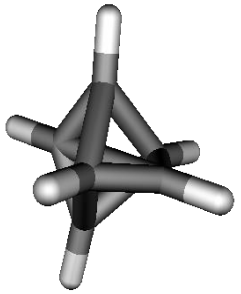
GDB and the Chemical Space

1. **GDB-17 and FDB-17**
2. Visualization
3. Virtual Screening
4. Extension to larger molecules

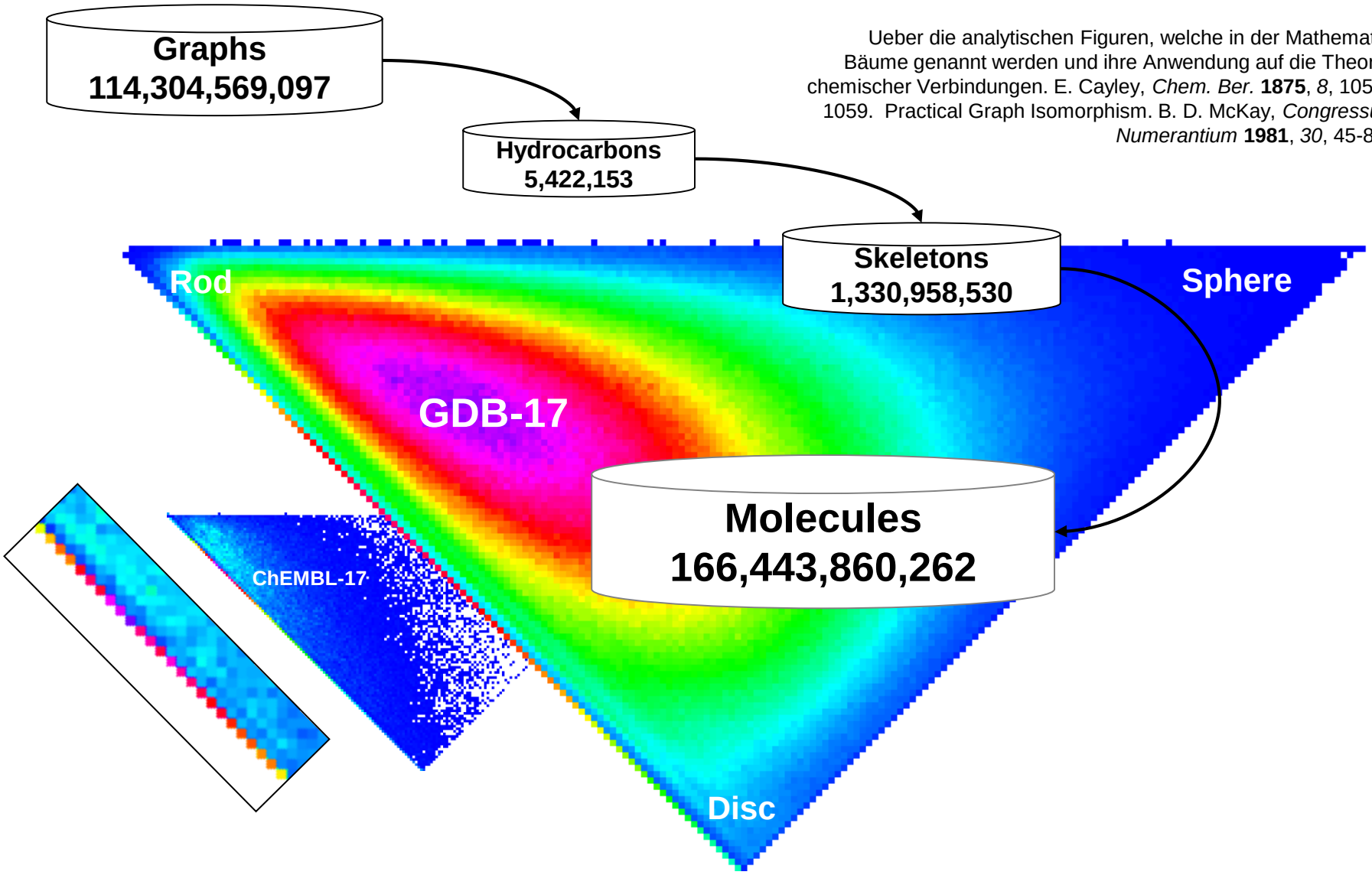
gdb.unibe.ch



Jean-Louis Reymond

19 April 2017, BigChem BCN Meeting

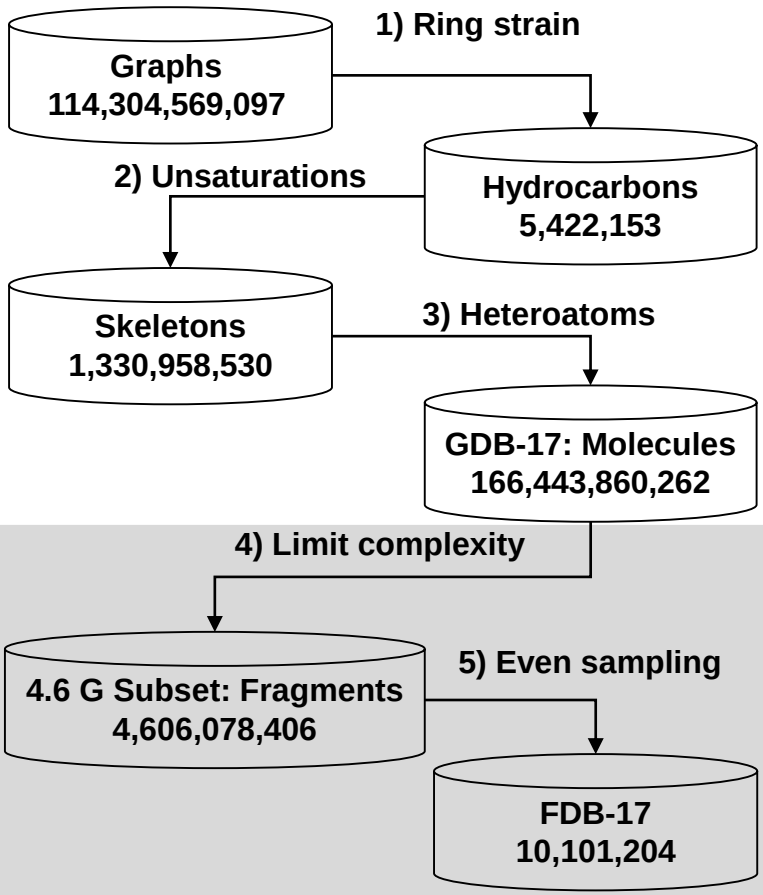
The Chemical Space Project J.-L. Reymond, *Acc. Chem. Res.* **2015**, *48*, 722-730



Ueber die analytischen Figuren, welche in der Mathematik Bäume genannt werden und ihre Anwendung auf die Theorie chemischer Verbindungen. E. Cayley, *Chem. Ber.* **1875**, *8*, 1056-1059. Practical Graph Isomorphism. B. D. McKay, *Congressus Numerantium* **1981**, *30*, 45-87.

T. Fink et al. *Angew. Chem. Int. Ed.* **2005**, *44*, 1504-1508,
J. Chem. Inf. Model. **2007**, *47*, 342-353 (GDB-11)
L. C. Blum, J.-L. Reymond, *J. Am. Chem. Soc.* **2009**, *131*, 8732-3 (GDB-13);
L. Ruddigkeit et al., *J. Chem. Inf. Model.* **2012,** *52*, 2864-2875 (GDB-17)

Fragment Database FDB-17



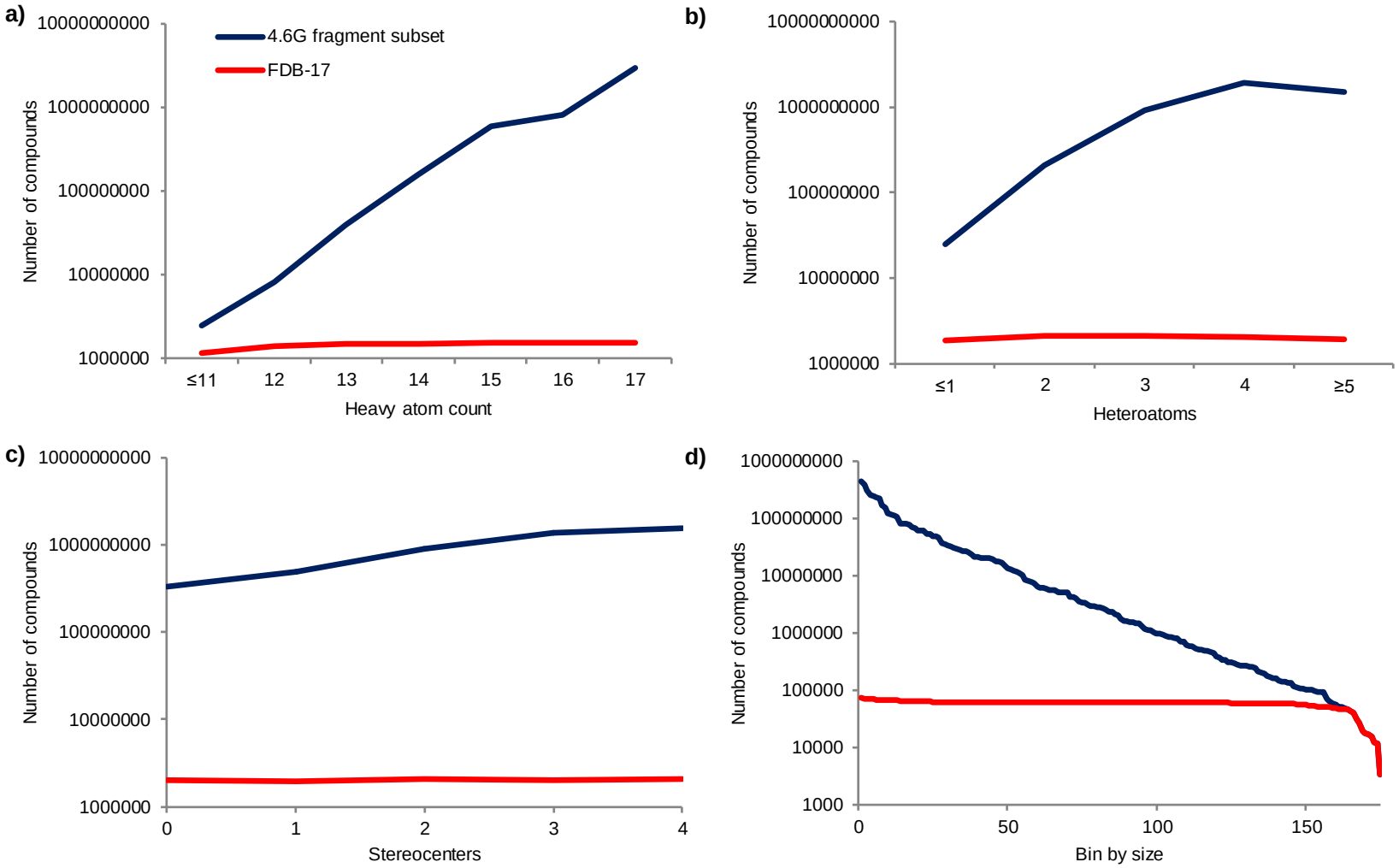
Limit Complexity

Table 1. Filtering criteria to reduce GDB-17 to its 4.6G fragment subset.

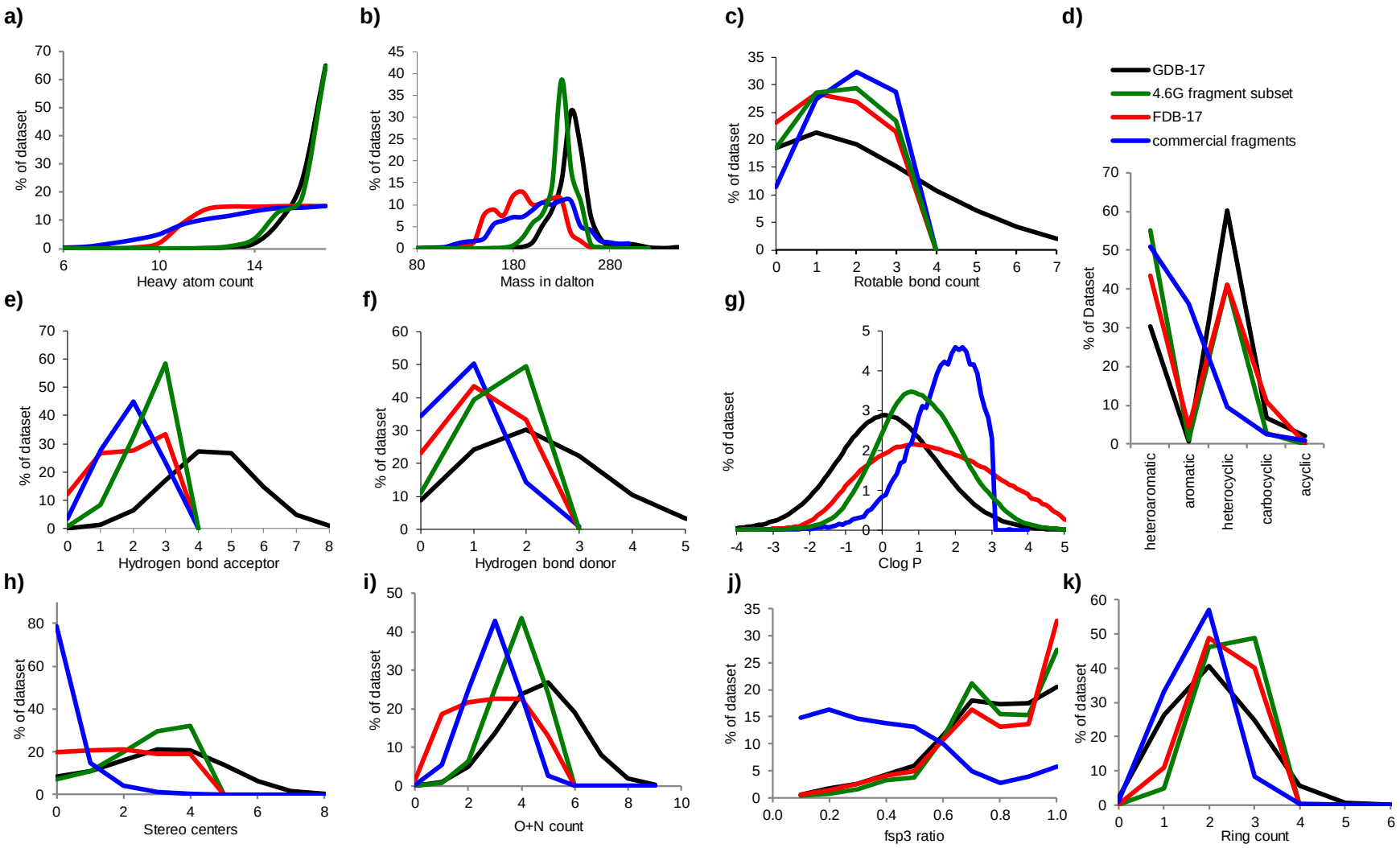
Scaffolds	**FG Density**	**Problematic/Superfluous FGs**	
≤ 3 rings	≤ 5 Nitrogen + Oxygen atoms	No aldehydes	No aromatic ring > 6 ato
≤ 2 small (3- or 4-membered) rings	≤ 1 positive charge at neutral pH	No epoxides, aziridines	≤ 1 C≡N (cyanide)
≤ 2 quaternary centers	≤ 1 negative charge at neutral pH	No O-(C=O)-O (carbonate)	No non-aromatic C=C
≤ 4 stereocenters	≤ 3 H-bond acceptor atoms	No O-C=N (imidate)	No C≡C (triple bonds)
≤ 3 rotatable bonds	≤ 2 H-bond donor atoms	No NO_2 (nitro)	No halogens

Even Sampling

value triplets (HAC, heteroatoms, stereocenters)

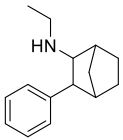


Property Profiles

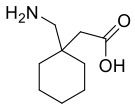


Nearest Neighbor Searches

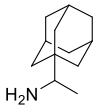
a)



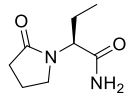
fencamfamine



gabapentin

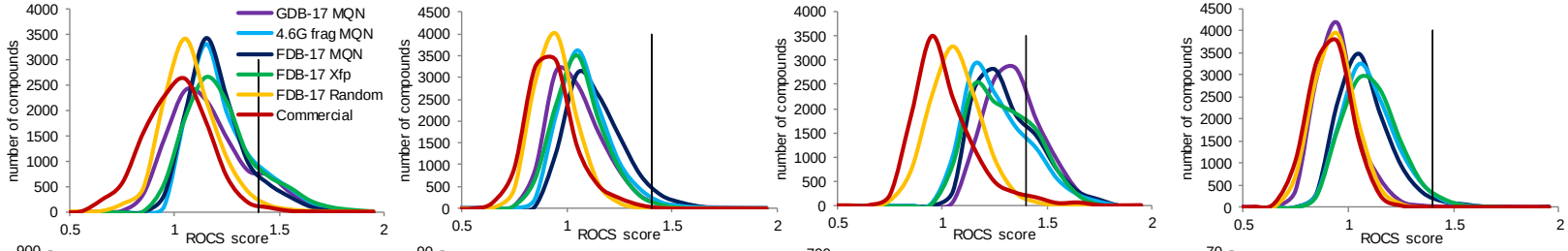


rimantadine

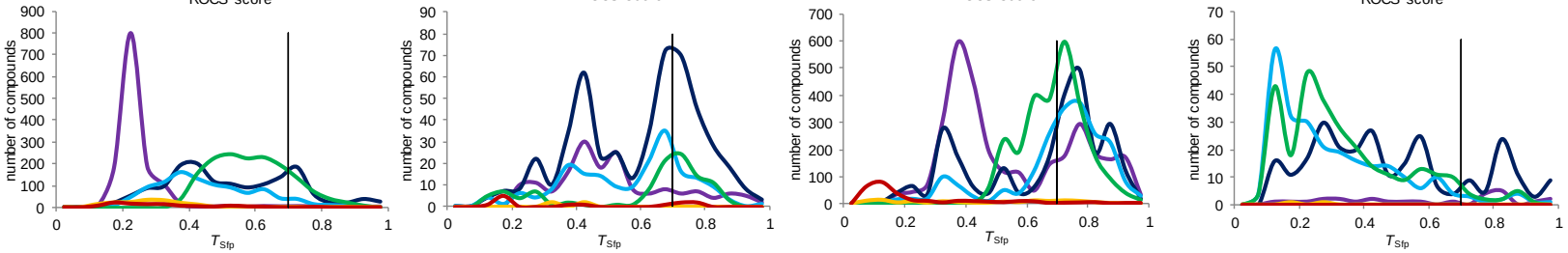


levetiracetam

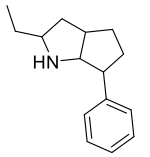
b)



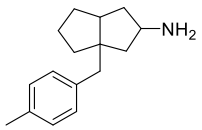
c)



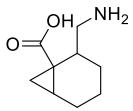
d)



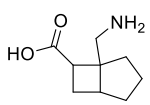
1 (T_{Sfp} = 0.88)



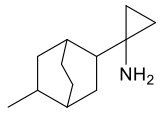
2 (T_{Sfp} = 0.71)



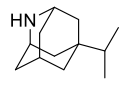
5 (T_{Sfp} = 0.82)



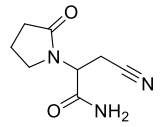
6 (T_{Sfp} = 0.74)



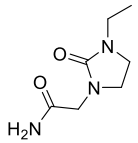
9 (T_{Sfp} = 0.86)



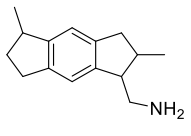
10 (T_{Sfp} = 0.69)



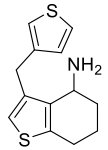
13 (T_{Sfp} = 0.81)



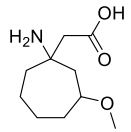
14 (T_{Sfp} = 0.43)



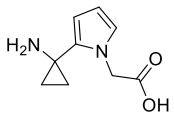
3 (T_{Sfp} = 0.5)



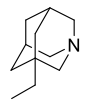
4 (T_{Sfp} = 0.21)



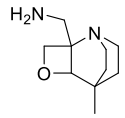
7 (T_{Sfp} = 0.53)



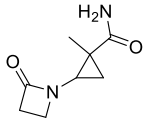
8 (T_{Sfp} = 0.12)



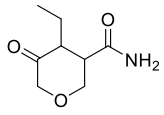
11 (T_{Sfp} = 0.56)



12 (T_{Sfp} = 0.25)



15 (T_{Sfp} = 0.46)



16 (T_{Sfp} = 0.20)

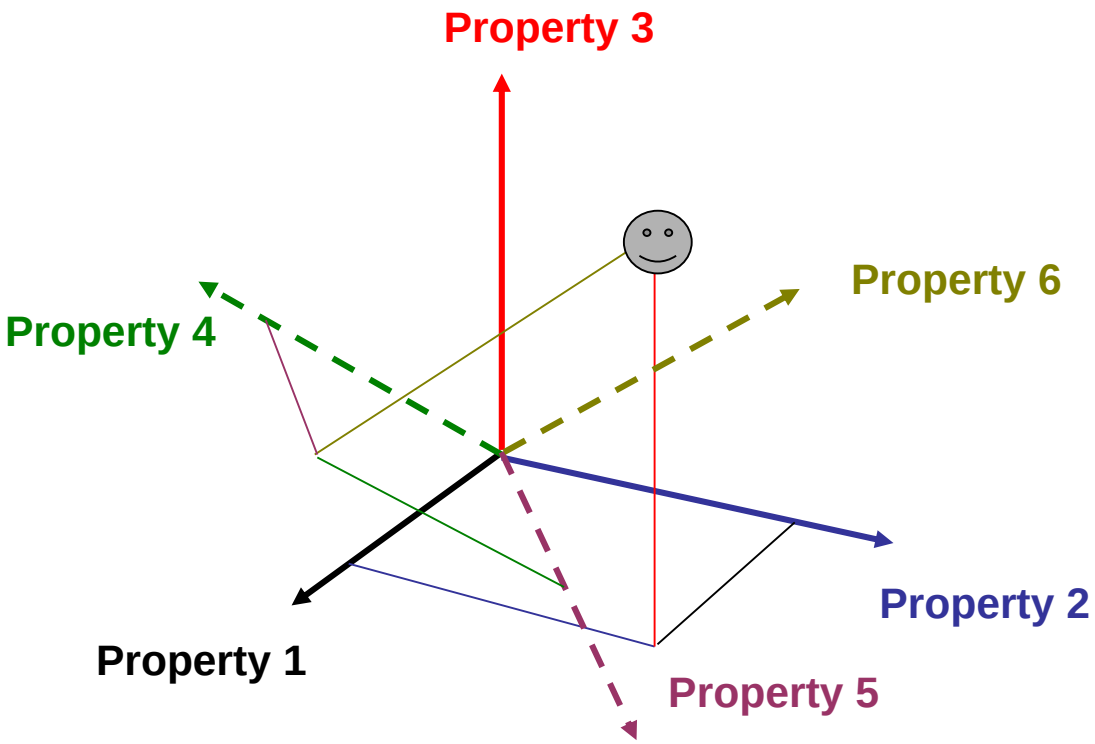
Understanding Molecular Diversity

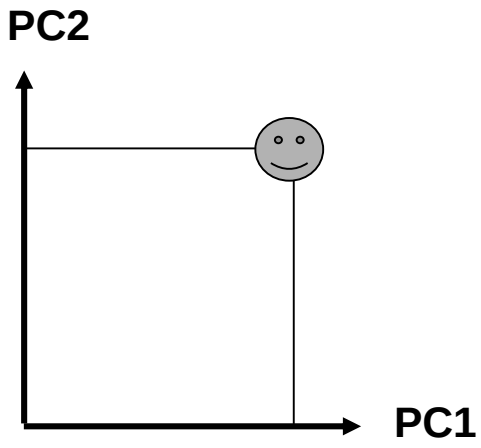
1. GDB-17 and FDB-17
2. **Visualization**
3. Virtual Screening
4. Extension to larger molecules

n-dimensional Space
search

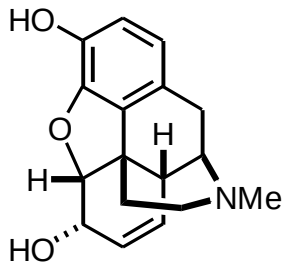
PCA

2D (or 3D) projection
Visualise

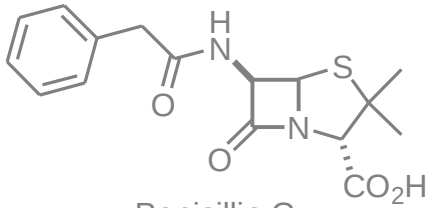




Molecular Quantum Numbers



Morphine



Penicillin G

Atoms		
Carbon	**17**	16
Fluorine	**0**	0
Chlorine	**0**	0
Bromine	**0**	0
Iodine	**0**	0
Sulphur	**0**	1
Phosphor	**0**	0
Acyclic nitrogen	**0**	1
Cyclic nitrogen	**1**	1
Acyclic oxygen	**2**	4
Cyclic oxygen	**1**	0
Heavy atom count	**21**	23

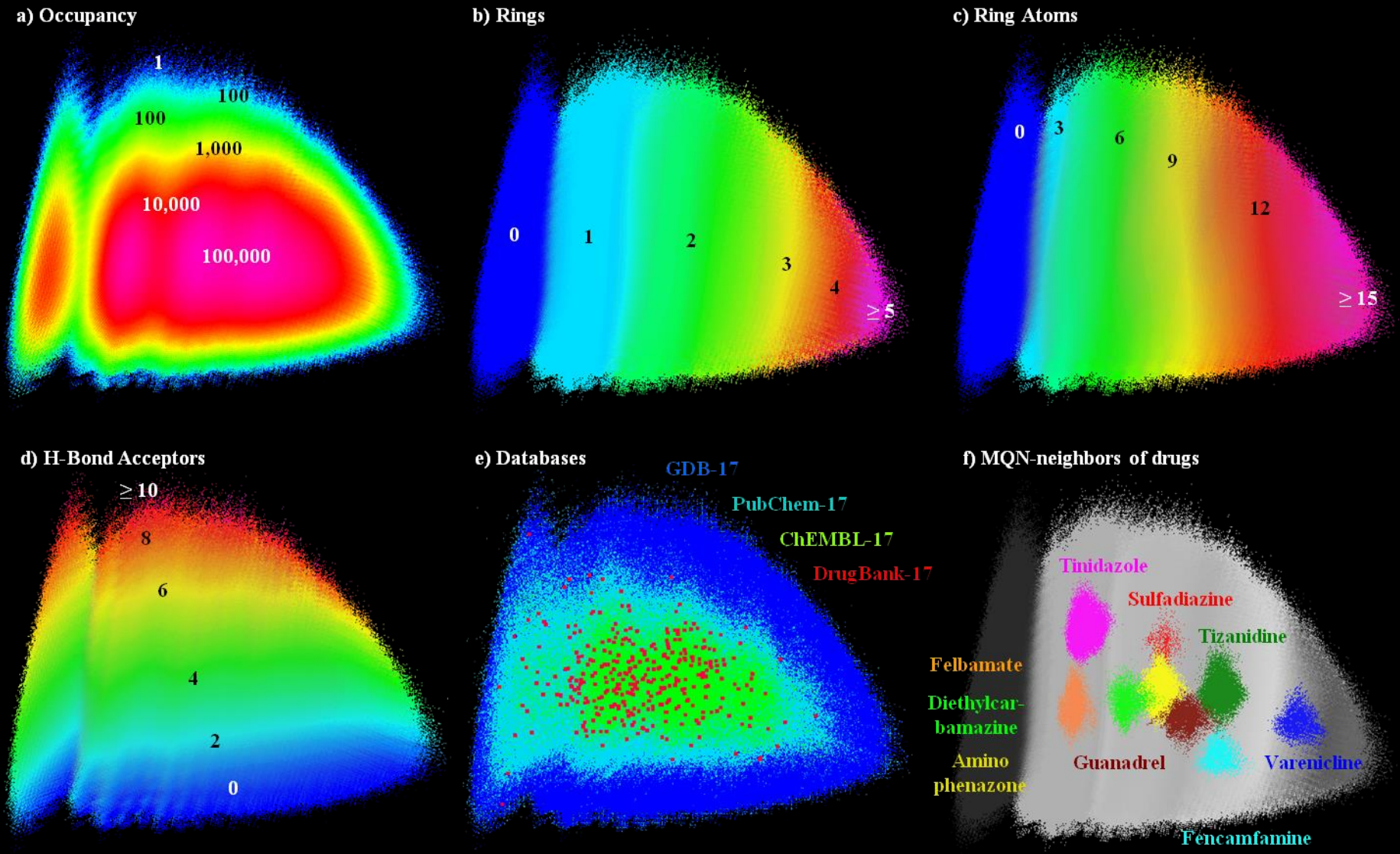
Polar groups		
H-Bond donor atoms	**3**	1
H-Bond donor sites	**3**	1
H-Bond acceptor atoms	**3**	4
H-Bond acceptor sites	**3**	7
Positive charges	**1**	0
Negative charges	**0**	1

Bonds		
Acyclic single bonds	**3**	8
Acyclic double bonds	**0**	3
Acyclic triple bonds	**0**	0
Cyclic single bonds	**18**	11
Cyclic double bonds	**4**	3
Cyclic triple bonds	**0**	0
Rotatable bonds	**0**	4

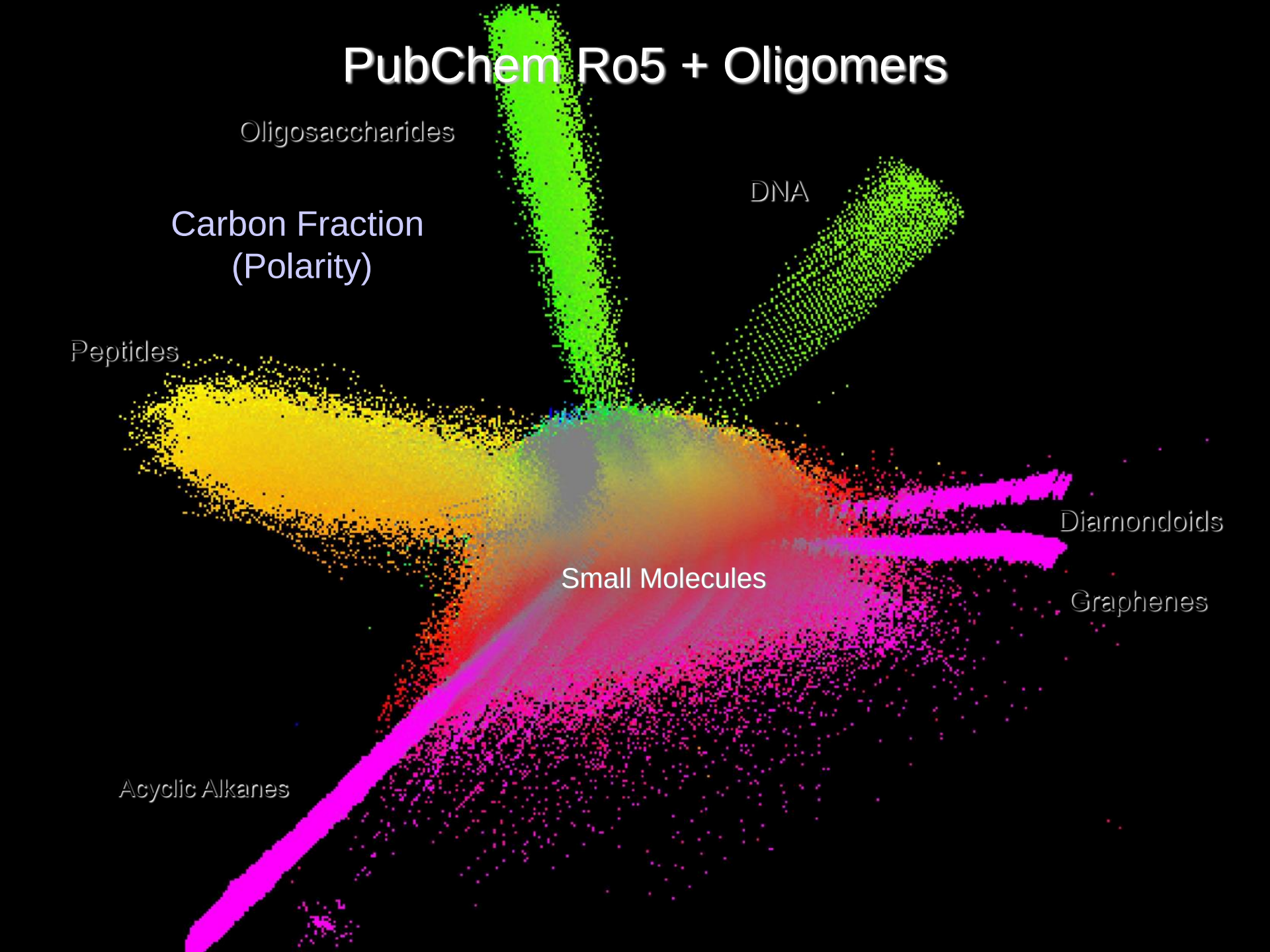
Topology		
Acyclic monovalent nodes	**3**	6
Acyclic divalent nodes	**0**	2
Acyclic trivalent nodes	**0**	2
Acyclic tetravalent nodes	**0**	0
Cyclic divalent nodes	**8**	6
Cyclic trivalent nodes	**9**	6
Cyclic tetravalent nodes	**1**	1
3-Membered rings	**0**	0
4-Membered rings	**0**	1
5-Membered rings	**1**	1
6-Membered rings	**4**	1
7-Membered rings	**0**	0
8-Membered rings	**0**	0
9-Membered rings	**0**	0
≥ 10 membered rings	**0**	0
Atoms shared by fused rings	**7**	2
Bonds shared by fused rings	**6**	1

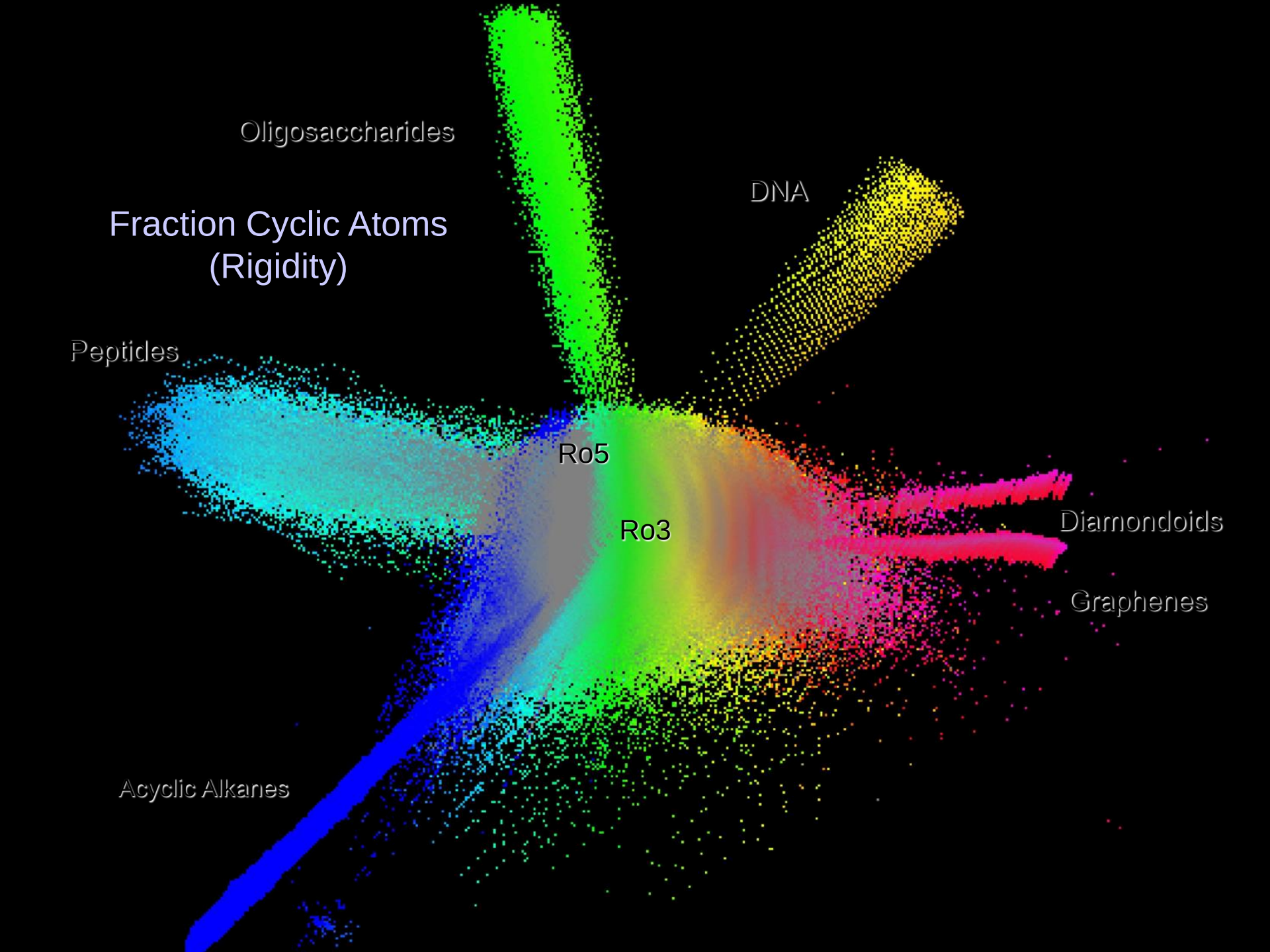
R. van Deursen et al. *J. Chem. Inf. Model.* **2010**, *50*, 1924-1934

GDB-17



Visualization and Virtual Screening of the Chemical Universe Database GDB-17. L. Ruddigkeit, L. C. Blum, J.-L. Reymond, *J. Chem. Inf. Model.* **2013**, *53*, 56-65.


PubChem Ro5 + Oligomers
Oligosaccharides
DNA
Carbon Fraction
(Polarity)
Peptides
Diamondoids
Small Molecules
Graphenes
Acyclic Alkanes


Oligosaccharides
DNA
Fraction Cyclic Atoms
(Rigidity)
Peptides
Ro5
Ro3
Diamondoids
Graphenes
Acyclic Alkanes

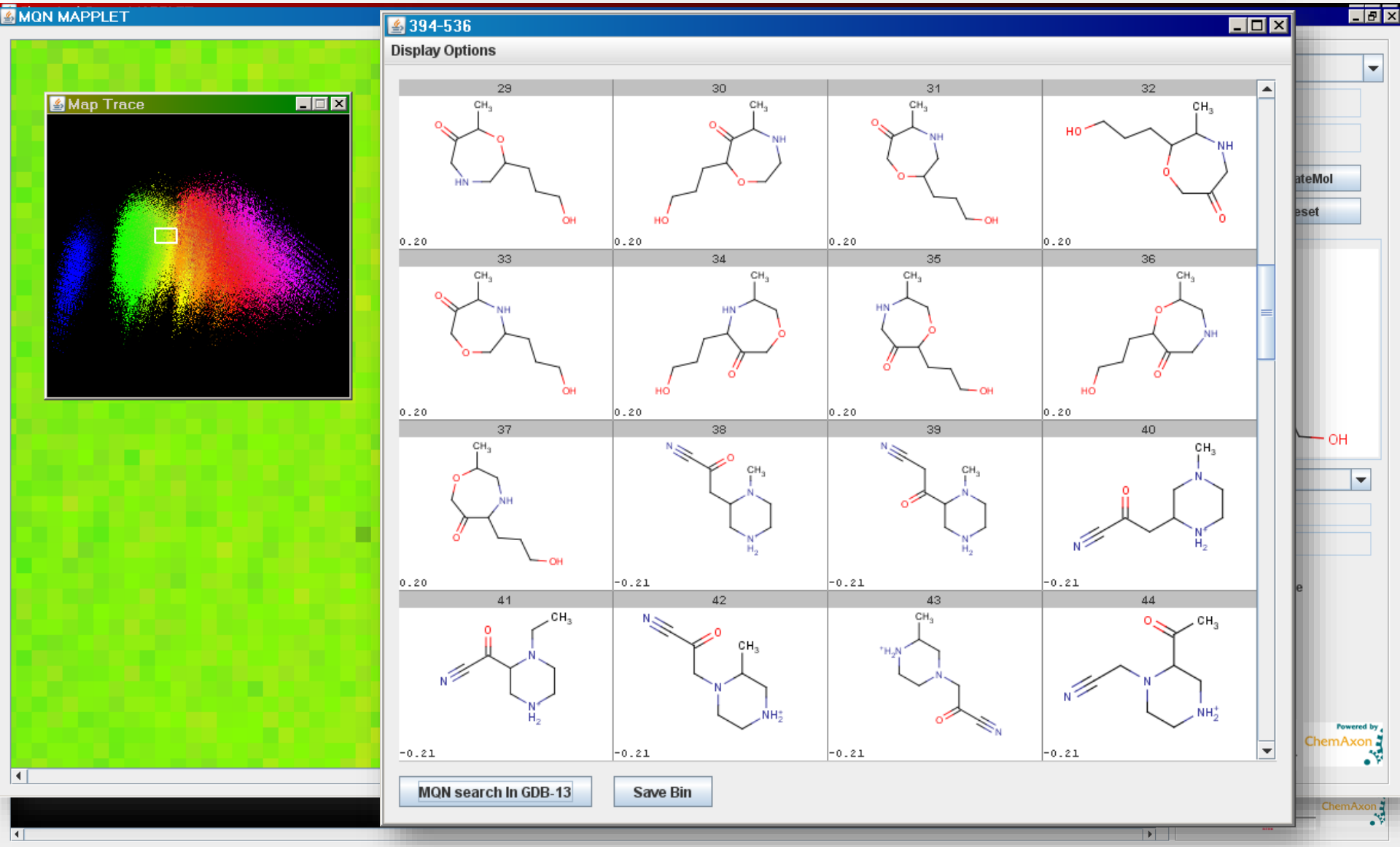
- N-dimensional fingerprints:
  - APfp: 20 counts for atom pairs (shape)
  - MQN: 42 molecular quantum numbers
  - SMIfp: 34 counts for SMILES characters
  - Xfp: 55 counts for atom pairs with properties (pharmacophore)
  - Sfp: 1024-bit binary substructure fp
  - ECfp4: 1024-bit binary extended connectivity fp

- Tools at gdb.unibe.ch
  - **Web-browsers** for DrugBank, ChEMBL, ZINC, PubChem, GDB-11, GDB-13, GDB-17
  - **Mapplets** (downloadable Java applications, ca. 100 Mb)
  - **Similarity Mapplets** (tailored mapplets by e-mail)
  - **WebDrugCS** (3D-viewer, platform independent, DrugBank)
  - **WebMolCS** (3D-viewer, platform independent, up to 5000 molecules from user)

Visualisation and subsets of the chemical universe database GDB-13 for virtual screening. L. C. Blum, R. van Deursen, J.-L. Reymond, *J. Comput. Aided Mol. Des.* **2011**, *25*, 637-647
A multi-fingerprint browser for the ZINC database. M. Awale, J.-L. Reymond, *Nucleic Acids Res.* **2014**, *42*, 234-239
The MQN-Mapplet: Visualization of Chemical Space with Interactive Maps of DrugBank, ChEMBL, PubChem, GDB-11 and GDB-13. M. Awale, R. van Deursen, J. L. Reymond, *J. Chem. Inf. Model.* **2013**, *53*, 509-518
Similarity Mapplet: Interactive Visualization of the Directory of Useful Decoys and ChEMBL in High Dimensional Chemical Spaces. Awale M, Reymond JL, *J. Chem. Inf. Model.*, **2015**, *55*, 1509-1516.

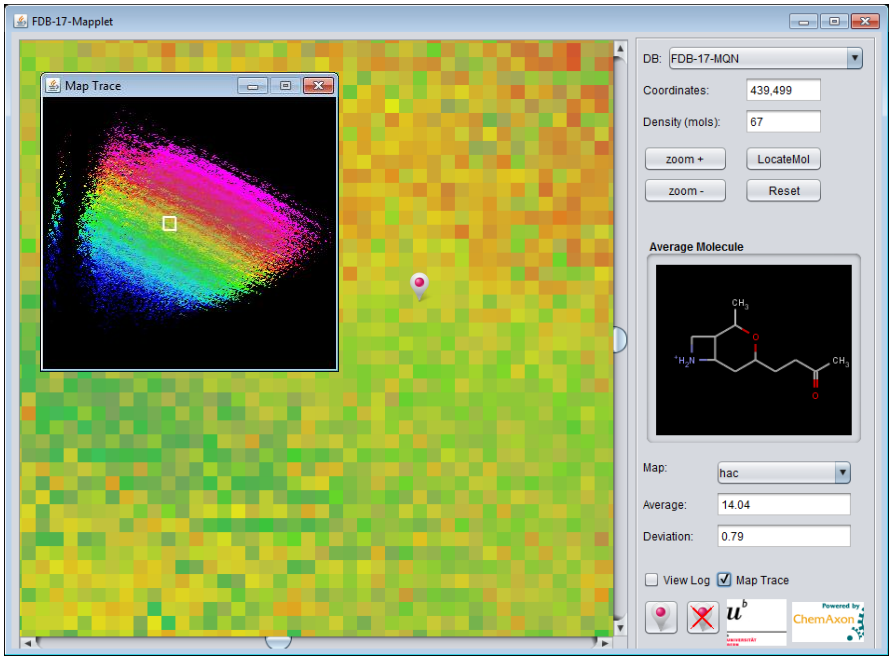
14

MQN-Mapplet

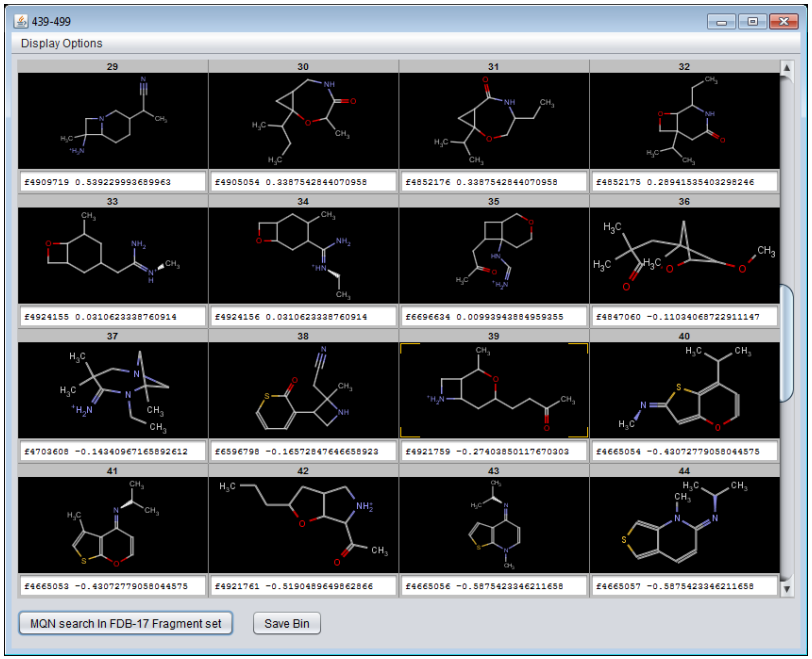


The MQN-Mapplet: Visualization of Chemical Space with Interactive Maps of DrugBank, ChEMBL, PubChem, GDB-11 and GDB-13. M. Awale, R. van Deursen, J. L. Reymond, *J. Chem. Inf. Model.* **2013**, *53*, 509-518

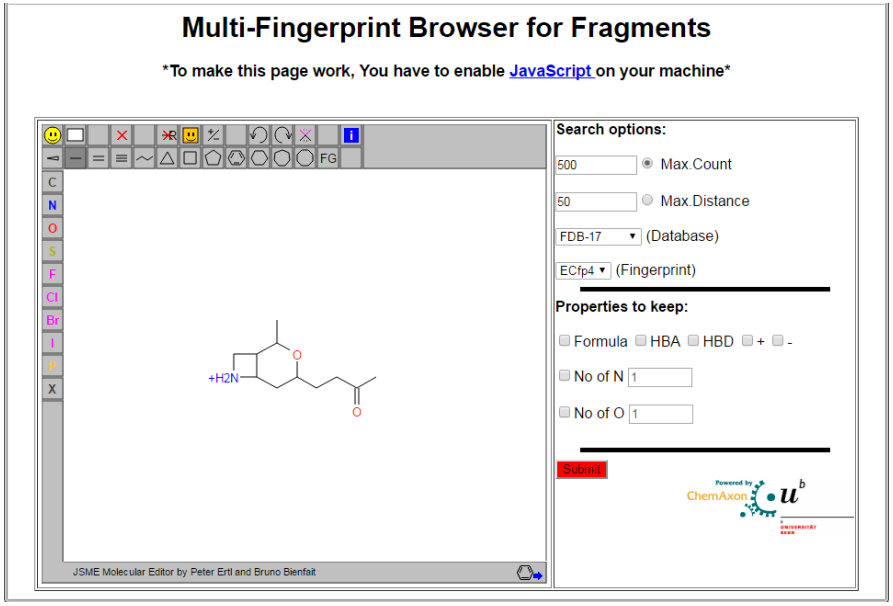
a)


FDB-17-Mapplet
Map Trace
DB: FDB-17-MQN
Coordinates: 439,499
Density (mols): 67
zoom +
LocateMol
zoom -
Reset
Average Molecule
Map: hac
Average: 14.04
Deviation: 0.79
View Log
Map Trace
Powered by ChemAxon

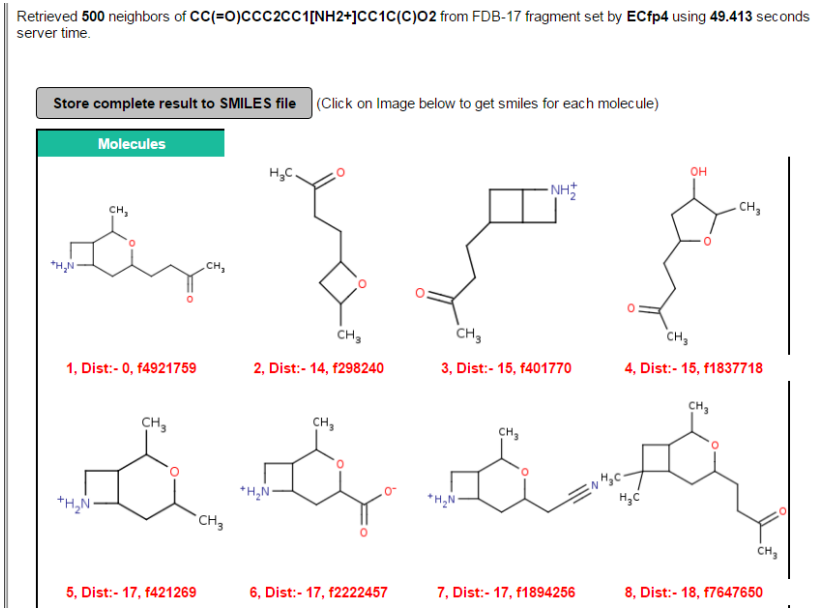
b)


439-499
Display Options
MQN search In FDB-17 Fragment set
Save Bin

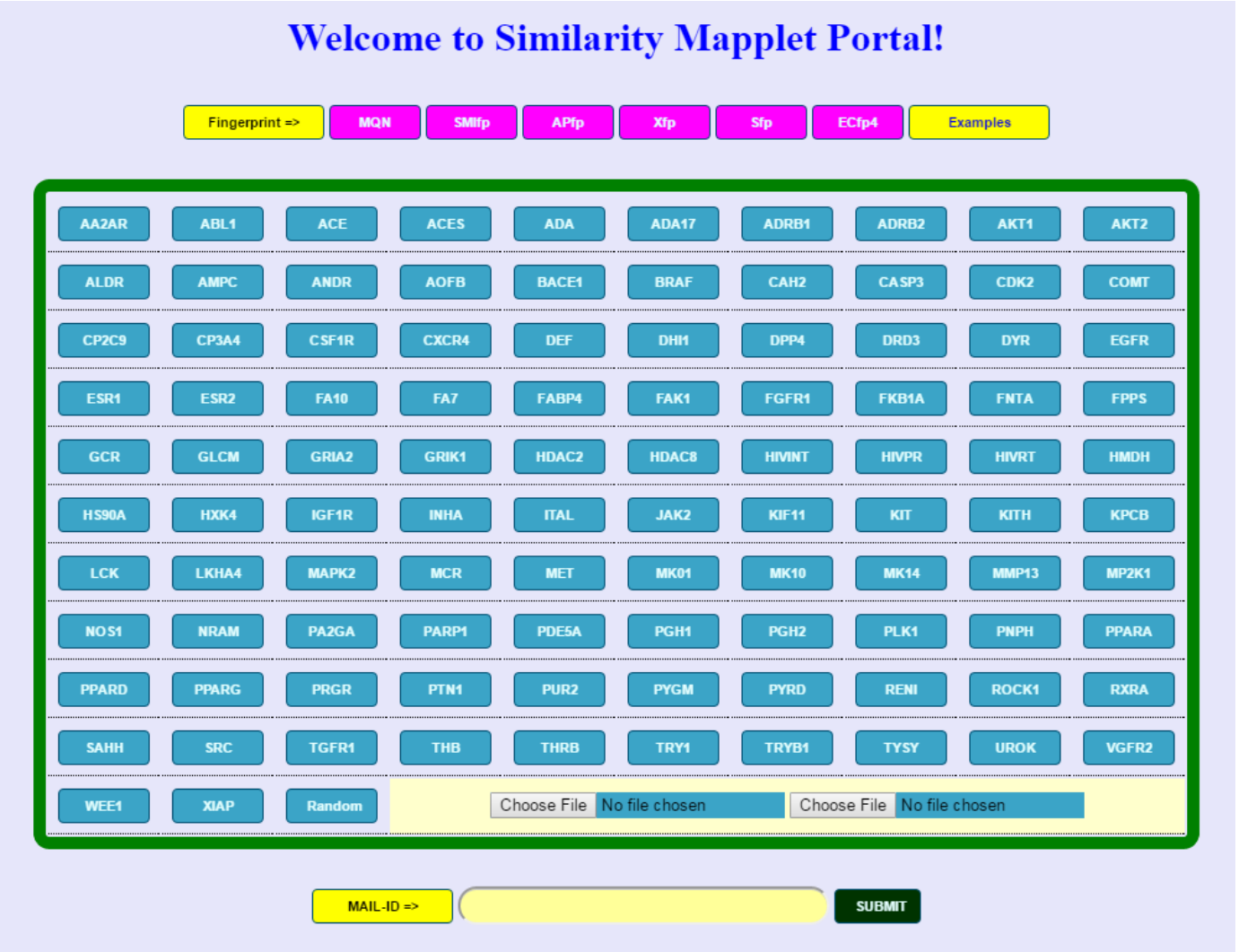
c)


Multi-Fingerprint Browser for Fragments
To make this page work, You have to enable JavaScript on your machine
Search options:
500
Max.Count
50
Max.Distance
FDB-17
(Database)
ECfp4
(Fingerprint)
Properties to keep:
Formula
HBA
HBD
+
-
No of N
1
No of O
1
Submit
JSME Molecular Editor by Peter Ertl and Bruno Bienfait
Powered by ChemAxon

d)

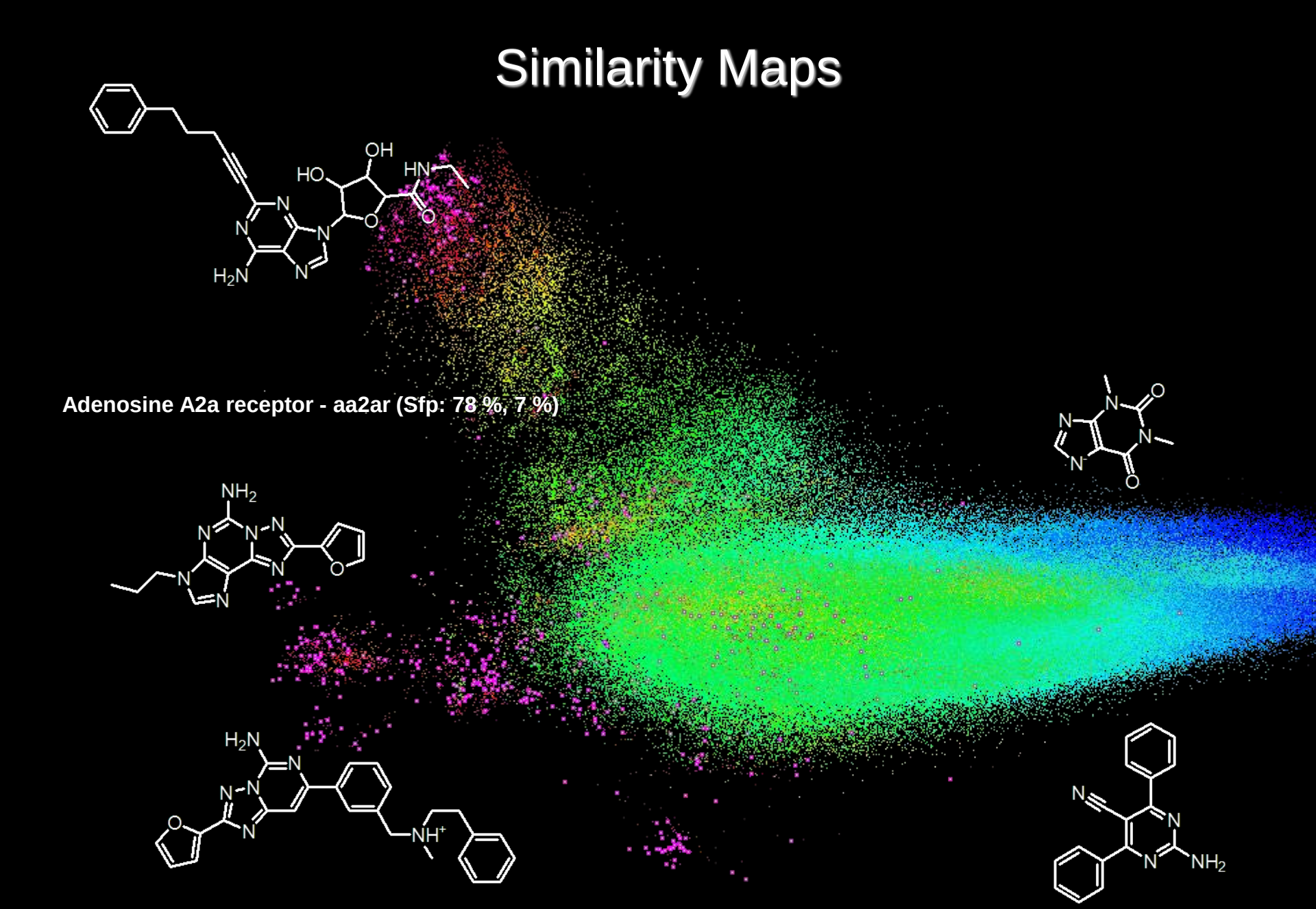

Retrieved 500 neighbors of CC(=O)CCC2CC1[NH2+]CC1C(C)O2 from FDB-17 fragment set by ECfp4 using 49.413 seconds server time.
Store complete result to SMILES file
(Click on Image below to get smiles for each molecule)
Molecules
1, Dist:- 0, f4921759
2, Dist:- 14, f298240
3, Dist:- 15, f401770
4, Dist:- 15, f1837718
5, Dist:- 17, f421269
6, Dist:- 17, f2222457
7, Dist:- 17, f1894256
8, Dist:- 18, f7647650

Similarity Maps



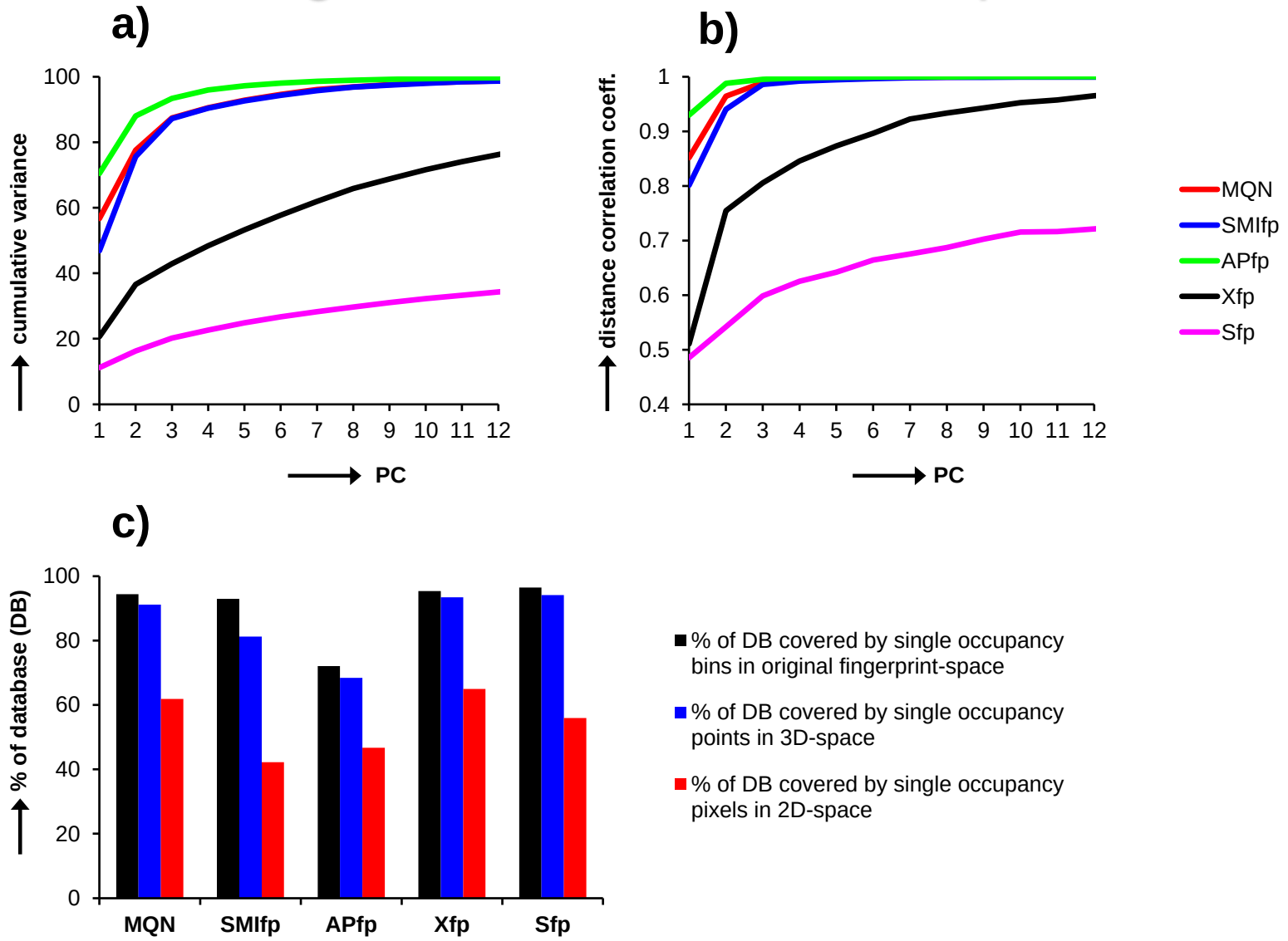
Similarity Mapplet: Interactive Visualization of the Directory of Useful Decoys and ChEMBL in High Dimensional Chemical Spaces. Awale M, Reymond JL, *J. Chem. Inf. Model.*, **2015**, *55*, 1509-1516

Similarity Maps

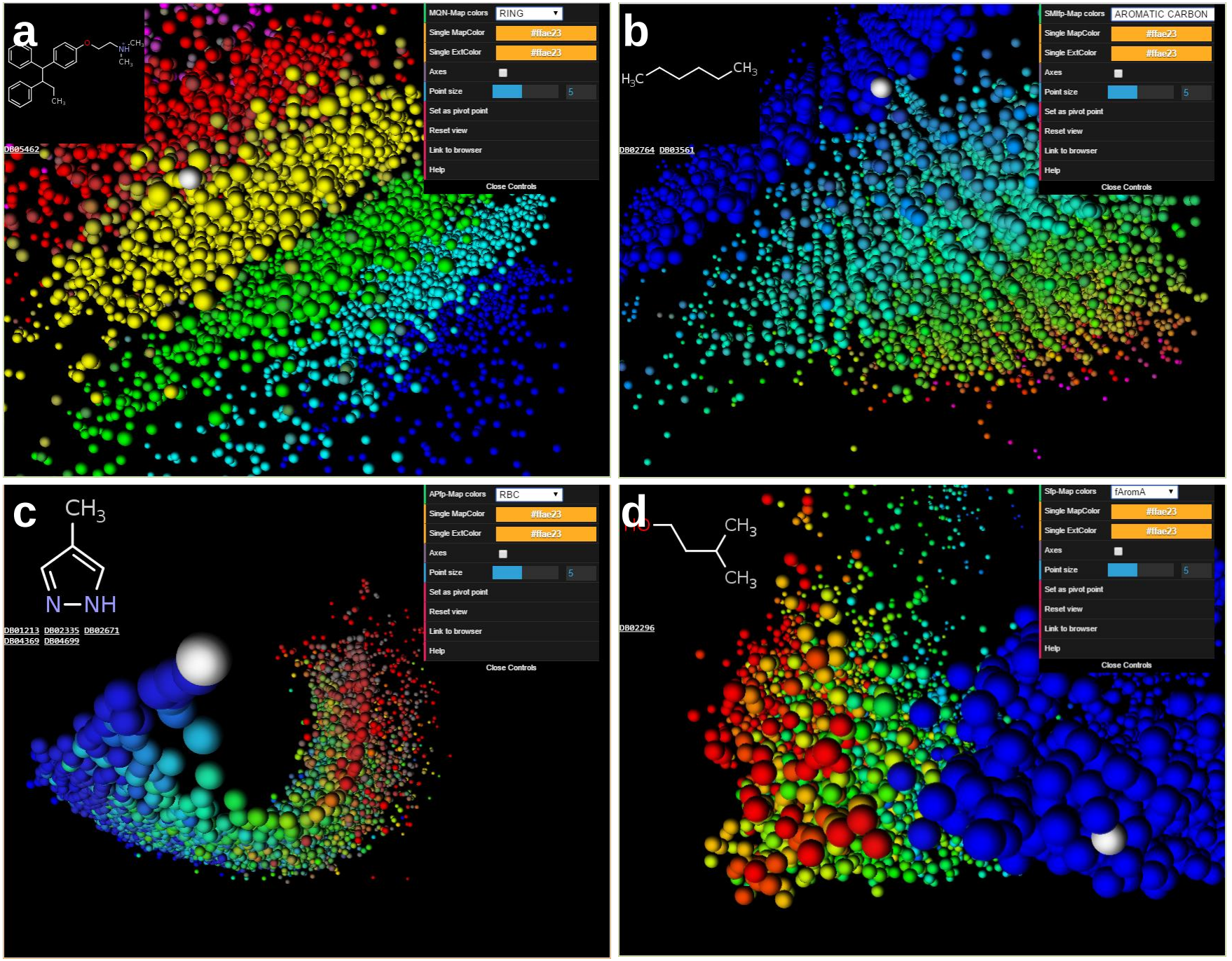


Similarity Mapplet: Interactive Visualization of the Directory of Useful Decoys and ChEMBL in High Dimensional Chemical Spaces. Awale M, Reymond JL, *J. Chem. Inf. Model.*, **2015**, *55*, 1509-1516

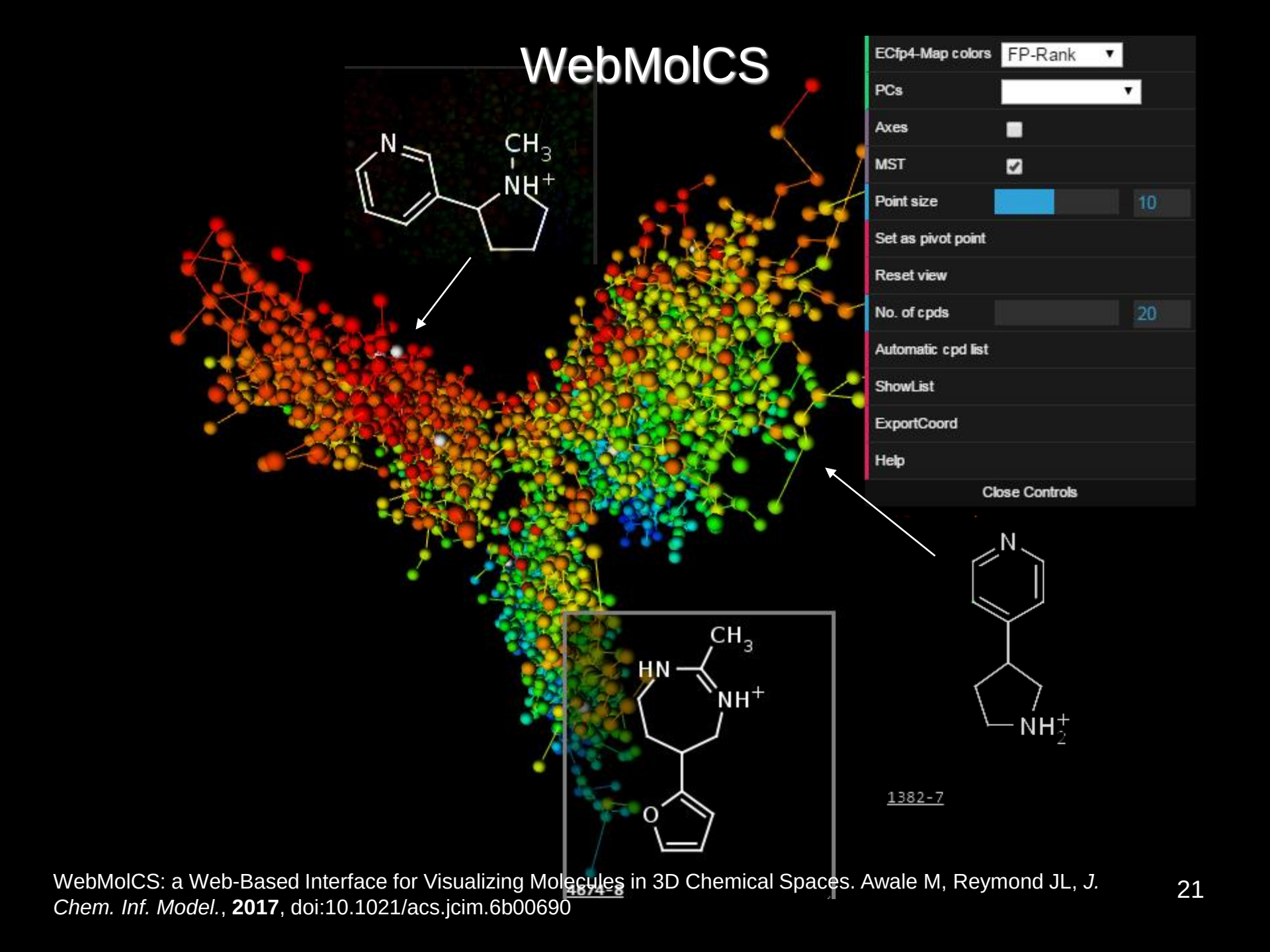
DrugBank in 3D Chemical Spaces



Web-based 3D-visualization of the DrugBank chemical space Awale M and Reymond JL, *J. Cheminform.*, **2016**, doi:10.1186/s13321-016-0138-2.


a
DB05462
MQN-Map colors
RING
Single MapColor
#ffae23
Single ExtColor
#ffae23
Axes
Point size
5
Set as pivot point
Reset view
Link to browser
Help
Close Controls
b
DB02764 DB03561
SMIfp-Map colors
AROMATIC CARBON
Single MapColor
#ffae23
Single ExtColor
#ffae23
Axes
Point size
5
Set as pivot point
Reset view
Link to browser
Help
Close Controls
c
DB01213 DB02335 DB02671
DB04369 DB04699
APfp-Map colors
RBC
Single MapColor
#ffae23
Single ExtColor
#ffae23
Axes
Point size
5
Set as pivot point
Reset view
Link to browser
Help
Close Controls
d
DB02296
Sfp-Map colors
fAromA
Single MapColor
#ffae23
Single ExtColor
#ffae23
Axes
Point size
5
Set as pivot point
Reset view
Link to browser
Help
Close Controls

WebMolCS



WebMolCS: a Web-Based Interface for Visualizing Molecules in 3D Chemical Spaces. Awale M, Reymond JL, *J. Chem. Inf. Model.*, **2017**, doi:10.1021/acs.jcim.6b00690

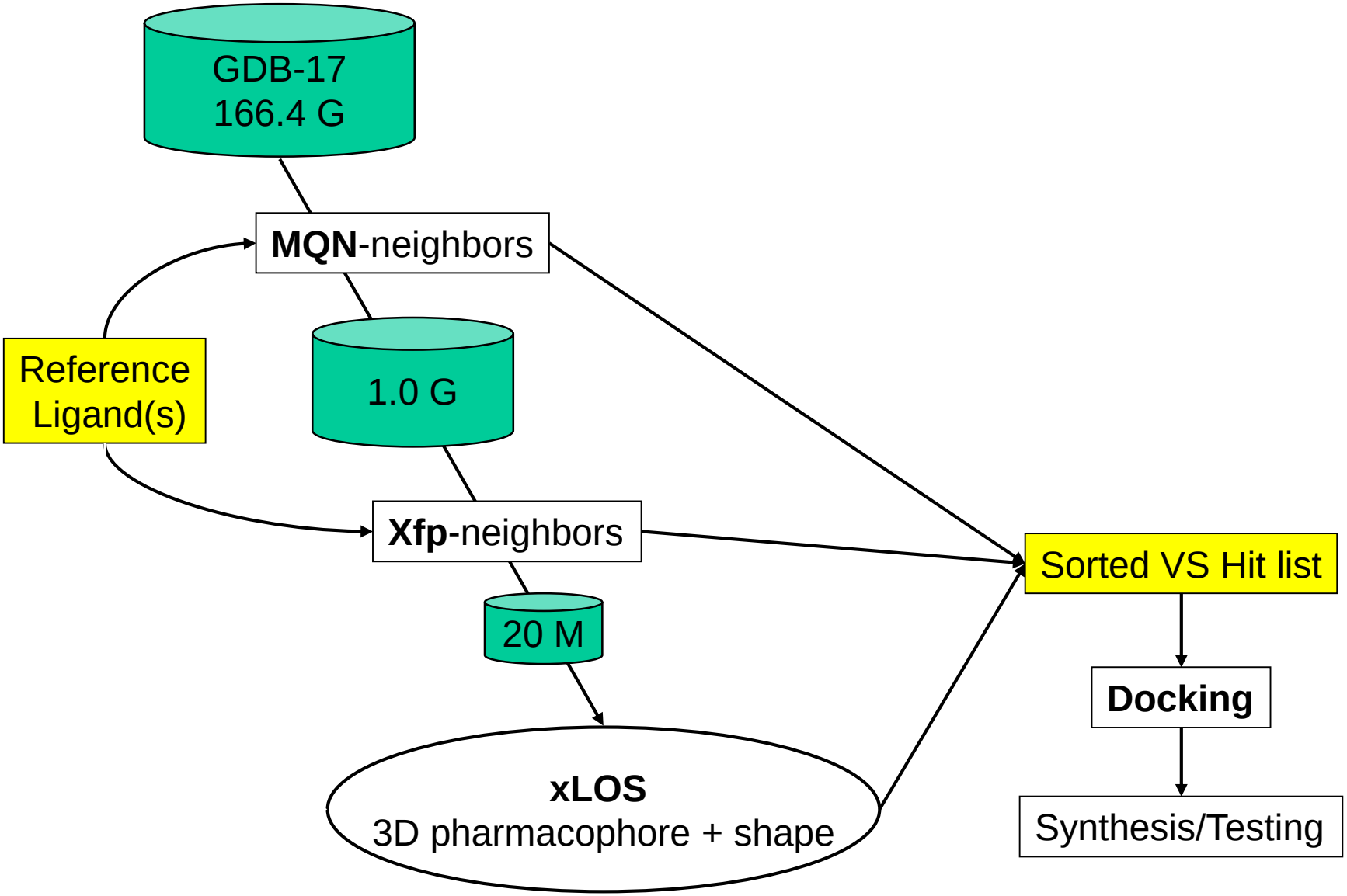
Does this work?

1. GDB-17 and FDB-17
2. Visualization
3. **Virtual Screening**
4. Extension to larger molecules

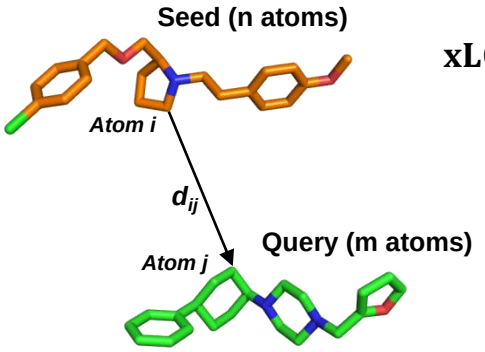
HiQSCREEN
High Quality Screening for Drug Discovery

NCCR TransCure
Excellence in Membrane Transport Research

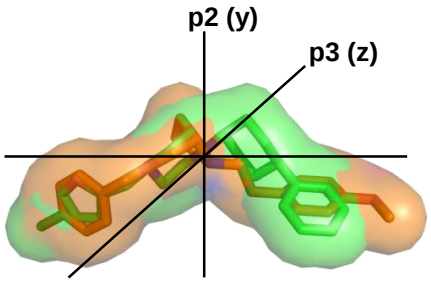
SWISS NATIONAL CENTRE OF COMPETENCE IN RESEARCH
CHEMICAL BIOLOGY
VISUALISATION AND CONTROL
OF BIOLOGICAL PROCESSES USING CHEMISTRY

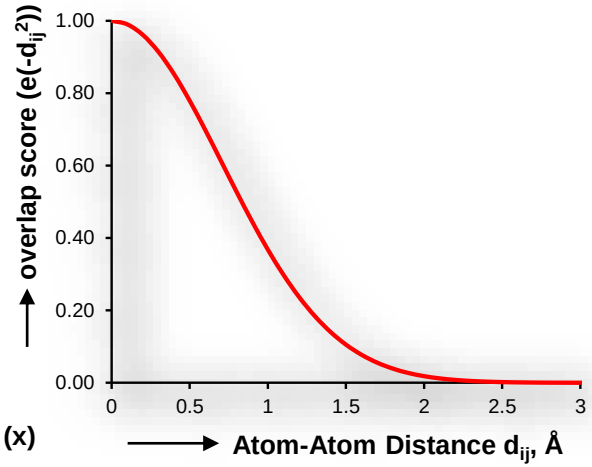

GDB-17
166.4 G
MQN-neighbors
Reference
Ligand(s)
1.0 G
Xfp-neighbors
20 M
xLOS
3D pharmacophore + shape
Sorted VS Hit list
Docking
Synthesis/Testing

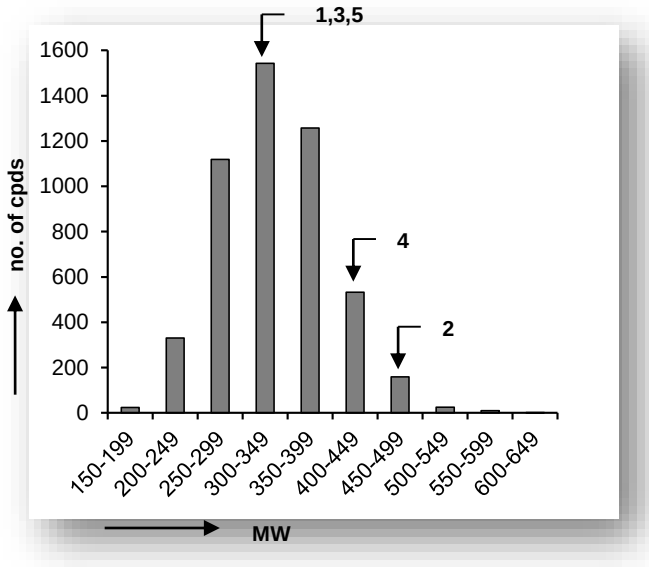
3D-Shape and pharmacophore (xLOS)



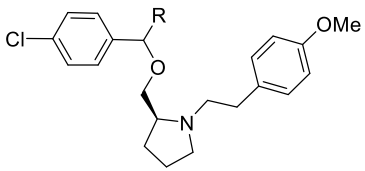
$$\text{xLOS score} = \frac{\left(\sum_{j=1}^{m_d}\sum_{i=1}^{n_d} e^{-d_{ij}^2}\right)}{m_d + n_d} + \frac{\left(\sum_{j=1}^{m_a}\sum_{i=1}^{n_a} e^{-d_{ij}^2}\right)}{m_a + n_a} + \frac{2 \times \left(\sum_{j=1}^{m_h}\sum_{i=1}^{n_h} e^{-d_{ij}^2}\right)}{m_h + n_h}$$



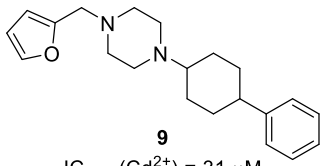




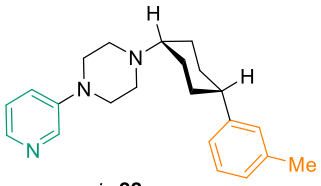
Optimization of TRPV6 Calcium Channel Inhibitors Using a 3D Ligand-Based Virtual Screening Method. C Simonin, M Awale et al., *Angew. Chem., Int. Ed. Engl.* **2015**, *54*, 14748-14752.



1 (R = H) IC_{50} = 90 μM
2 (R = Ph) IC_{50} = 675 μM



9
IC_{50} (Cd^{2+}) = 31 μM
IC_{50} (Ca^{2+}) = 26 μM



cis-**22a**
IC_{50} = 0.32 ± 0.12 μM

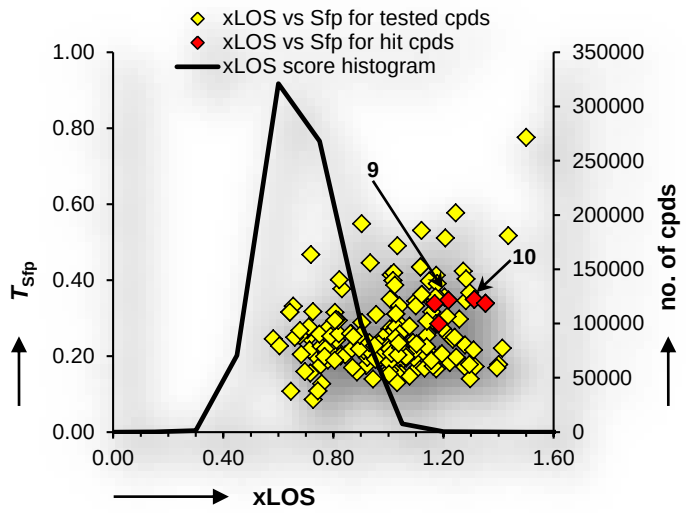
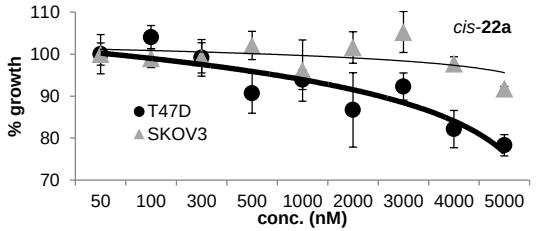
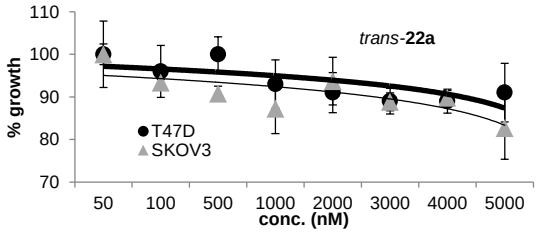
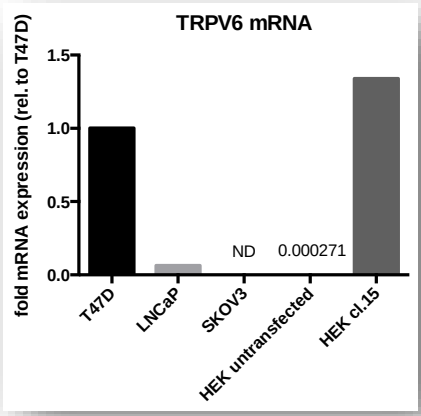


Table S1: *In vitro* pharmacology profile of *cis*-**22a** on several ion channels.[a]

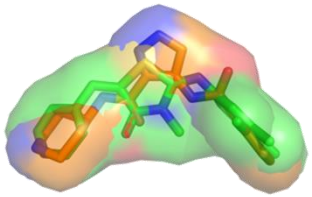
Channel	% Inh. @ 10 μM
TRP channels	
TRPV1 (agonist effect) (*h*)	-15.5 ± 0.7[c]
TRPV1 (antagonist effect) (*h*)	29.6 ± 8.6
TRPV3 (antagonist effect) (*h*)	-11.5 ± 2.0
TRPV5 (antagonist effect) (*r*)	79.9 ± 1.4 (IC_{50} = 2.4 μM)
TRPM8 (agonist effect) (*h*)	-7.4 ± 1.4[c]
TRPM8 (antagonist effect) (*h*)	21.2 ± 7.5
Voltage-gated ion channels	
Ca^{2+} channel L-type, dihydropyridine (*r*)	-1.6 ± 20.5
Ca^{2+} channel L-type, diltiazem (*r*)	18.4 ± 13.0
Ca^{2+} channel L-type, verapamil (*r*)	31.9 ± 3.0
Ca^{2+} channels, N-type (*r*)	4.4 ± 8.1
K^+ channel [K_v] (*r*)	-15.9 ± 3.8
hERG (*h*)	82.3 ± 4.0[d]
Na^+ channels, site 2 (*r*)	52.1 ± 4.2
Ligand-gated ion channels	
5-HT_3 (*h*)	-1.3 ± 5.9
GABA, central benzodiazepine (*r*)[b]	-10.7 ± 8.7
Glutamate, NMDA (*r*)	8.2 ± 1.5
Store-operated Ca^{2+} channels	-10.1 ± 6.0





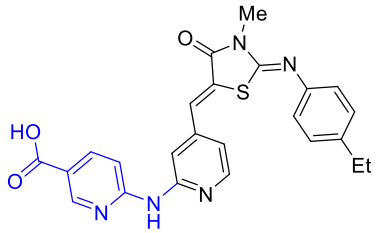


Optimization of TRPV6 Calcium Channel Inhibitors Using a 3D Ligand-Based Virtual Screening Method. C Simonin, M Awale et al., *Angew. Chem., Int. Ed. Engl.* **2015**, *54*, 14748-14752.

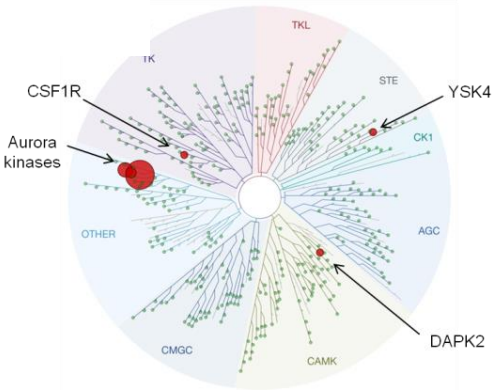


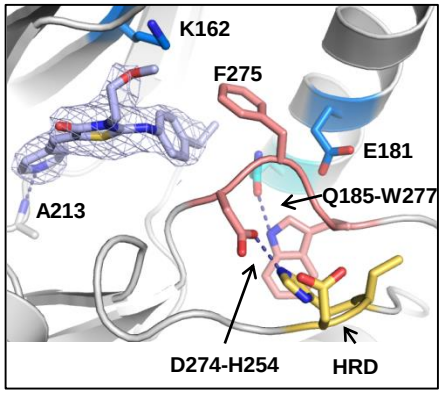
1) xLOS
2) Biochemical screen
3) Structure-based optimization

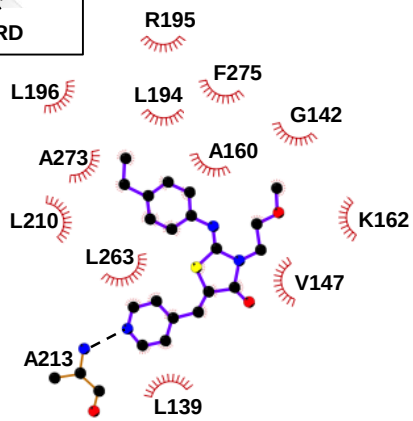
Kinase Inhibitors

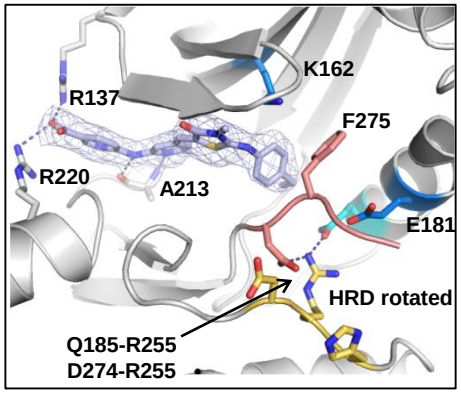


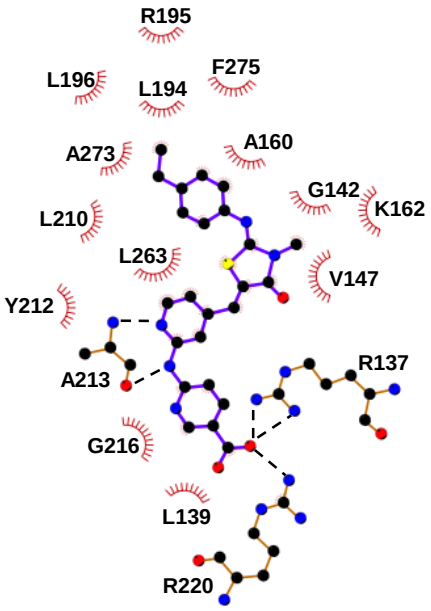
Aurora A
IC_{50} = 2 nM



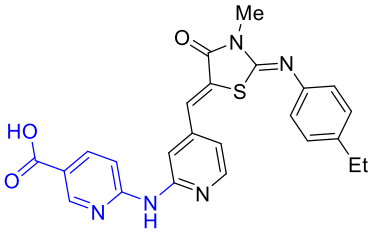


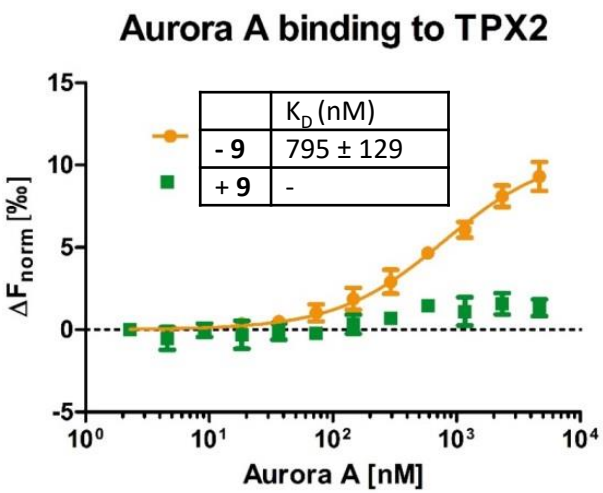


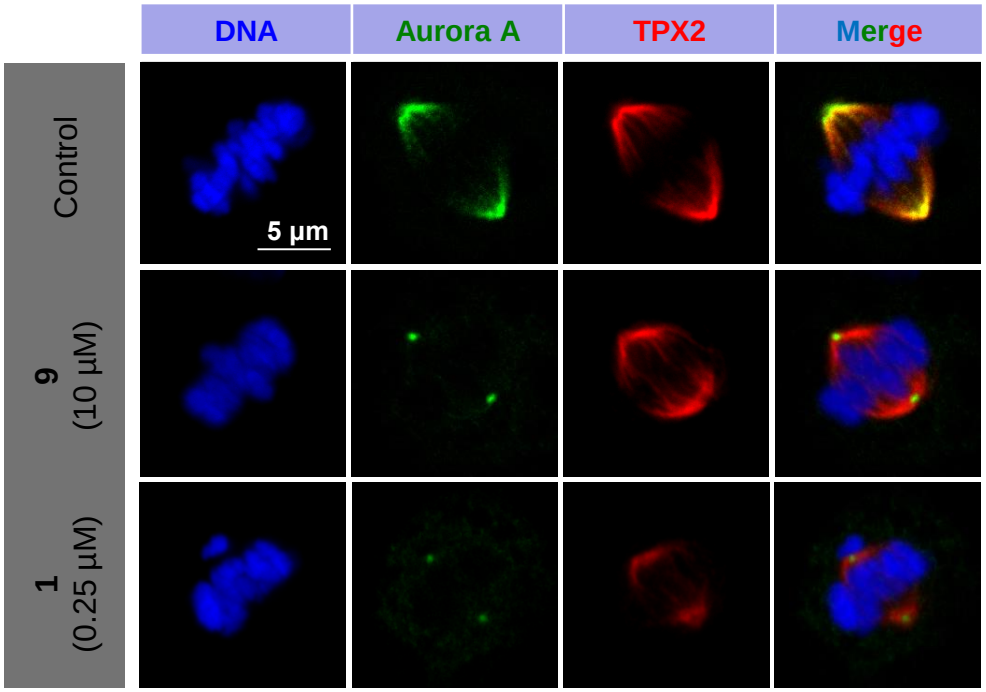


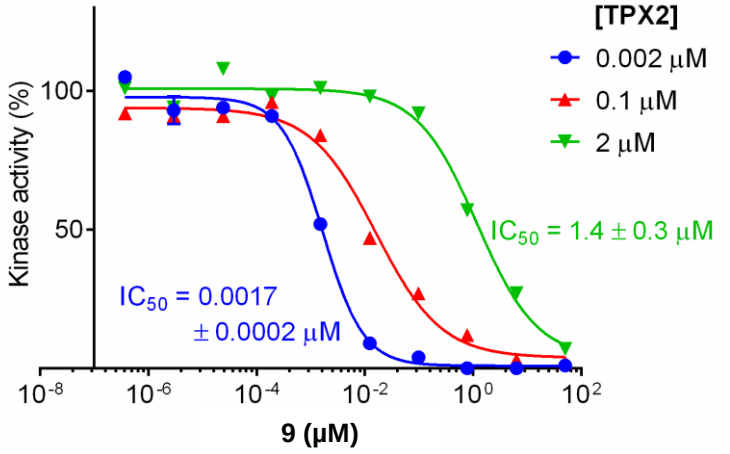


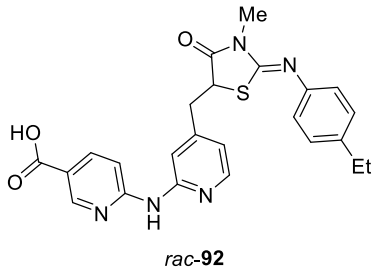
Discovery of a Selective Aurora A Kinase Inhibitor by Virtual Screening. Falco Kilchmann et al., *J. Med. Chem.*, **2016**, doi:10.1021/acs.jmedchem.6b00709


Me
O
N
S
N
HO
O
N
H
N
Et

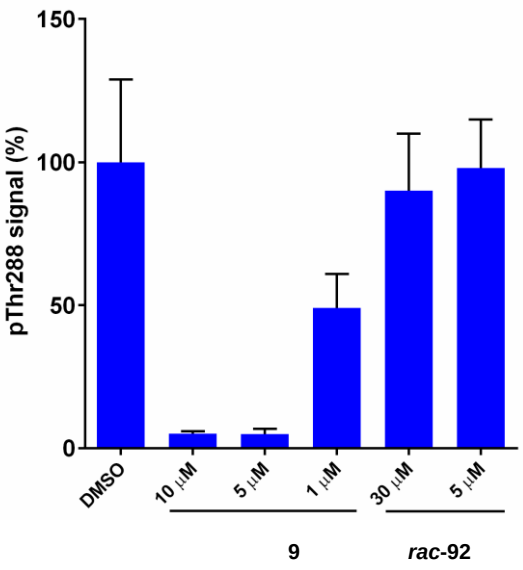

Aurora A binding to TPX2
K_D (nM)
- 9
795 ± 129
+ 9
-
ΔF_norm [‰]
Aurora A [nM]


DNA
Aurora A
TPX2
Merge
Control
9 (10 μM)
1 (0.25 μM)
5 μm


[TPX2]
0.002 μM
0.1 μM
2 μM
Kinase activity (%)
IC50 = 1.4 ± 0.3 μM
IC50 = 0.0017
± 0.0002 μM
9 (μM)


Me
O
N
S
N
HO
O
N
H
N
Et

rac-**92**


pThr288 signal (%)
DMSO
10 μM
5 μM
1 μM
30 μM
5 μM
9
rac-92

Awale and Reymond *J Cheminform* (2017) 9:11
DOI 10.1186/s13321-017-0199-x

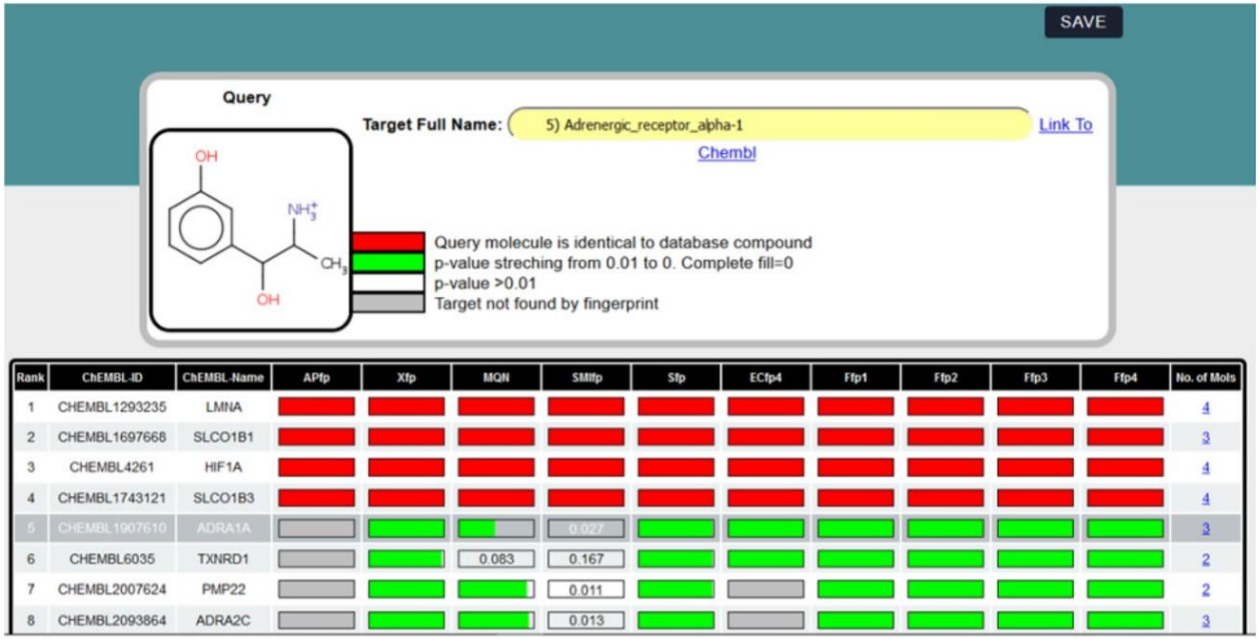
Journal of Cheminformatics

METHODOLOGY

Open Access

The polypharmacology browser: a web-based multi-fingerprint target prediction tool using ChEMBL bioactivity data

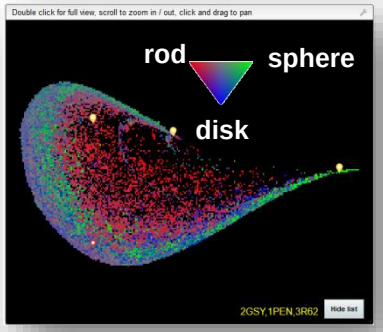
Mahendra Awale and Jean-Louis Reymond*



Rank	ChEMBL-ID	ChEMBL-Name	APfp	Xfp	MQN	SMIfp	Sfp	ECfp4	Ffp1	Ffp2	Ffp3	Ffp4	No. of Mols
1	CHEMBL1293235	LMNA											4
2	CHEMBL1697668	SLCO1B1											3
3	CHEMBL4261	HIF1A											4
4	CHEMBL1743121	SLCO1B3											4
5	CHEMBL1907610	ADRA1A				0.027							3
6	CHEMBL6035	TXNRD1			0.083	0.167							2
7	CHEMBL2007624	PMP22				0.011							2
8	CHEMBL2093864	ADRA2C				0.013							3

New Molecules

1. GDB-17 and FDB-17
2. Visualization
3. Virtual Screening
4. **Extension to larger molecules**



www.cheminfo.org/pdbexplorer

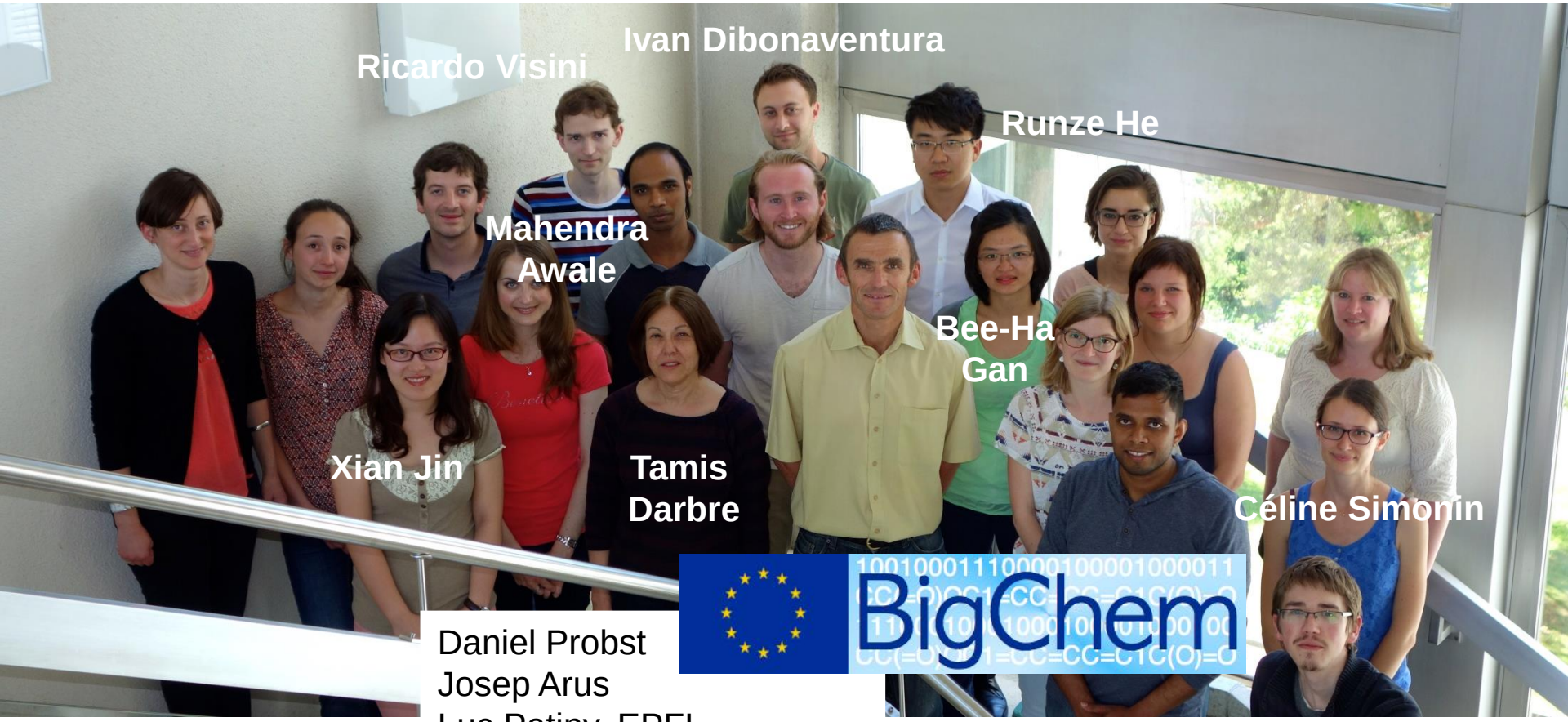
Help Page
Select image
Heavy atom count
Drop your pdb file here
Double click for full view, scroll to zoom in / out, click and drag to pan
2GSY,1PEN,3R62
Hide list
2A18
View in jsmol
mode
cartoon
Sort by similarity
Enter pdb code
2A18
Max Count
100
Max Distance
10000
Search by similarity
List of proteins in clicked cell (2A18)
Download zip
Rank
Code
pdb
pdb....
Jsmol
Distance
2A18
4A61
3BLH
Hovered
Name: 2A18
Position: 79, 195
Avg, stdev: 4760.76, 314.68
Lookup a list of PDBs
Enter a list of PDBs
1PEN,2GSY,3R62

Jin X, Awale M, Zasso M, Kostro D, Patiny L, Reymond JL: PDB-Explorer: a web-based interactive map of the protein data bank in shape space. BMC Bioinformatics **2015**, *16*, 339.

FNSNF
Fonds national suisse
Schweizerischer Nationalfonds
Fondo nazionale svizzero
Swiss National Science Foundation

gdb.unibe.ch

b UNIVERSITÄT BERN



BigChem

Daniel Probst
Josep Arus
Luc Patiny, EPFL
Daniel Bertrand, HiQscreen
Matthias Hediger, UniBE

HiQSCREEN
High Quality Screening for Drug Discovery

NCCR TransCure
Excellence in Membrane Transport Research