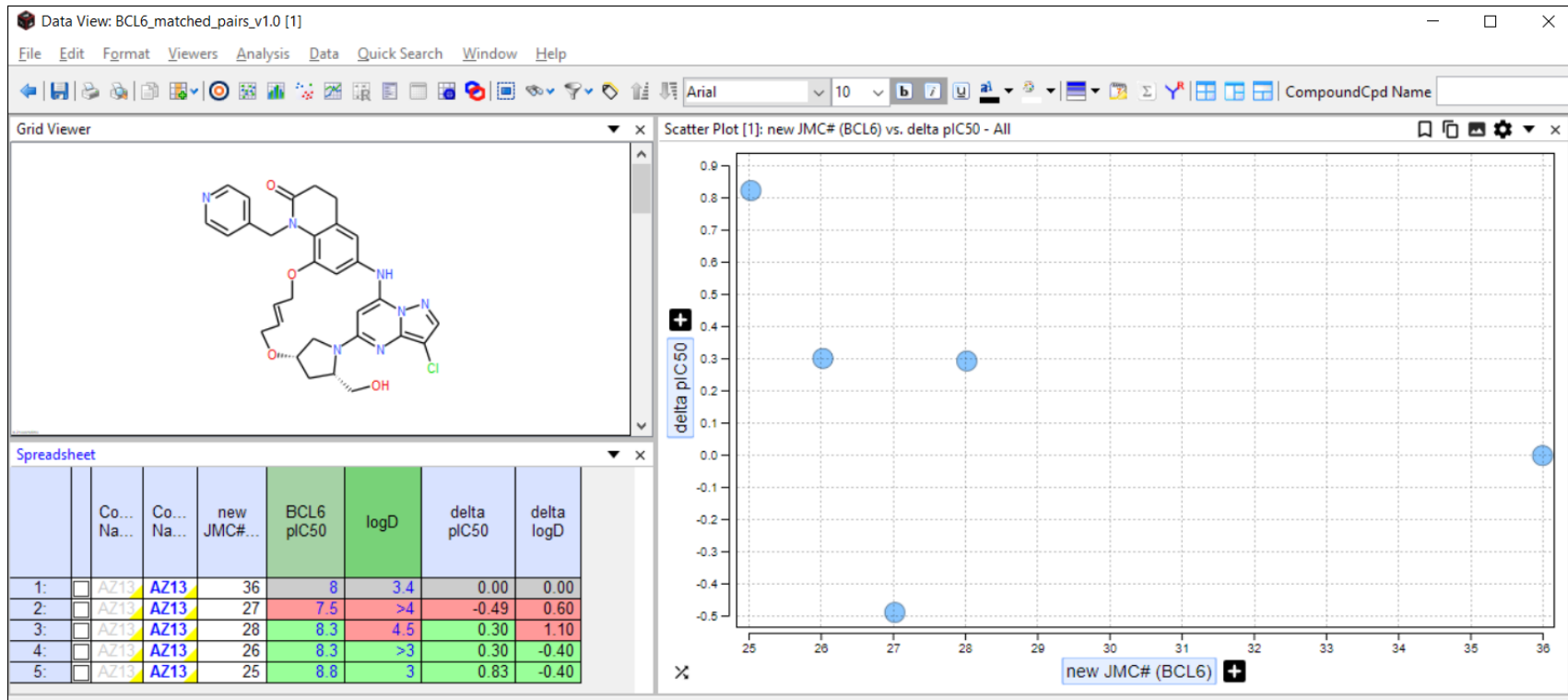
Data Comparison

Property	AZ13	AZ13	Delta (C - R)	Ratio (C / R)
BCL6 FRET - pIC50	8	7.5	-0.49	0.94
logD	3.4	>4	>0.60	>1.18
Delta - alogP	0	-0.19	-0.19	N/A

Settings Copy Table



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 Lewis
- Certara D360 team

Confidentiality Notice

This file is private and may contain confidential and proprietary information. If you have received this file in error, please notify us and remove it from your system and note that you must not copy, distribute or take any action in reliance on it. Any unauthorized use or disclosure of the contents of this file is not permitted and may be unlawful. AstraZeneca PLC, 1 Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge, CB2 0AA, UK, T: +44(0)203 749 5000, www.astrazeneca.com

