A Picture Paints a Thousand Words – Visualisation of SAR

John Cumming, AstraZeneca, Alderley Park, UK

AstraZeneca

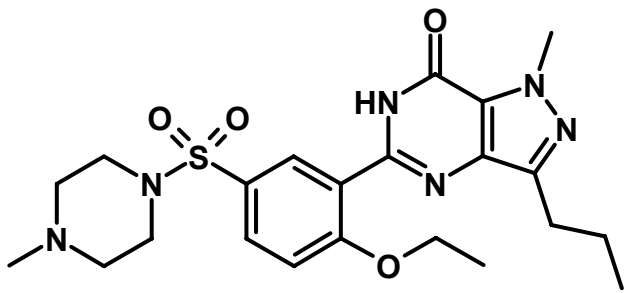
Outline

- Why visualisation?
- Multi-parameter visualisation
- Structure fragmentation
- Matched pair analysis
- Visualisation in library design
- Conclusions

AstraZeneca

A Picture Paints a Thousand Words

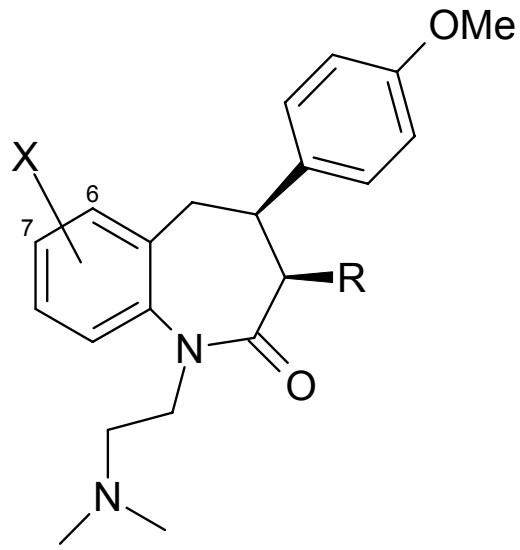
1-[4-ethoxy-3-(6,7-dihydro-1-methyl-7-oxo-3-propyl-1*H*-pyrazolo[4,3-*d*]pyrimidin-5-yl)phenylsulfonyl]-4-methylpiperazine



AstraZeneca

Potency of benzazepinone calcium channel blockers

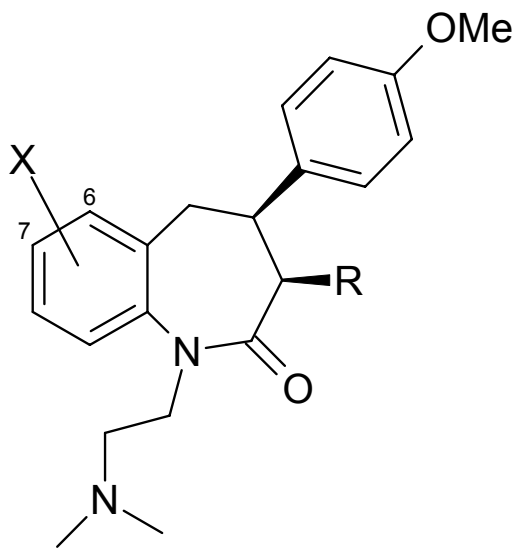
X	R	pIC_{50}
H	$OCOCH_3$	5.1
6-Cl	$OCOCH_3$	5.8
6-CH_3	$OCOCH_3$	5.7
6-CN	$OCOCH_3$	6.5
6-$CONH_2$	$OCOCH_3$	4.5
6-OCH_3,7-Br	CH_3	7.2
7-OC_6H_5	$OCOCH_3$	6.6
7-$OCONHCH_3$	$OCOCH_3$	5.1

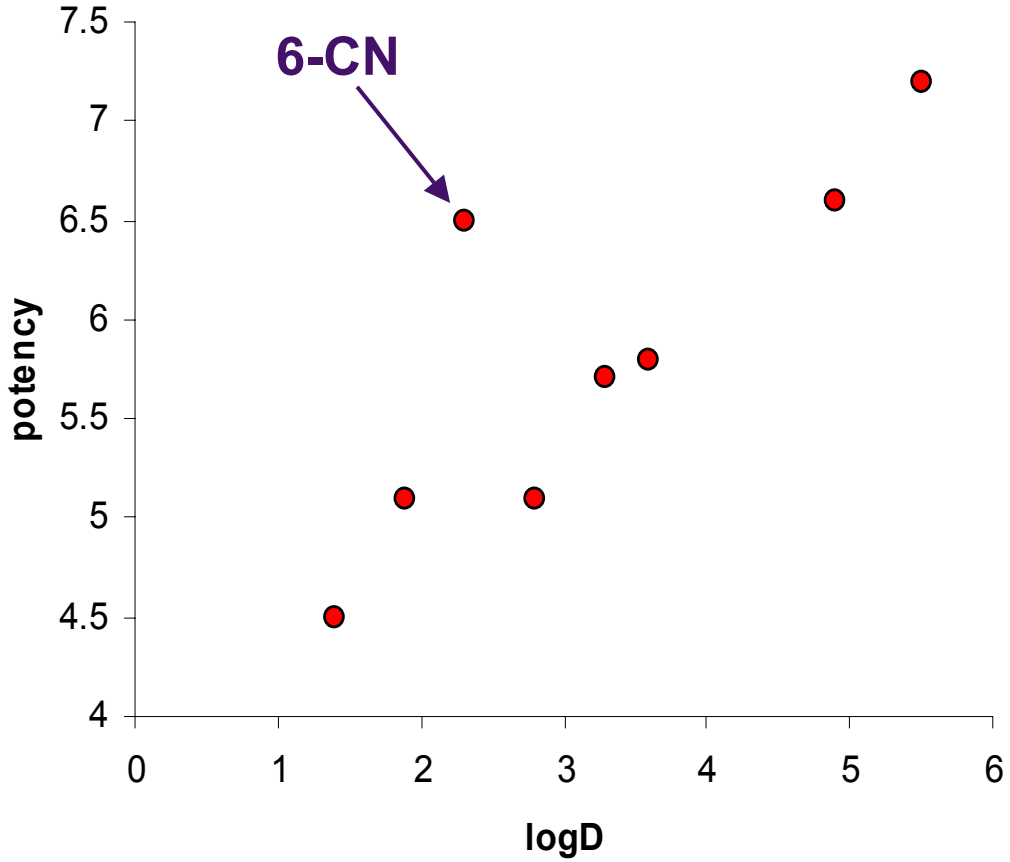


Which is the most interesting compound?

AstraZeneca

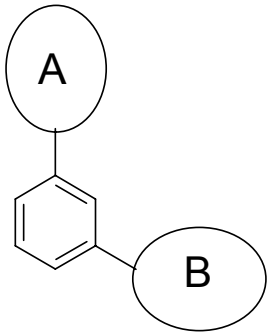
Potency of benzazepinone calcium channel blockers

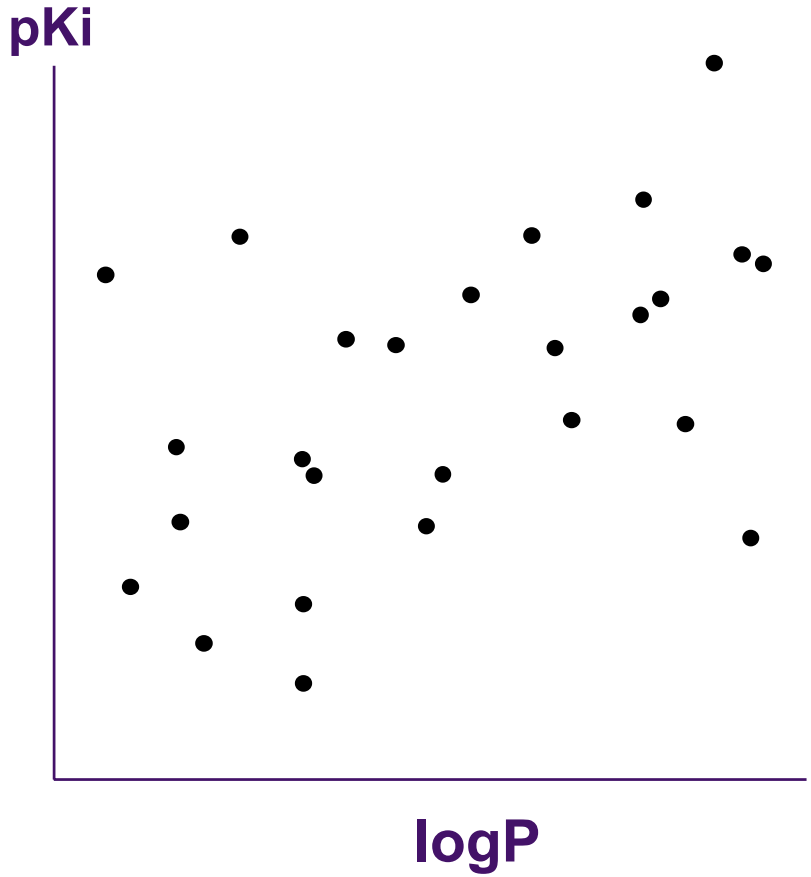




AstraZeneca

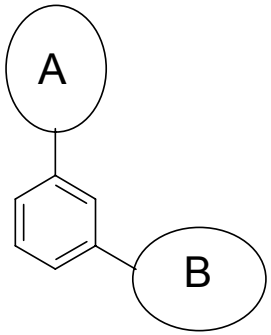
QSAR using Biological Neural Nets !

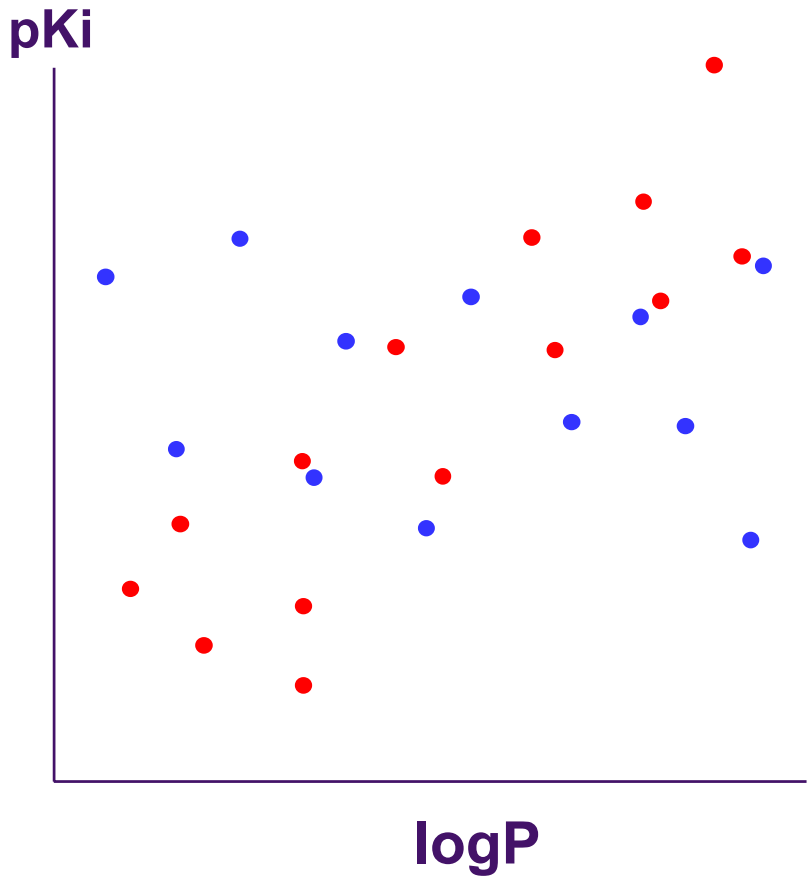




AstraZeneca

QSAR using Biological Neural Nets !

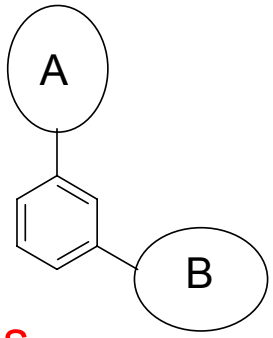




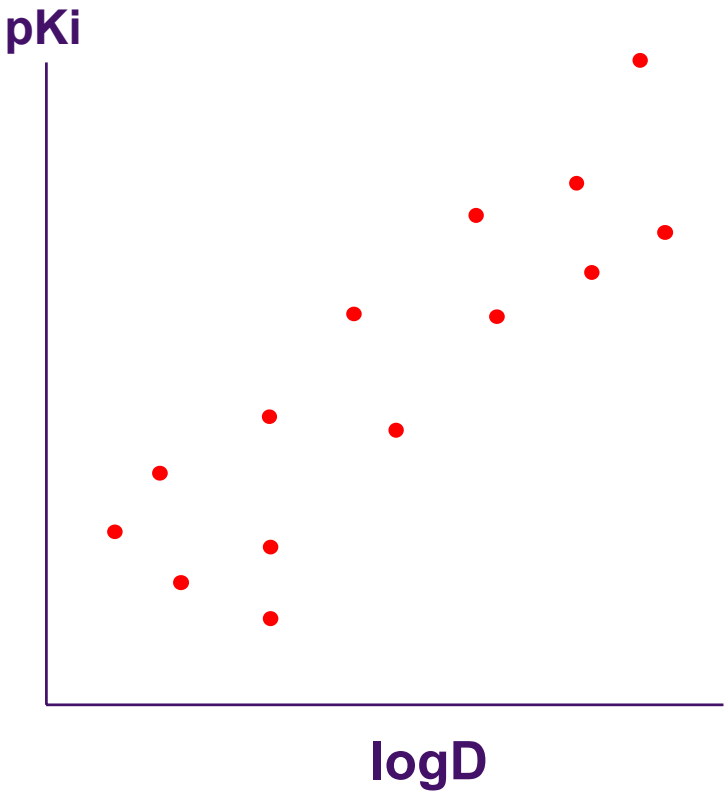
QSAR of A-Ring Variations

QSAR of B-Ring Variations

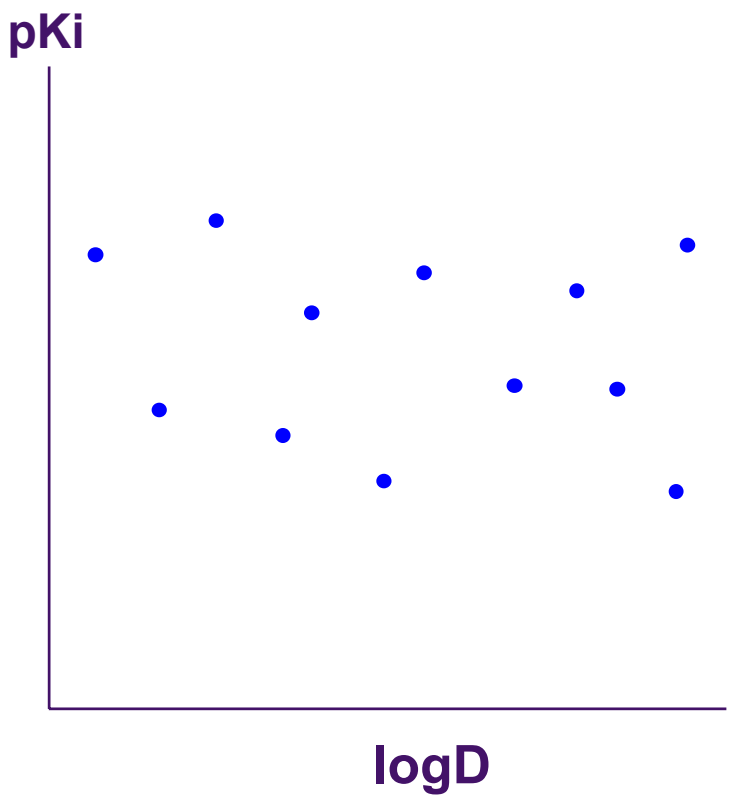
AstraZeneca



QSAR of A-Ring Variations



QSAR of B-Ring Variations



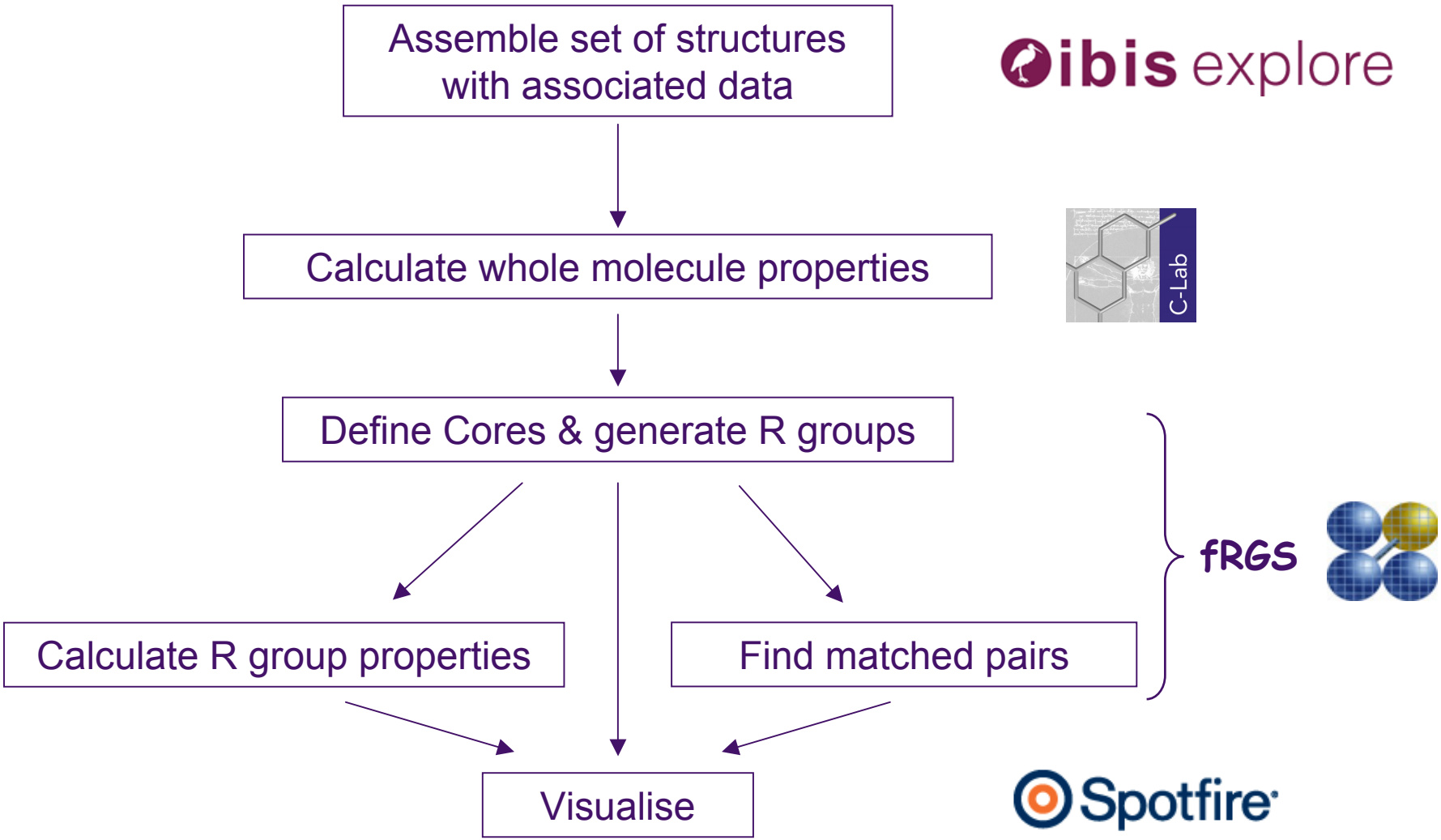
AstraZeneca

Multi-parameter Visualisation

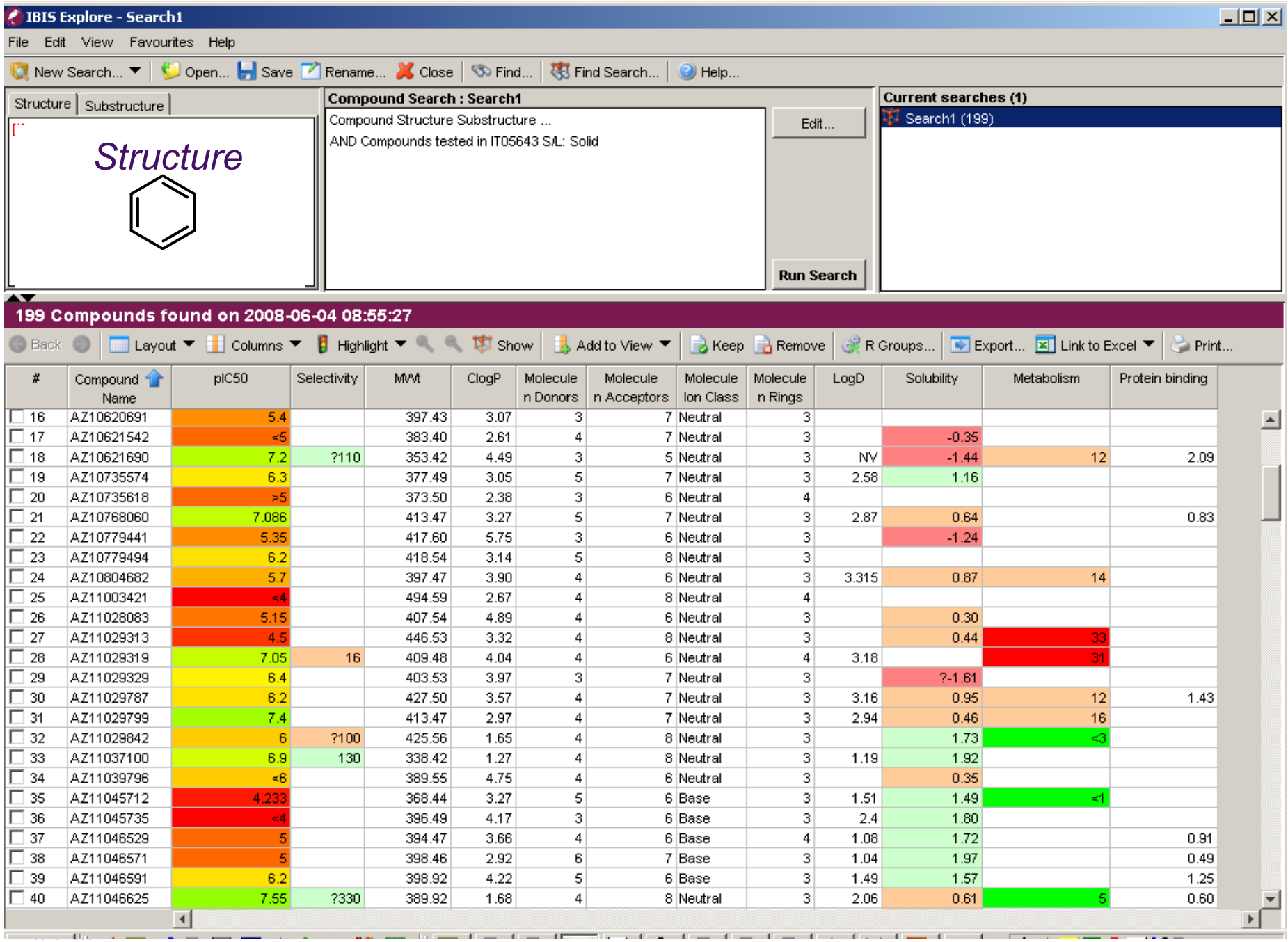
- Lead series require optimisation against many parameters
  - Potency
  - Selectivity
  - Physical properties e.g. solubility, protein binding
  - DMPK e.g. metabolism, absorption, Cyp inhibition
  - Toxicity e.g. hERG activity, genotox, off-target activity
- Activities can be related to multiple physicochemical descriptors
  - Lipophilicity – logP, logD
  - Size, shape, flexibility
  - H-bonding
  - Electronics e.g. sigma values, pKa
  - Sub-structural features, chemical series

AstraZeneca

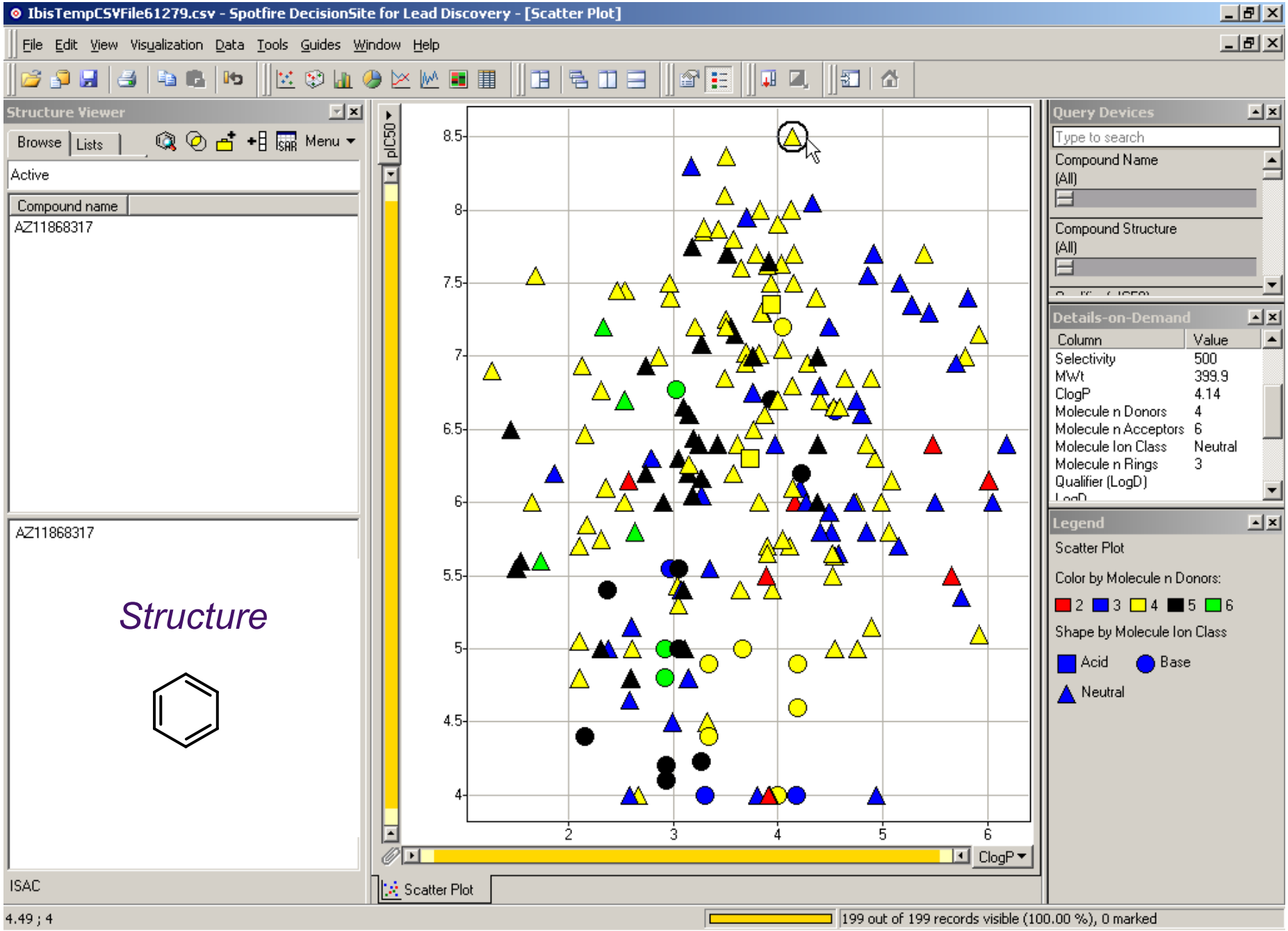
AZ SAR/Data Analysis Workflow



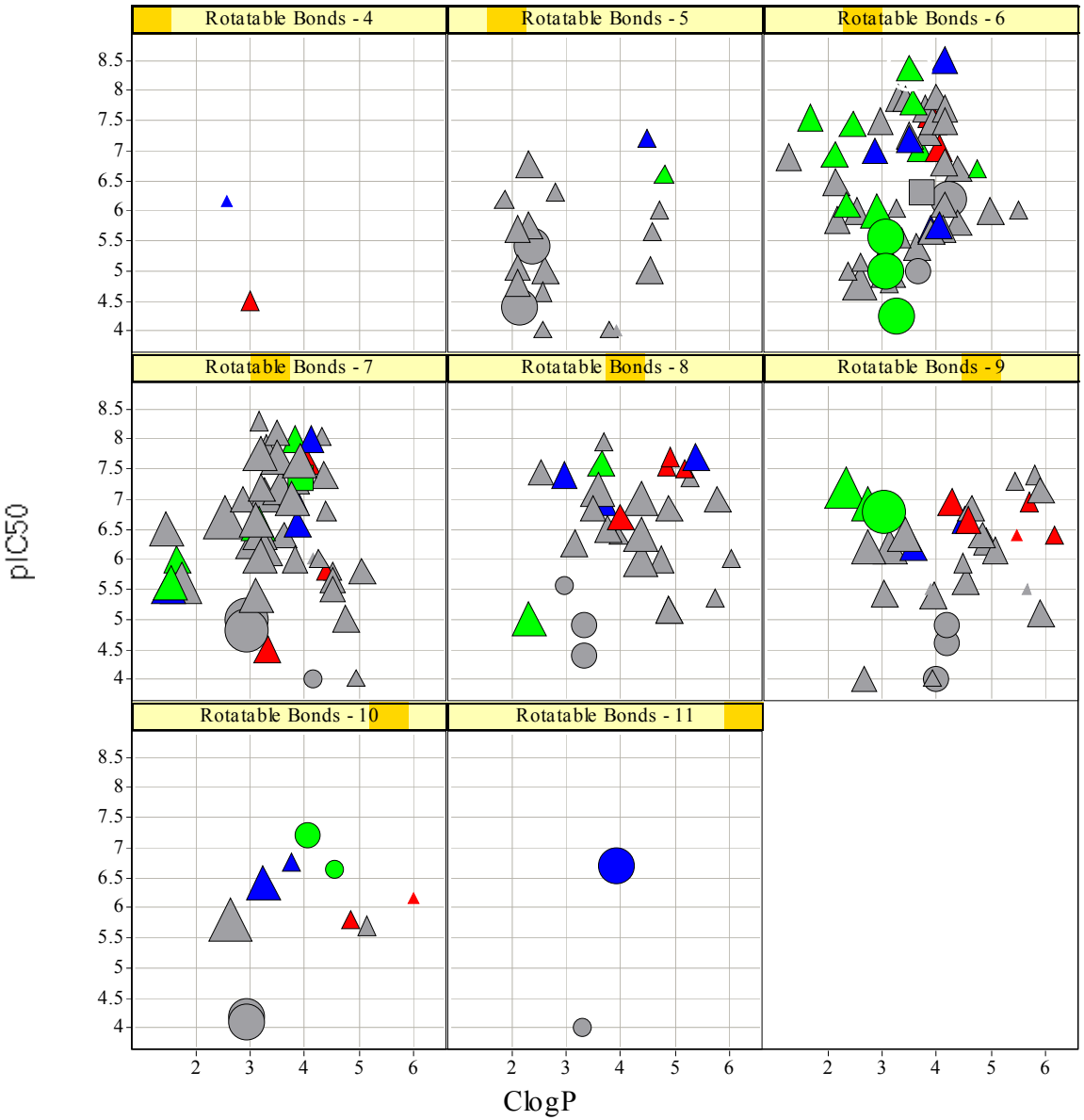
AstraZeneca



#	Compound Name	pIC50	Selectivity	MWt	ClogP	Molecule n Donors	Molecule n Acceptors	Molecule Ion Class	Molecule n Rings	LogD	Solubility	Metabolism	Protein binding
16	AZ10620691	5.4		397.43	3.07	3	7	Neutral	3				
17	AZ10621542	<5		383.40	2.61	4	7	Neutral	3		-0.35		
18	AZ10621690	7.2	?110	353.42	4.49	3	5	Neutral	3	NV	-1.44	12	2.09
19	AZ10735574	6.3		377.49	3.05	5	7	Neutral	3	2.58	1.16		
20	AZ10735618	>5		373.50	2.38	3	6	Neutral	4				
21	AZ10768060	7.086		413.47	3.27	5	7	Neutral	3	2.87	0.64		0.83
22	AZ10779441	5.35		417.60	5.75	3	6	Neutral	3		-1.24		
23	AZ10779494	6.2		418.54	3.14	5	8	Neutral	3				
24	AZ10804682	5.7		397.47	3.90	4	6	Neutral	3	3.315	0.87	14	
25	AZ11003421	<4		494.59	2.67	4	8	Neutral	4				
26	AZ11028083	5.15		407.54	4.89	4	6	Neutral	3		0.30		
27	AZ11029313	4.5		446.53	3.32	4	8	Neutral	3		0.44	33	
28	AZ11029319	7.05	16	409.48	4.04	4	6	Neutral	4	3.18		31	
29	AZ11029329	6.4		403.53	3.97	3	7	Neutral	3		?-1.61		
30	AZ11029787	6.2		427.50	3.57	4	7	Neutral	3	3.16	0.95	12	1.43
31	AZ11029799	7.4		413.47	2.97	4	7	Neutral	3	2.94	0.46	16	
32	AZ11029842	6	?100	425.56	1.65	4	8	Neutral	3		1.73	<3	
33	AZ11037100	6.9	130	338.42	1.27	4	8	Neutral	3	1.19	1.92		
34	AZ11039796	<6		389.55	4.75	4	6	Neutral	3		0.35		
35	AZ11045712	4.233		368.44	3.27	5	6	Base	3	1.51	1.49	<1	
36	AZ11045735	<4		396.49	4.17	3	6	Base	3	2.4	1.80		
37	AZ11046529	5		394.47	3.66	4	6	Base	4	1.08	1.72		0.91
38	AZ11046571	5		398.46	2.92	6	7	Base	3	1.04	1.97		0.49
39	AZ11046591	6.2		398.92	4.22	5	6	Base	3	1.49	1.57		1.25
40	AZ11046625	7.55	?330	389.92	1.68	4	8	Neutral	3	2.06	0.61	5	0.60


IbisTempCSVFile61279.csv - Spotfire DecisionSite for Lead Discovery - [Scatter Plot]
File Edit View Visualization Data Tools Guides Window Help
Structure Viewer
Browse Lists Menu
Active
Compound name
AZ11868317
AZ11868317
Structure
ISAC
pIC50
8.5
8
7.5
7
6.5
6
5.5
5
4.5
4
2
3
4
5
6
ClogP
Scatter Plot
Query Devices
Type to search
Compound Name
(All)
Compound Structure
(All)
Details-on-Demand
Column Value
Selectivity 500
MWt 399.9
ClogP 4.14
Molecule n Donors 4
Molecule n Acceptors 6
Molecule Ion Class Neutral
Molecule n Rings 3
Qualifier (LogD)
Legend
Scatter Plot
Color by Molecule n Donors:
2 3 4 5 6
Shape by Molecule Ion Class
Acid Base
Neutral
4.49 ; 4
199 out of 199 records visible (100.00 %), 0 marked

Visualisation of 6 Parameters



pIC50	Y axis
ClogP	X axis
Ion Class	Shape
N Donors	Size
N Rotatable Bonds	Trellis
Binned metabolism	Colour
	Orientation
	Labels
	Z axis

Color by Binned Metabolism:

Low Moderate High

(Empty)

Size by Molecule n Donors

2 6

Shape by Molecule Ion Class

Acid Base Neutral

The labels show Compound Name.

AstraZeneca

Hierarchical Clustering on Structure

- Clustering techniques group similar objects together for visualisation or analysis
- Structural similarity calculated by comparing structural fingerprint keys
- Ward's hierarchical agglomerative clustering produces a dendrogram visualisation

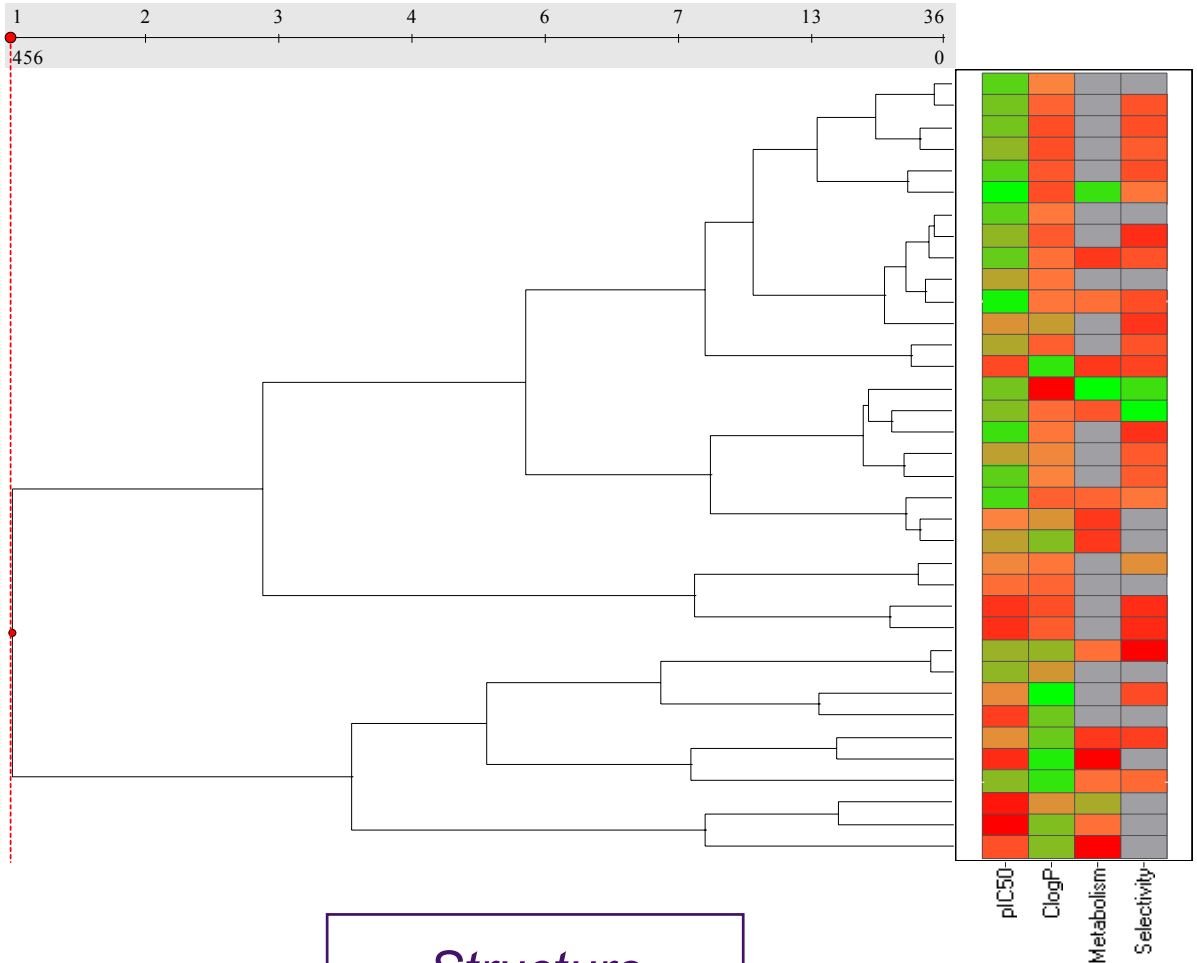
AstraZeneca

A Fingerprint Key

10,13,16,19,20,22,29,34,36,37,38,39,47,49,50,58,59,63,64,68,69,70,76,80,87,91,95,96,98,106,110,113,116,120,121,128,131,132,135,136,138,140,143,146,150,151,154,155,159,162,164,165,166,168,169,170,174,176,177,178,185,186,187,200,202,203,208,209,210,212,213,215,216,217,220,226,232,234,240,242,243,247,256,261,262,265,266,273,274,275,276,281,283,286,290,291,293,294,295,296,297,299,300,302,303,307,308,310,312,326,330,332,333,334,337,339,340,341,344,346,347,350,351,352,353,357,361,363,365,368,369,371,372,375,383,389,390,391,392,394,395,397,399,400,405,406,407,412,415,418,420,421,426,428,430,432,441,452,453,454,455,456,457,462,467,469,471,474,475,476,479,480,483,484,488,489,496,497,498,501,503,505,506,507,508,510,512,514,516,517,530,531,532,533,537,539,540,542,545,550,552,555,556,558,559,560,563,564,565,568,570,575,577,585,589,593,597,601,602,604,606,607,608,610,620,622,623,630,632,635,636,637,638,643,644,646,647,654,655,658,661,666,669,671,675,677,678,682,690,691,694,696,698,699,703,705,707,710,712,713,718,719,720,723,729,740,745,746,751,752,754,756,763,764,766,769,772,773,776,778,783,785,792,794,795,799,802,804,805,806,807,808,810,811,814,815,818,820,822,823,824,825,826,827,832,835,837,847,850,853,854,857,858,860,867,869,874,878,883,885,886,887,890,892,893,895,899,901,904,913,914,917,930,933,936,939,940,941,943,944,945,946,956,961,963,965,967,969,971,974,975,978,979,981,989,991,995,996,999,1001,1006,1015,1017,1018,1021,1026,1027,1032,1033,1034,1038,1044,1049,1051,1054,1059,1066,1068,1071,1072,1073,1076,1079,1083,1085,1091,1092,1094,1096,1101,1104,1107,1109,1111,1113,1115,1117,1122,1123,1124,1130,1131,1132,1136,1137,1150,1160,1163,1169,1170,1172,1173,1175,1178,1181,1186,1187,1190,1191,1192,1194,1200,1207,1208,1209,1210,1212,1213,1216,1222,1228,1230,1232,1233,1236,1239,1244,1245,1247,1248,1249,1253,1255,1256,1264,1265,1273,1279,1282,1289,1291,1293,1294,1301,1314,1315,1317,1318,1320,1323,1336,1337,1338,1342,1343,1344,1346,1352,1355,1356,1358,1359,1360,1363,1364,1367,1368,1371,1373,1375,1376,1377,1380,1383,1384,1385,1388,1389,1402,1406,1408,1411,1413,1414,1415,1416,1419,1422,1423,1424,1425,1427,1430,1433,1435,1440,1441,1444,1447,1448,1449,1455,1459,1466,1471,1474,1475,1477,1479,1481,1483,1486,1490,1491,1494,1497,1498,1499,1500,1505,1513,1519,1520,1522,1528,1530,1536,1541,1546,1549,1550,1552,1554,1564,1567,1574,1576,1578,1580,1583,1587,1590,1593,1594,1600,1608,1615,1618,1619,1621,1625,1630,1631,1632,1635,1637,1638,1642,1645,1646,1648,1649,1656,1662,1664,1666,1669,1675,1677,1678,1679,1680,1681,1682,1683,1684,1687,1689,1693,1696,1697,1701,1702,1703,1704,1705,1712,1713,1715,1716,1720,1722,1725,1728,1730,1734,1736,1741,1744,1745,1746,1752,1753,1754,1760,1762,1764,1765,1769,1772,1774,1777,1779,1780,1781,1785,1786,1787,1790,1796,1800,1801,1807,1809,1810,1813,1821,1822,1825,1831,1832,1834,1839,1842,1846,1848,1852,1855,1858,1861,1862,1866,1867,1868,1871,1893,1897,1898,1900,1901,1902,1904,1905,1910,1913,1921,1923,1925,1927,1929,1933,1935,1937,1940,1941,1945,1959,1961,1964,1974,1975,1977,1980,1982,1994,1995,1998,2002,2005,2006,2007,2008,2011,2015,2018,2020,2022,2024,2026,2029,2032,2034,2035,2038,2040,2041,2043,2044,2047,2048

AstraZeneca

Dendrogram and Heat Map



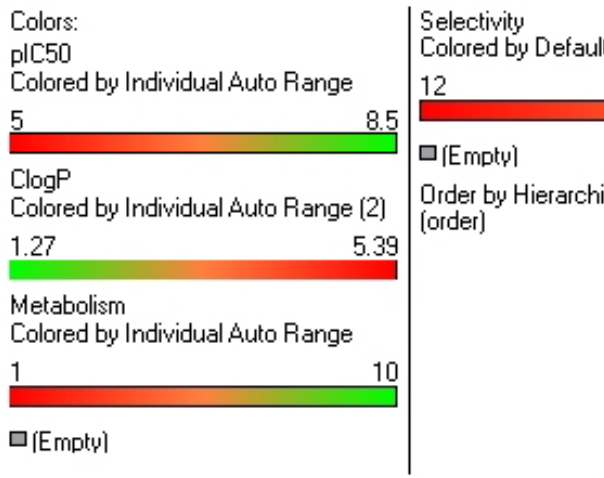
(Clustered on keys)

This visualization shows records in a dendrogram (a tree graph) based on the similarity between them

Calculation settings:

Total no of records: 36 (All records)
Key column: KEYS

Clustering method: Ward's method
Similarity measure: Half square euclidean distance
Ordering function: Average value



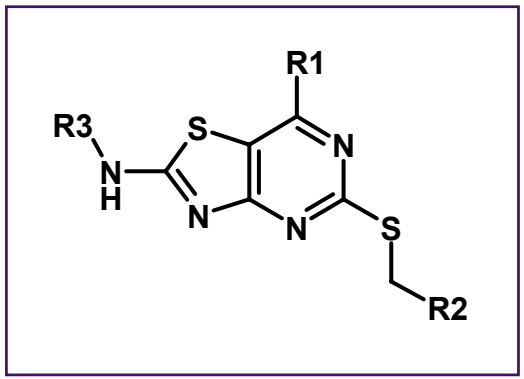
Structure

What is fRGS?

- fRGS: **f**or **R**-**G**roup **S**tripping
- Chemical deconvolution and subsequent analysis
- Used to discover SAR between, and within, chemical series
- User defines Core structures and R group positions
- User may define multiple cores, program tries to match each compound against each core in turn
- Stand-alone application and integrated in IBIS Explore

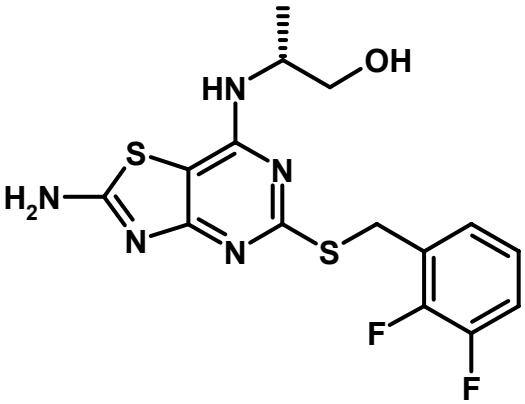
AstraZeneca

fRGS Output



Core

[R1]c1c2c(nc(n1)SC[R2])nc(s2)N[R3]



Compound

[R1]N[C@H](C)CO
[(1R)-2-hydroxy-1-methyl-ethyl]amino

[R2]c1cccc(c1F)F
2,3-difluorophenyl

[R3][H]
hydrogen

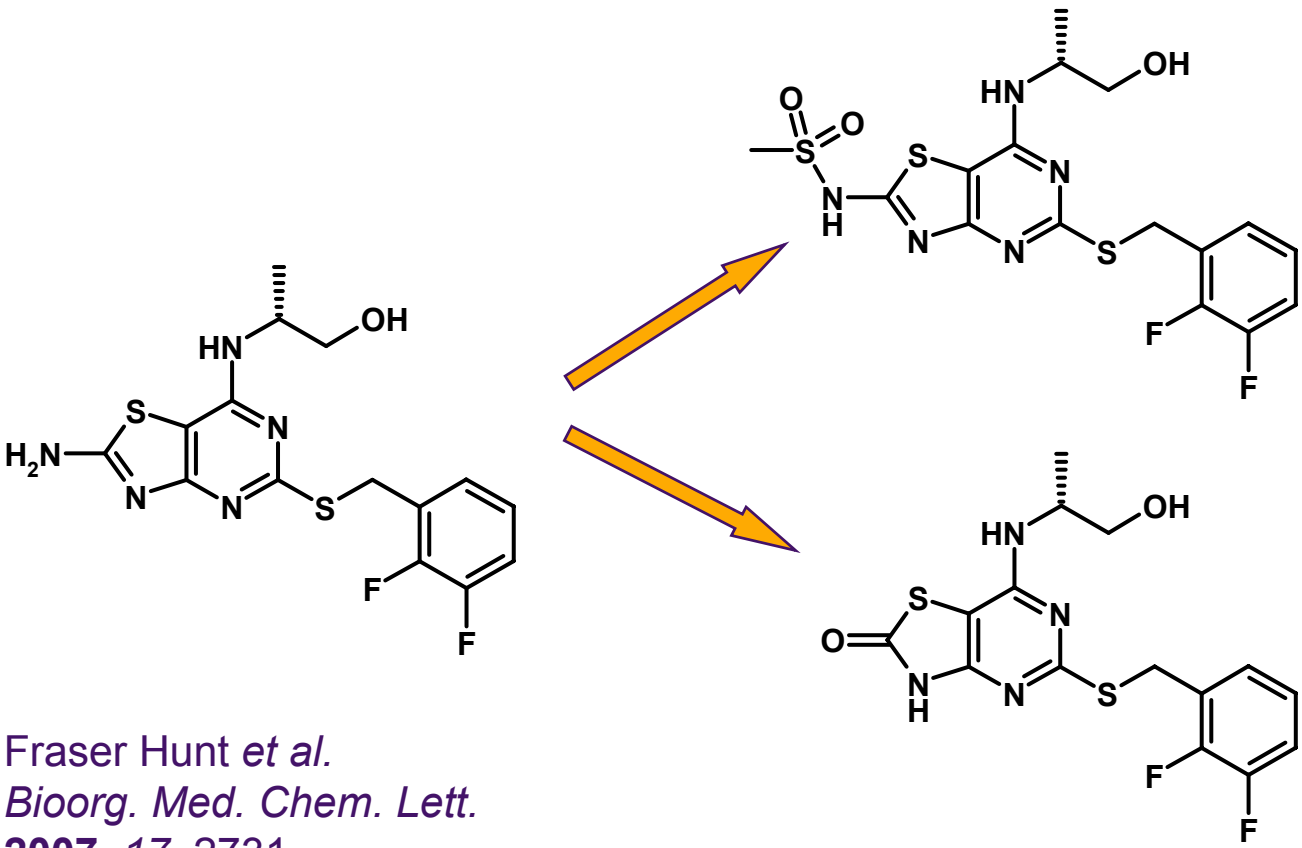
AstraZeneca

Using R Groups

- Visualise R group structures
- Plot activity vs R
- Visualise R group properties (σ, π etc.)
- Compare matched pairs/groups
  - If 2 compounds differ only by Core they will have the same combination of R groups: R1+R2+…
  - If 2 compounds differ only by R1 they will have the same combination of Core+R2+R3…
- Generate SAR matrix (R1 by R2)

CXCR2 Antagonist Project

CXCR2 – a chemokine receptor expressed on neutrophils
Potential role in arthritis, COPD, asthma, UC and psoriasis



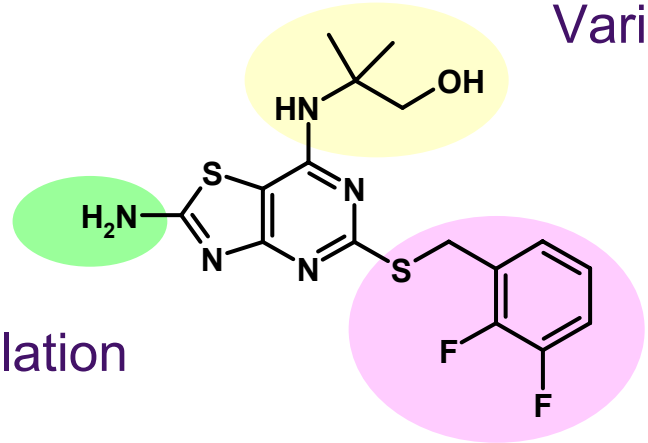
Fraser Hunt *et al.*
Bioorg. Med. Chem. Lett.
2007, *17*, 2731

Iain Walters *et al.*
Bioorg. Med. Chem. Lett.
2008, *18*, 798

AstraZeneca

Substituent SAR

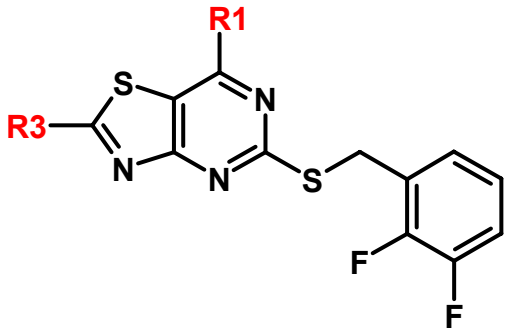
Various hydroxyethylamines



Mono and di-alkylation

Fixed

24 compounds in 2-dimensional matrix synthesised



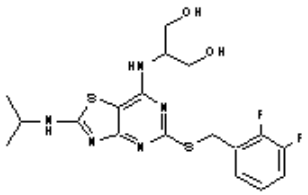
fRGS Core

AstraZeneca

fRGS : C:\DOCUME~1\m238495\LOCALS~1\Temp\IbisTempSDFile61281.sdf

File Edit Export Structures RGroup Properties Help

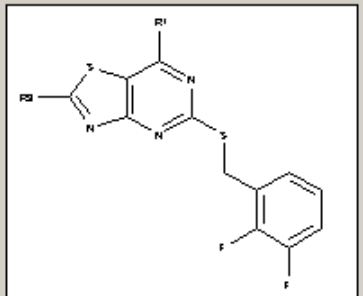
Open Save Paste Strip RGroups Properties Match Matrix Spotfire Cores



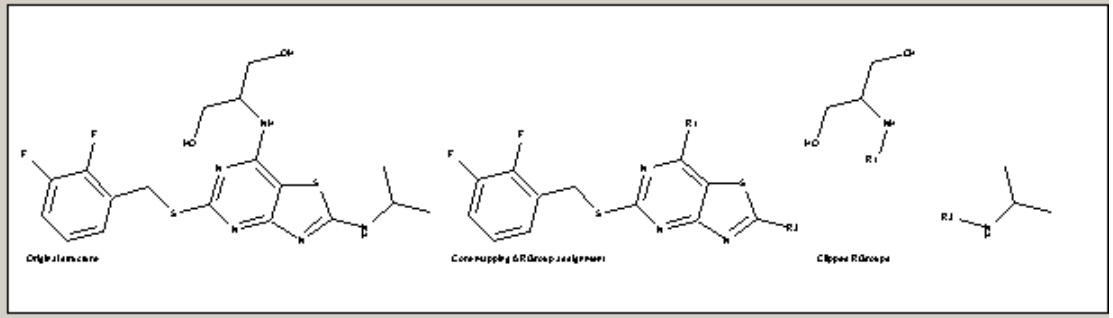
ISAC details
R Groups

RGroup Mapping

First detected core



Chosen mapping to compound

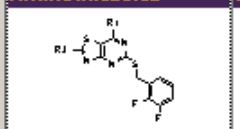
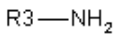
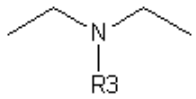
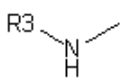
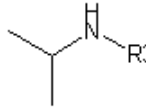
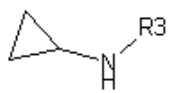
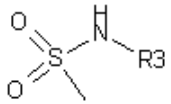
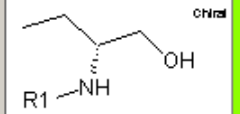
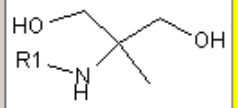
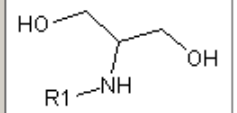
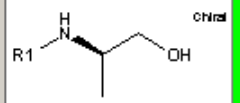
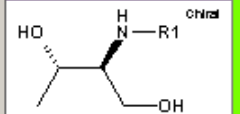
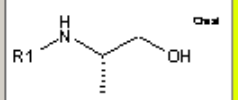


	core(1)-R1	core(1)-R1_Name	core(1)-R3	core(1)-R3_Name
	core(1)-R1_Name : [(1R)-1-(hydroxymethyl)propyl]amino - 6 item(s)			
1	[R1]N[C@H](CC)CO	[(1R)-1-(hydroxymethyl)propyl]amino	[R3]N	amino
2	[R1]N[C@H](CC)CO	[(1R)-1-(hydroxymethyl)propyl]amino	[R3]N(CC)CC	diethylamino
3	[R1]N[C@H](CC)CO	[(1R)-1-(hydroxymethyl)propyl]amino	[R3]NS(=O)(=O)C	methanesulfonamido
4	[R1]N[C@H](CC)CO	[(1R)-1-(hydroxymethyl)propyl]amino	[R3]NC	methylamino
5	[R1]N[C@H](CC)CO	[(1R)-1-(hydroxymethyl)propyl]amino	[R3]NC(C)C	isopropylamino
6	[R1]N[C@H](CC)CO	[(1R)-1-(hydroxymethyl)propyl]amino	[R3]NC1CC1	cyclopropylamino
	core(1)-R1_Name : [2-hydroxy-1-(hydroxymethyl)-1-methyl-ethyl]amino - 1 item(s)			
1	[R1]NC(C)(CO)CO	[2-hydroxy-1-(hydroxymethyl)-1-methyl-ethyl]amino	[R3]N	amino
	core(1)-R1_Name : [2-hydroxy-1-(hydroxymethyl)ethyl]amino - 5 item(s)			
1	[R1]NC(CO)CO	[2-hydroxy-1-(hydroxymethyl)ethyl]amino	[R3]N(CC)CC	diethylamino
2	[R1]NC(CO)CO	[2-hydroxy-1-(hydroxymethyl)ethyl]amino	[R3]NC(C)C	isopropylamino
3	[R1]NC(CO)CO	[2-hydroxy-1-(hydroxymethyl)ethyl]amino	[R3]NC1CC1	cyclopropylamino
4	[R1]NC(CO)CO	[2-hydroxy-1-(hydroxymethyl)ethyl]amino	[R3]NS(=O)(=O)C	methanesulfonamido
5	[R1]NC(CO)CO	[2-hydroxy-1-(hydroxymethyl)ethyl]amino	[R3]NC	methylamino
	core(1)-R1_Name : [(1R)-2-hydroxy-1-methyl-ethyl]amino - 6 item(s)			
1	[R1]N[C@H](C)CO	[(1R)-2-hydroxy-1-methyl-ethyl]amino	[R3]NC	methylamino
2	[R1]N[C@H](C)CO	[(1R)-2-hydroxy-1-methyl-ethyl]amino	[R3]NS(=O)(=O)C	methanesulfonamido
3	[R1]N[C@H](C)CO	[(1R)-2-hydroxy-1-methyl-ethyl]amino	[R3]NC(C)C	isopropylamino
4	[R1]N[C@H](C)CO	[(1R)-2-hydroxy-1-methyl-ethyl]amino	[R3]N(CC)CC	diethylamino
5	[R1]N[C@H](C)CO	[(1R)-2-hydroxy-1-methyl-ethyl]amino	[R3]NC1CC1	cyclopropylamino
6	[R1]N[C@H](C)CO	[(1R)-2-hydroxy-1-methyl-ethyl]amino	[R3]N	amino

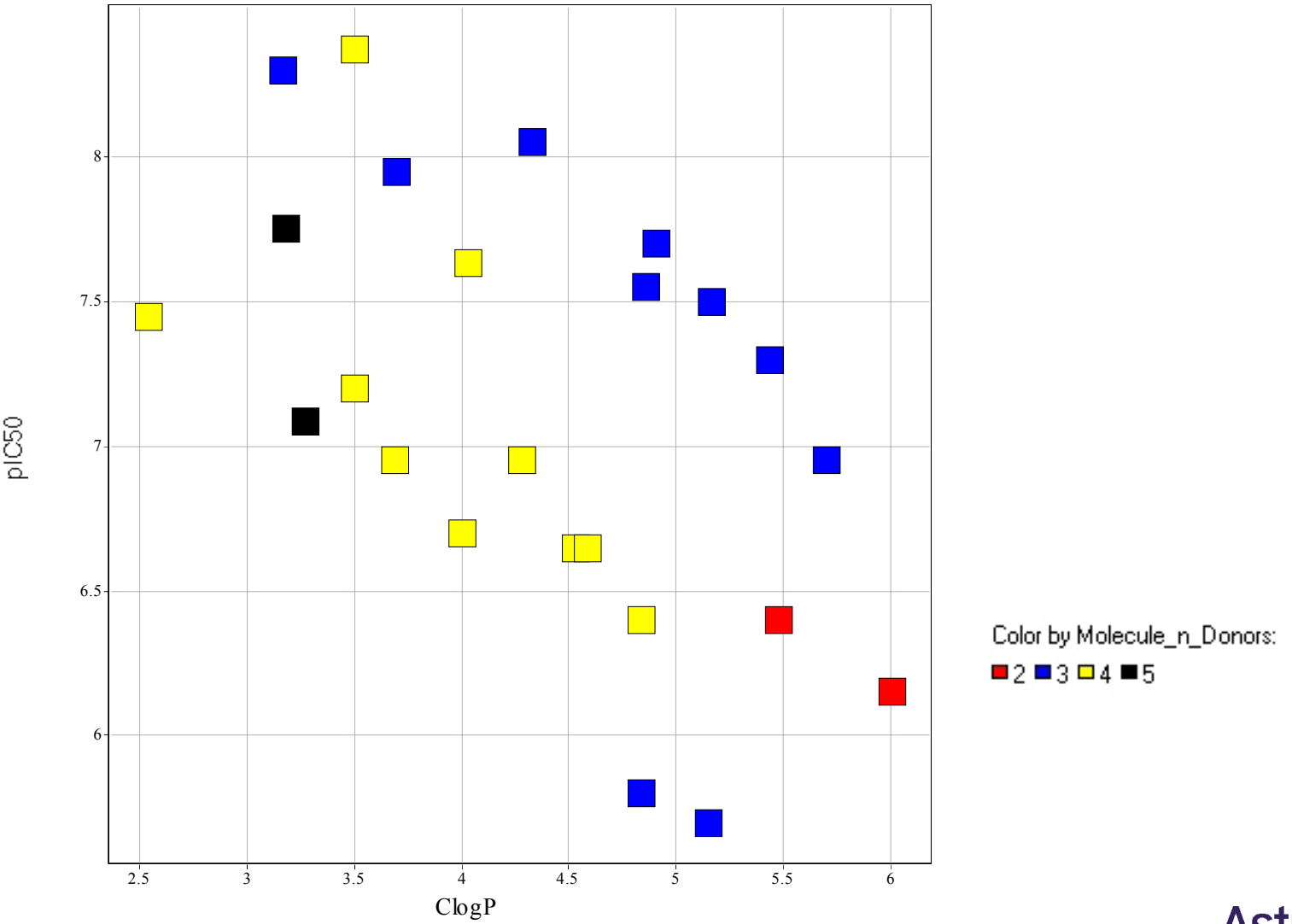
RGroup Stripping Complete | 2,23 | FileImport

MatrixView

Close

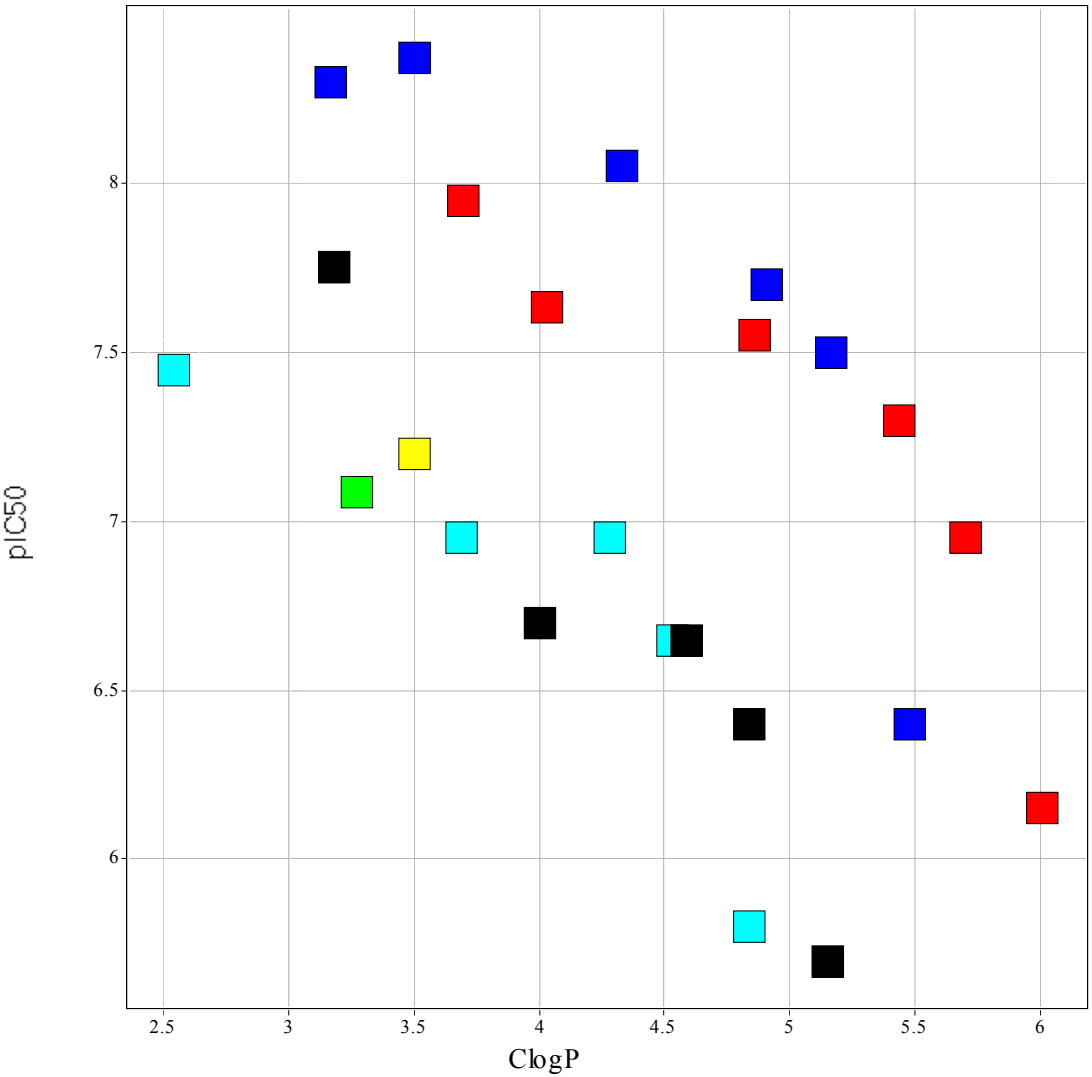
Aminothiazole2	R3 : amino	R3 : diethylamino	R3 : methylamino	R3 : isopropylamino	R3 : cyclopropylamino	R3 : methanesulfonami
						
R1 : [(1R)-1-(hydroxym	pIC50: 7.633	pIC50: 6.15	pIC50: 7.55	pIC50: 6.95	pIC50: 7.3	pIC50: 7.95
						
R1 : [2-hydroxy-1-(hydr	pIC50: 7.086					
						
R1 : [2-hydroxy-1-(hydr		pIC50: 5.8	pIC50: 6.95	pIC50: 6.65	pIC50: 6.95	pIC50: 7.45
						
R1 : [(1R)-2-hydroxy-1-	pIC50: 8.37	pIC50: 6.4	pIC50: 8.05	pIC50: 7.5	pIC50: 7.7	pIC50: 8.3
						
R1 : [(1S,2S)-2-hydroxy	pIC50: 7.75	pIC50: 5.7	pIC50: 6.7	pIC50: 6.4	pIC50: 6.65	
						
R1 : [(1S)-2-hydroxy-1-	pIC50: 7.2					
						

Potency *vs* clogP



AstraZeneca

Coloured by R1



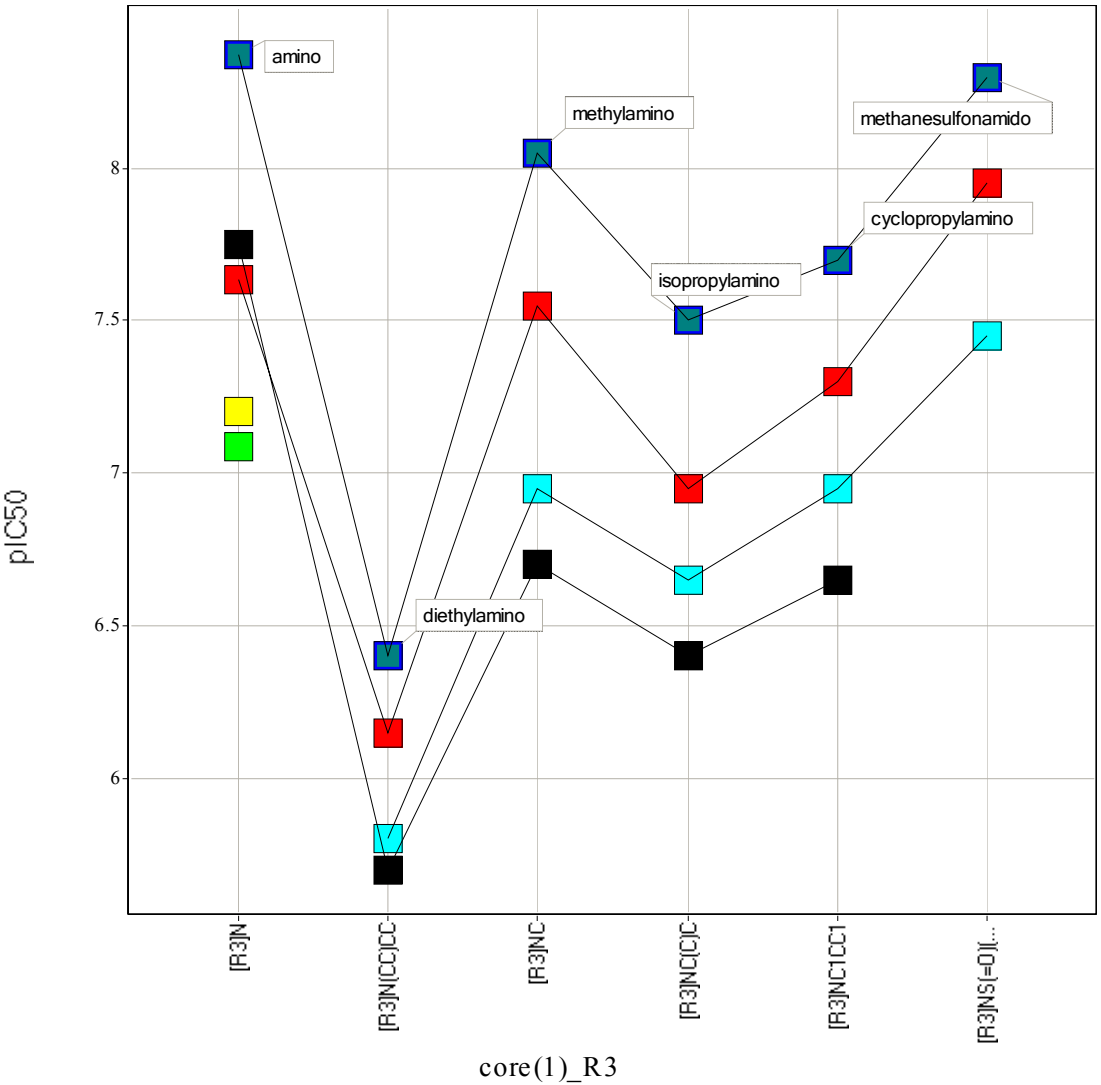
Inverse dependency on lipophilicity of R3 apparent

Color by core(1)-R1_Name:

- [(1R)-1-(hydroxymethyl)propyl]amino
- [(1R)-2-hydroxy-1-methyl-ethyl]amino
- [(1S)-2-hydroxy-1-methyl-ethyl]amino
- [(1S,2S)-2-hydroxy-1-(hydroxymeth...
- [2-hydroxy-1-(hydroxymethyl)-1-me...
- [2-hydroxy-1-(hydroxymethyl)ethyl]...

AstraZeneca

Potency *vs* R3, matched by R1



Clear SAR for R1 and R3

Parallel SAR across matched pairs

Color by core(1)-R1_Name:

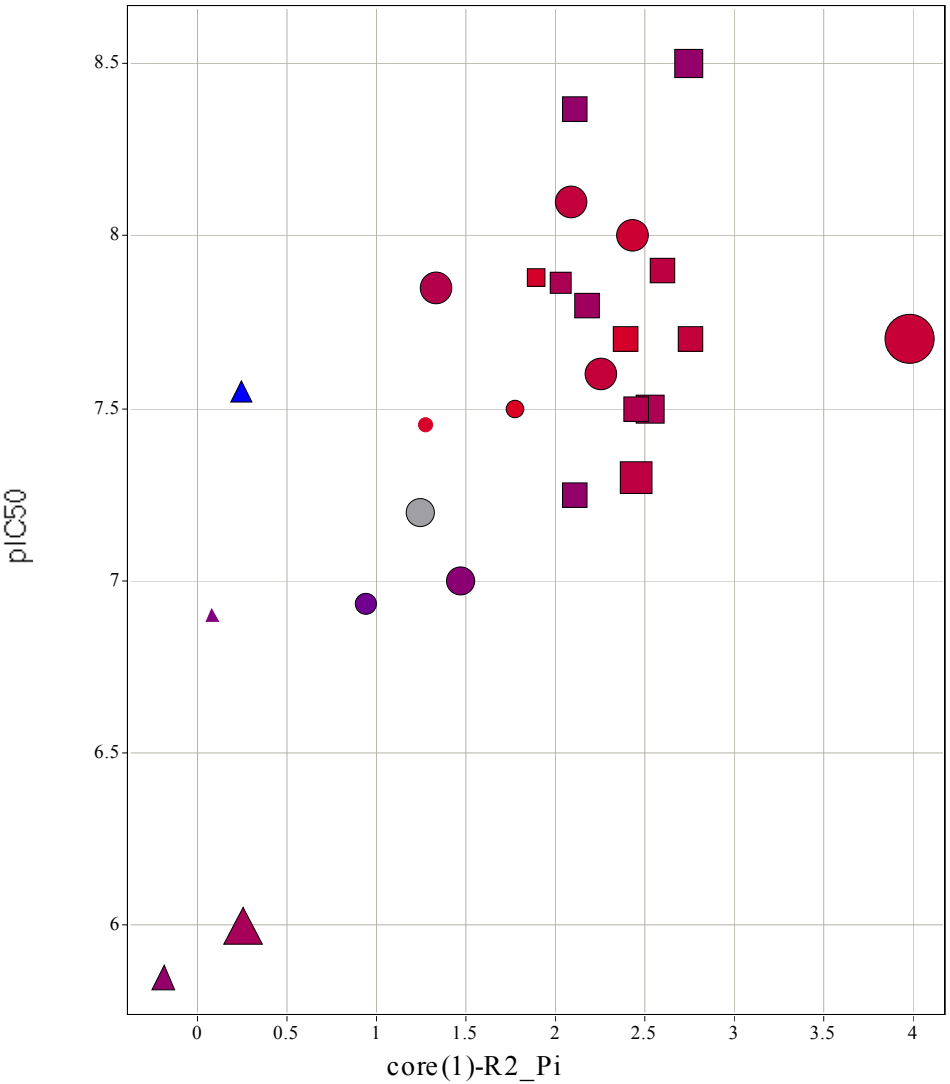
- [(1R)-1-(hydroxymethyl)propyl]amino
- [(1R)-2-hydroxy-1-methyl-ethyl]amino
- [(1S)-2-hydroxy-1-methyl-ethyl]amino
- [(1S,2S)-2-hydroxy-1-(hydroxymeth...
- [2-hydroxy-1-(hydroxymethyl)-1-me...
- [2-hydroxy-1-(hydroxymethyl)ethyl]...

Markers are connected by core(1)_R1, and ordered by core(1)_R3.

The labels show core(1)-R3_Name.

AstraZeneca

Potency *vs* R2 Calculated Properties



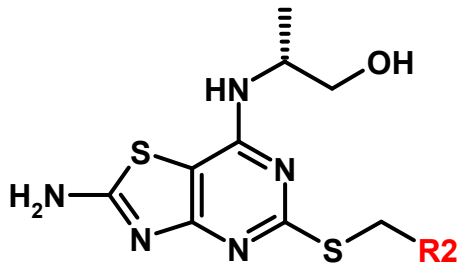
A general trend with lipophilicity but more subtle SAR apparent

Color by
core(1)-R2_ACD_Sigma-para:
-0.16 0.85

(Empty)

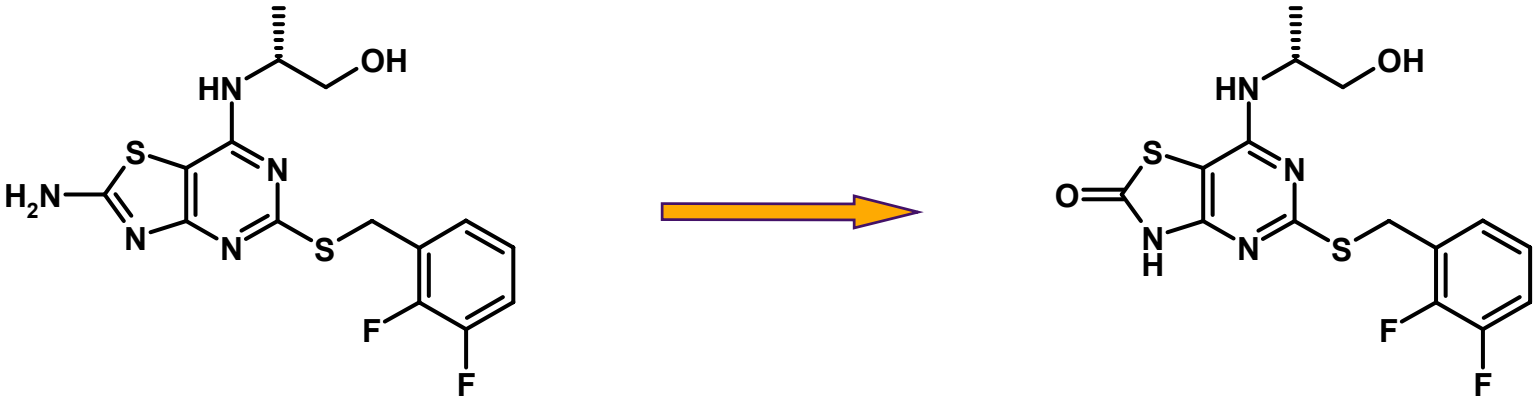
Size by core(1)-R2_Volume
58.872 151.482

Shape by core(1)-R2_H_Acceptors
0 1 2
(Empty)

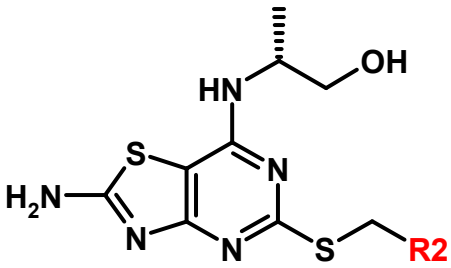


AstraZeneca

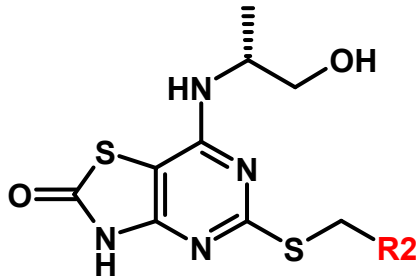
Comparison of Cores



Is the new core more or less potent than the old core?
Is the substituent SAR the same for the new core?



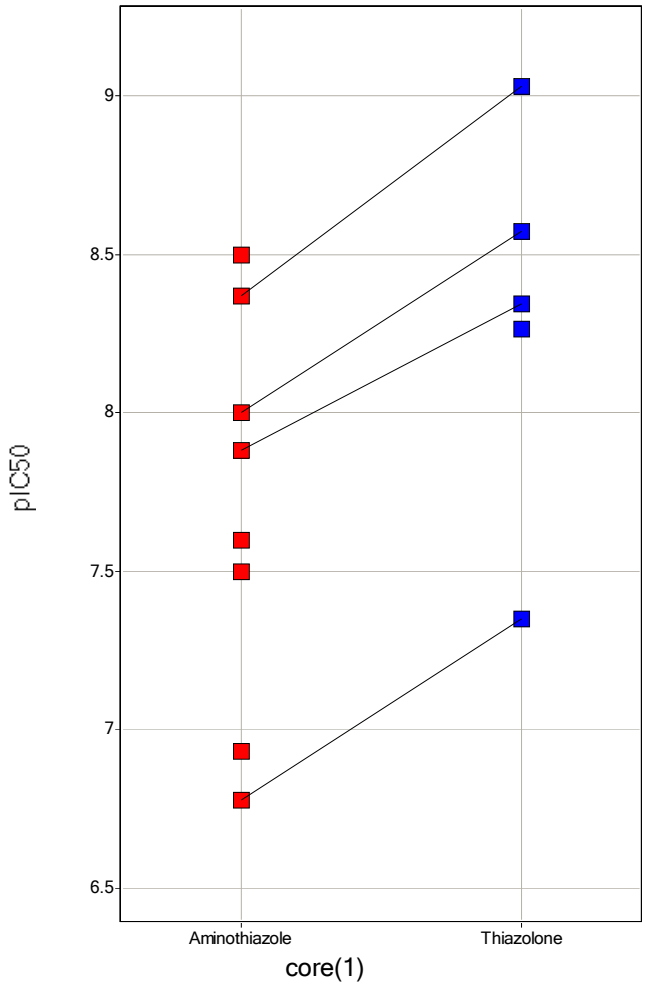
"Aminothiazole"



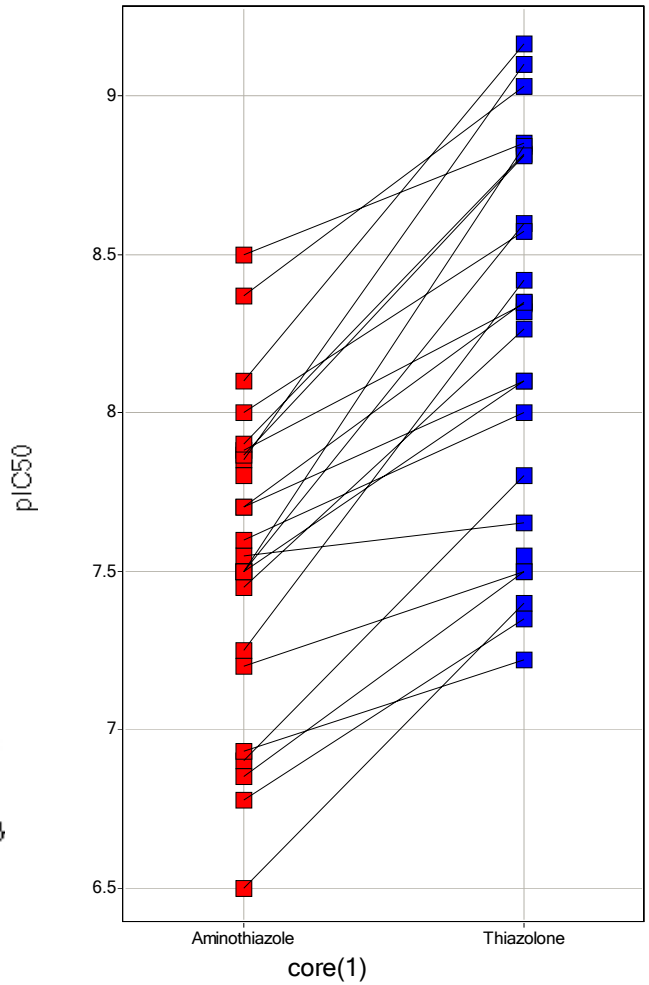
"Thiazolone"

AstraZeneca

Comparison of Cores



Color by core(1):
Aminothiazole Thiazolone
Markers are connected by core(1)_R1&R2, and ordered by core(1)_Core.



Thiazolone core more potent than Aminothiazole

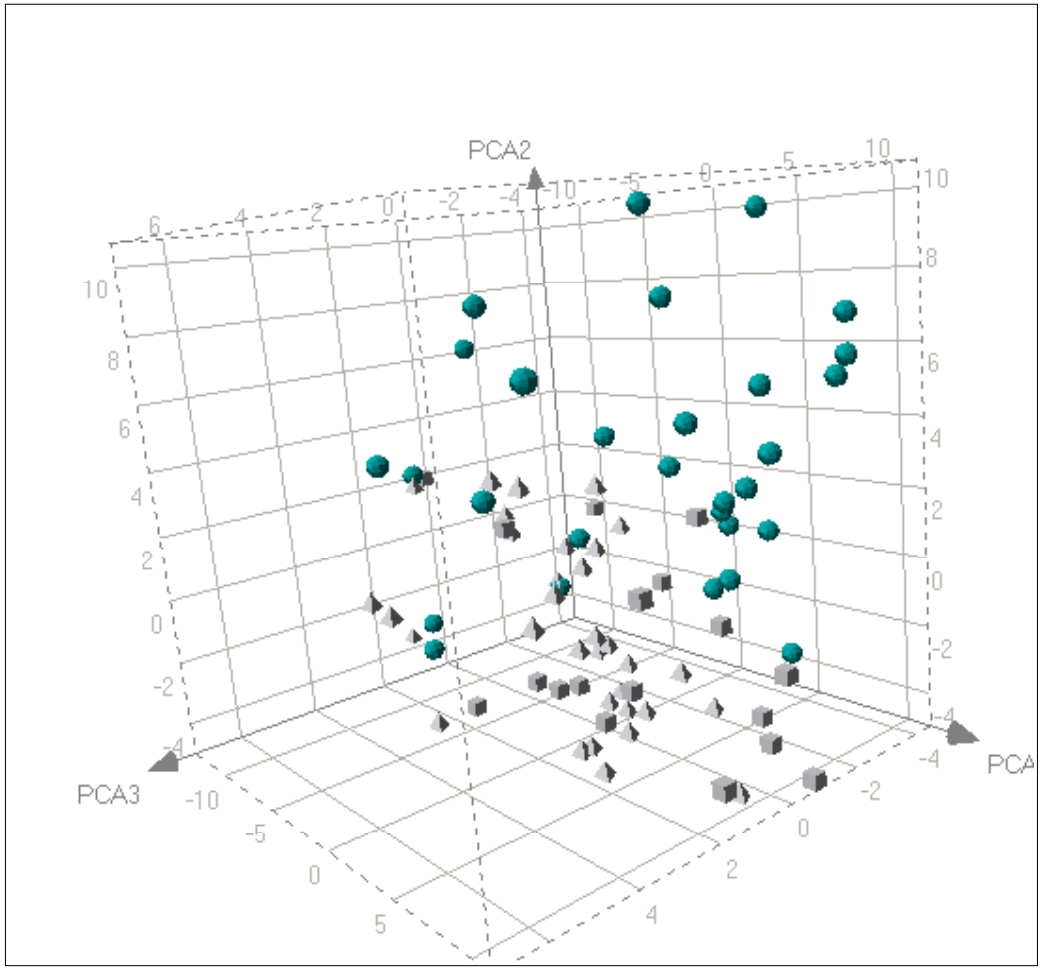
Substituent SAR parallel across cores

AstraZeneca

Visualisation in Library Design

- Visualisation of property space
  - What range of property space is covered?
  - How close are virtual compounds to previously made compounds in property space?
  - Manual selection spanning property space
- Hierarchical clustering on structure keys
  - Manual structural diversity selection

Visualisation of Property Space



Principal Components Analysis on calculated properties/descriptors
Plot first 3 PC's in 3D

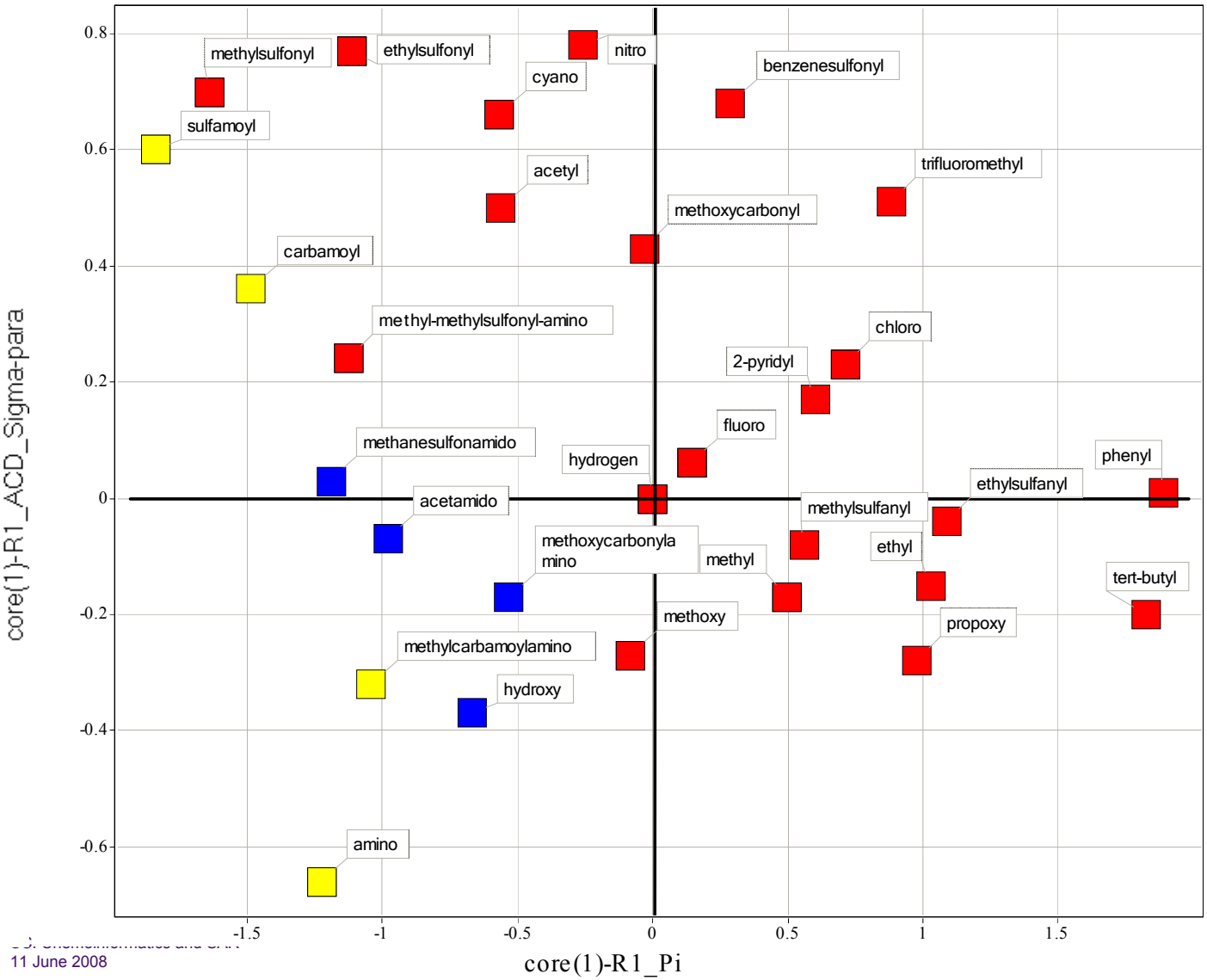
Or plot MWt *vs* ClogP (*vs* PSA)

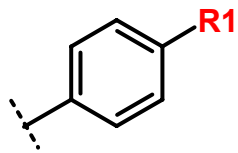
Superimpose real and virtual compounds

Spheres = Virtual
Cubes = Active
Pyramids = Inactive

AstraZeneca

Substituent Property Space





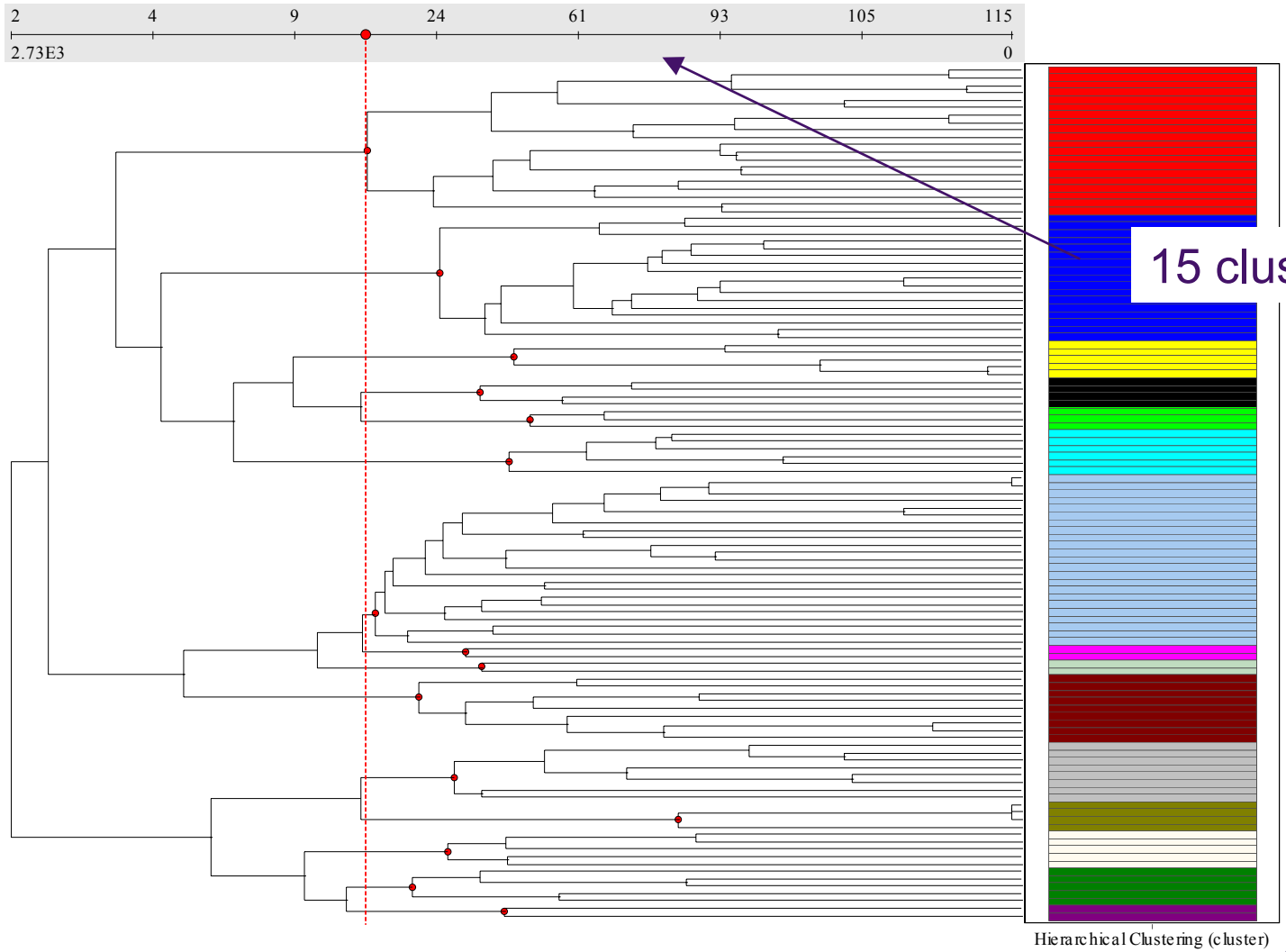
σ *vs* π
"Craig plot"

Color by Molecule_n_Donors:
1 2 3
Curve: y(x) = 0
Curve: y(x) = x*1000
The labels show core(1)-R1_Name.

AstraZeneca

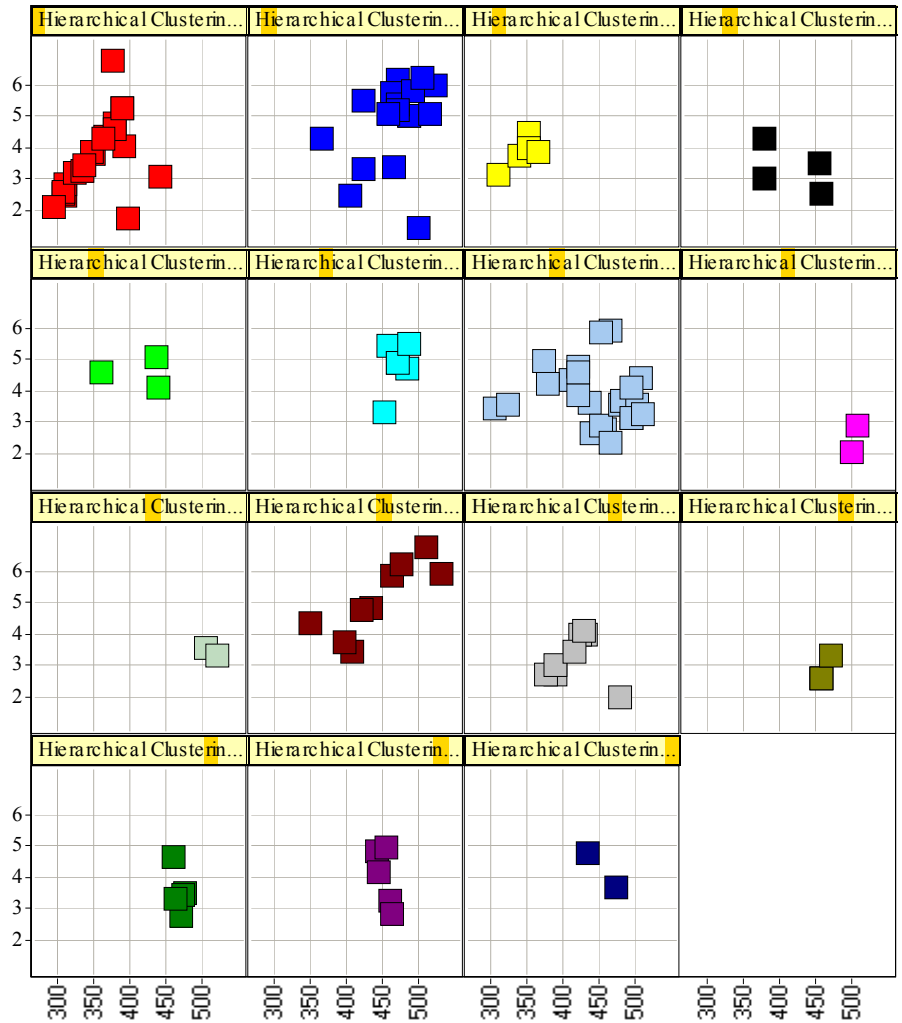
Hierarchical Clustering on Keys

115 virtual compounds



AstraZeneca

Hierarchical Clustering on Keys



Can manually pick compounds from clusters for synthesis

AstraZeneca

Summary

- Visualisation is an important technique for analysing and understanding SAR
- Visualisation techniques complement QSAR
- Structure fragmentation algorithm fits with how medicinal chemists think
- Establishing parallel SAR within one series or across related series enables prediction of biological (and other) activity of new combinations
- Visualisation of small virtual library property or structural diversity space facilitates manual selection

AstraZeneca

Acknowledgements

- Andrew Poirrette
- Justin Morley
- Linda Townsend
- Rupert Austin
- Andy Davis
- Dave Cosgrove

AstraZeneca