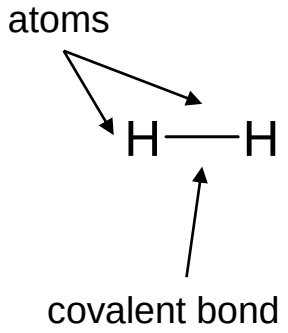
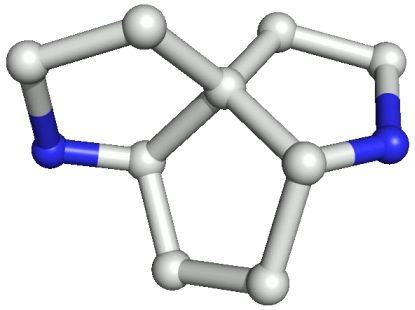
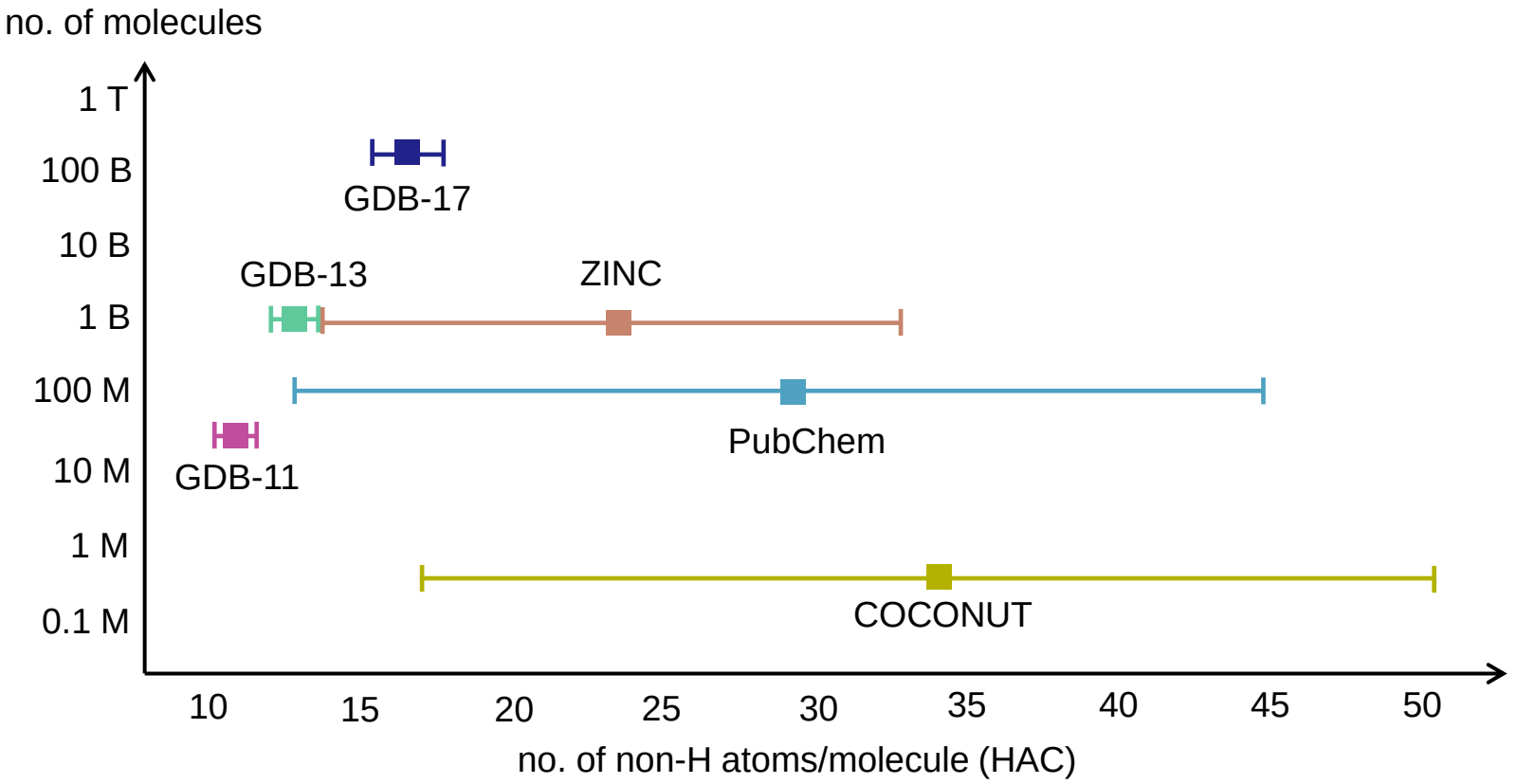
Artificial Intelligence and Chemical Space

Jean-Louis Reymond https://gdb.unibe.ch

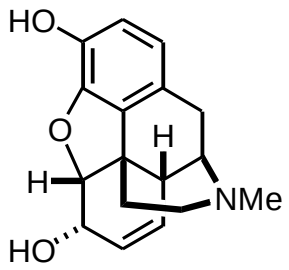
12 May 2022, second School of the AIDD, IDISA, Lugano



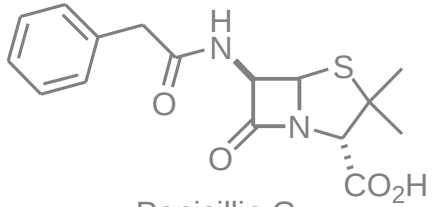




Molecular quantum numbers (MQN, 42D)



Morphine



Penicillin G

City-Block **Distance** (CBD) **= 64**

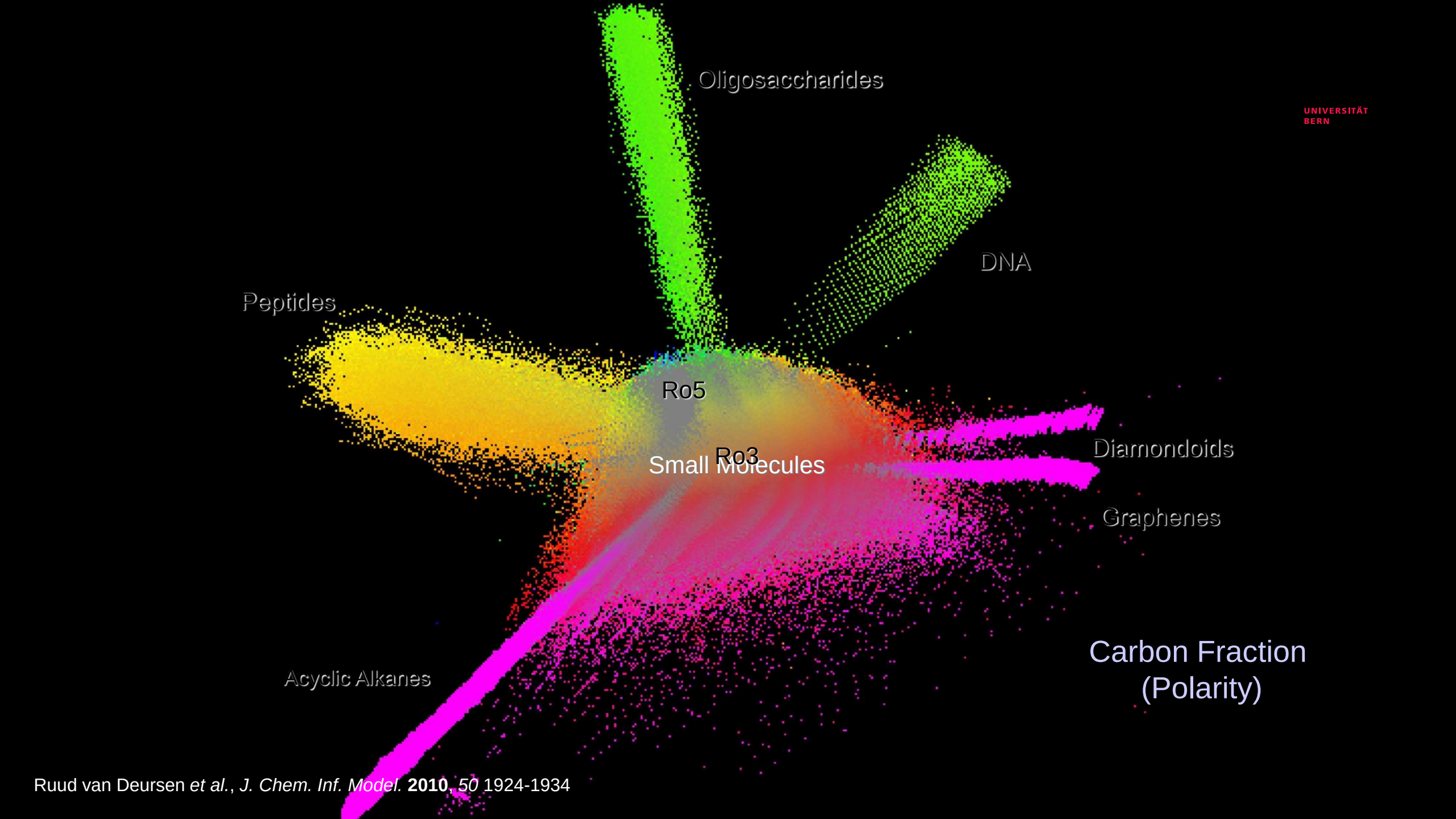
Atoms			
Carbon	**17**	16	**1**
Fluorine	**0**	0	
Chlorine	**0**	0	
Bromine	**0**	0	
Iodine	**0**	0	
Sulphur	**0**	1	**1**
Phosphor	**0**	0	
Acyclic nitrogen	**0**	1	**1**
Cyclic nitrogen	**1**	1	
Acyclic oxygen	**2**	4	**2**
Cyclic oxygen	**1**	0	**1**
Heavy atom count	**21**	23	**2**

Bonds			
Acyclic single bonds	**3**	8	**5**
Acyclic double bonds	**0**	3	**3**
Acyclic triple bonds	**0**	0	
Cyclic single bonds	**18**	11	**7**
Cyclic double bonds	**4**	3	**1**
Cyclic triple bonds	**0**	0	
Rotatable bonds	**0**	4	**4**

Polar groups			
H-Bond donor atoms	**3**	1	**2**
H-Bond donor sites	**3**	1	**2**
H-Bond acceptor atoms	**3**	4	**1**
H-Bond acceptor sites	**3**	7	**4**
Positive charges	**1**	0	**1**
Negative charges	**0**	1	**1**

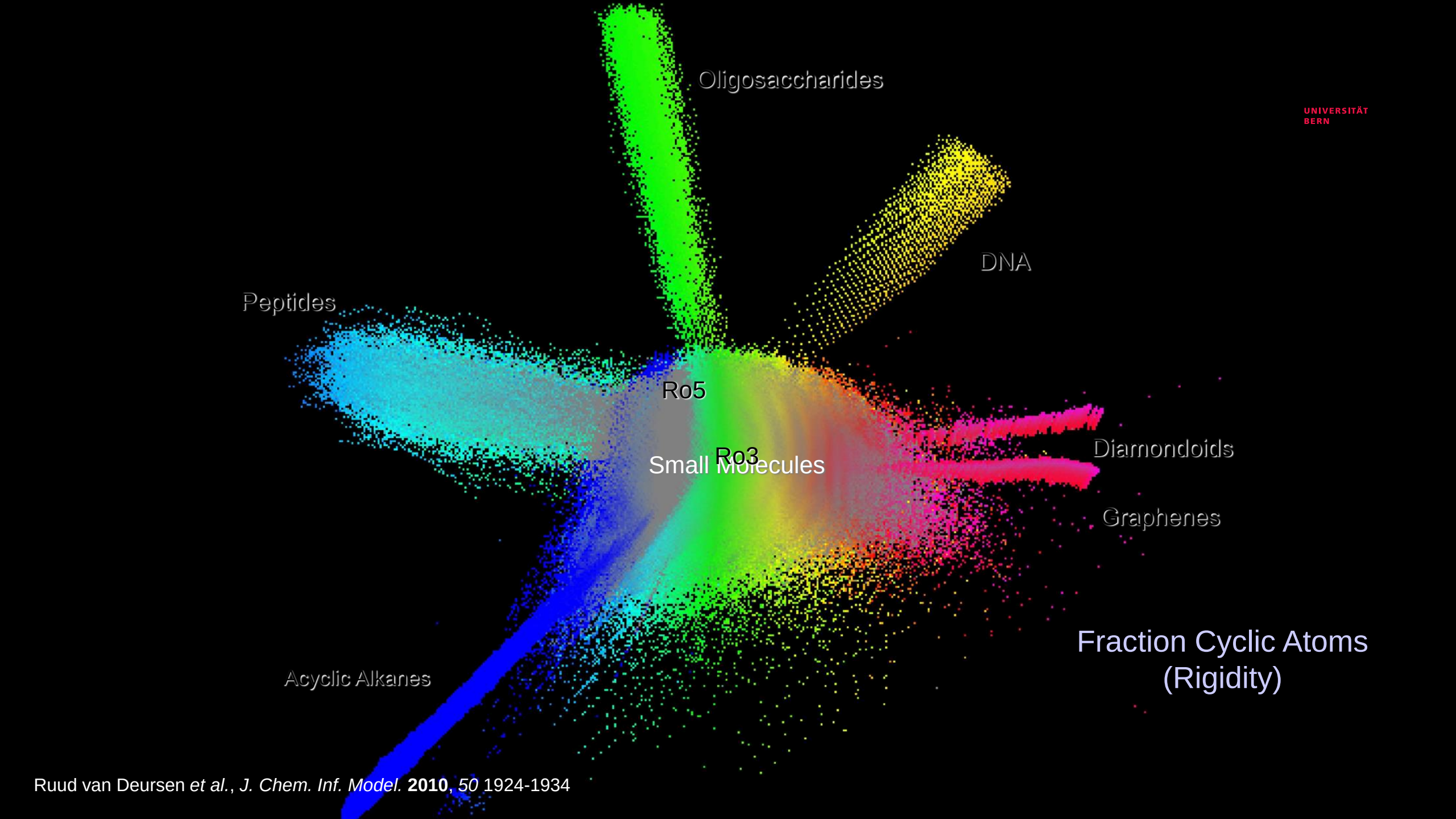
Topology			
Acyclic monovalent nodes	**3**	6	**3**
Acyclic divalent nodes	**0**	2	**2**
Acyclic trivalent nodes	**0**	2	**2**
Acyclic tetravalent nodes	**0**	0	
Cyclic divalent nodes	**8**	6	**2**
Cyclic trivalent nodes	**9**	6	**3**
Cyclic tetravalent nodes	**1**	1	
3-Membered rings	**0**	0	
4-Membered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

K. T. Nguyen et al., *ChemMedChem* **2009**, *4*, 1803-1805



UNIVERSITÄT
BERN

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



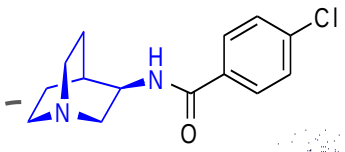
UNIVERSITÄT
BERN

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

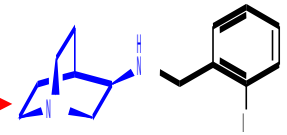
HiQSCREEN
High Quality Screening for Drug Discovery

UNIVERSITÄT BERN

City-Block **Distance** (CBD) **= 11**



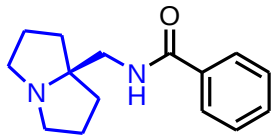
2 (PNU-282,987)



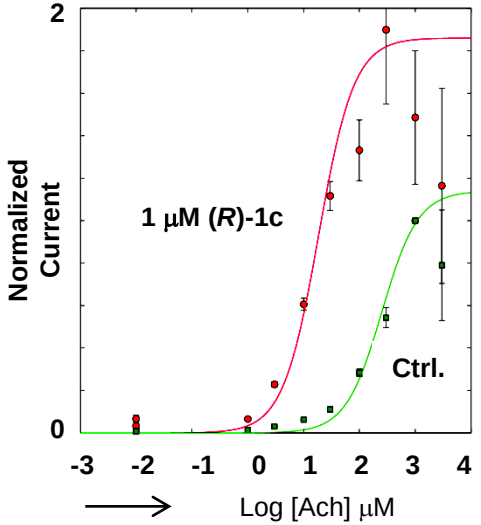
(*R*)-**1c**

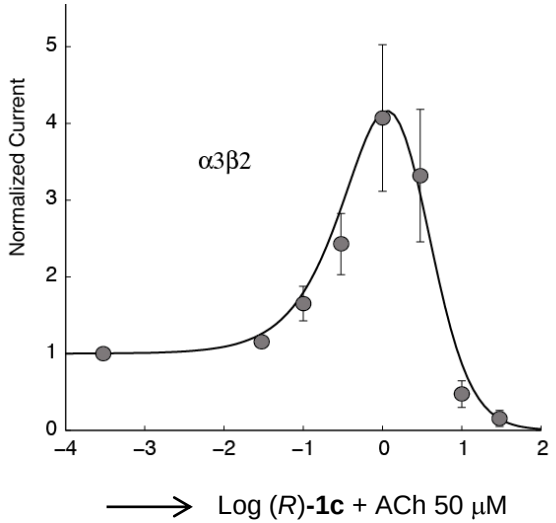
Positive allosteric modulator of α3β2 nAChR

GDB-11

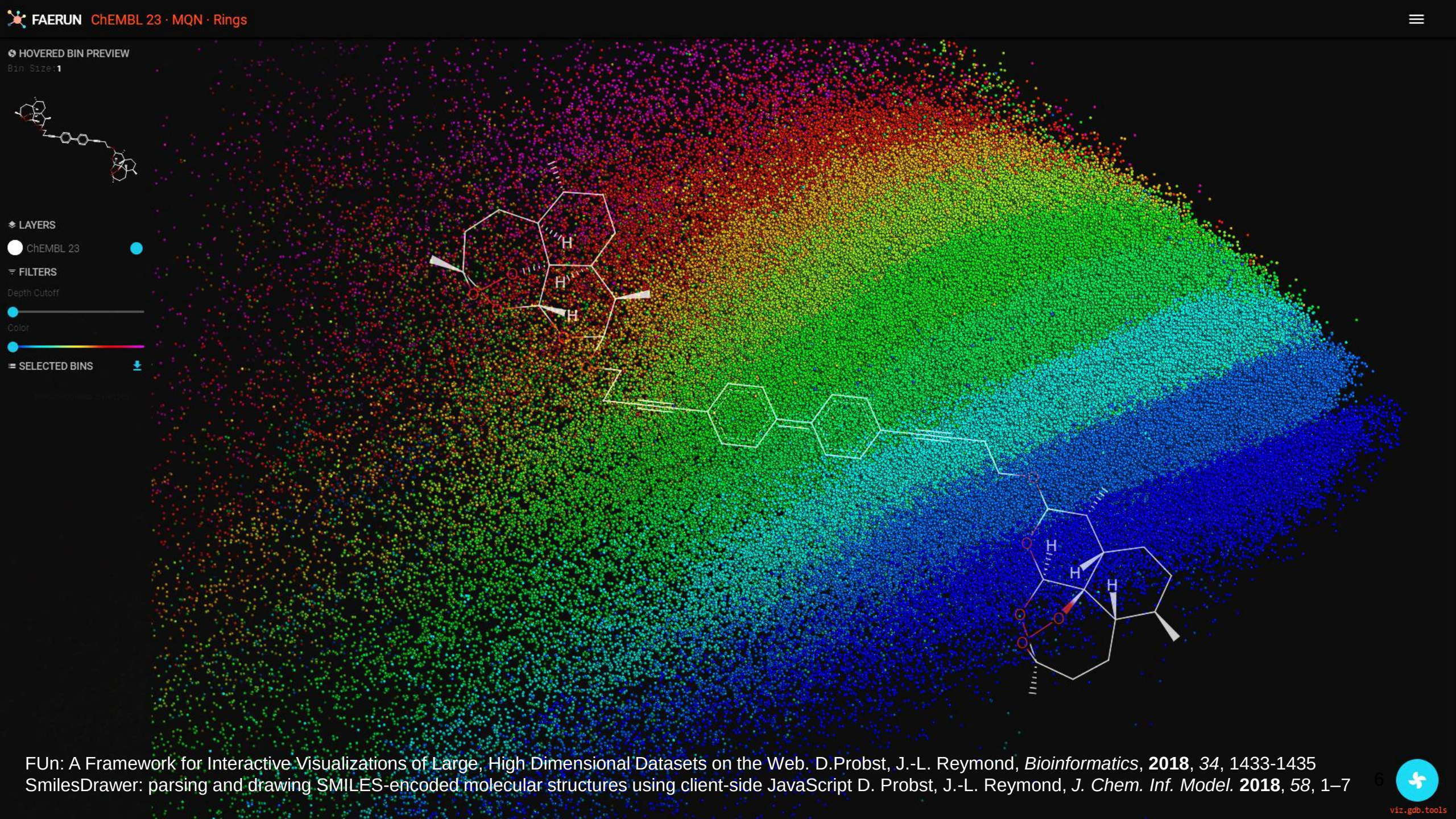


ChEMBL
MQN-Space



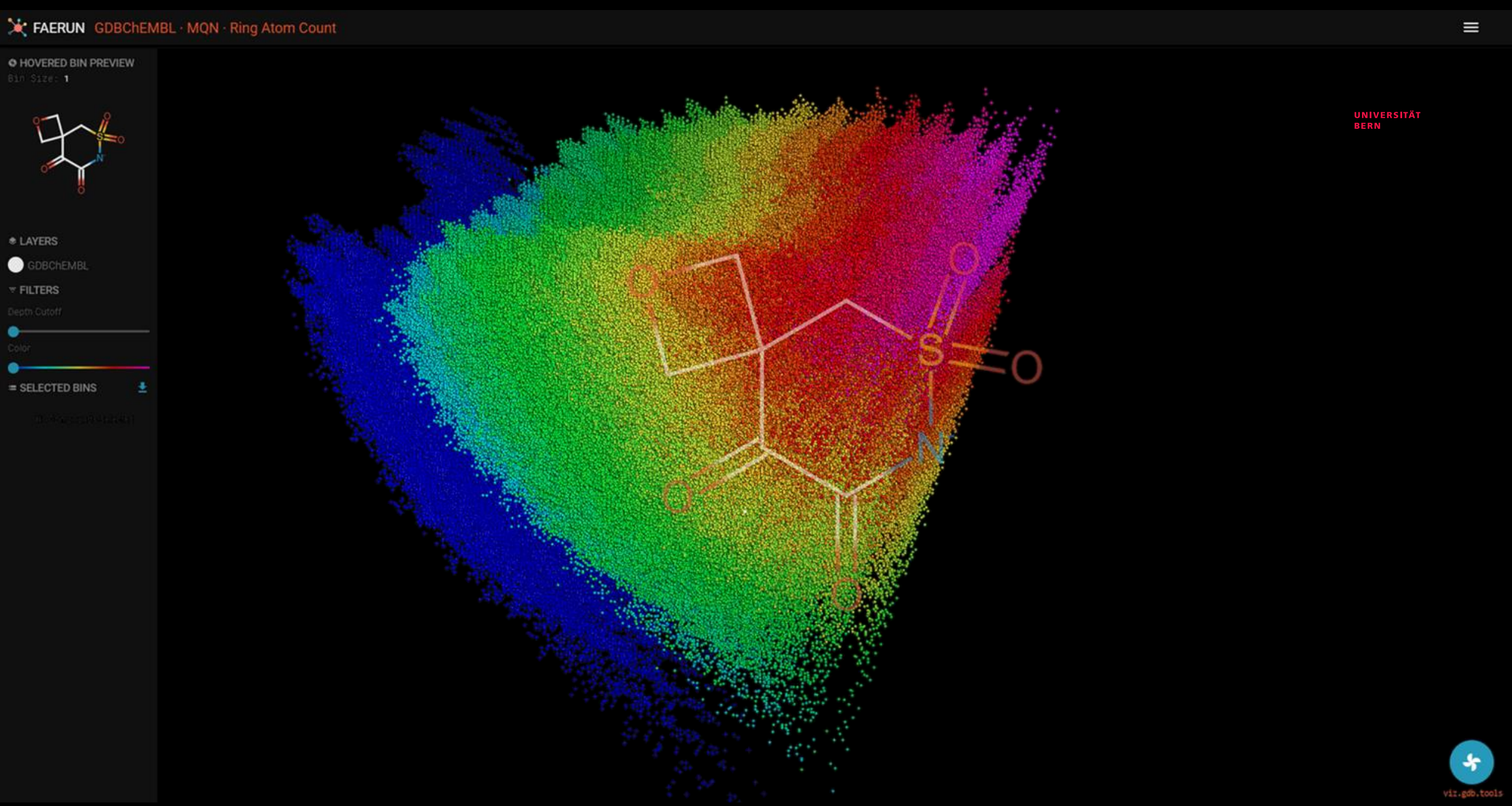


Garcia Delgado et al. *ACS Med. Chem. Lett.* **2010**, *1*, 422-426
Bréthous L. *et al*. *J. Med. Chem.* **2012,** *55,* 4605-4518
J. Bürgi *et al. ACS Chem. Neurosci.* **2014**, *5*, 346-359.

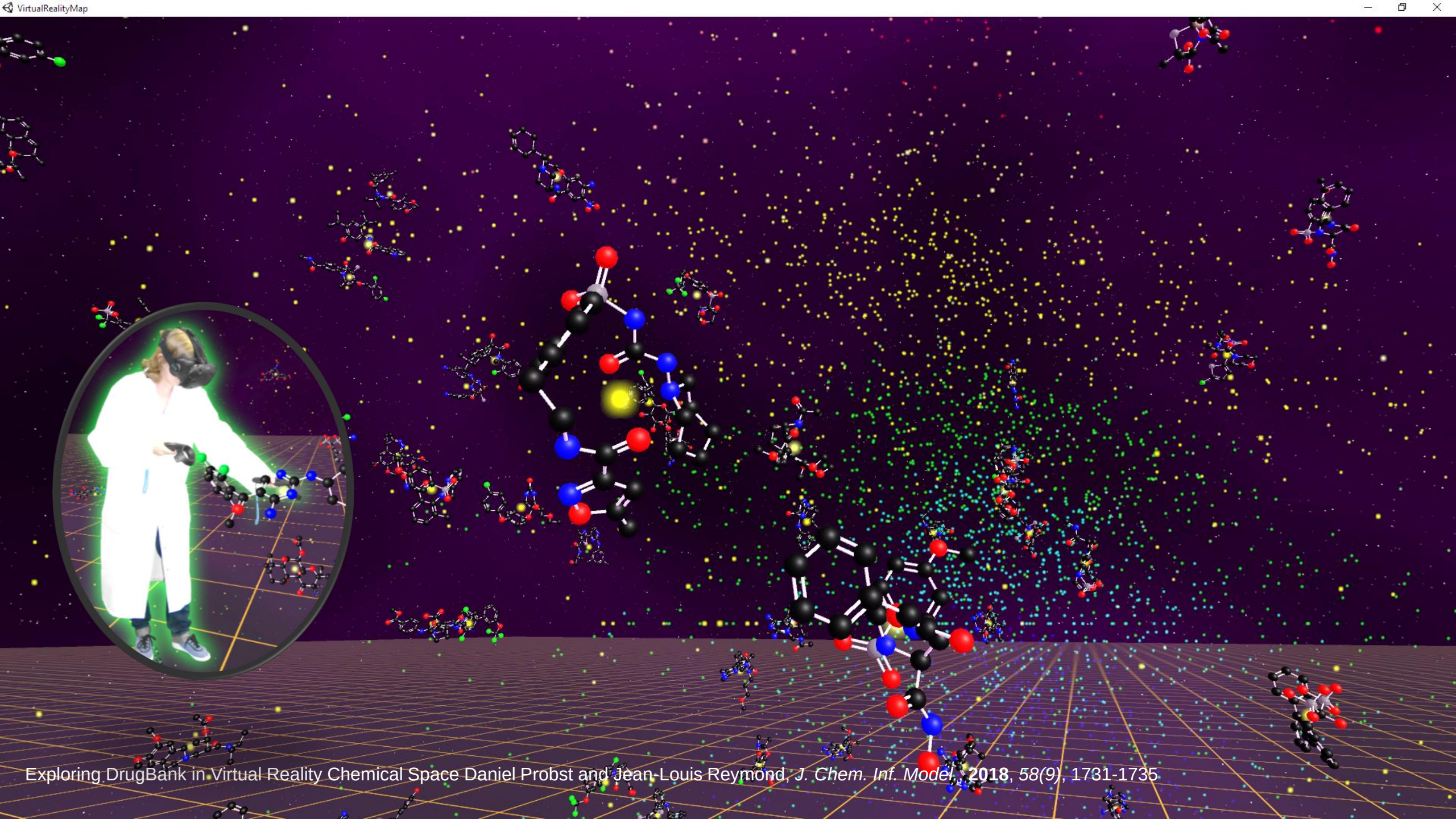


FUn: A Framework for Interactive Visualizations of Large, High Dimensional Datasets on the Web. D.Probst, J.-L. Reymond, *Bioinformatics*, **2018**, *34*, 1433-1435
SmilesDrawer: parsing and drawing SMILES-encoded molecular structures using client-side JavaScript D. Probst, J.-L. Reymond, *J. Chem. Inf. Model.* **2018**, *58*, 1–7

viz.gdb.tools

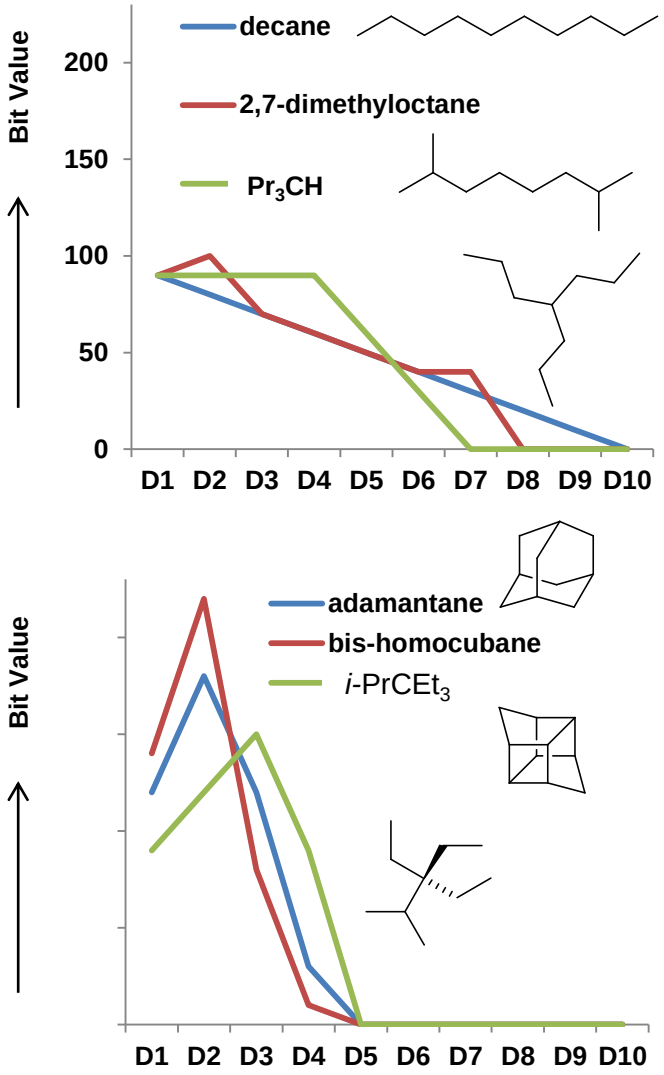


Sven Bühlmann *et al.*, *Front. Chem,* **2020**, *8, 46,* doi:10.3389/fchem.2020.00046



Exploring DrugBank in Virtual Reality Chemical Space Daniel Probst and Jean-Louis Reymond, *J. Chem. Inf. Model.*, **2018**, *58(9)*, 1731-1735

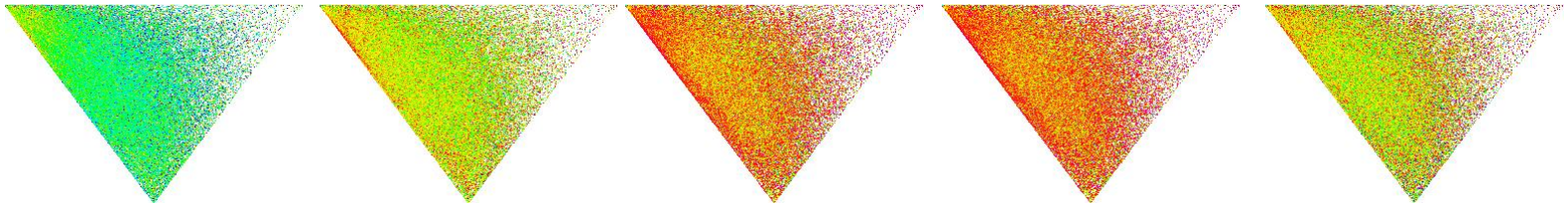
Atom-pairs with topological distances in bonds

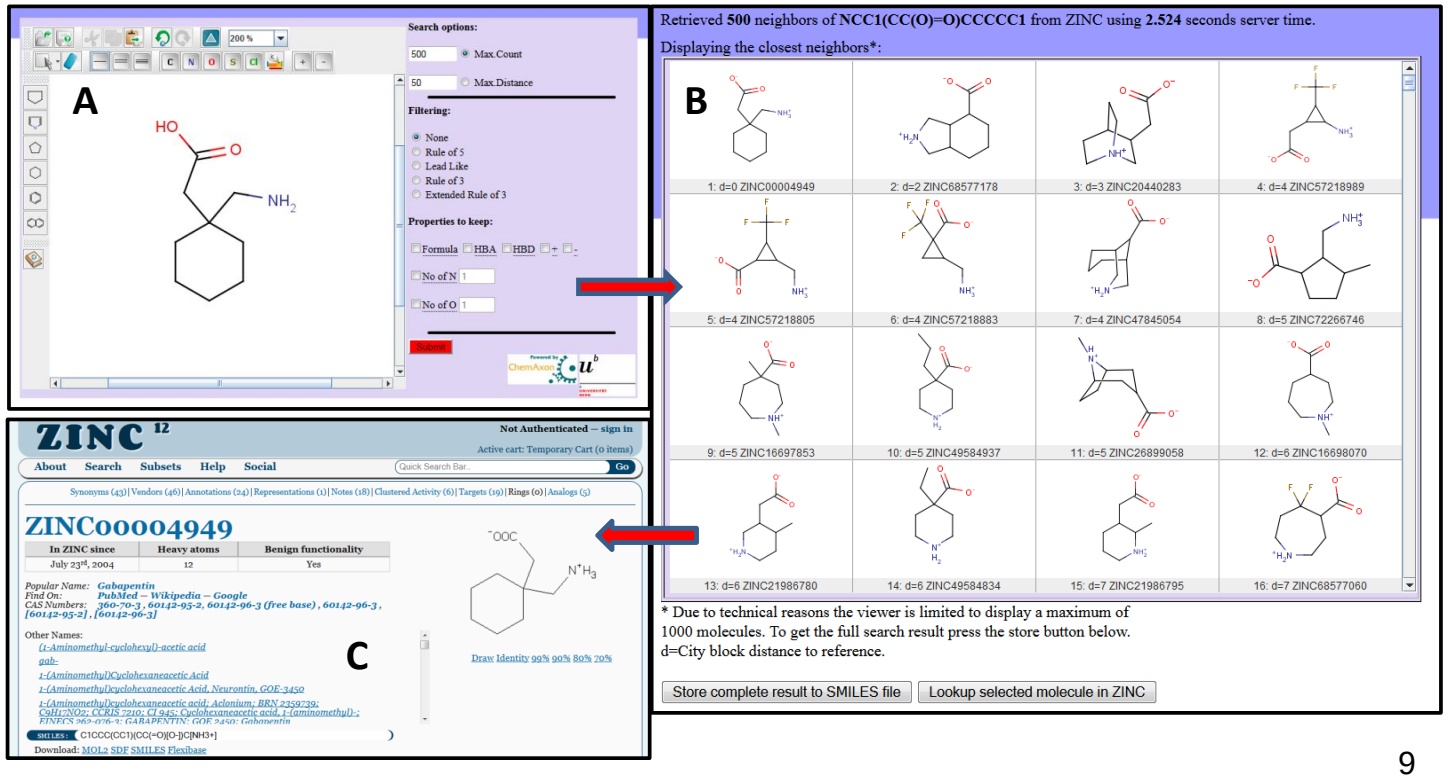


AUC value

≤ 50 % 60 % 75 % ≥ 90 %

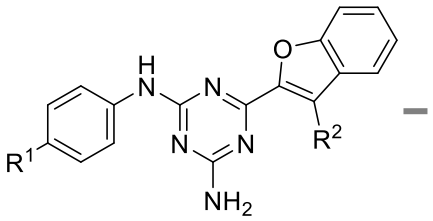
AUC for recovery of ROCS Color Tanimoto analogues





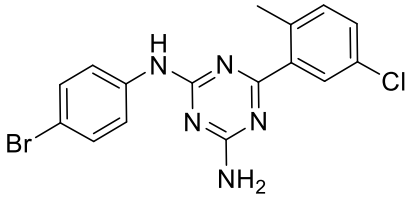
Mahendra Awale et al., *J. Chem. Inf. Model.* **2014**, *54*, 1892-1907

Target prediction



angiogenesis inhibitor from phenotypic screen

PPB2 with Xfp similiarity



43

known LPAAT-β inhibitor

Table 1. Cytotoxicity of triazines on HeLa cells.

Compd[a]	R^1	R^2	IC_{50} [nM][b]
1	Et	Me	111 ± 12
2	Cl	Me	252 ± 7
3	OMe	Me	190 ± 3
4	*i*Pr	Me	814 ± 20
5	*n*Pr	Me	> 10 000
6	*n*Hex	Me	> 10 000
7	Ph	Me	> 10 000
8	F	Me	3060 ± 110
9	Br	Me	264 ± 5
10	CF_3	Me	263 ± 6
11	NHAc	Me	2920 ± 60
12	Et	Et	51 ± 1
13	Et	*n*Bu	780 ± 30
14	OMe	Et	12.4 ± 0.2
15	F	Et	4550 ± 940
16	Cl	Et	140 ± 8
17	Br	Et	69 ± 13
18	CF_3	Et	110 ± 56

[a] See Scheme 1. [b] HeLa cells (10 000 cells per well) were incubated with various concentrations of compounds for 96 h at 37 °C and 5% CO_2. Cell viability was determined after addition of WST-8 working solution and phenazine ethosulfate. Final absorbance was measured at 450 nm. Results were normalized to control values; data are the mean ± SD (standard deviation) from triplicate measurements.

Table 2. Kinome scan with triazines **1** and **14**.[a]

Compd	Kinase	Percent of control at 10 μM	K_D [nM]
1	JAK1 (JH2 domain pseudopocket)	4.5	n.d.
	PIP5K2C	16	n.d.
	TYK2 (JH2 domain pseudopocket)	31	n.d.
14	JAK1 (pseudokinase)	4.2	380
	EPHB6	5.3	860
	PIKFYVE	6	3100
	GRK7	6	> 60 000
	TAK1	7.2	> 60 000
	IRAK3	12	3500
	GAK	15	490
	TYK2 (JH2 domain pseudokinase)	15	1800

[a] The kinome scan was performed for binding of the ATP binding pockets on 442 kinases by DiscoveRx (San Diego, CA, USA); the test compound concentration was 10 μM.

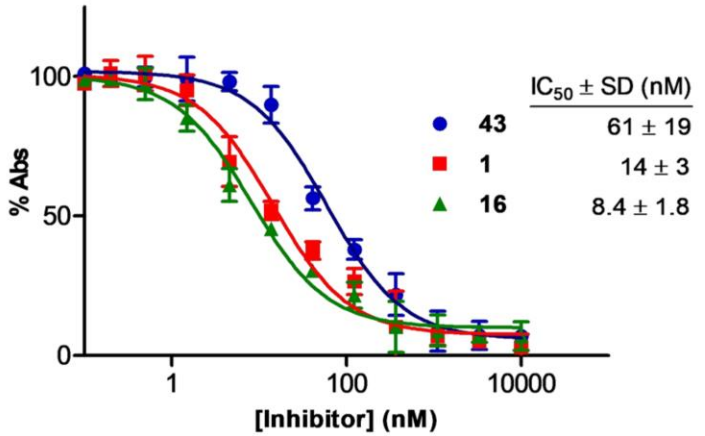
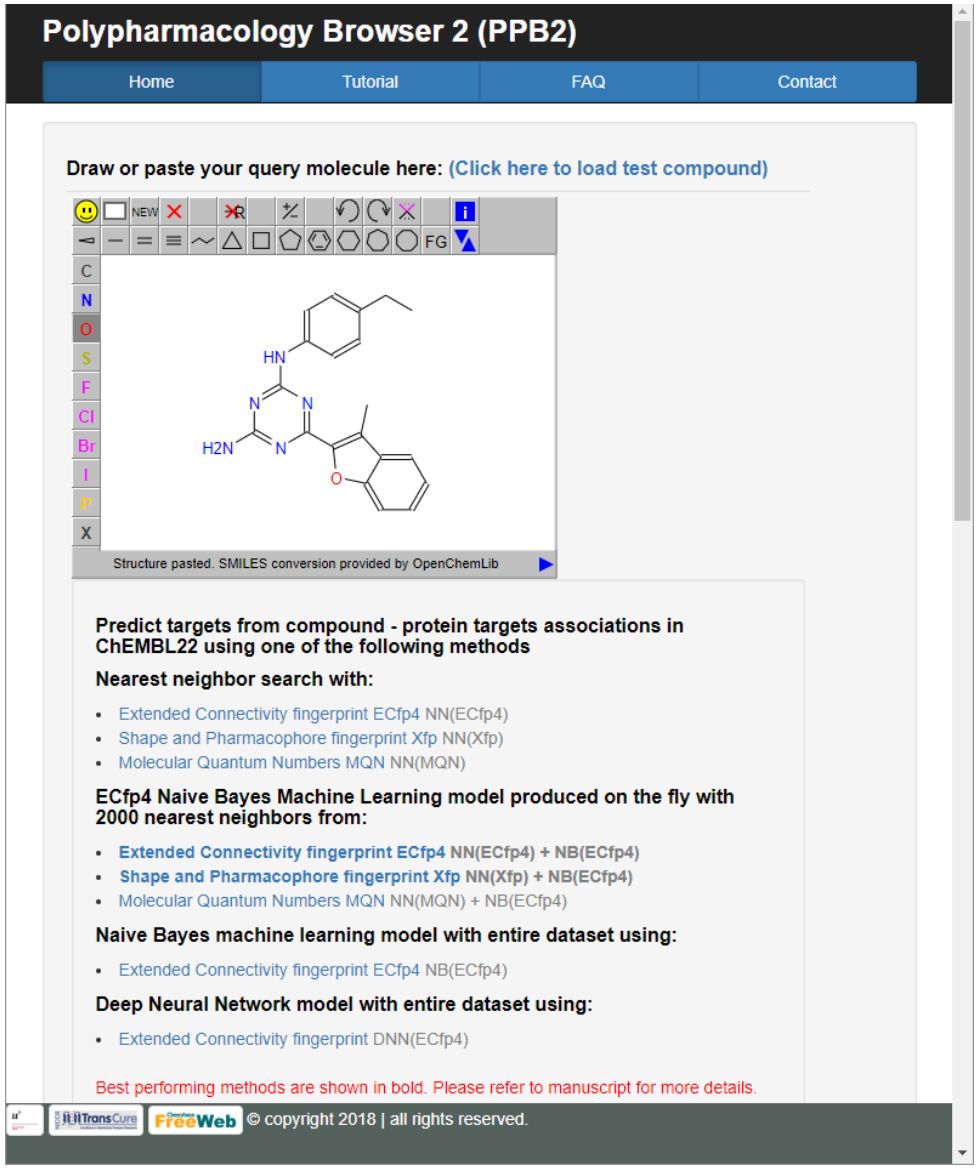
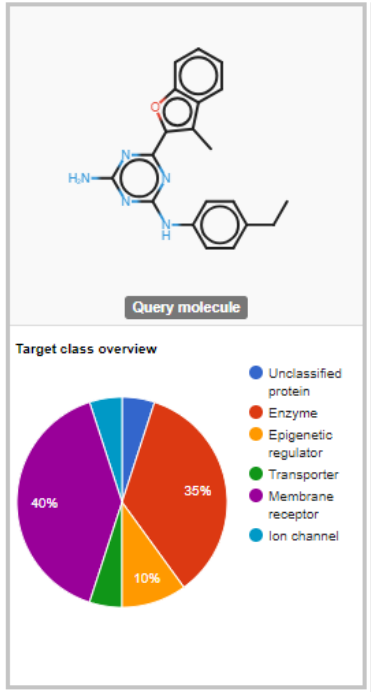


Figure 3. Activity of triazines **1**, **16** and **43** on LPAAT-β by quantifying the release of free CoA-SH in the presence of DTNB monitored by the change in absorbance at 405 nm over 3 min at room temperature (200 μM of 18:1-CoA, 200 μM sn-1-18:1 lysoPA, 500 μM DTNB, 25 mM HEPES·HCl, pH 7.5, 1 mM EDTA, 100 mM NaCl and various concentrations of inhibitors). SD = standard deviation from triplicate measurements.

Marion Poirier et al. *ChemMedChem* **2019**, *14*, 224-236

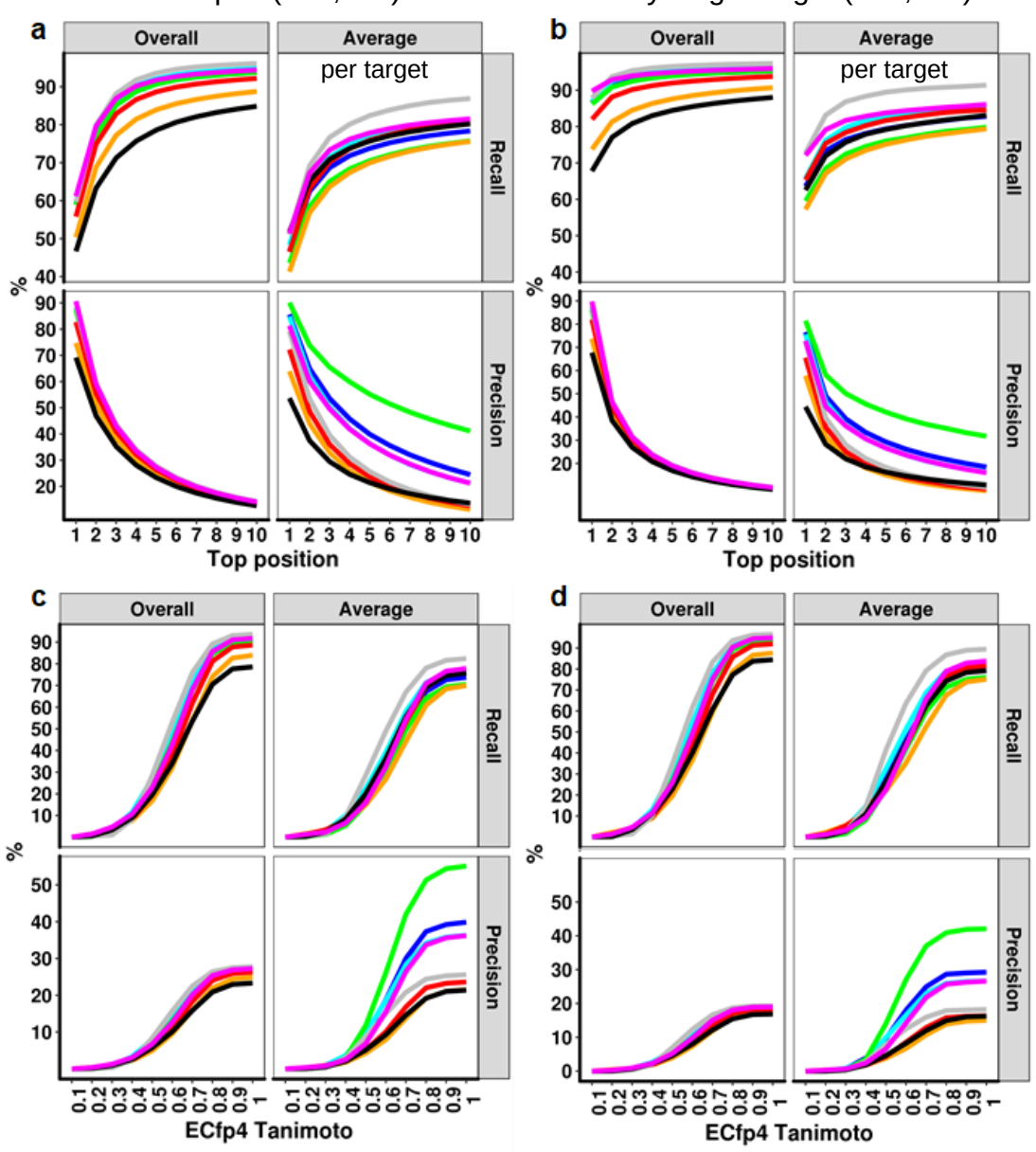


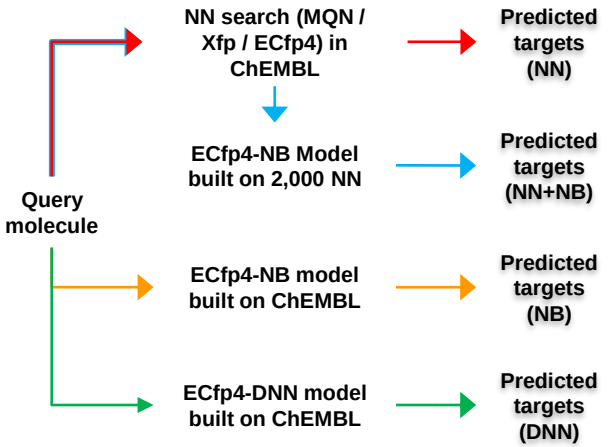


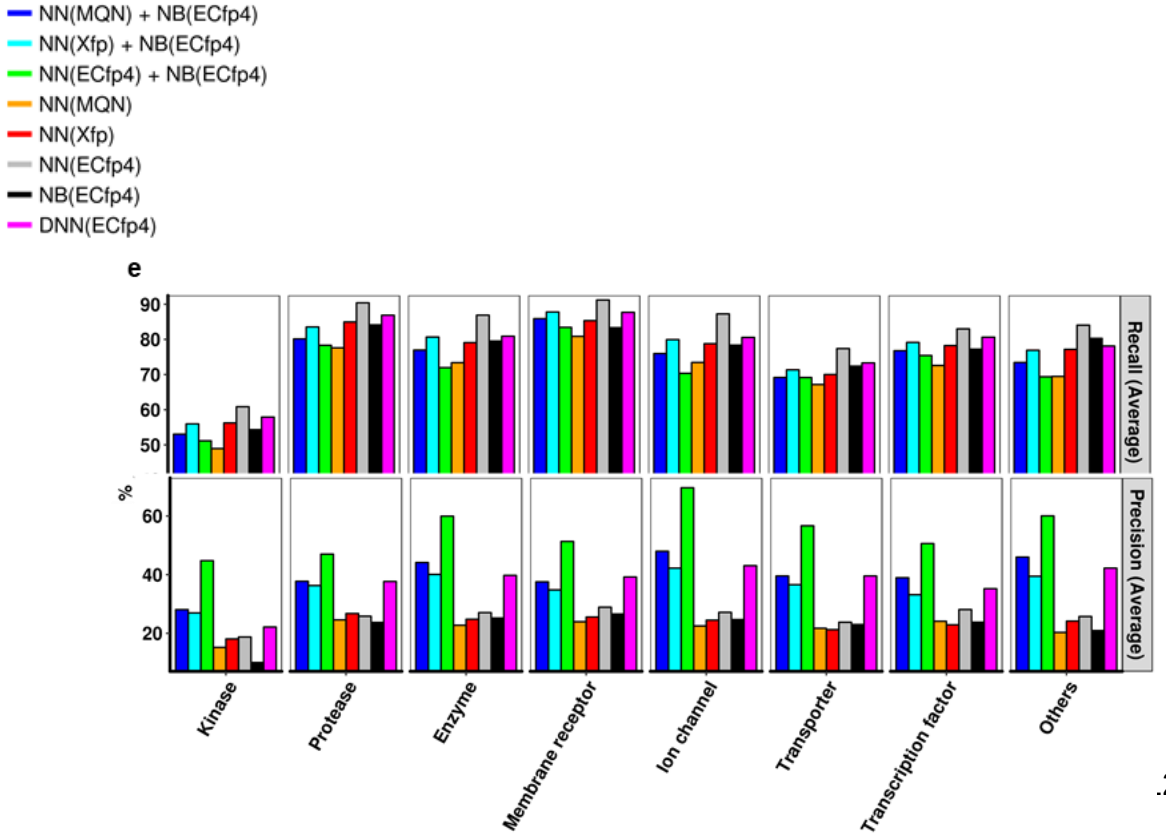
Targets predicted using NN(Xfp). Save Table

Rank	ChEMBL ID	Common name	Nearest neighbours
1	CHEMBL4772	1-acylglycerol-3-phosphate O-acyltransferase beta	Show NN
2	CHEMBL253	Cannabinoid CB2 receptor	Show NN
3	CHEMBL1904	Glutamate [NMDA] receptor subunit epsilon 2	Show NN
4	CHEMBL1800	Corticotropin releasing factor receptor 1	Show NN
5	CHEMBL224	Serotonin 2a (5-HT2a) receptor	Show NN
6	CHEMBL228	Serotonin transporter	Show NN
7	CHEMBL2916	Telomerase reverse transcriptase	Show NN
8	CHEMBL1293222	Nucleotide-binding oligomerization domain-containing protein 1	Show NN
9	CHEMBL2971	Tyrosine-protein kinase JAK2	Show NN
10	CHEMBL3764	Urotensin II receptor	Show NN
11	CHEMBL3833	Trace amine-associated receptor 1	Show NN
12	CHEMBL204	Thrombin	Show NN
13	CHEMBL3018	Matriptase	Show NN
14	CHEMBL3286	Urokinase-type plasminogen activator	Show NN
15	CHEMBL244	Coagulation factor X	Show NN
16	CHEMBL287	Sigma opioid receptor	Show NN
17	CHEMBL1867	Alpha-2a adrenergic receptor	Show NN
18	CHEMBL1615382	Nuclear receptor coactivator 3	Show NN
19	CHEMBL5524	Protein-arginine N-methyltransferase 1	Show NN
20	CHEMBL251	Adenosine A2a receptor	Show NN

M. Awale, J.-L. Reymond, *J. Chem. Inf. Model.* **2019,** *59*, 10-17

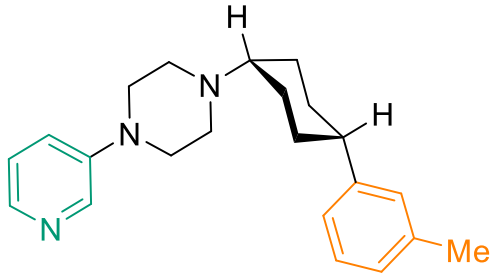






M. Awale, J.-L. Reymond, *J. Chem. Inf. Model.* **2019,** *59*, 10-17

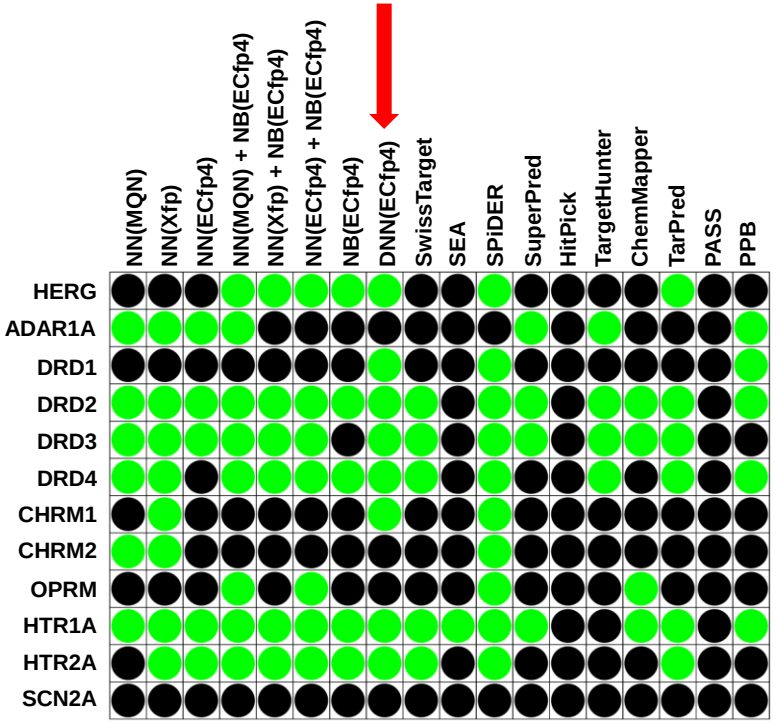
TRPV6 inhibitor:



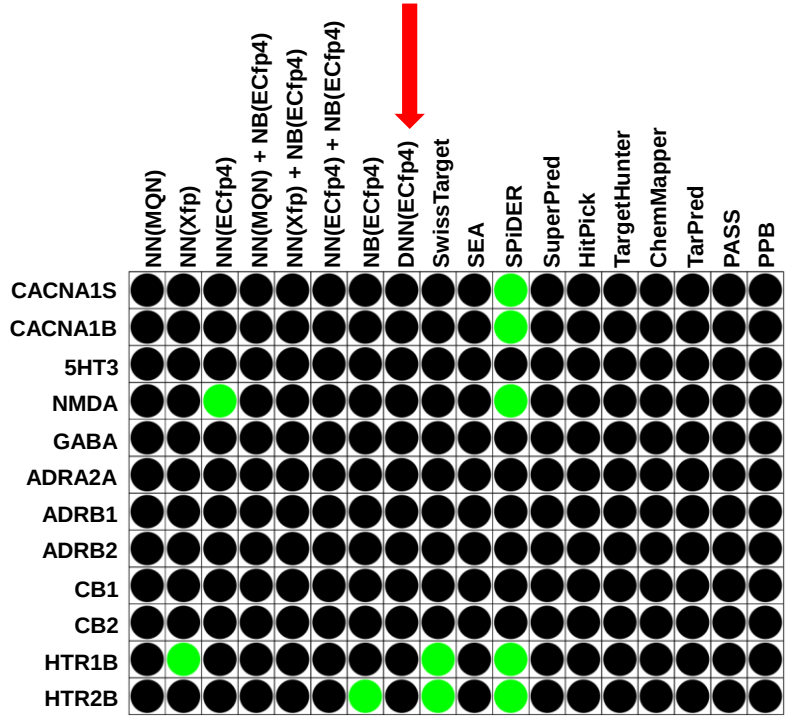
cis-**22a**

IC_{50} = 0.32 ± 0.12 µM

active off-targets (≥ 50 % @ 10 µM)



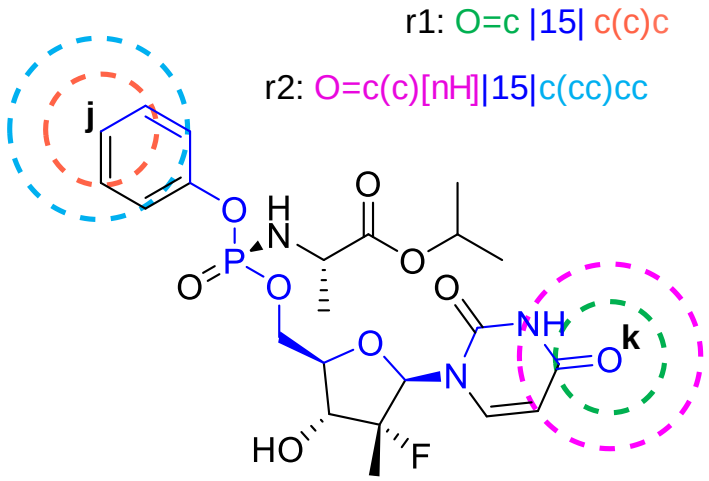
inactive off-targets

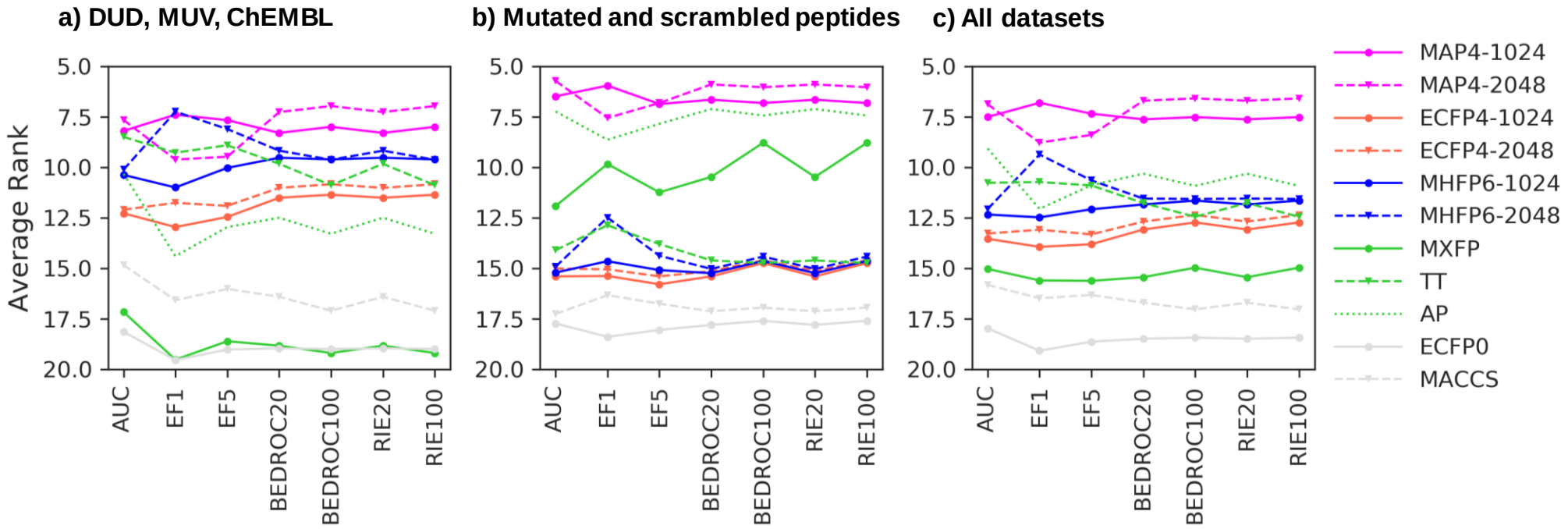


M. Awale, J.-L. Reymond, *J. Chem. Inf. Model.* **2019,** *59*, 10-17

MAP4

MAP4 encoding of jk

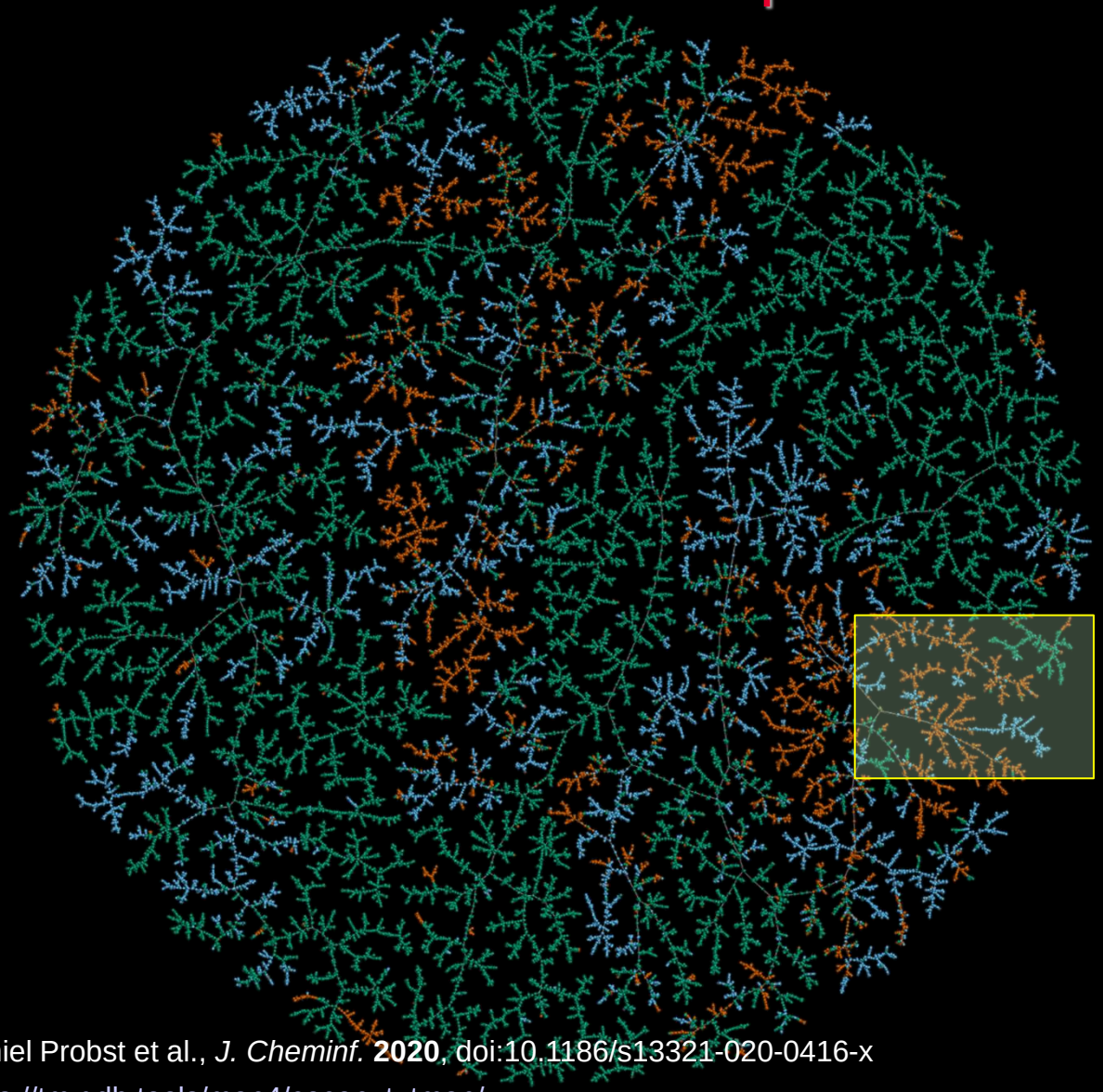


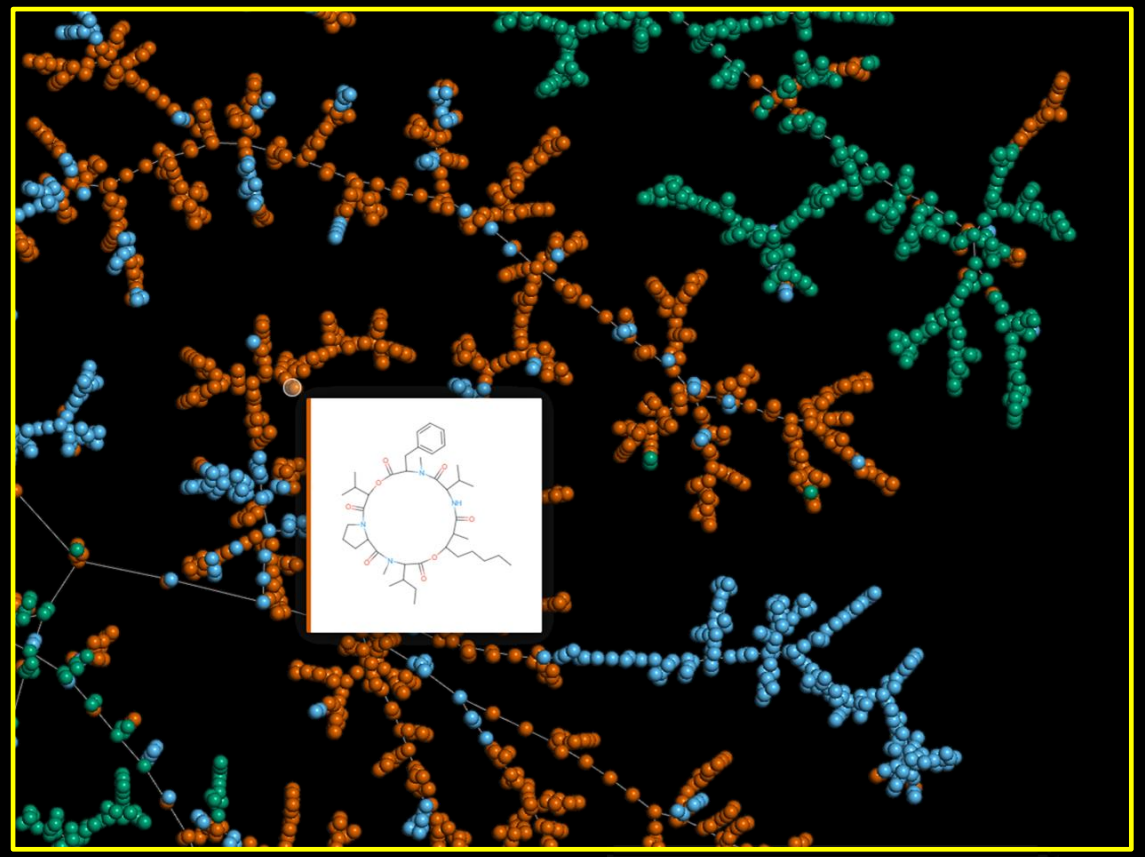


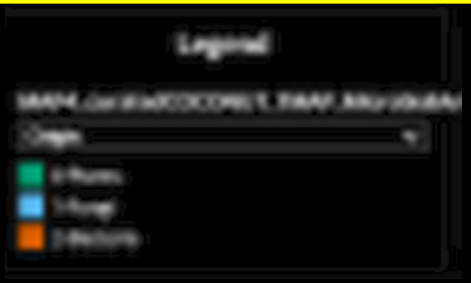
Alice Capecchi *et al*, *J. Cheminf.* **2020**, *12*, 43

MAP4 TMAP of natural products

UNIVERSITÄT BERN







Daniel Probst et al., *J. Cheminf.* **2020**, doi:10.1186/s13321-020-0416-x

https://tm.gdb.tools/map4/coconut_tmap/

Alice Capecchi et al., *J. Cheminf.* **2021**, *13*, 82, doi.org/10.1186/s13321-021-00559-3

Table 2 SVM evaluation with balanced accuracy, MCC, and F1 score

	Balanced acc.	MCC	F1
MAP4 SVM [a,b]	0.919 ± 0.005	0.879 ± 0.005	0.929 ± 0.003
ECFP4 SVM [a,b]	0.890 ± 0.005	0.827 ± 0.006	0.897 ± 0.003
RDKit AP SVM [a,b]	0.735 ± 0.005	0.592 ± 0.006	0.752 ± 0.004
Properties SVM [a,c]	0.758 ± 0.005	0.613 ± 0.007	0.761 ± 0.004

[a] Mean value and standard deviation (σ) of the five different test/training sets split of the fivefold cross-validation

[b] 1024 dimensions

[c] 11 properties: MW, Fsp3, HBD) and HBA, calculated logP with the Crippen method (AlogP), number of carbons, oxygen, and nitrogen, the total number of atoms, number of bonds, and topological polar surface area (TPSA)

Table 3 MAP4 SVM origin prediction for 20 recently published microbial and plants NPs that are not present in COCONUT

Natural product	Origin	MAP SVM prediction[a]
Epicospirocin 1	Fungal	Fungal (97%)
Penicimeroterpenoid A	Fungal	Fungal (82%)
Beetleane A	Fungal	Fungal (97%)
Funiculolide D	Fungal	Fungal (85%)
Rhizolutin	Fungal	Plant (55%, fungal: 29%)
Fusoxypenes A	Fungal	Fungal (69%)
Myxadazoles A	Fungal	Bacterial (74%, fungal: 16%)
Vertirhodin A	Bacterial	Bacterial (88%)
Marinoterpin A	Bacterial	Plant (44%, bacterial: 37%)
Bosamycin A	Bacterial	Bacterial (94%)
Dumulmycin	Bacterial	Bacterial (80%)
Fortuneicyclidin A	Plant	Plant (98%)
Meloyunnanine A	Plant	Plant (99%)
Hyperfol B	Plant	Plant (93%)
Pgaharmol A	Plant	Plant (77%)
Hunzeylanine A	Plant	Plant (95%)
Mucroniferal A	Plant	Plant (73%)
Perovsfolin A	Plant	Plant (92%)
Horienoid A	Plant	Plant (95%)
Erythrivarine J	Plant	Plant (91%)

[a] Predicted using the MAP4 SVM available online at https://np-svm-map4.gdb.tools/

epicospirocin 1: 1a (1*R*,8*R*), 1b (1*S*,8*S*)
(origin: fungal)

penicimeroterpenoid A
(origin: fungal)

beetleane A
(origin: fungal)

funiculolide D
(origin: fungal)

pegaharmol A
(origin: plant)

mucroniferal A
(origin: plant)

rhizolutin
(origin: fungal
prediction: plant)

fusoxypenes A
(origin: fungal)

myxadazoles A
(origin: fungal
prediction: bacterial)

vertirhodin A
(origin: bacterial)

hunzeylanine A
(origin: plant)

erythrivarine J
(origin: plant)

marinoterpin A
(origin: bacterial
prediction: plant)

bosamycin A
(origin: bacterial)

perovsfolin A
(origin: plant)

dumulmycin
(origin: bacterial)

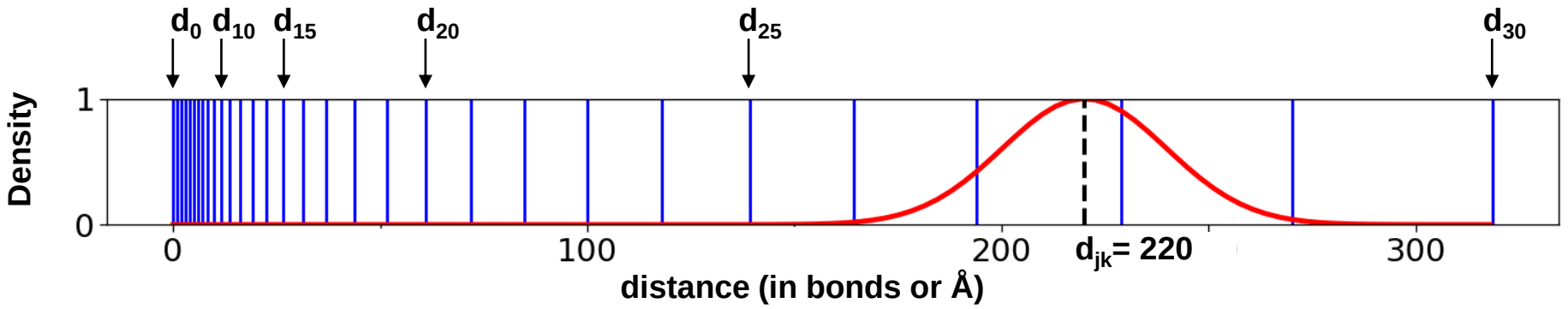
fortuneicyclidin A
(origin: plant)

meloyunnanine A
(origin: plant)

hyperfol B
(origin: plant)

horienoid A
(origin: plant)

Fuzzy atom-pairs



$$g_{jk}(d_i) = e^{-\frac{1}{2}\cdot\left(\frac{d_i - d_{jk}}{d_{jk}\cdot 0.09}\right)^2}$$

$$s_{jk} = \sum_{i=0}^{30} g_{jk}(d_i) \quad ,$$

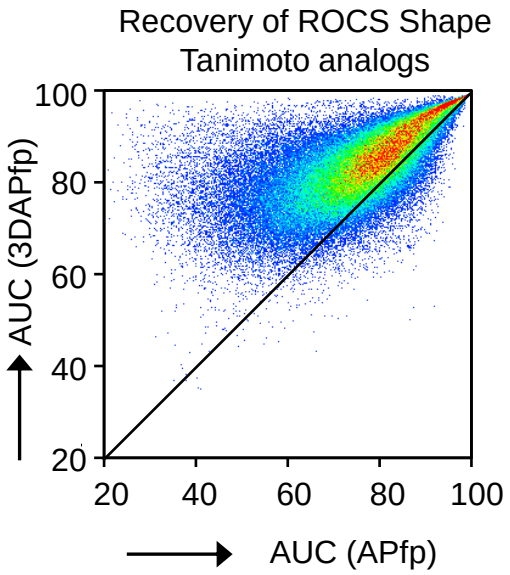
$$v_{Ci} = \frac{100}{{N_C}^{1.5}} \sum_{j=1} \sum_{k=1} \frac{g_{jk}(d_i)}{s_{jk}}$$

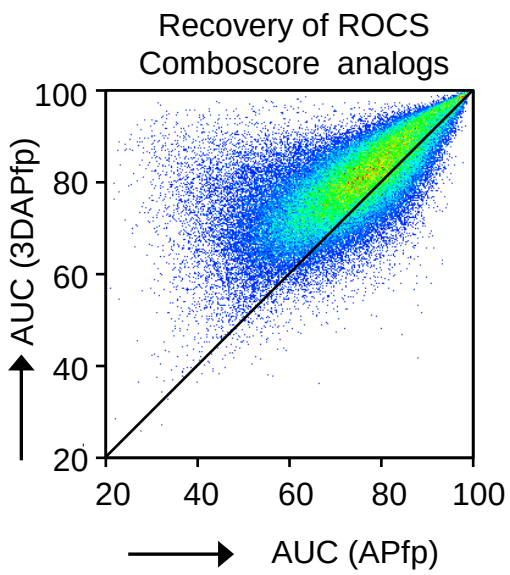
v = MXFP bin value
C = category
$C \epsilon \begin{Bmatrix} HA, HY, AR, HBA, HBD, \\ POS, NEG \end{Bmatrix}$
$i = \{i | i \epsilon \mathbb{N} \wedge 0 \leq i \leq 30\}$
N_C = total number of atoms in category C
$d_i = \{i | 0 \leq i \leq 6\}$
$d_i = \{d_{i-1} \cdot 1.18 | 7 \leq i \leq 30\}$
d_{jk}: topological distance between atoms j and k

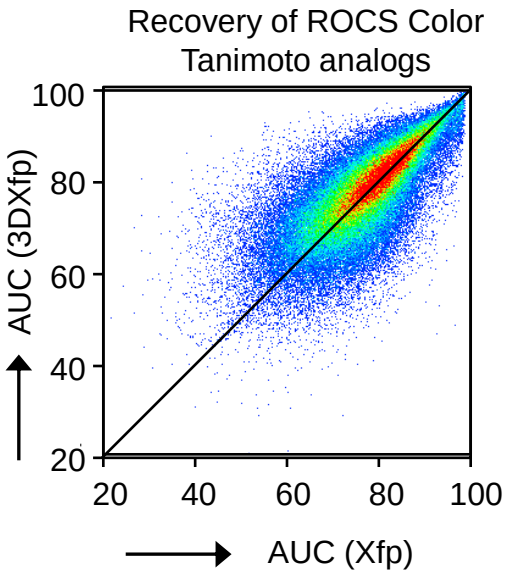
(1)

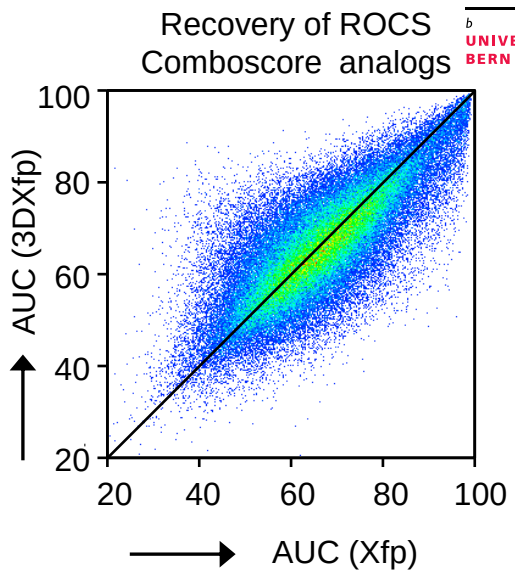
A. Capecchi et al., *Mol. Inform.*, **2019**, *38*, 1900016

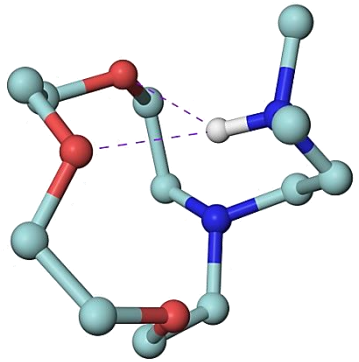
Atom-pair with 3D-distance in Å



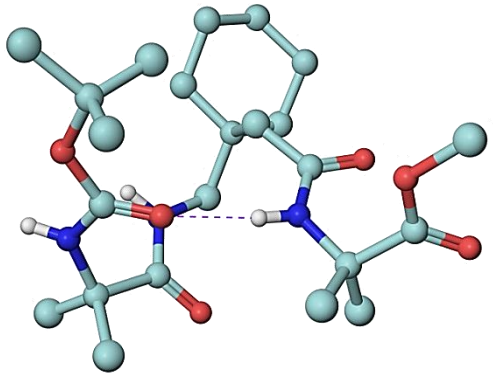




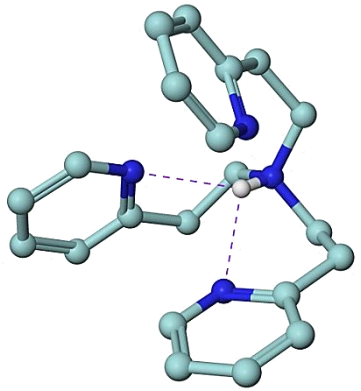




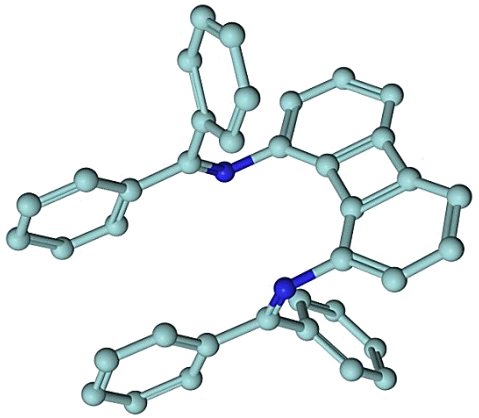
1 (VILGAY)
AUC_{3DAPfp} = 93%
AUC_{APfp} = 23%



2 (DOXOO)
AUC_{3DAPfp} = 89%
AUC_{APfp} = 27%



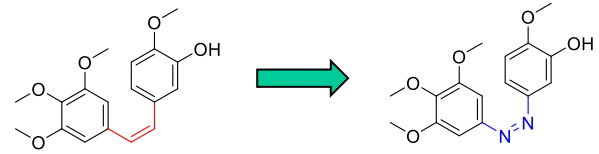
3 (QULPAO)
AUC_{3DAPfp} = 96%
AUC_{APfp} = 22%



4 (GUYVOM)
AUC_{3DAPfp} = 89%
AUC_{APfp} = 24%

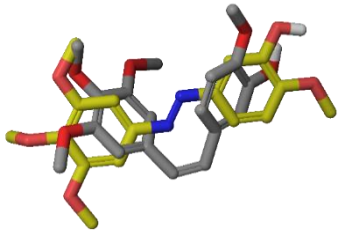
Mahendra Awale et al., *J. Cheminform.* **2015**, *7*, 3

Photopharmacology

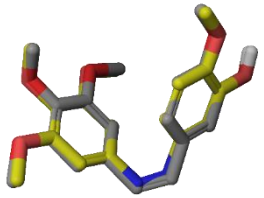


Combretastatin A-4

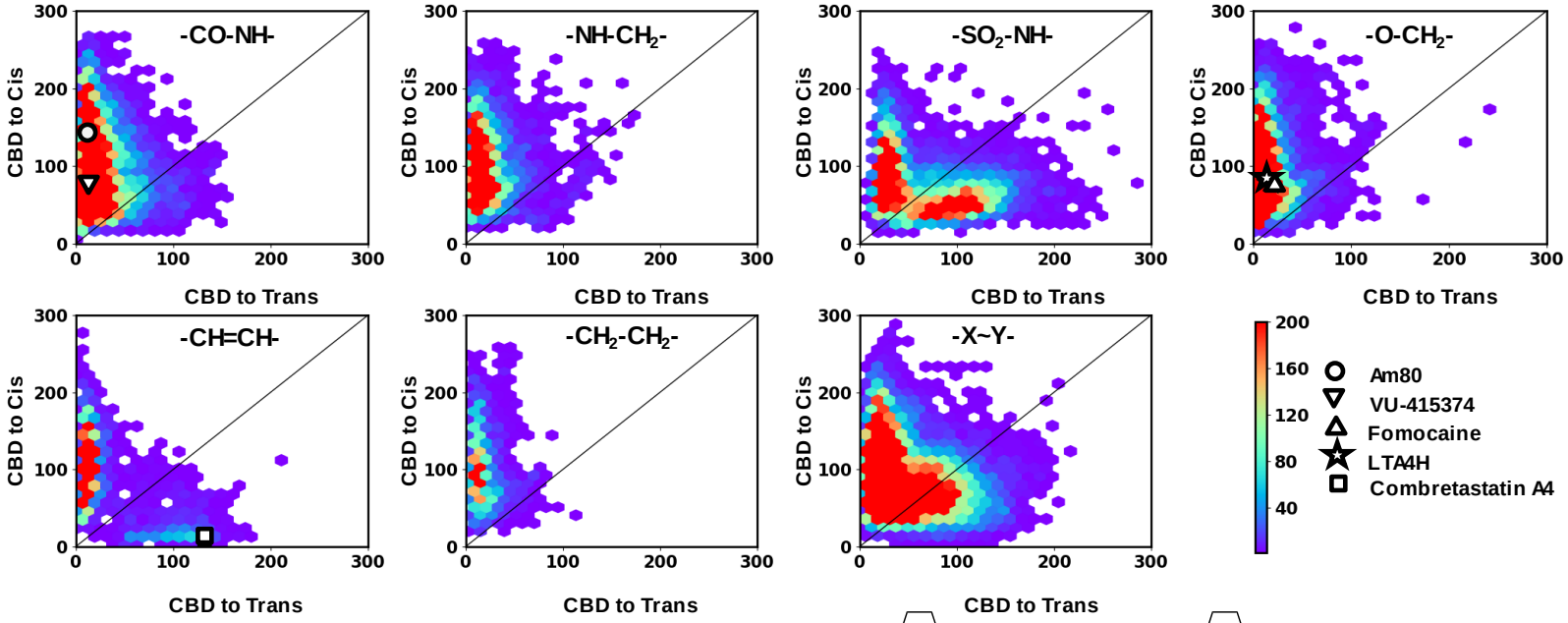
Photostatin



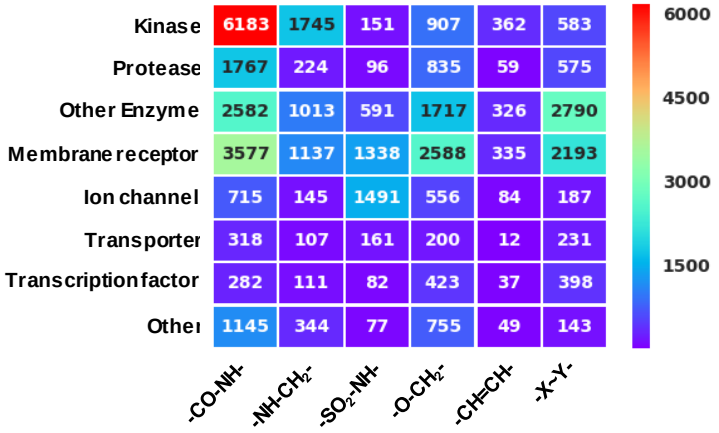
3DAPfp CBD: 132



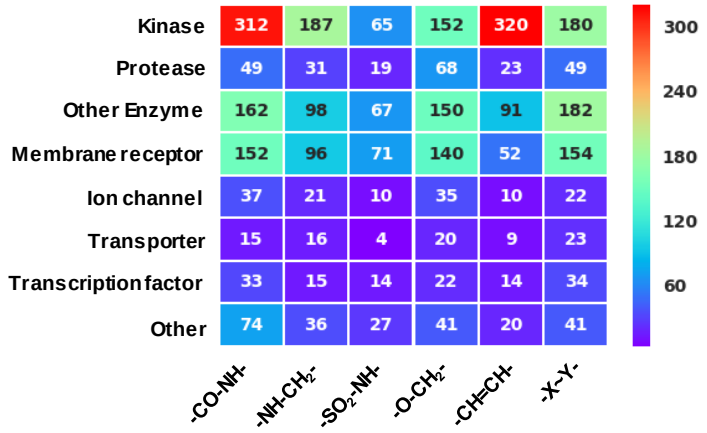
3DAPfp CBD: 9

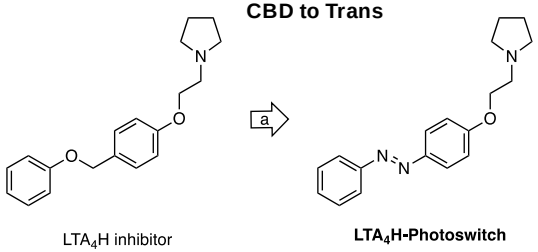


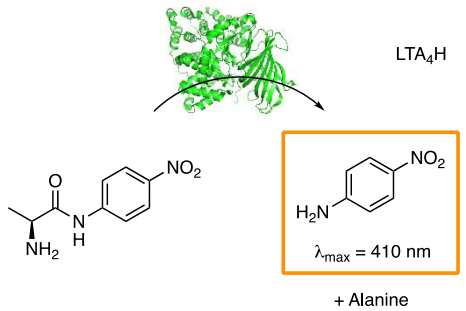
azologable ChEMBL cpds

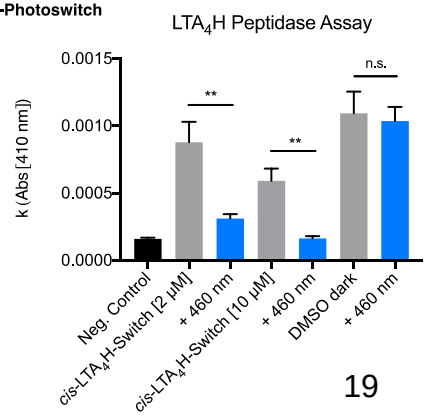


azologable ChEMBL targets



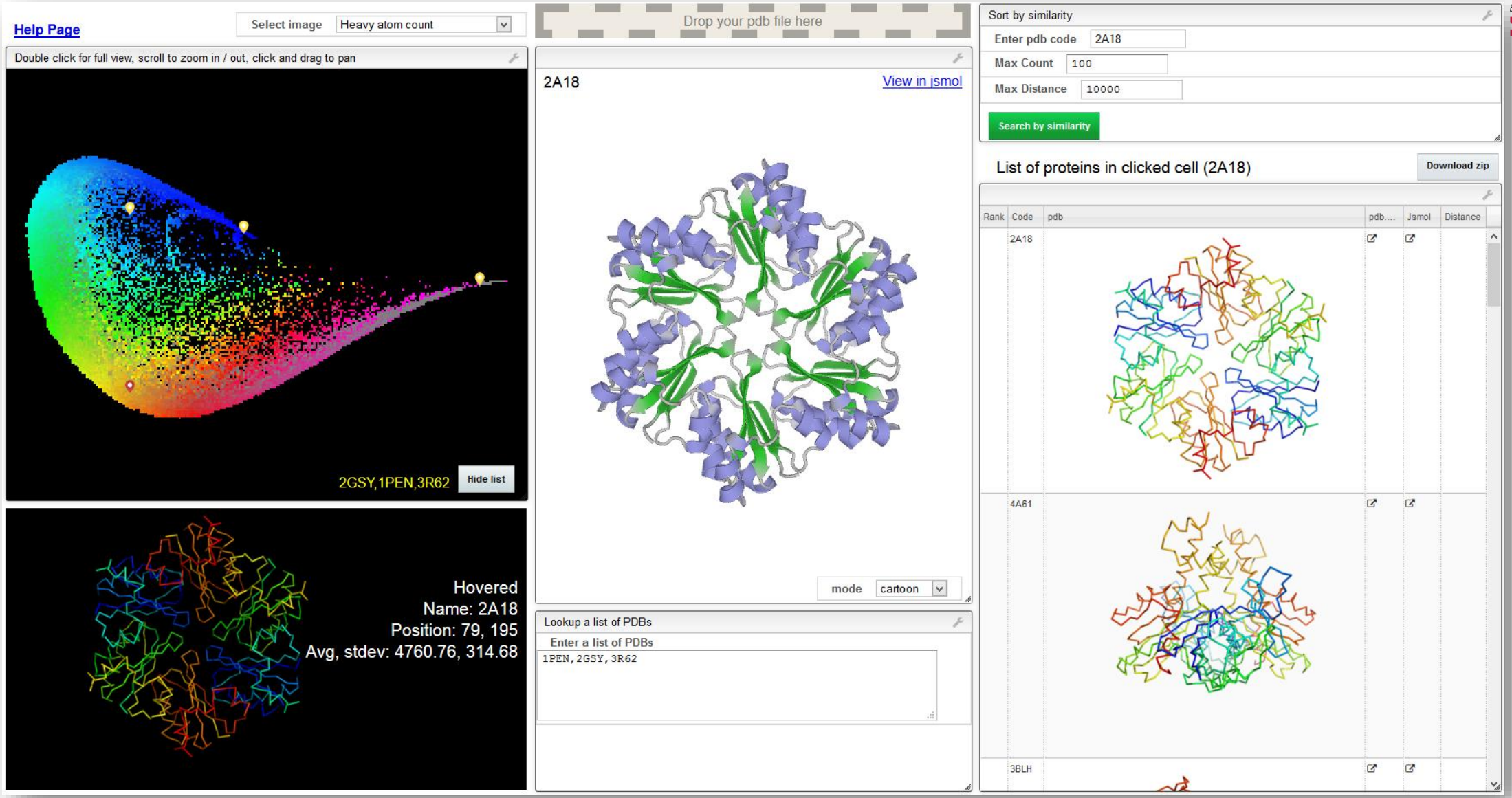






Johannes Morstein, Mahendra Awale *et al.*, *ACS Central Science* **2019**, *5*, 607-618

PDB explorer

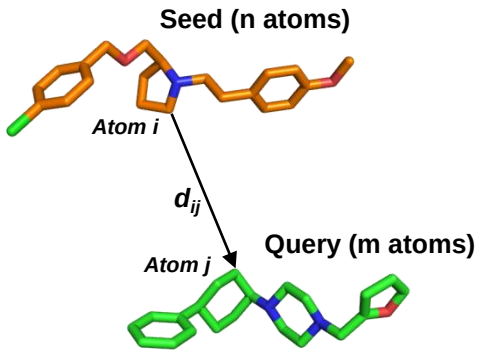


www.cheminfo.org/pdbexplorer

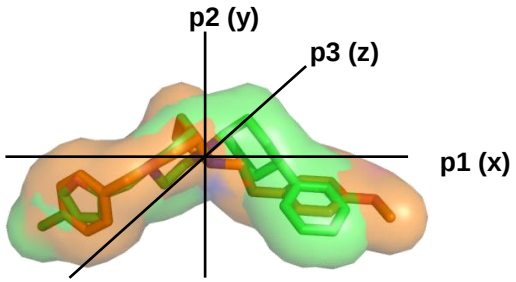
Xian Jin et al. *BMC Bioinformatics* **2015**, *16*, 339

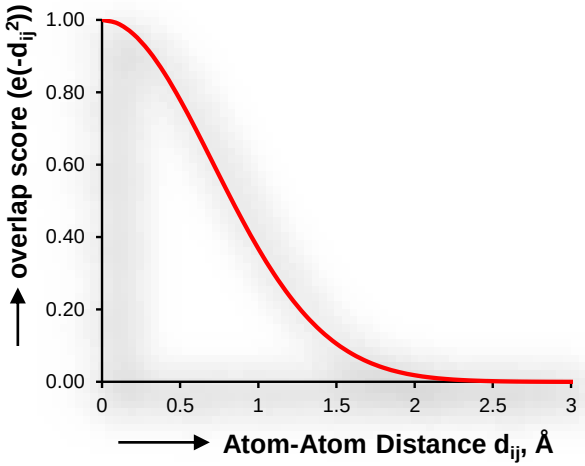
xLOS: matching pairs of atoms in 3D

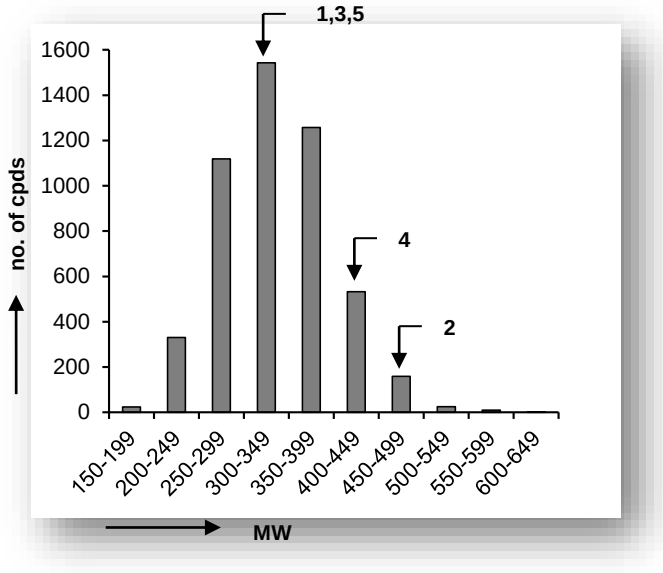
comparing molecular shapes and pharmacophores



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

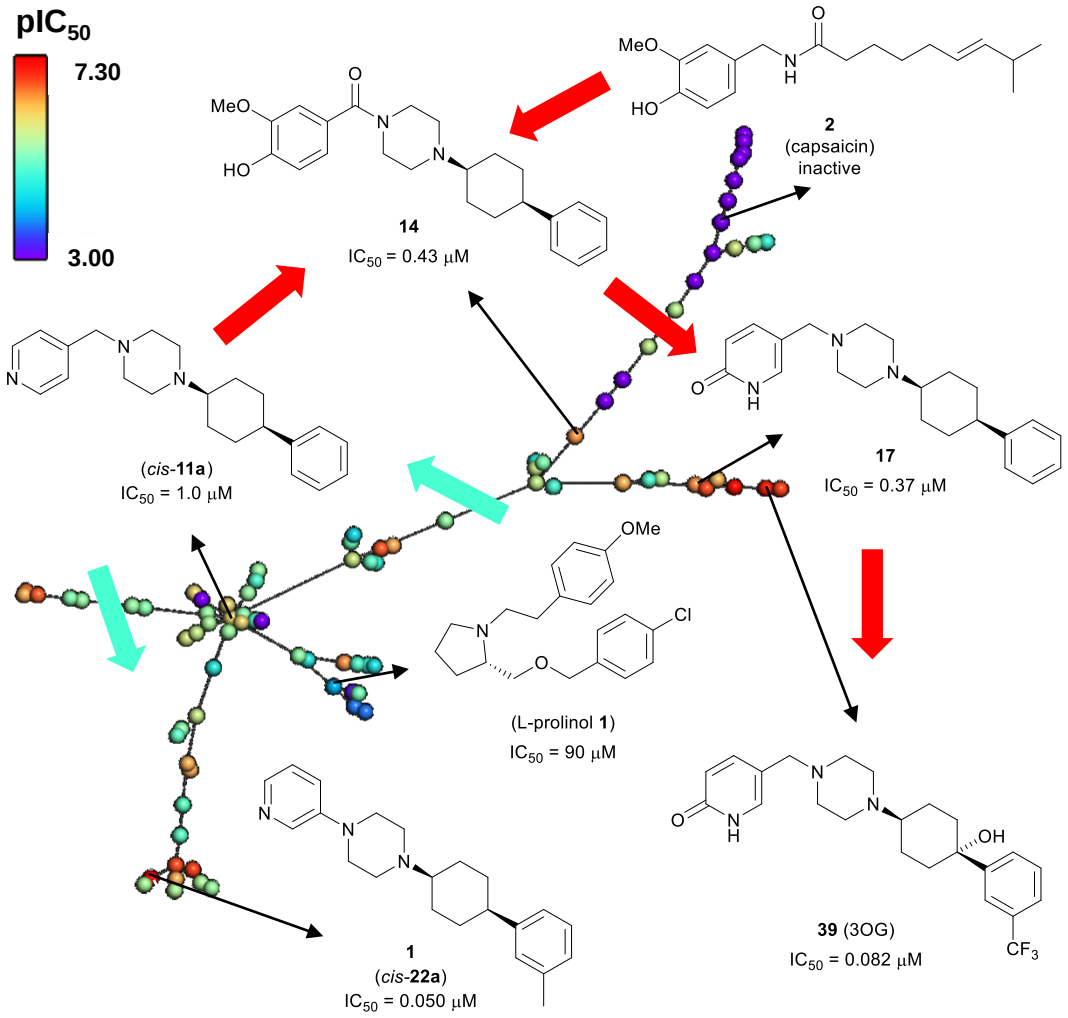


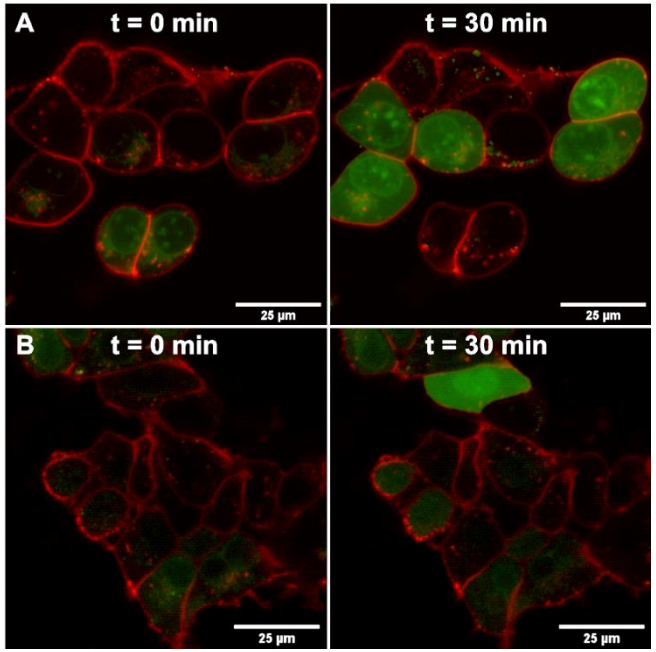


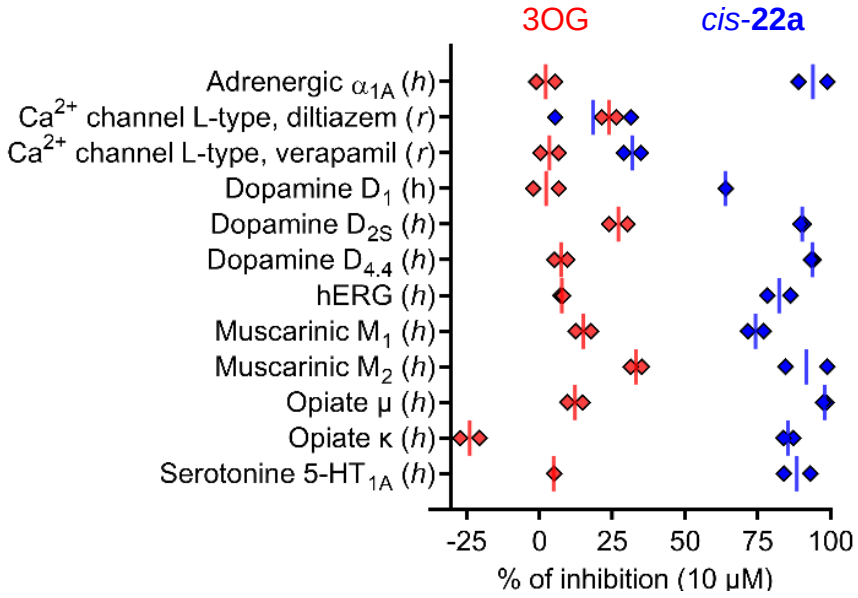


Céline Simonin et al., *Angew. Chem., Int. Ed. Engl.* **2015**, *54*, 14748-14752

3OG is a selective TRPV6 inhibitor

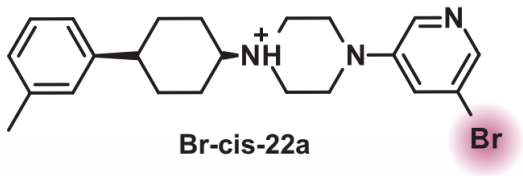


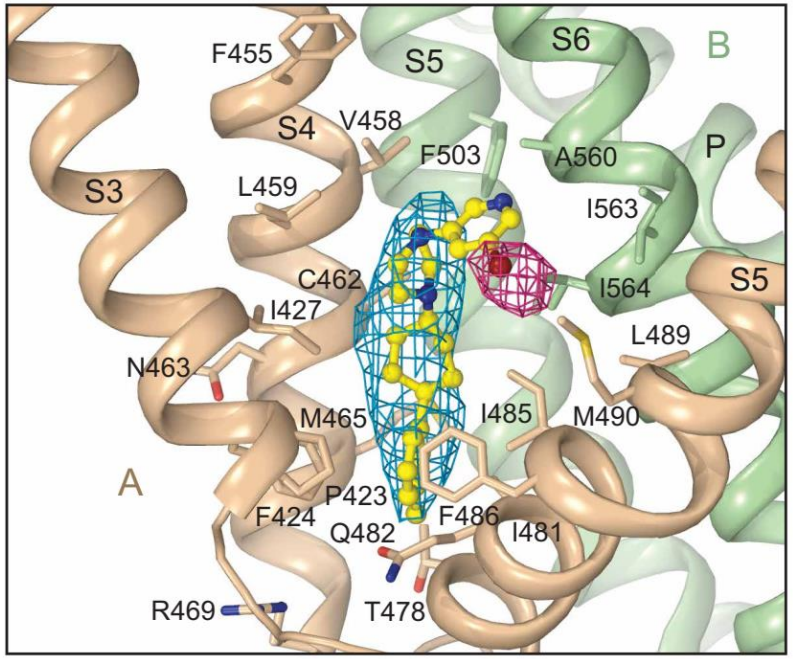


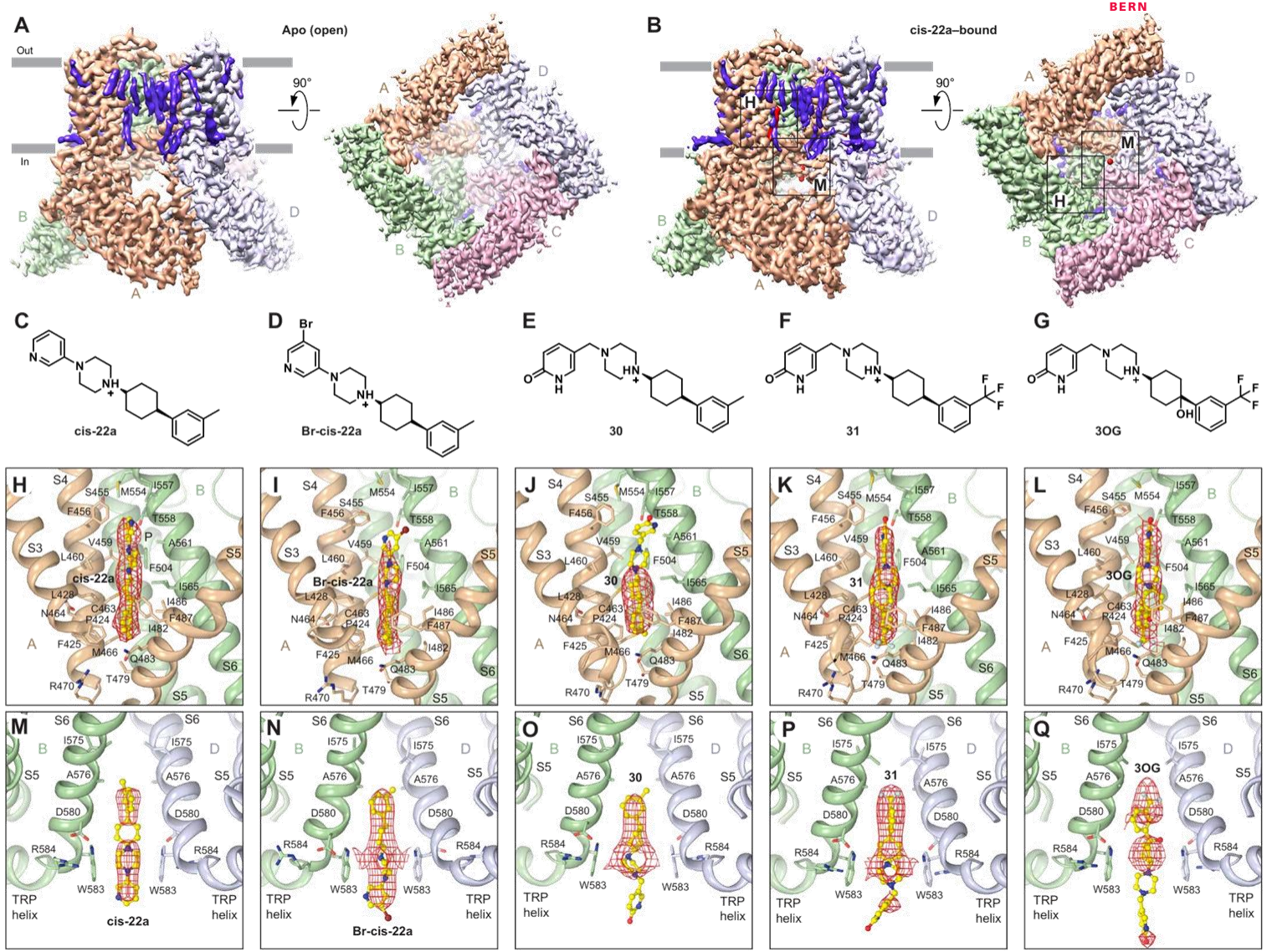


Micael R. Cunha *et al. RSC Med. Chem.* **2020**, *1*, 1032-1040

X-Ray and cryo-EM structures of TRPV6 bound inhibitors

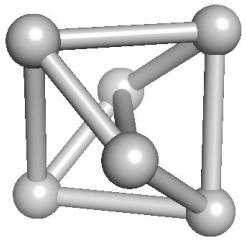




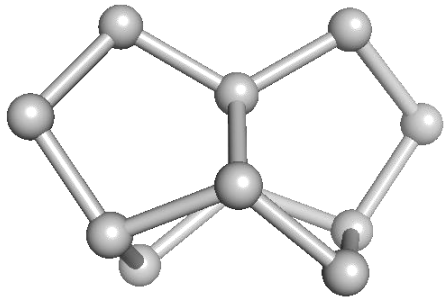


Rajesh Bhardwaj *et al. Sci. Adv.* **2020**, *6*, eabe1508

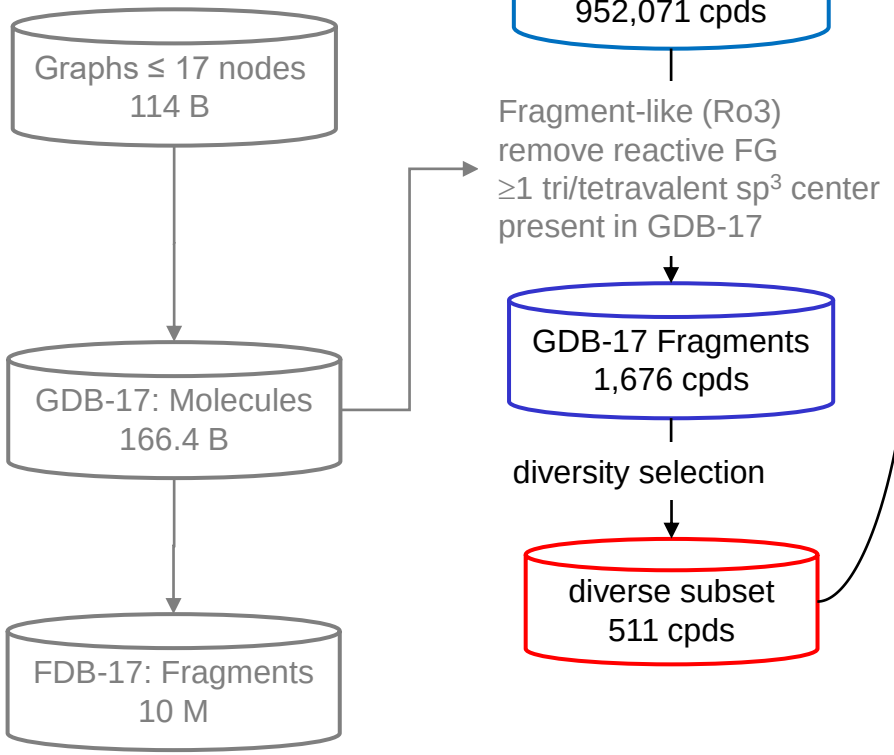
GDB

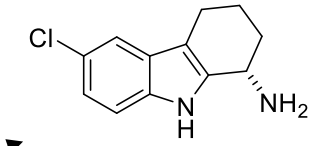


Claus Benzol
(1867)

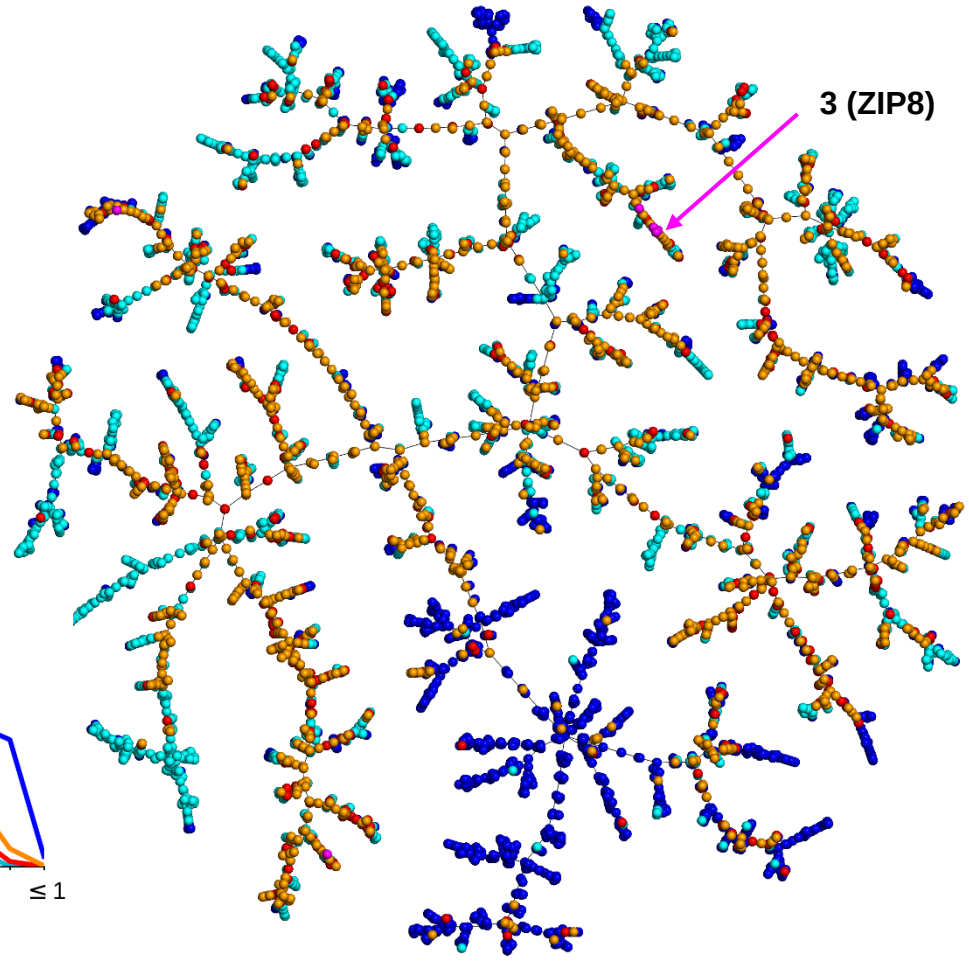


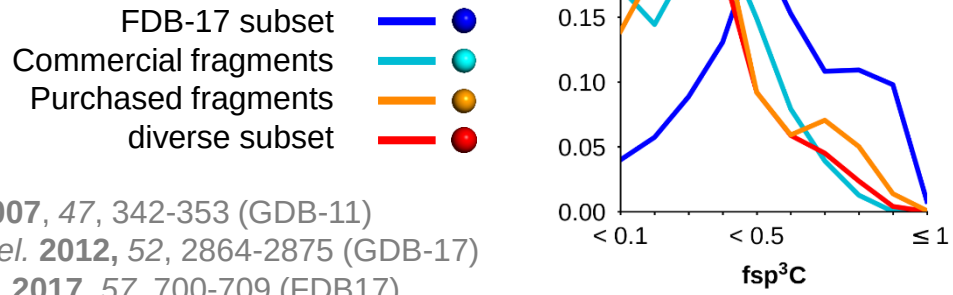
Trinorbornane
(2007, 2017)





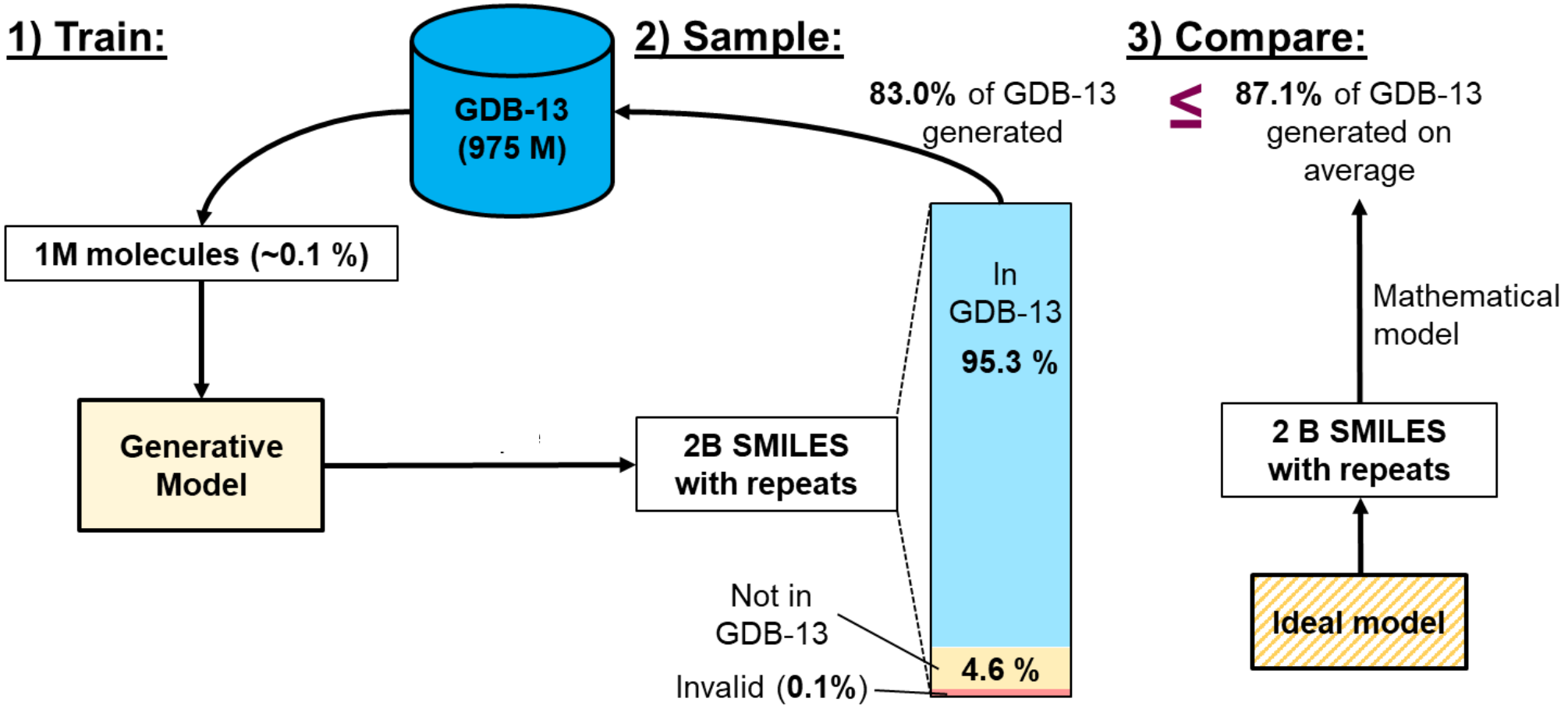
(*S*)-**3**
ZIP8
IC_{50} = 17.2 ± 3.8 μM





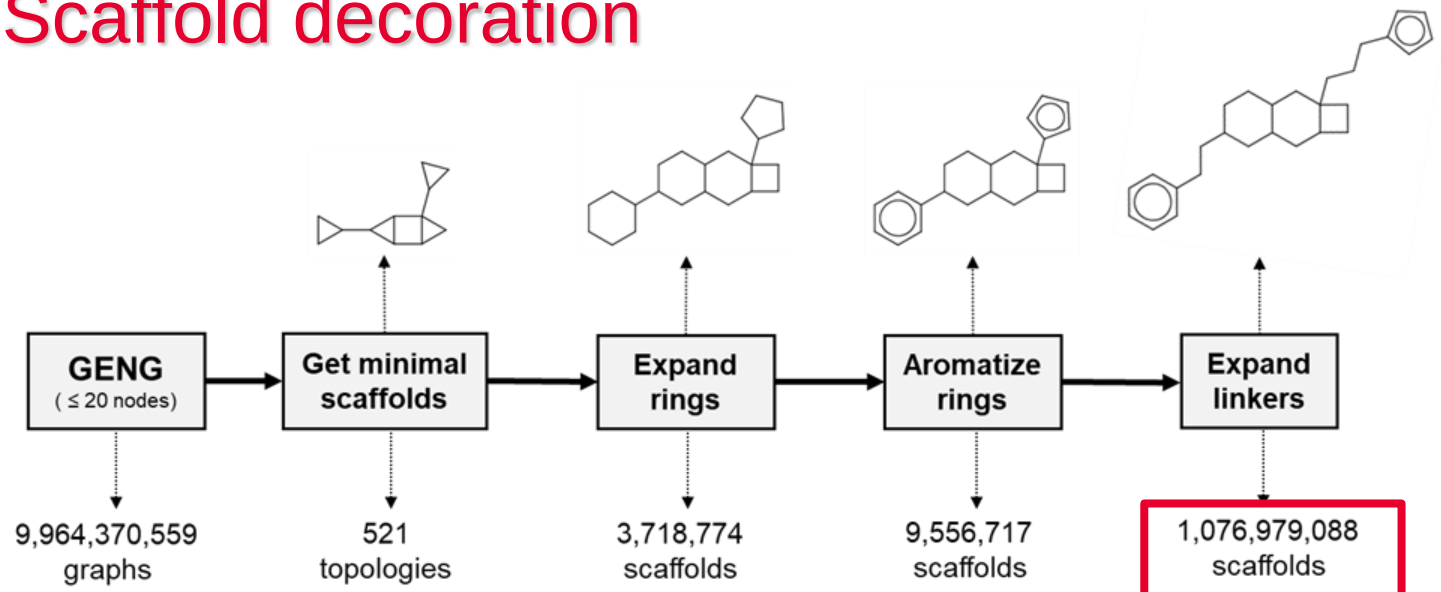
Tobias Fink et al. *J. Chem. Inf. Model.* **2007**, *47*, 342-353 (GDB-11)
Lars Ruddigkeit et al., *J. Chem. Inf. Model.* **2012,** *52*, 2864-2875 (GDB-17)
Ricardo Visini et al., *J. Chem. Inf. Model.* **2017,** *57*, 700-709 (FDB17)
Jonai Pujol-Giménez, Marion Poirier et al., *ChemMedChem* **2021**, *16*, 3306-3314

Sampling

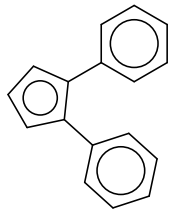


Josep Arus-Pous et al., *J. Cheminf.* **2019**, *11*, 20

Scaffold decoration

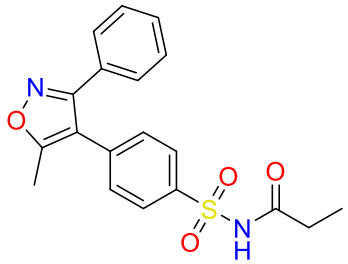


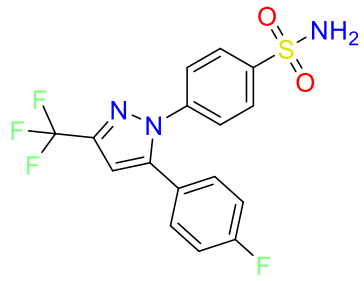
GDBscaffolds



c1(c2c(c3ccccc3)ccc2)ccccc1

LSTM or Transformer





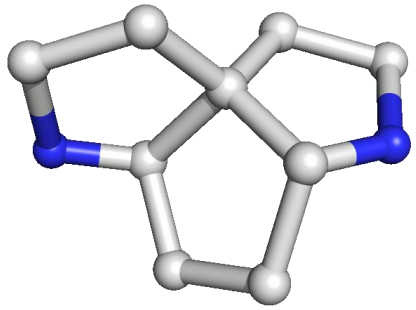
O=S(c1ccc(n2c(c3ccc(F)cc3)cc(C(F)(F)F)n2)cc1)(N)=O

Training set	Architecture	% Invalid	% Correct	% Recovered	% Unique
ChEMBL	LSTM	8.66 %	61.22 %	16.65 %	54.03 %
	Transformer	1.73 %	87.87 %	36.25 %	58.33 %
ZINC	LSTM	2.62 %	88.10 %	25.94 %	59.56 %
	Transformer	0.52 %	96.67 %	31.17 %	51.94 %

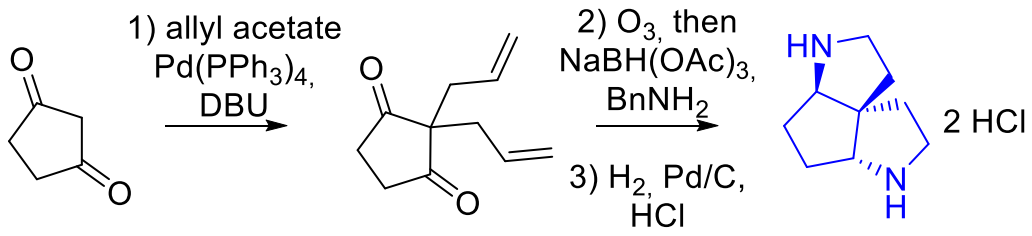
Josep Arus-Pous et al., *J. Cheminformatics* **2020**, *12*, 1-18, doi:10.1186/s13321-020-00441-8 + PhD thesis, **2020**

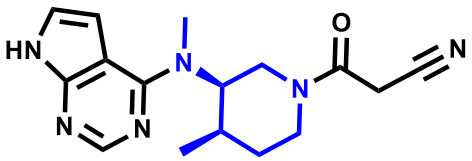
Ring systems

Ring Systems	
GDB4c	916,130
aliphatic	728,238
5-membered	181
no spiro	102
no norbornane	20

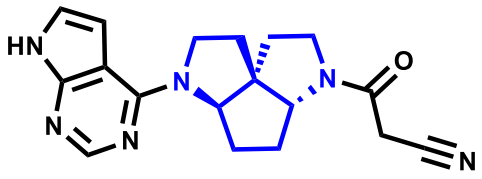


Triquinazine

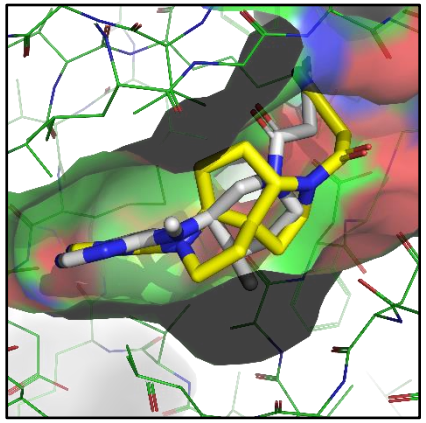




Tofacitinib



9 (KMC420)



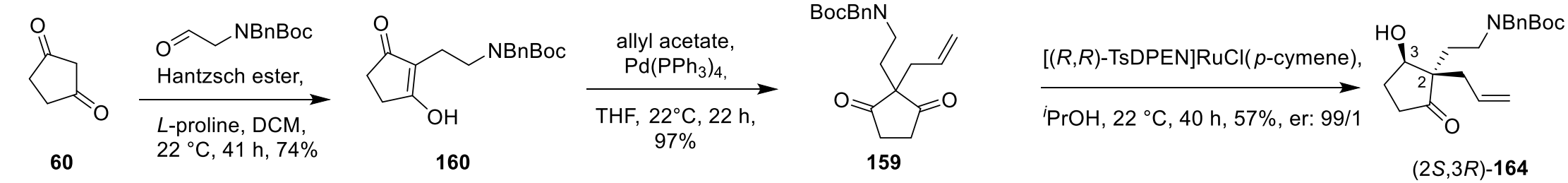
PDB: 3LXK (JAK3)

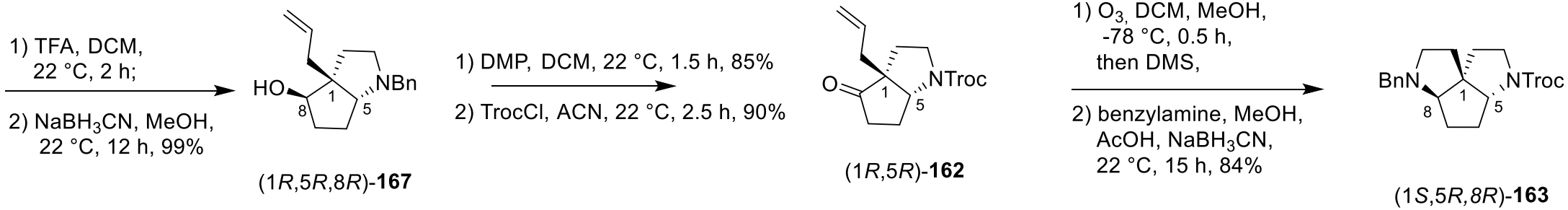
Table 1: JAK kinase inhibition profiles.

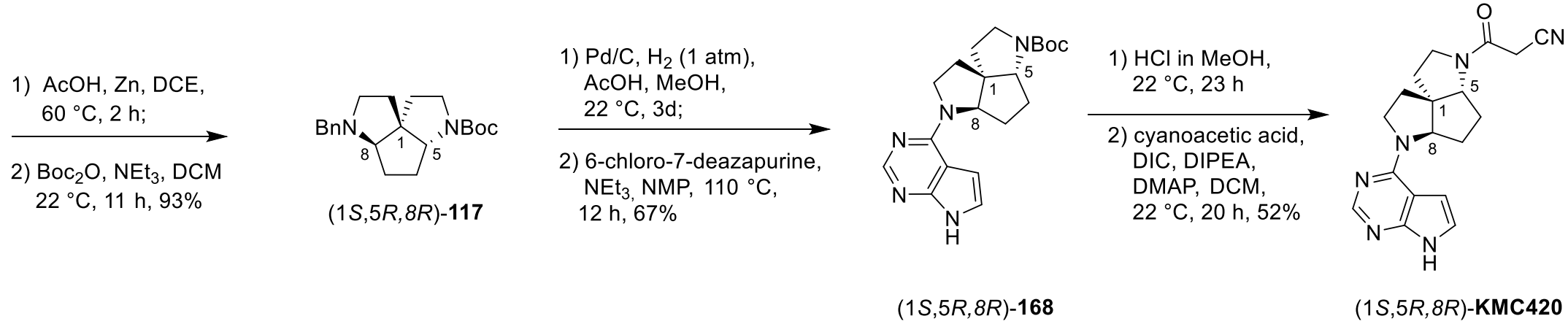
IC_{50} (nM)	JAK1	JAK2	JAK3	TYK2
Delgocitinib[45]	2.8	2.6	13	58
Tofacitinib[45]	2.9	1.2	1.1	42
PF-06651600[41]	>10,000	>10,000	33 nM	>10,000
rac-**9**[a)]	1.0	13	29	26
rac-**10**[a)]	30	53	9	183
(*R*)-**9**[a)]	0.8	5.4	14	52
(*S*)-**9** [b)]	>1,000	>1,000	>1,000	>1,000

Ricardo Visini, Josep Arùs-Pous et al., *J. Chem. Inf. Model.* **2017,** *57*, 2707-2718

Kris Meier et al., *Angew. Chem., Int. Ed. Engl.* **2020**, *60*, 2074–2077



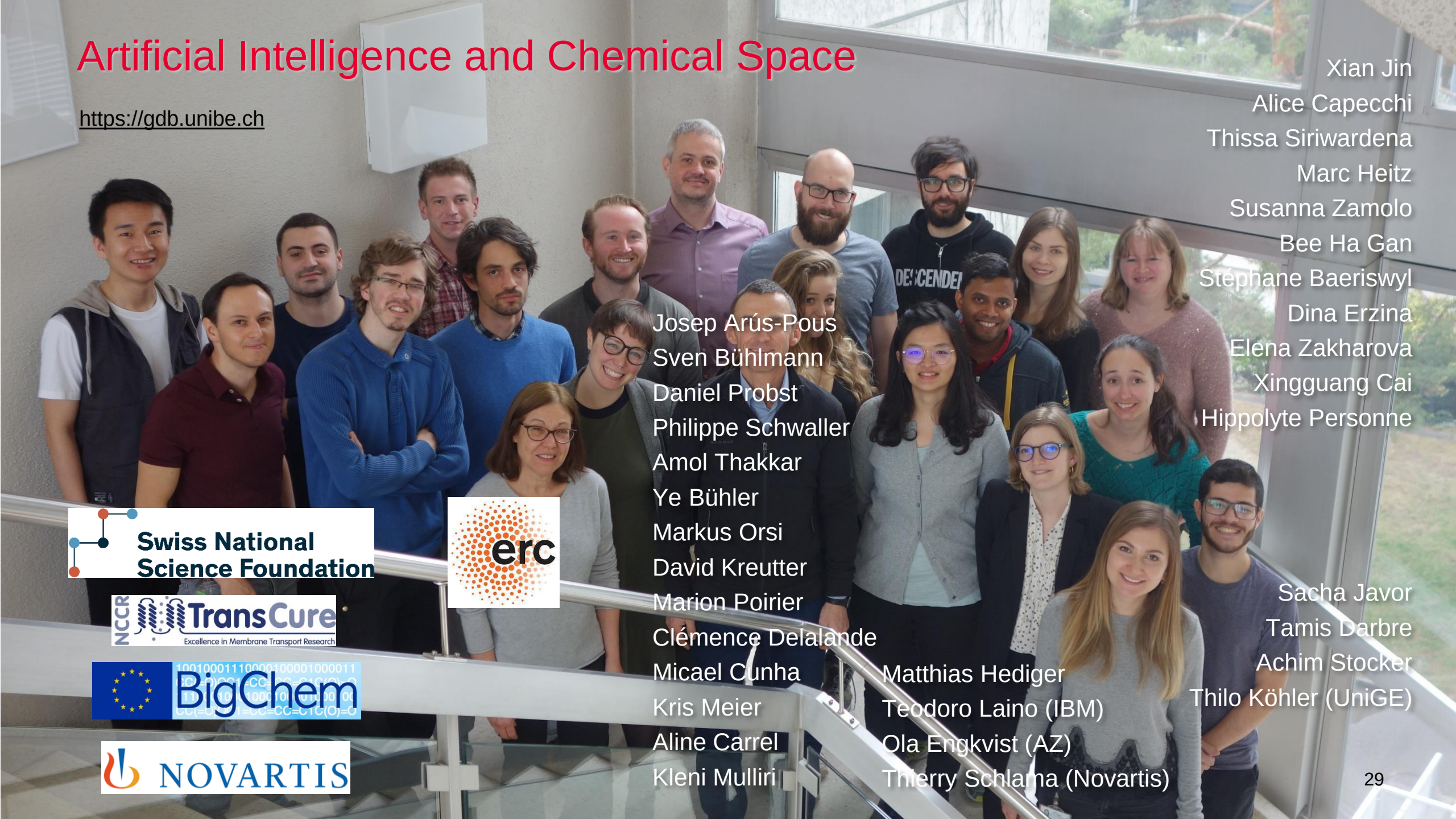




Kris Meier et al., *Angew. Chem., Int. Ed. Engl.* **2020**, *60*, 2074–2077

Artificial Intelligence and Chemical Space

https://gdb.unibe.ch



Swiss National Science Foundation

NCCR TransCure Excellence in Membrane Transport Research

BigChem

NOVARTIS

erc

Josep Arús-Pous
Sven Bühlmann
Daniel Probst
Philippe Schwaller
Amol Thakkar
Ye Bühler
Markus Orsi
David Kreutter
Marion Poirier
Clémence Delalande
Micael Cunha
Kris Meier
Aline Carrel
Kleni Mulliri

Matthias Hediger
Teodoro Laino (IBM)
Ola Engkvist (AZ)
Thierry Schlama (Novartis)

Xian Jin
Alice Capecchi
Thissa Siriwardena
Marc Heitz
Susanna Zamolo
Bee Ha Gan
Stéphane Baeriswyl
Dina Erzina
Elena Zakharova
Xingguang Cai
Hippolyte Personne

Sacha Javor
Tamis Darbre
Achim Stocker
Thilo Köhler (UniGE)