Structural Cheminformatics Workflows for Computer-Aided Drug Discovery

Chris de Graaf

VU UNIVERSITY AMSTERDAM

UK QSAR – MGMS

© Heptares Therapeutics 2018
The HEPTARES name, the logo and STAR are trade marks of Heptares Therapeutics Ltd

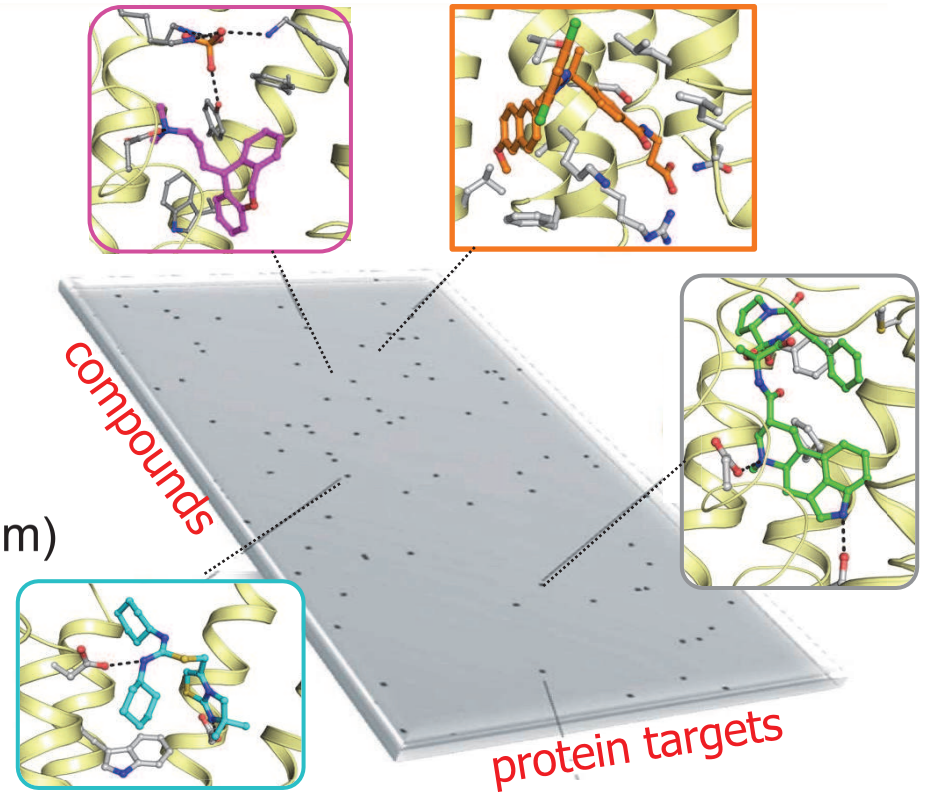
Heptares Therapeutics is a wholly owned subsidiary of Sosei Group Corporation

Cardiff, UK
2018-04-11

Pieces of Protein-Ligand Interaction Data

Protein structures

137.322 structures (PDB): 42.120 distinct proteins, 24.678 ligands



Compounds

94.202.854 (PubChem)

10^{60} (theoretical)

Bioactivity data

14.675.320 (ChEMBL)

11.538 proteins

1.735.442 compounds

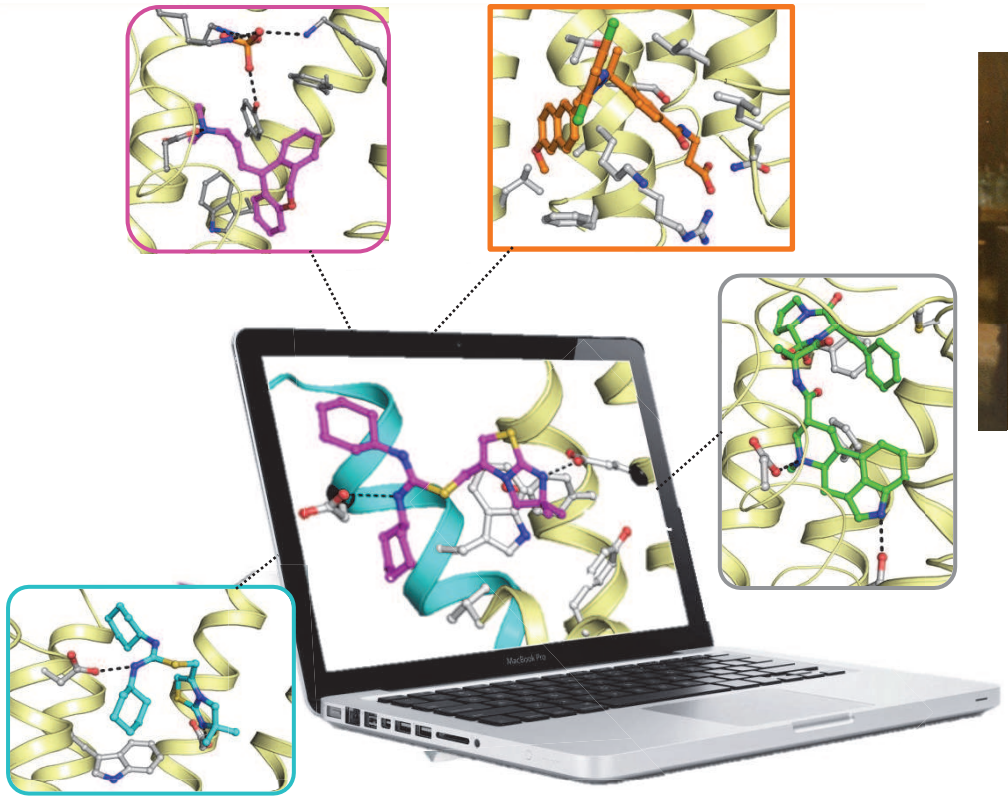
Protein sequences

102.248.261 (TrEMBL), 556.388 (SwissProt), 161.521 (human)

Data: Jan 2018

Predicting Structural Protein-Ligand Interactions?

Structural chemogenomics databases to **navigate** and **predict** ...



Computer-aided drug discovery tools to **predict** and **design** ....

New free nodes for structural protein-ligand interaction analysis!

KNIME – ChEMBL

small molecule cheminformatics

RCSB PDB PROTEIN DATA BANK · 3D e-Chem

structural cheminformatics

RDKit Open-Source Cheminformatics and Machine Learning

CDK

Indigo

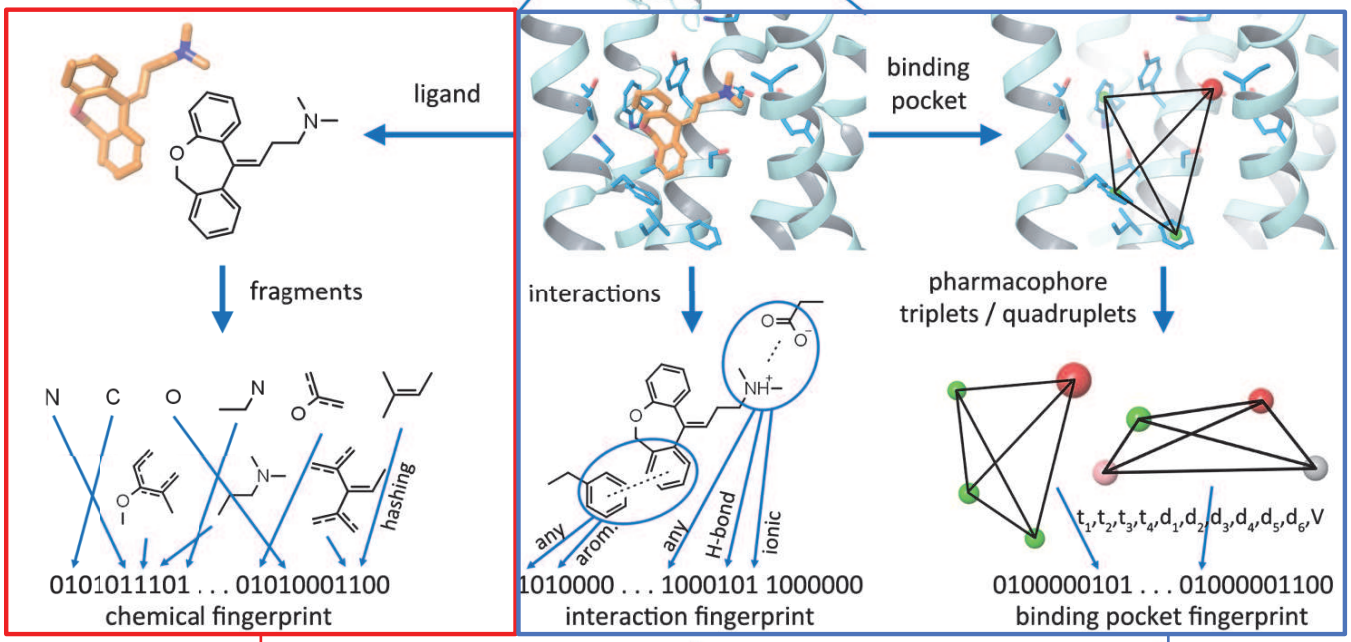
Open Babel

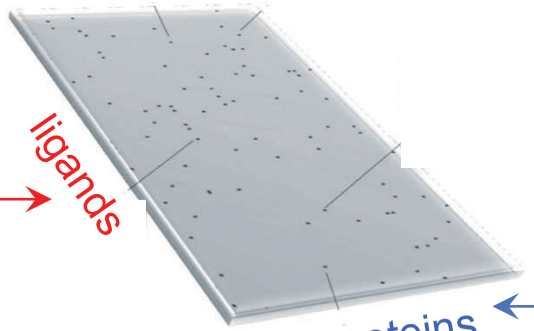
Erl Wood Cheminformatics

SyGMa

ChemAxon

OpenPHACTS





GPCRdb

ss-TEA score

KLIFS

KRIPO Key Representation of Interactions in POckets

Ligands and Proteins Viewer

Ph4 aligner

silicos it

PLANTS

https://knime.org

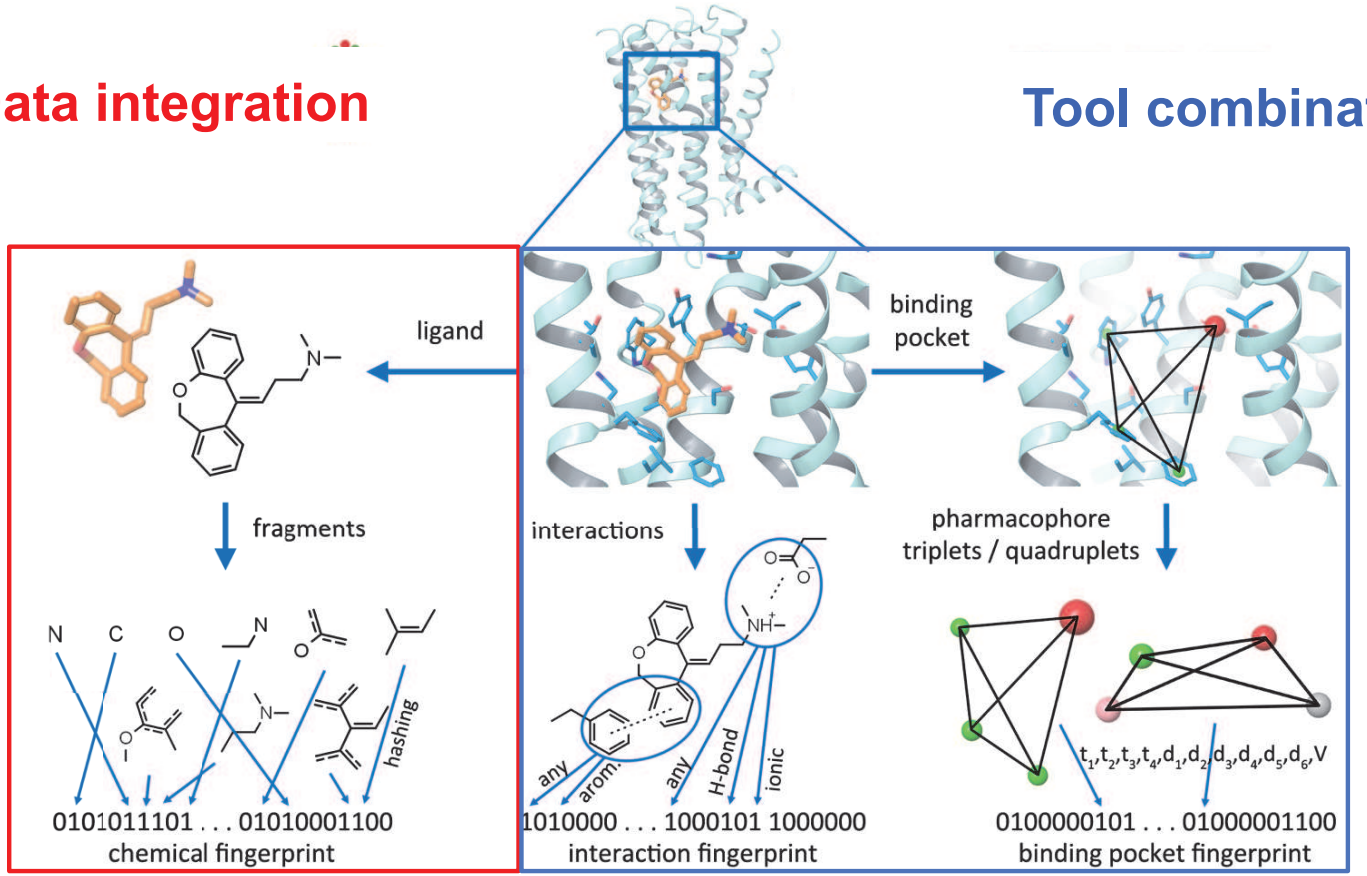
Vass (2016) Curr Opin Pharm 30:59

https://3d-e-chem.github.io

New nodes for computer-aided drug discovery

Data integration

Tool combination



ChEMBL RDKit Open-Source Cheminformatics and Machine Learning SyGMa silicos it KLIFS GPCRdb ss-TEA score KRIPO Key Representation of Interactions in PDockets PLANTS Ph4 aligner Ligands and Proteins Viewer

target prediction ligand repurposing scaffold hopping bioisoteric replacement

https://3d-e-chem.github.io

3D-e-Chem toolkit

Web applications

KNIME nodes to extract and combine data

KLIFS	Kinase structure/ligand interaction database	9 nodes
GPCRdb	GPCR structure/sequence database	7 nodes
KRIPO (Key Representation of Interactions in POckets)	Ligand substructure annotated protein pharmacophore	2 nodes

Executables

KNIME nodes to configure, run, and analyze results

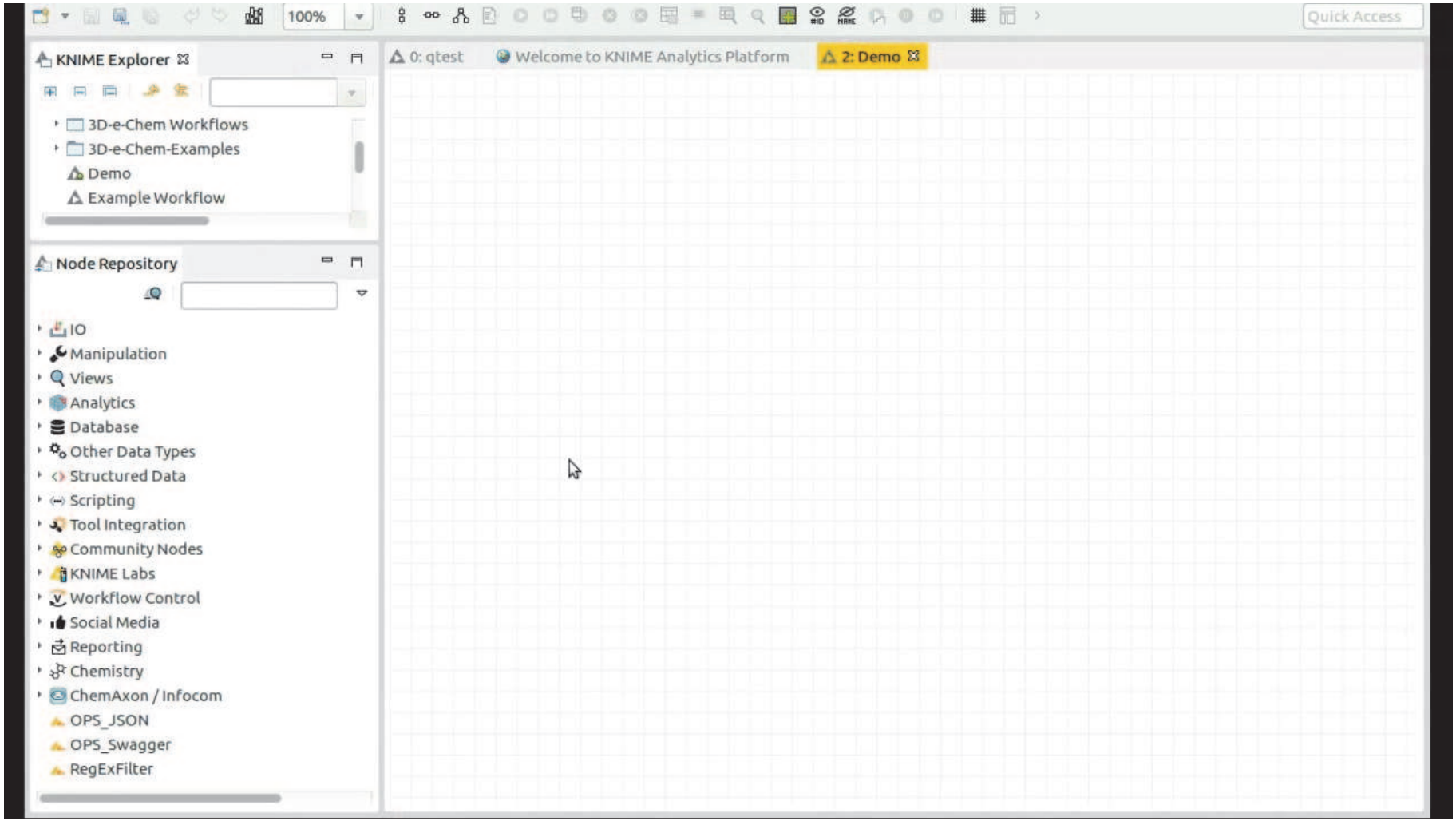
SyGMa	Site of metabolism prediction	1 node
silicos it	Ligand-based shape and pharmacophore alignment	7 nodes
PLANTS	Molecular docking simulation	7 nodes

Standalone

KNIME nodes

ss-TEA score	Amino acid sequence entropy analysis	1 node
Ph4 Aligner	Align, read, write pharmacophores	7 nodes
Molviewer	3D Viewer of protein-ligand complexes/Ph4s	4 nodes

KNIME: from nodes to cheminformatics workflows



GPCRdb

ss-TEA score

KLIFS

KRIPO
Key Representation of Interactions in POckets

silicos it

PLANTS

Ligands and Proteins Viewer

https://tech.knime.org/3d-e-chem-nodes-for-knime

https://3d-e-chem.github.io

Design your own drug discovery workflow!

https://3d-e-chem.github.io

Radboudumc

netherlands eScience center

VU UNIVERSITY AMSTERDAM

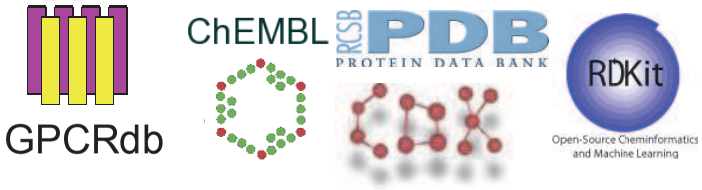
M. Vass
S. Verhoeven
R. McGuire
A.J. Kooistra
C. De Graaf

KNIME

McGuire, Verhoeven, Vass, et al. (2017) *J Chem Info Model* 57: 115; Kooistra, Vass et al. (2018) *ChemMedChem* 13, 614

Structural Cheminformatics Workflows

- Structural coverage of chemical GPCR ligand space



- Structure-based GPCR mutation data mapping
- Sequence-based GPCR binding site prediction ligand repurposing

- Structural interaction-based ligand scaffold hopping

- Structure-based kinase-GPCR ligand repurposing

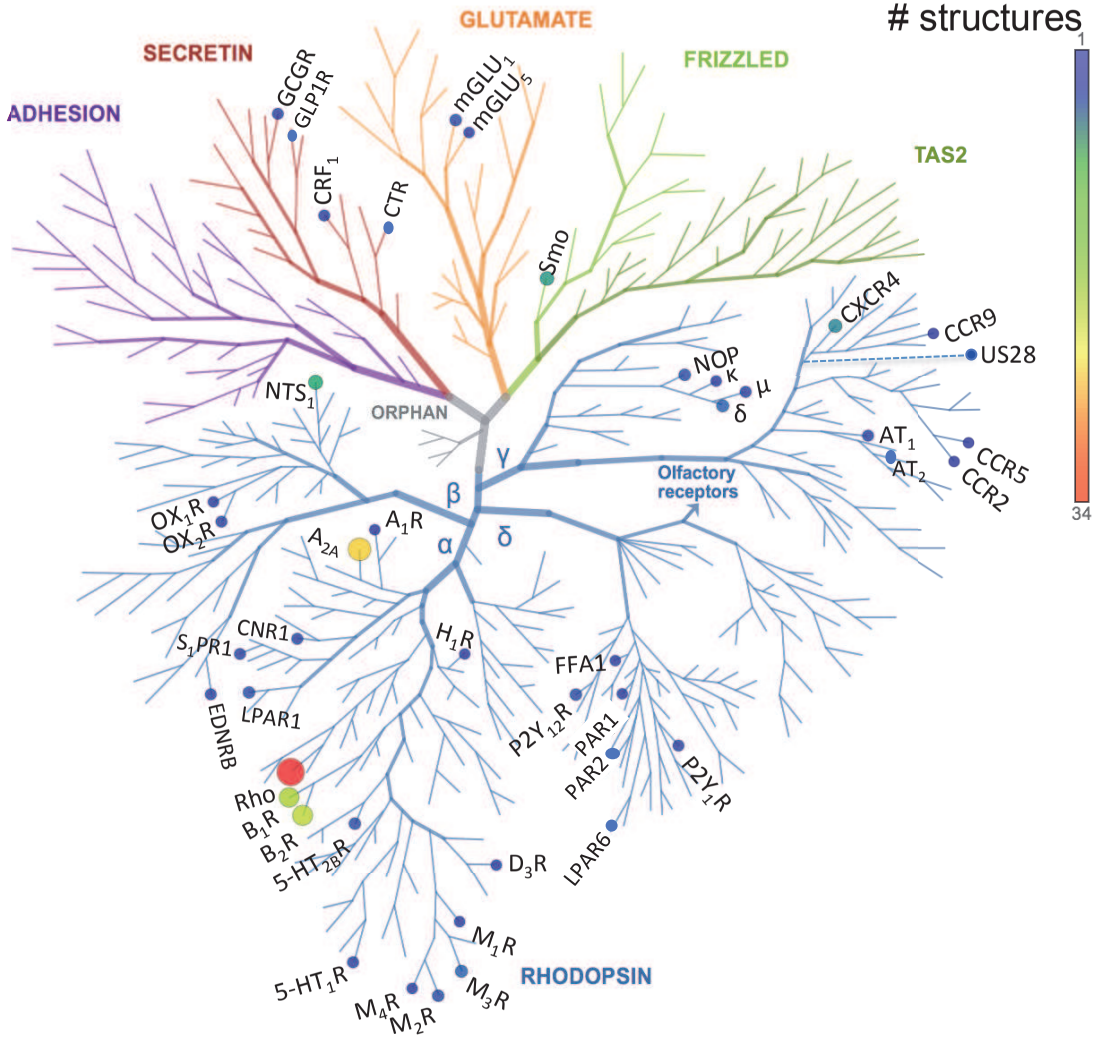
Vass et al. *Tr Pharm Sci* 2018, doi:10.1016/j.tips.2018.02.004
Kooistra et al. *ChemMedChem*, 2018, 13: 614

Vass et al. *In preparation* 2018
McGuire et al. *J Chem Info Model* 2017

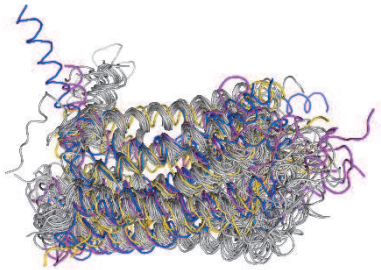
- Structure-based virtual screening - SAR - SKR

docking - IFP - RAMD

Exciting era of GPCR crystal structure data crunching!



GPCRs: Transmembrane Drug Targets in 3D



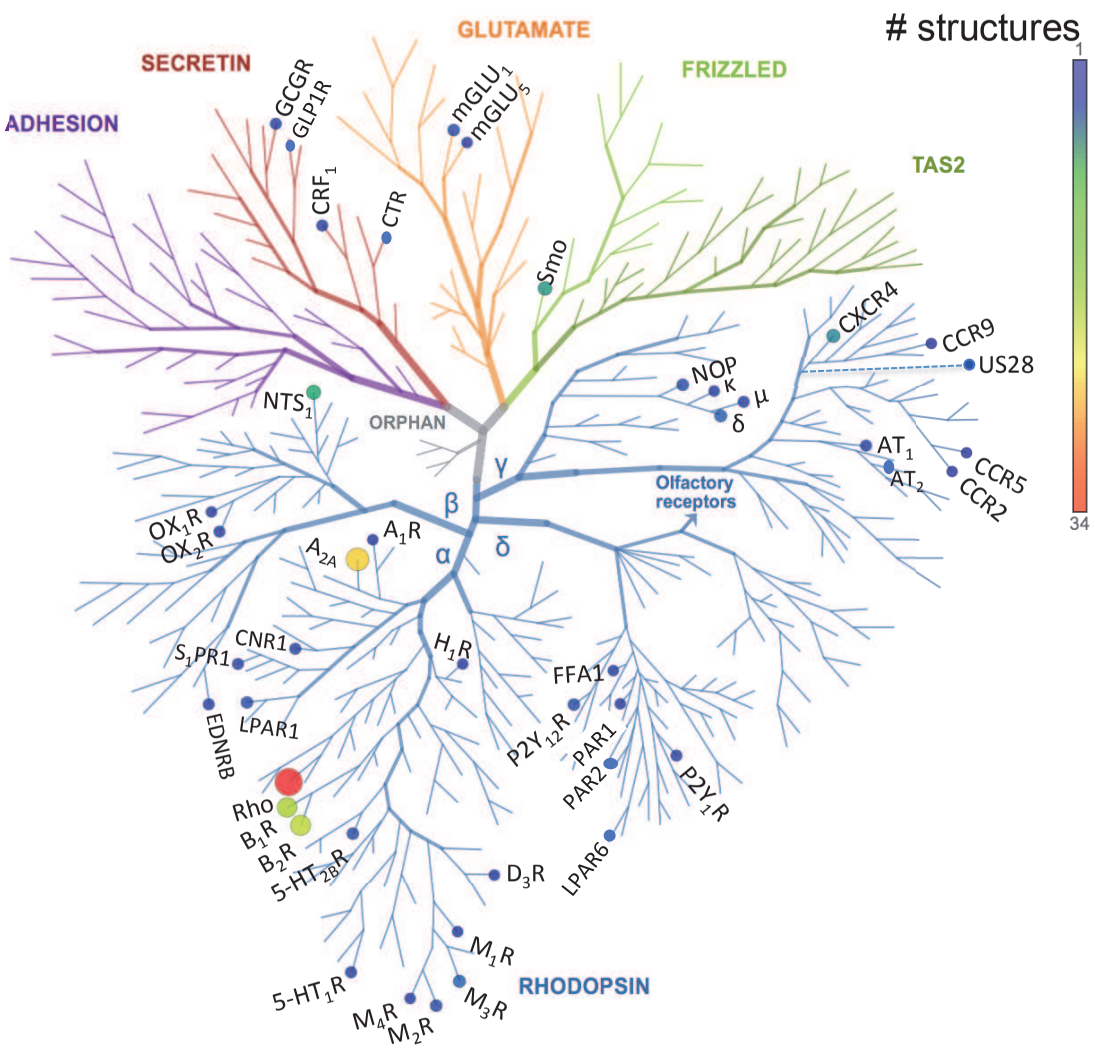
468 drugs:
33% of all small-molecule drugs target a GPCR

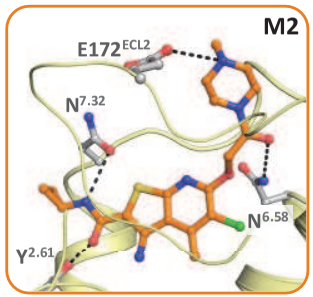
214 GPCR crystal structures:

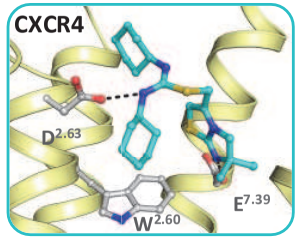
- 48 different GPCRs
- 125 unique GPCR-ligand complexes

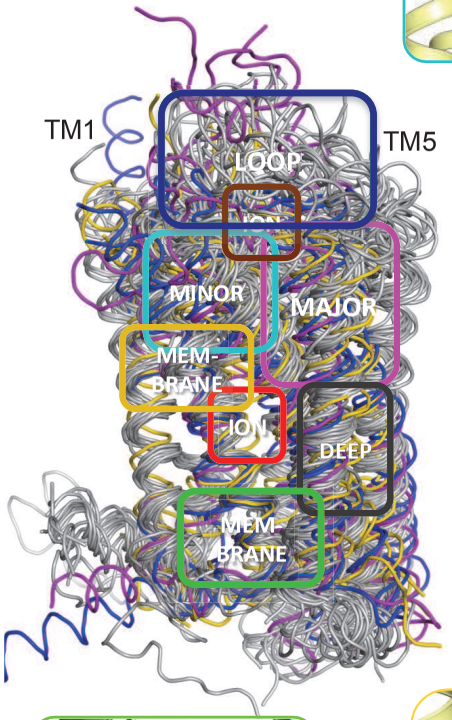
16 drugs
10 (former) drug candidates

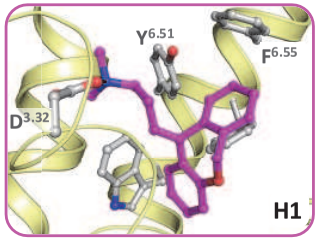
Exciting era of GPCR crystal structure data crunching!

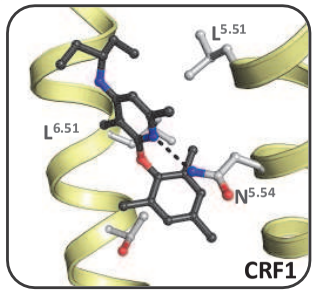


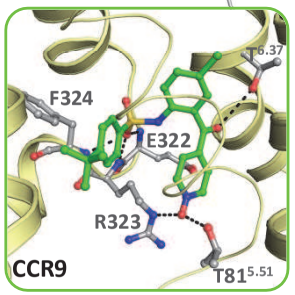


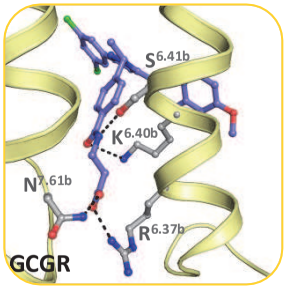








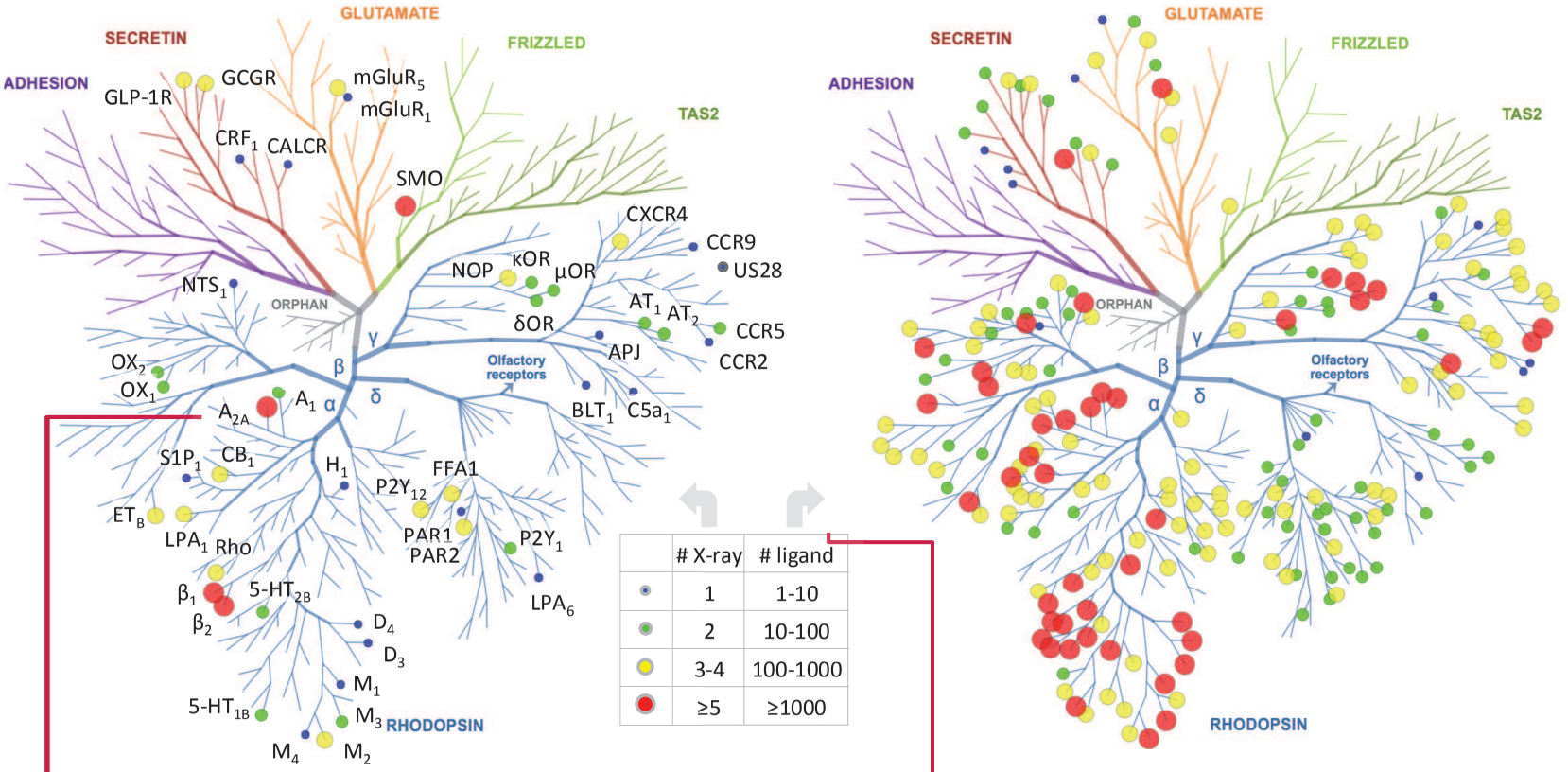




214 GPCR structures:

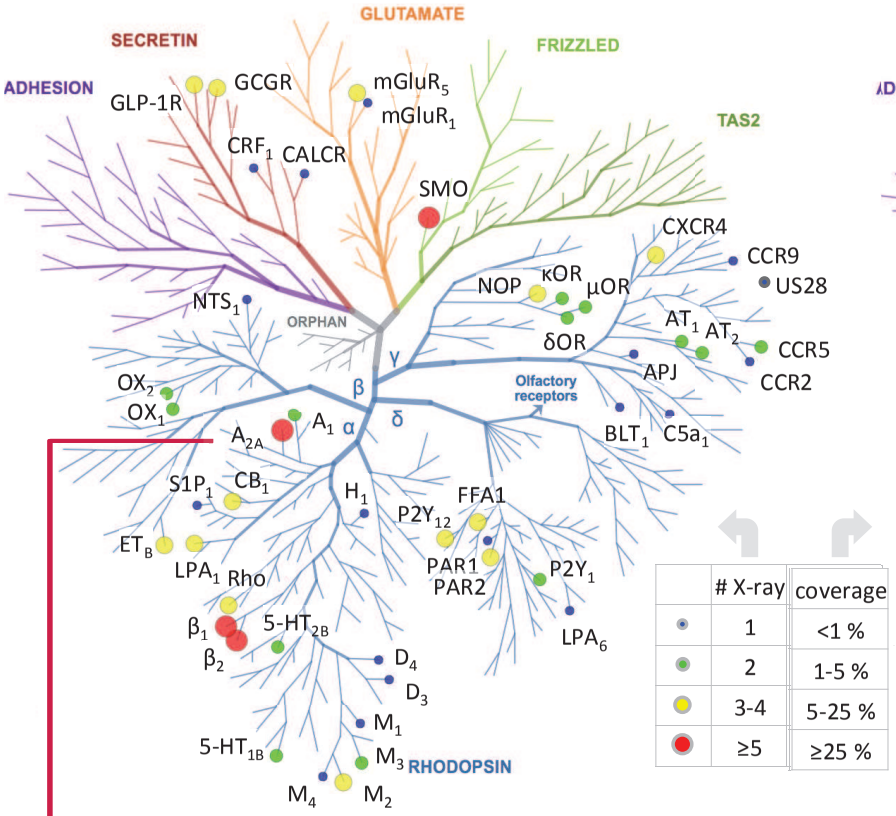
- 125 GPCR-ligand complexes
- 48 different GPCRs
- 111 small molecule ligands

Structural coverage of chemical GPCR ligand space?

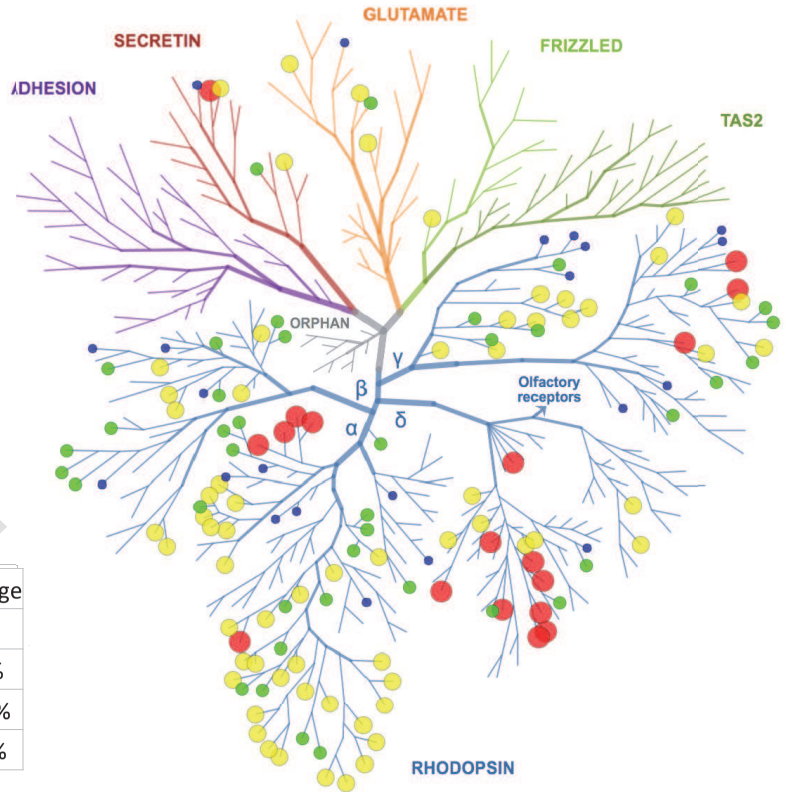


Vass, Kooistra, de Graaf et al. *TiPS*

Structural coverage of chemical GPCR ligand space?



	# X-ray	coverage
	1	<1 %
	2	1-5 %
	3-4	5-25 %
	≥5	≥25 %



414 human non-olfactory GPCRs

111 ligands
48 GPCRs
136 GPCR-ligand complexes

GPCR X-ray
Small molecules

122.873 ligands
321 GPCRs
174.345 GPCR-ligand complexes

ChEMBL
GPCR bioactivities

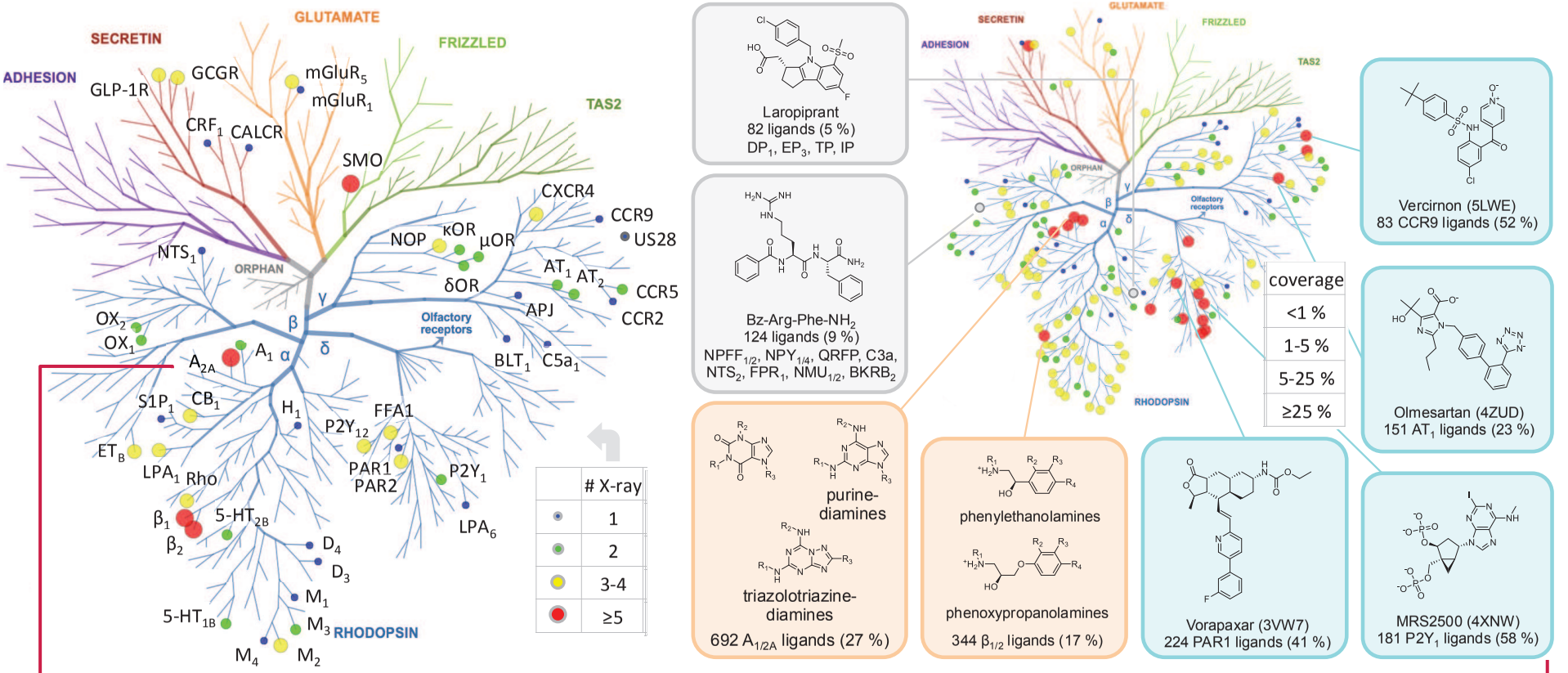
ECFP-4
MACCS

GPCRome

158 GPCRs
9.600 ligands
15.504 GPCR-ligand complexes

Vass, Kooistra, de Graaf et al. *TiPS*

Structural coverage of chemical GPCR ligand space?



414 human non-olfactory GPCRs

111 ligands
48 GPCRs
136 GPCR-ligand complexes

GPCR X-ray
Small molecules

122.873 ligands
321 GPCRs
174.345 GPCR-ligand complexes

ChEMBL
GPCR bioactivities

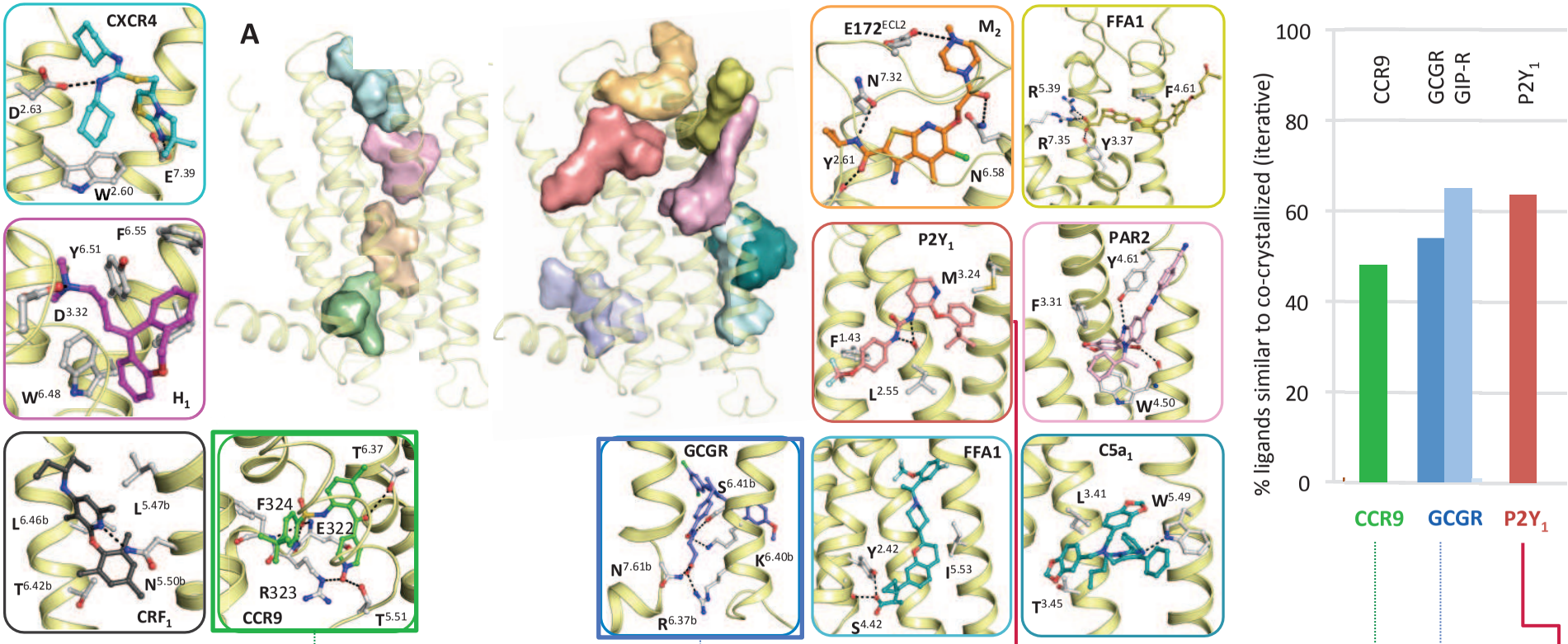
ECFP-4
MACCS

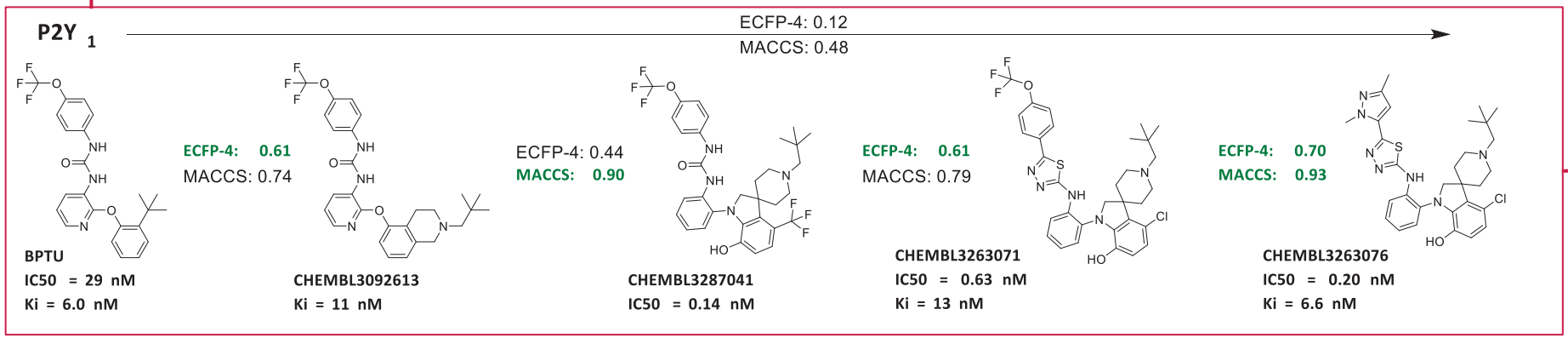
GPCRome

158 GPCRs
9.600 ligands
15.504 GPCR-ligand complexes

Vass, Kooistra, de Graaf et al. *TiPS*

Allosteric extrahelical and intracellular GPCR ligands





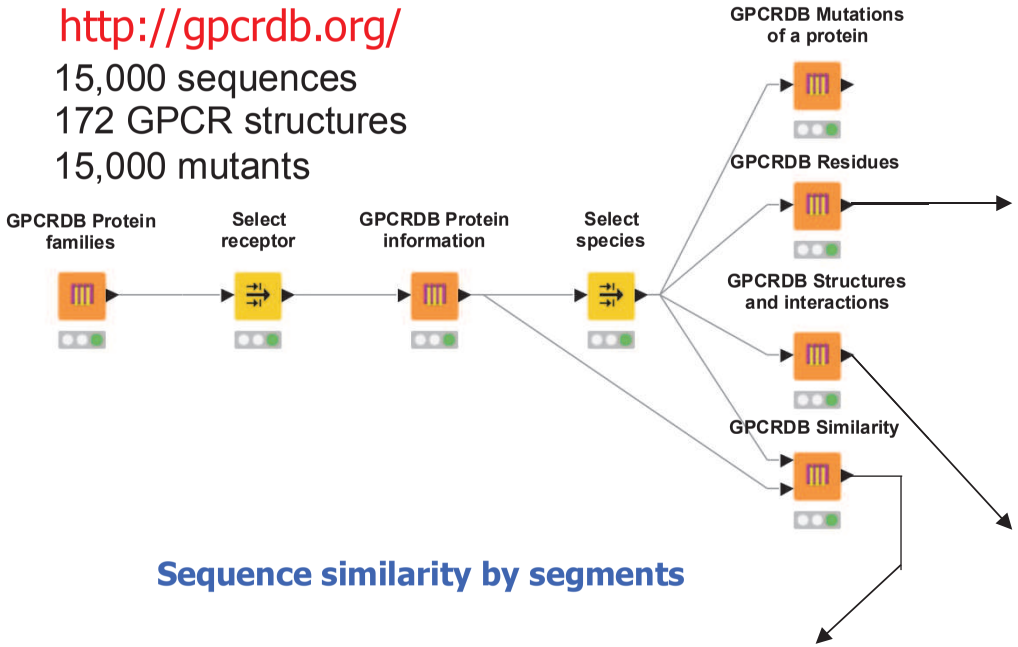
Vass, Kooistra, Stevens, Wang, de Graaf et al. ***TiPS*** 2018, doi: 10.1016/j.tips.2018.02.004

GPCRdb nodes: GPCR-ligand interaction data analysis

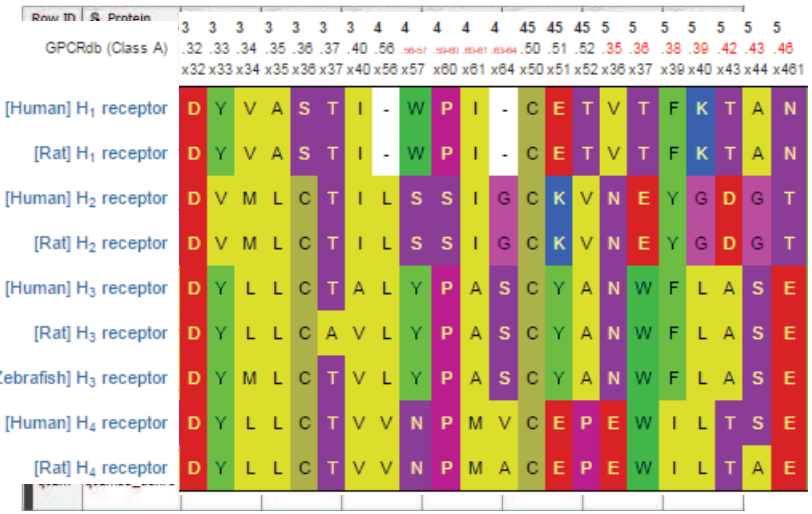
GPCRdb nodes to integrate structural GPCR data

http://gpcrdb.org/

15,000 sequences
172 GPCR structures
15,000 mutants



residue information and sequence alignment



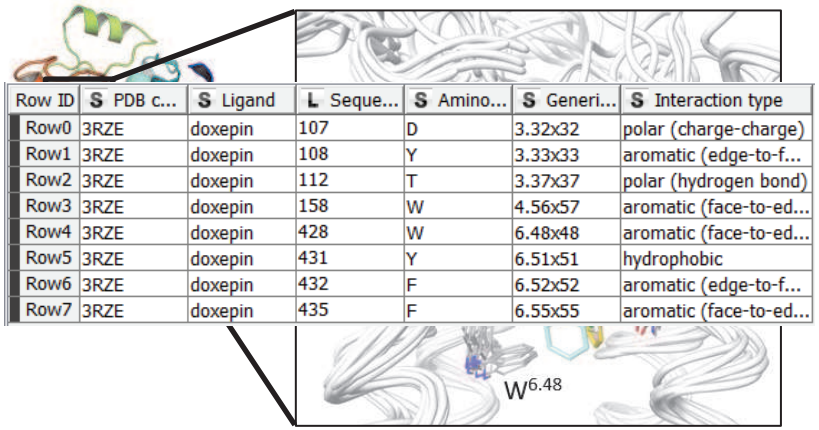
Sequence similarity by segments

Row ID	Query	Subject	Similar...	Identity
adrb2_hu...	adrb2_human	g5cvw2_phoph	81	78

	1	2	3	4	5	6	7	8	9
1. [Human] α_{1A}-adrenoceptor	-	73	63	41	40	40	40	38	41
2. [Human] α_{1B}-adrenoceptor	89	-	69	41	42	41	41	40	42
3. [Human] α_{1D}-adrenoceptor	81	83	-	38	40	39	41	40	43
4. [Human] α_{2A}-adrenoceptor	63	66	61	-	71	73	41	36	36
5. [Human] α_{2B}-adrenoceptor	62	65	62	84	-	73	39	35	37
6. [Human] α_{2C}-adrenoceptor	62	64	62	83	84	-	37	34	37
7. [Human] β_1-adrenoceptor	61	62	64	61	62	60	-	65	61
8. [Human] β_2-adrenoceptor	62	62	63	56	57	56	78	-	55
9. [Human] β_3-adrenoceptor	60	60	61	55	57	58	74	69	-

adrb2_hu...	adrb2_human	ada2c_human	56	34
adrb2_hu...	adrb2_human	ada2a_human	56	36

X-ray structures and protein-ligand interactions

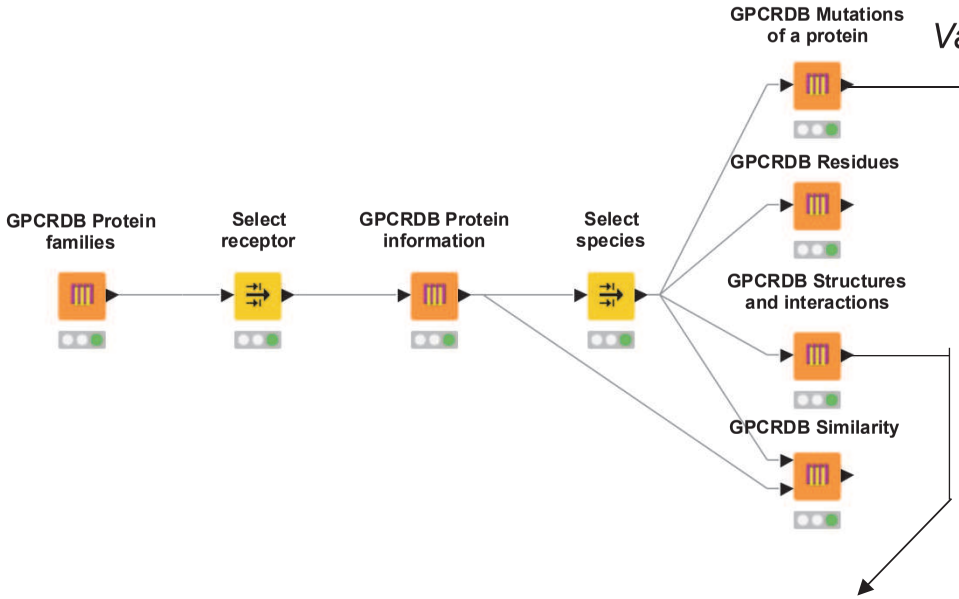


Row ID	PDB c...	Ligand	Seque...	Amino...	Generi...	Interaction type
Row0	3RZE	doxepin	107	D	3.32x32	polar (charge-charge)
Row1	3RZE	doxepin	108	Y	3.33x33	aromatic (edge-to-f...
Row2	3RZE	doxepin	112	T	3.37x37	polar (hydrogen bond)
Row3	3RZE	doxepin	158	W	4.56x57	aromatic (face-to-ed...
Row4	3RZE	doxepin	428	W	6.48x48	aromatic (face-to-ed...
Row5	3RZE	doxepin	431	Y	6.51x51	hydrophobic
Row6	3RZE	doxepin	432	F	6.52x52	aromatic (edge-to-f...
Row7	3RZE	doxepin	435	F	6.55x55	aromatic (face-to-ed...

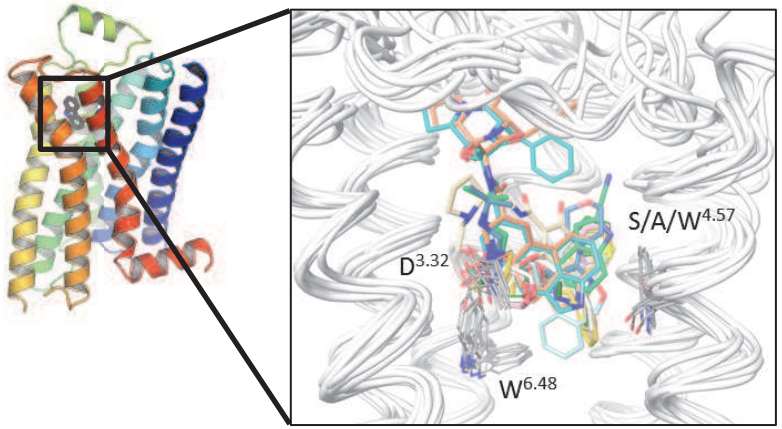
Vass et al. (2017) *J Chem Info Model* 57: 115

GPCRdb nodes: Structural mapping of GPCR mutation data

GPCRdb nodes to integrate structural GPCR data



X-ray structures and protein-ligand interactions

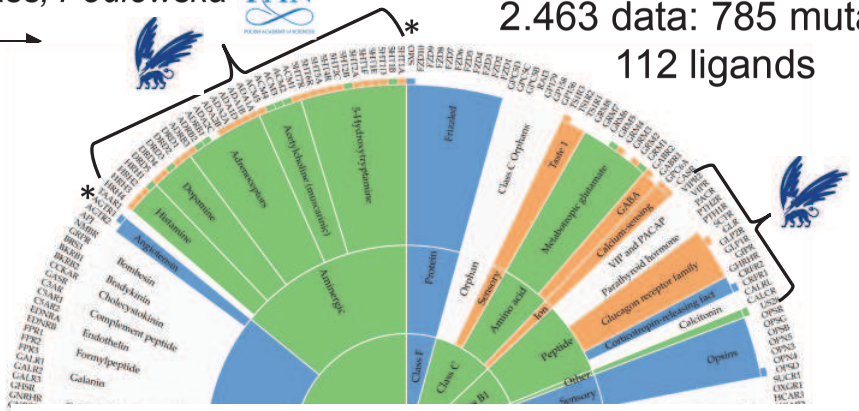


35 aminergic receptors:
6.553 data: 1.284 mutants, 538 ligands

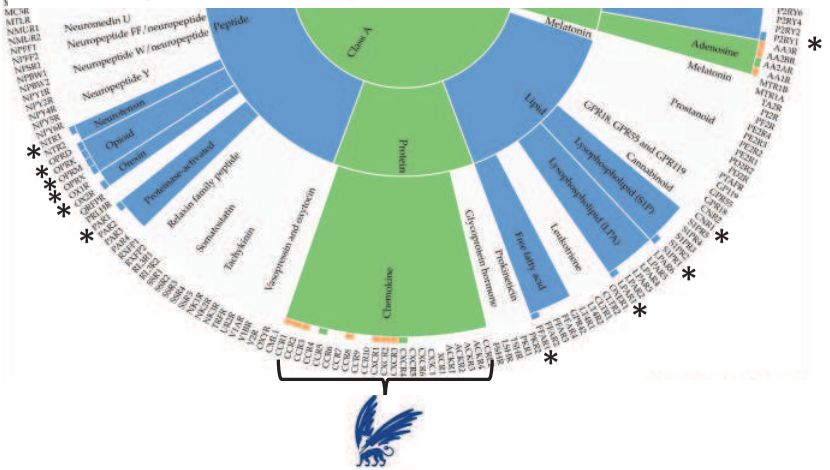
de Graaf (2016) Pharmacol Rev 68: 954

Vass, Podlewska PAN

14 class B1 GPCRs:
2.463 data: 785 mutants
112 ligands



http://gpcrdb.org/mutations/statistics



12 chemokine receptors:
2.704 data: 338 mutants, 76 ligands

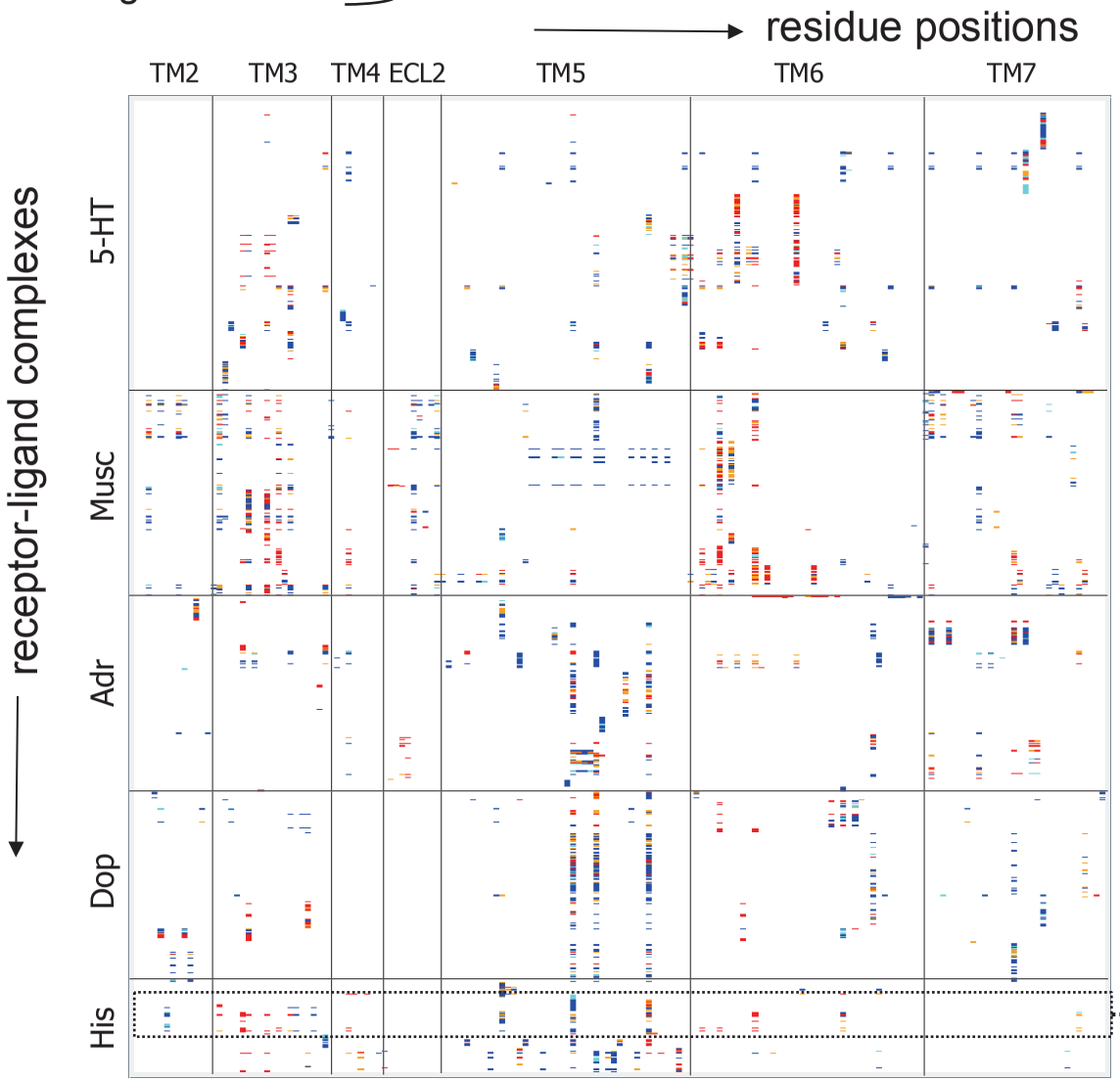
Arimont, Sun et al. (2017) J Med Chem

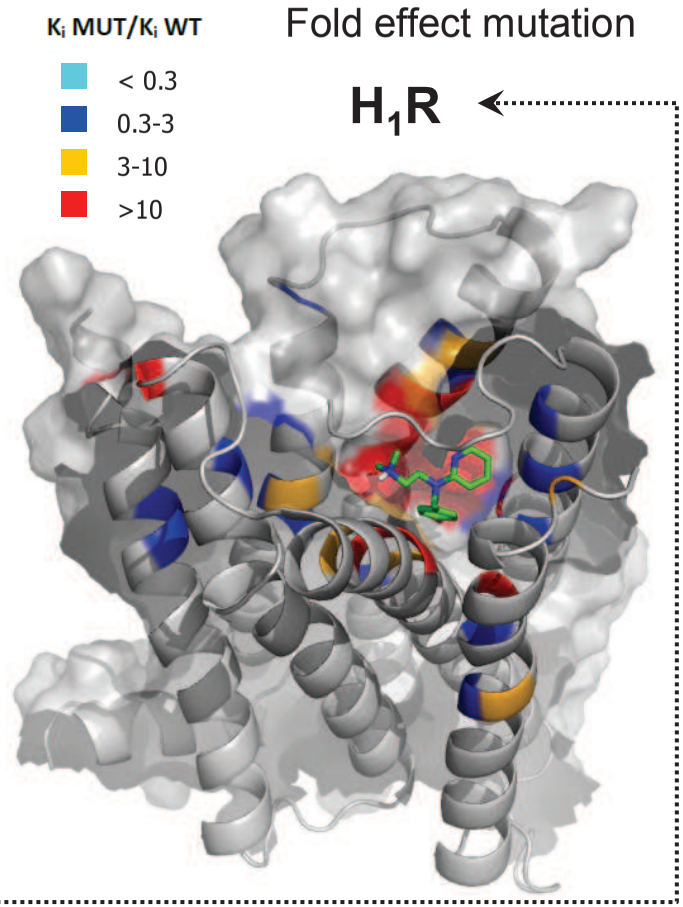
Sparse mutant-ligand matrix to 3D binding mode map?

6.553 mutation data
1.281 mutants
35 receptors
538 ligands

1-15% coverage
mutant-ligand matrix

M. Vass





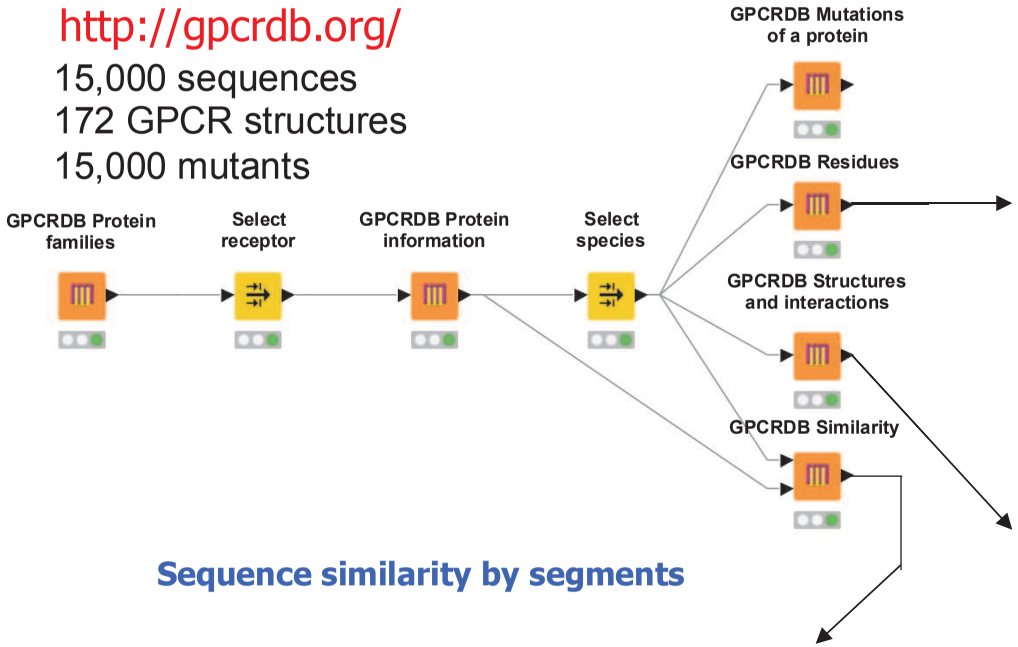
Vass, Podlewska et al., *In preparation*

GPCRdb nodes: GPCR-ligand binding site prediction

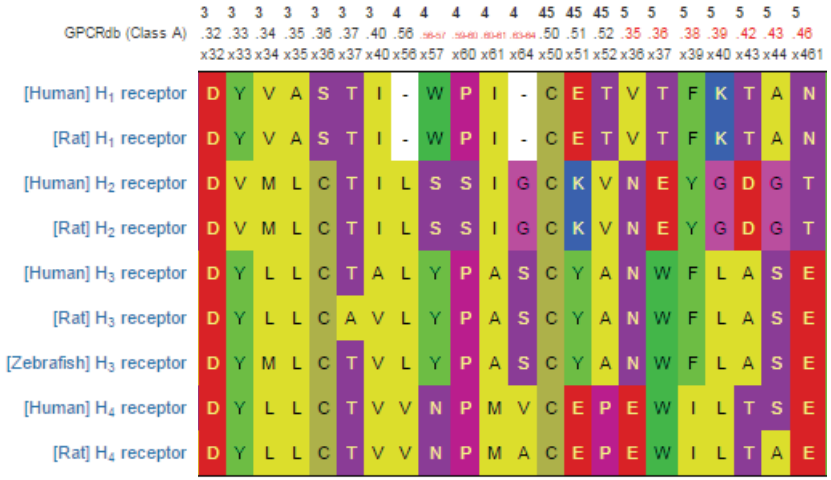
GPCRdb nodes to integrate structural GPCR data

http://gpcrdb.org/

15,000 sequences
172 GPCR structures
15,000 mutants



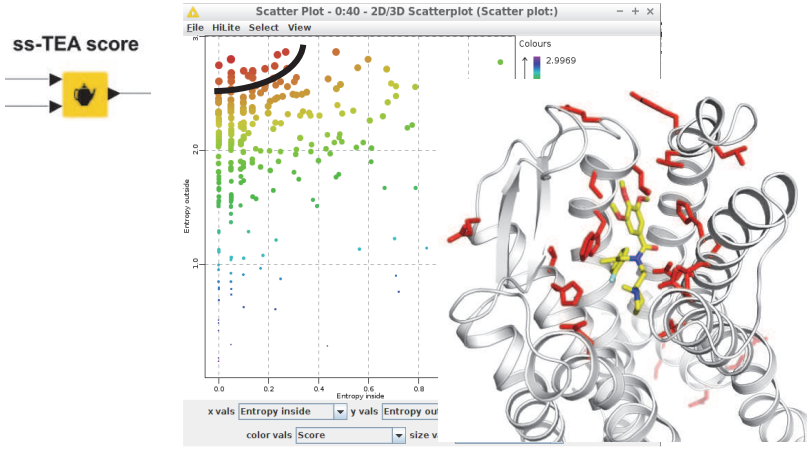
residue information and sequence alignment



Sequence similarity by segments

	1	2	3	4	5	6	7	8	9
1. [Human] α_{1A}-adrenoceptor	-	73	63	41	40	40	40	38	41
2. [Human] α_{1B}-adrenoceptor	89	-	69	41	42	41	41	40	42
3. [Human] α_{1D}-adrenoceptor	81	83	-	38	40	39	41	40	43
4. [Human] α_{2A}-adrenoceptor	63	66	61	-	71	73	41	36	36
5. [Human] α_{2B}-adrenoceptor	62	65	62	84	-	73	39	35	37
6. [Human] α_{2C}-adrenoceptor	62	64	62	83	84	-	37	34	37
7. [Human] β_1-adrenoceptor	61	62	64	61	62	60	-	65	61
8. [Human] β_2-adrenoceptor	62	62	63	56	57	56	78	-	55
9. [Human] β_3-adrenoceptor	60	60	61	55	57	58	74	69	-

ss-TEA node: GPCR binding site prediction



McGuire (2017) *J Chem Info Model* 57: 115

Sanders (2011) *BMC Bioinformatics* 12: 332

Amino acid entropy based CXCR7 binding site prediction

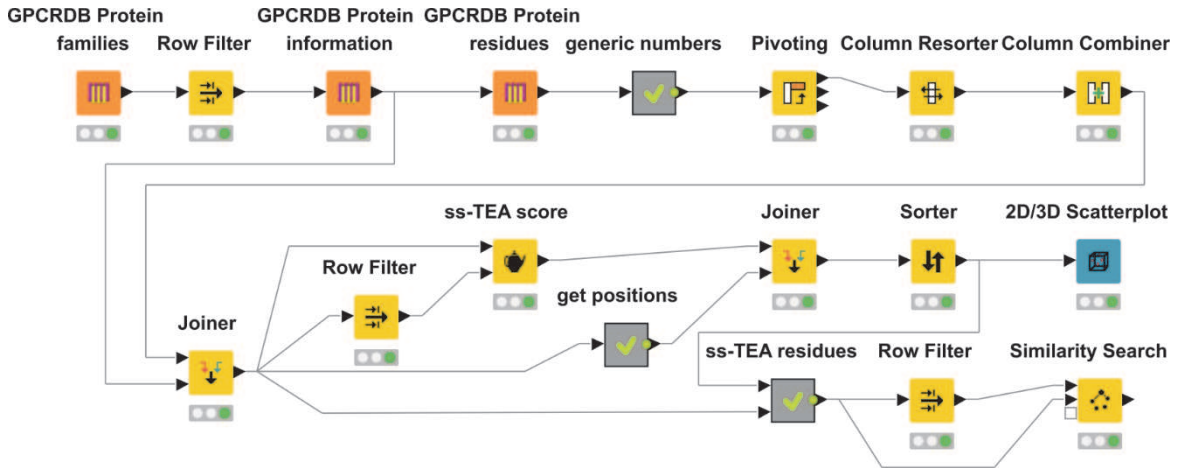
Sub-family specific two-entropy analysis (ssTEA):

- low entropy within a subfamily
- high entropy other subfamilies

} sub-family specific {

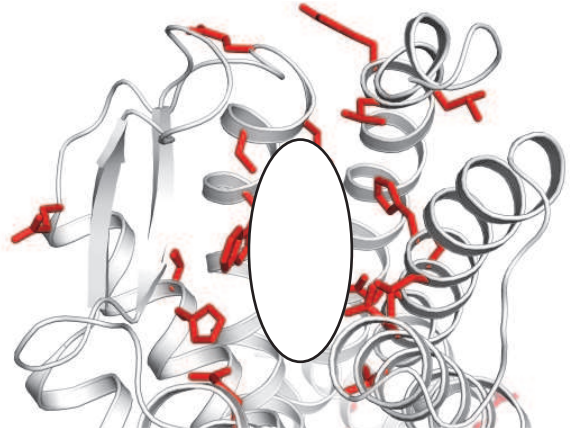
- functional role?
- **ligand binding**?

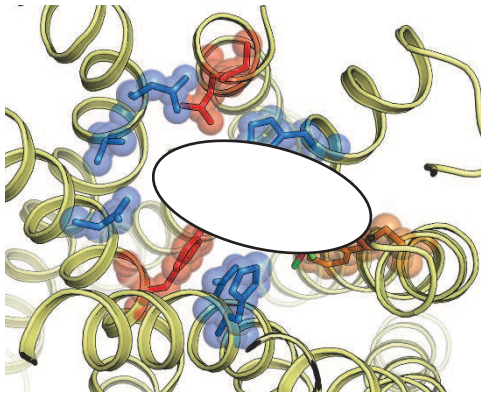
M. Arimont

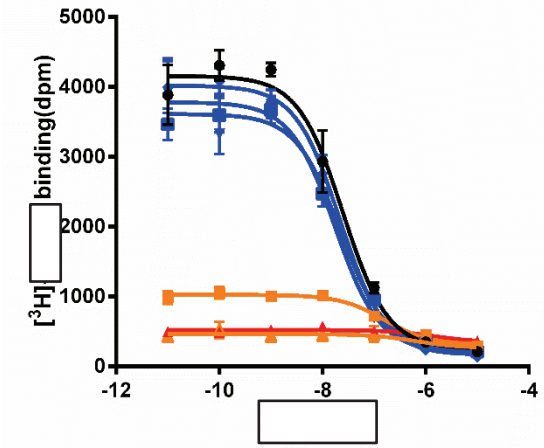


Receptor **binding residues?**

Exp. validated receptor **radioligand** binding mode

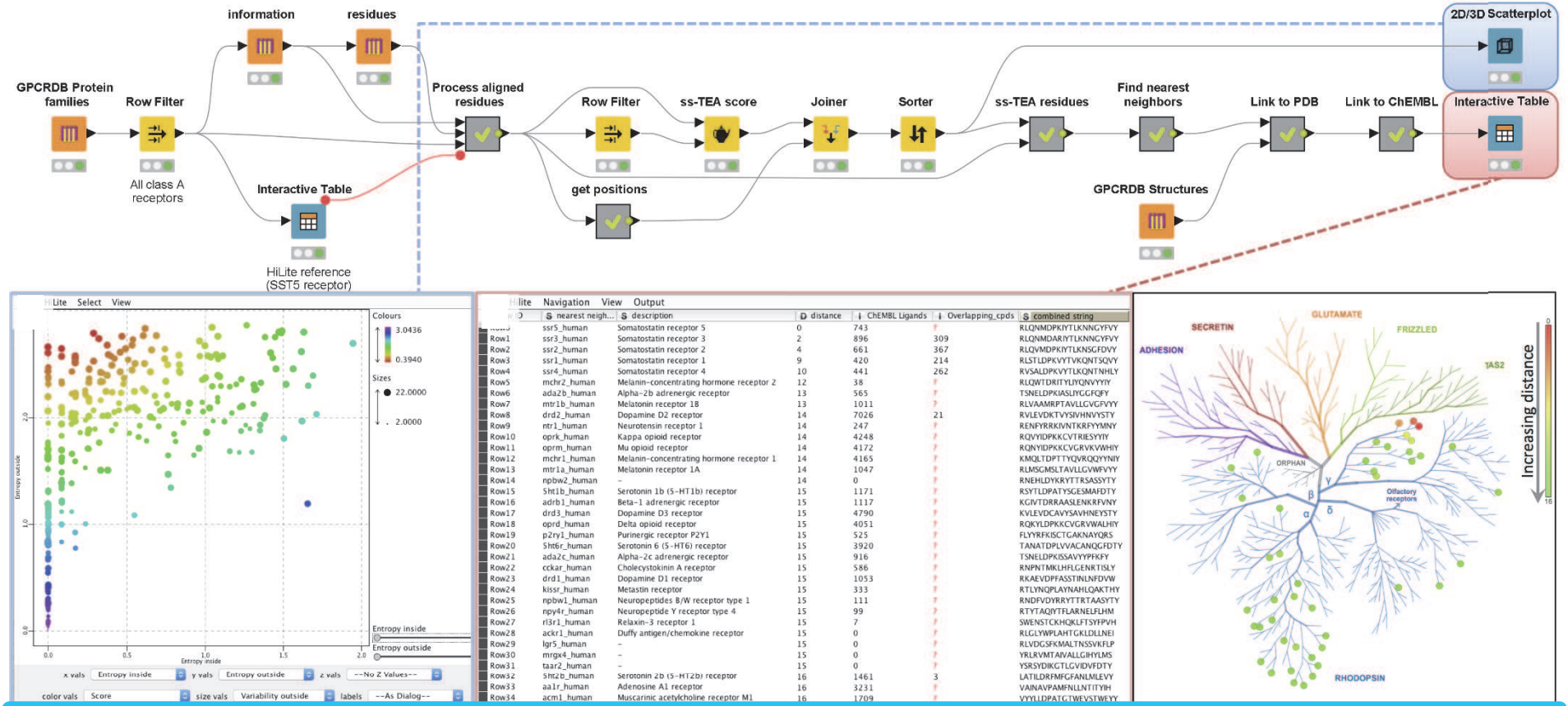




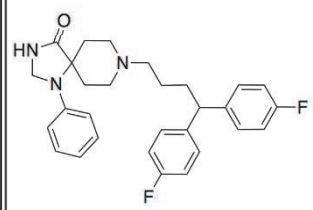


Arimont, Vass et al. (2018) *In preparation*

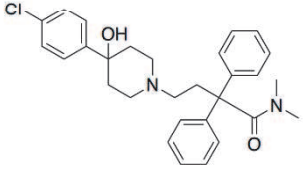
Ligand repurposing using hot spot residue similarities



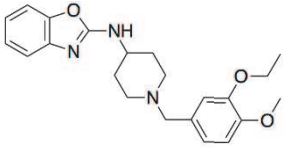
	1x59	2x60	2x63	2x64	3x29	3x32	4x39	4x43	5x40	5x42	5x43	5x59	5x64	6x55	6x58	7x32	7x34	7x35	7x38	7x42
SST_5R	R	L	Q	N	M	D	P	K	I	Y	T	L	K	N	N	G	Y	F	V	Y
SST_3R	R	L	Q	N	M	D	A	R	I	Y	T	L	K	N	N	G	Y	F	V	Y
SST_2R	R	L	Q	V	M	D	P	K	I	Y	T	L	K	N	S	G	F	D	V	Y
SST_1R	R	L	S	T	L	D	P	K	V	Y	T	V	K	Q	N	T	S	Q	V	Y
SST_4R	R	V	S	A	L	D	P	K	V	Y	T	L	K	Q	N	T	N	H	L	Y
D_2R	R	V	L	E	V	D	K	T	V	Y	S	I	V	H	N	V	Y	S	T	Y
μ opioid R	R	Q	N	Y	I	D	P	K	K	C	V	G	R	V	K	V	W	H	I	Y
$5\text{-}HT_{2B}R$	L	A	T	I	L	D	R	F	M	F	G	F	A	N	L	M	L	E	V	Y



K_i SST_5R = 4510 nM
K_i D_2R = 0.67 nM



K_i SST_5R = 4520 nM
K_i μ opioid R = 0.53 nM

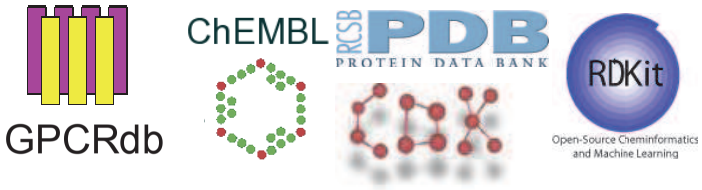


K_i SST_5R = 121 nM
K_i $5\text{-}HT_{2B}R$ = 41 nM

Kooistra, Vass, Verhoeven, de Graaf et al. ***ChemMedChem***, 2018, 13: 614

Structural Cheminformatics Workflows

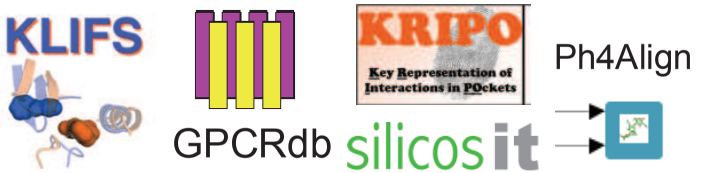
- Structural coverage of chemical GPCR ligand space
- Structure-based GPCR mutation data mapping



- Sequence-based GPCR binding site prediction
  ligand repurposing

- Structural interaction-based ligand scaffold hopping

- Structure-based kinase-GPCR ligand repurposing



Vass et al. *Tr Pharm Sci* 2018, doi:10.1016/j.tips.2018.02.004
Kooistra et al. *ChemMedChem*, 2018, 13: 614

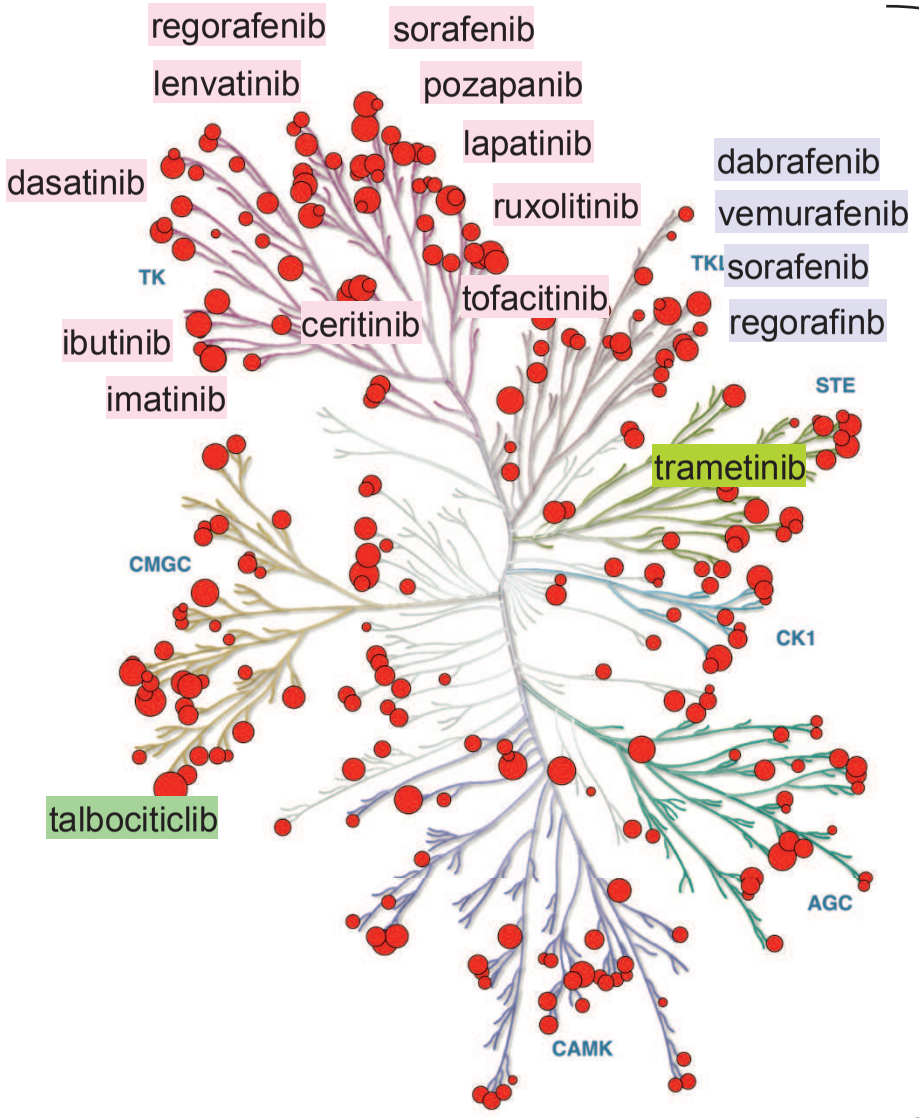
Vass et al. *In preparation* 2018
McGuire et al. *J Chem Info Model* 2017

- Structure-based virtual screening - SAR - SKR

docking - IFP - RAMD

Navigating structural protein-drug interaction space

Marketed drugs targeting the catalytic domain kinase



Kooistra et al. (2016) Nucl. Acids Res. 44: D365

KLIFS

klifs.vu-compmedchem.nl

Kinase
Ligand
Interaction
Fingerprint
Structure
database

A.J. Kooistra

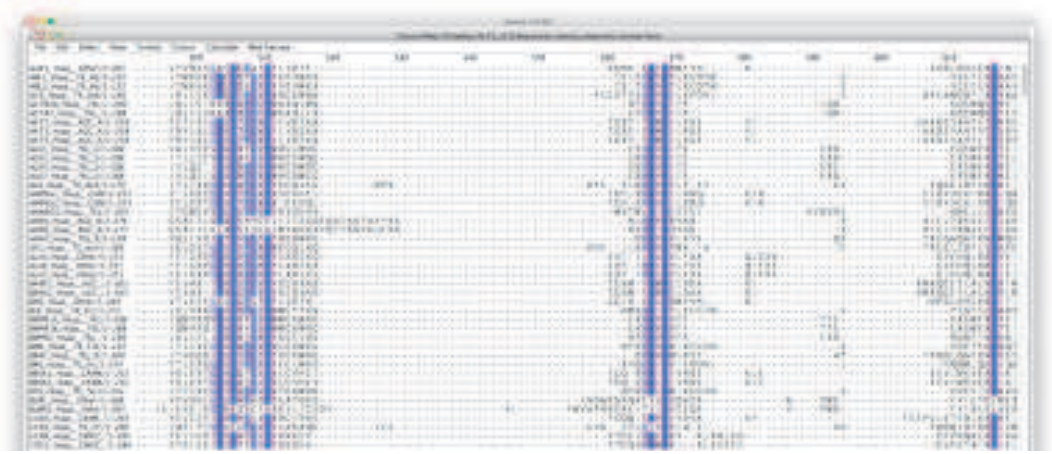
G.K. Kanev

O.P.J. van Linden

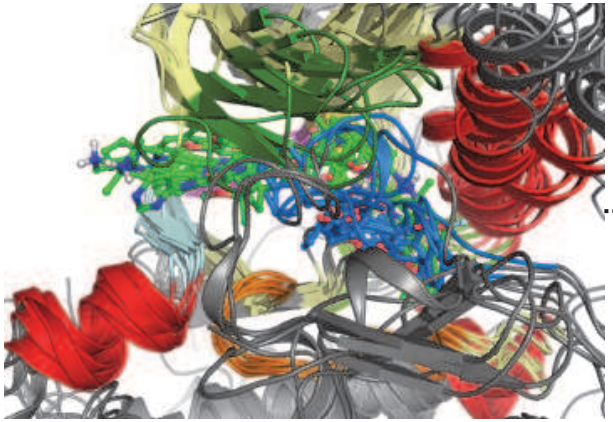
4.107 PDB entries
282 of 518 kinases
3.088 unique complexes
2.601 unique ligands

Van Linden & Kooistra et al. (2014) *J Med Chem* 57: 249

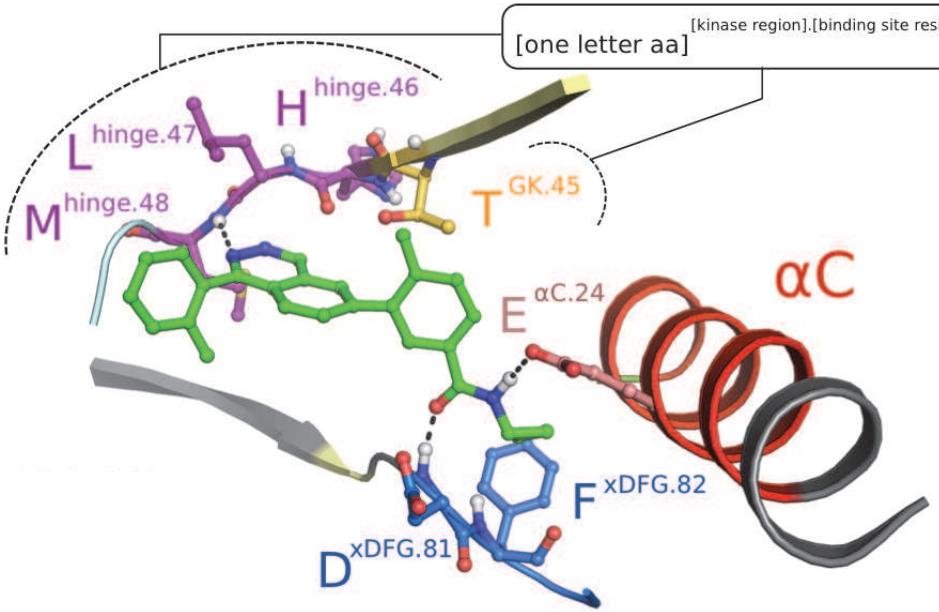
From sequence to structure to binding site alignment ...



Sequence alignment



Structural alignment



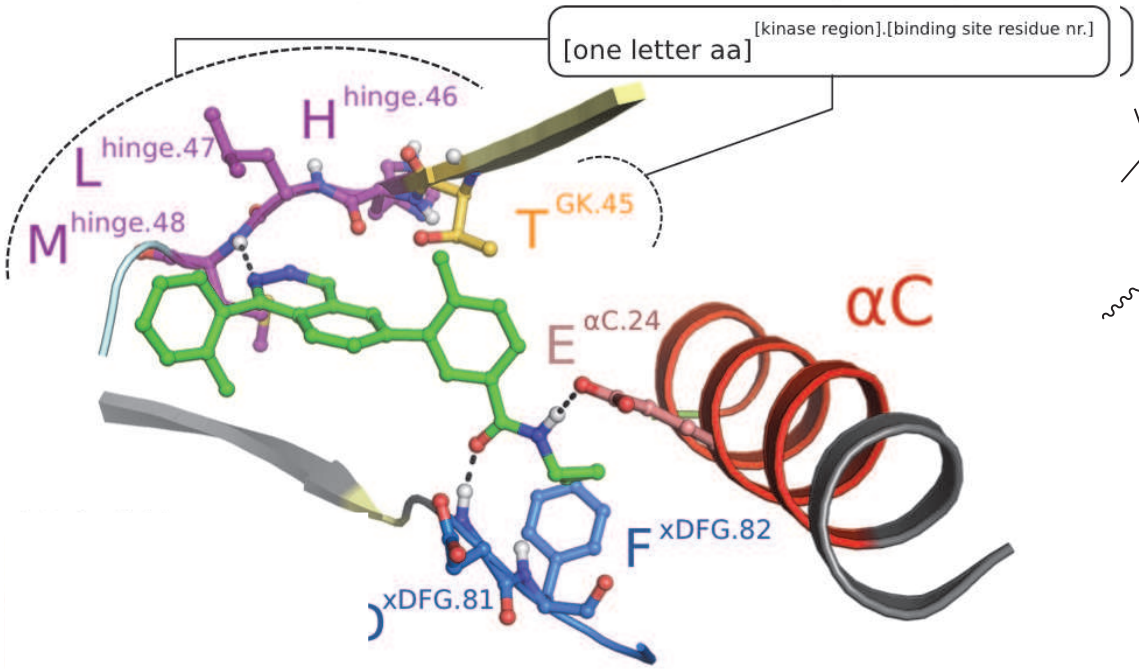
Consistent **binding site residue annotation** based on alignment reference kinase structures

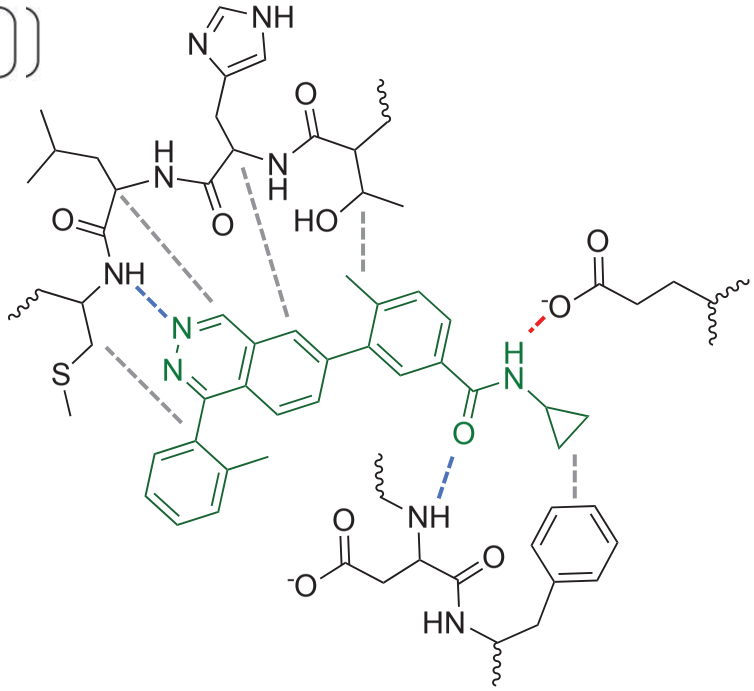
Kooistra et al. (2016) Nucl. Acids Res. 44: D365

Van Linden & Kooistra et al. (2014) *J Med Chem* 57: 249

From alignment to comparative interaction analysis

Consistent **binding site residue annotation** based on alignment 7.339 kinase structures





85 binding site residues interacting with ligand (in ≥1 kinase structure)

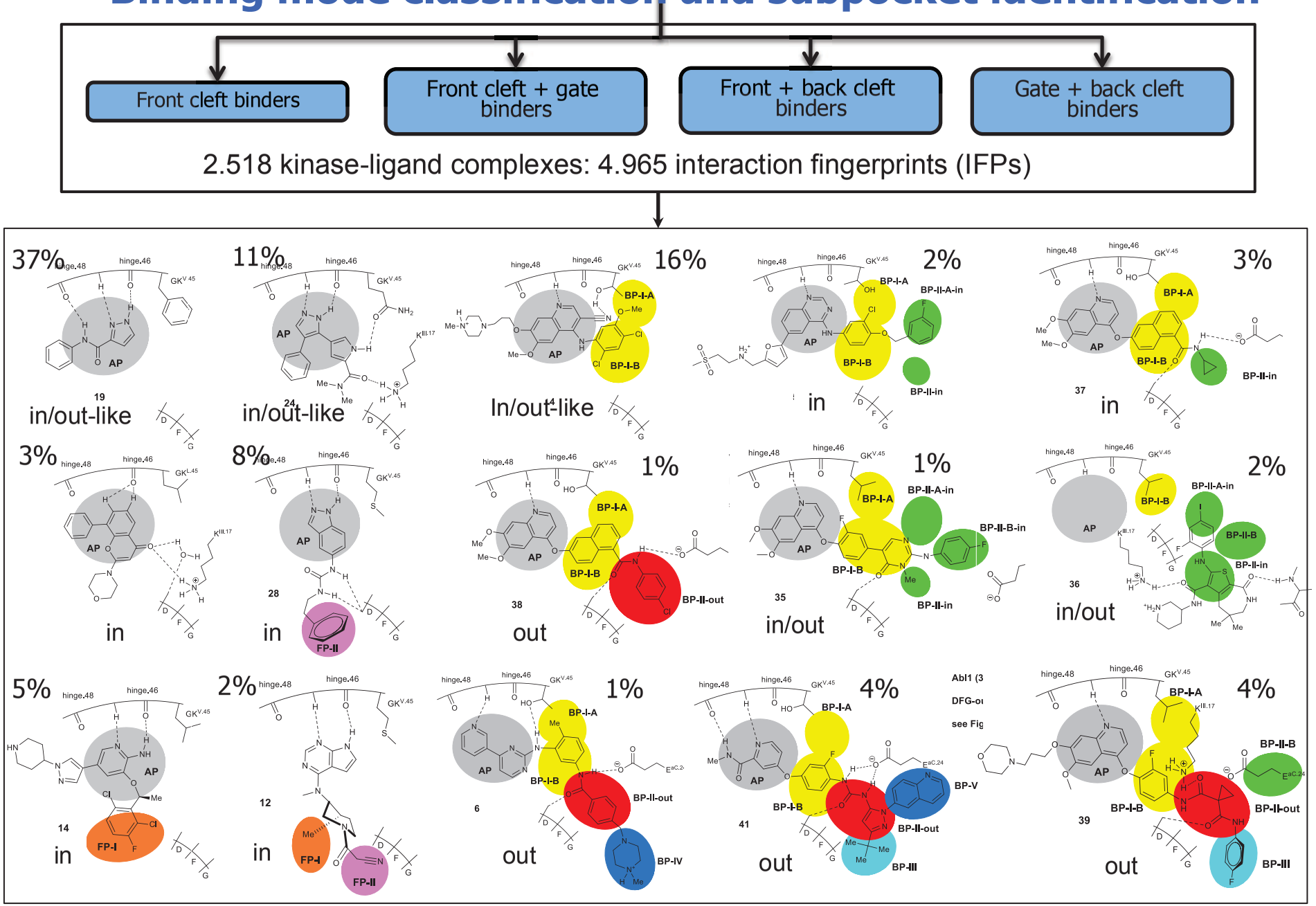
1. Hydrophobic 2. aromatic face to face 3. aromatic face to edge 4. H-bond (ligand HBD-protein HBA)
5. H-bond (ligand HBA-protein HBD) 6. ionic (ligand cation-protein anion) 7. ionic (ligand anion–protein cation)

6.200 kinase-ligand Interaction FingerPrints (**IFPs**) derived from 3.531 PDB entries

Kooistra et al. (2016) Nucl. Acids Res. 44: D365

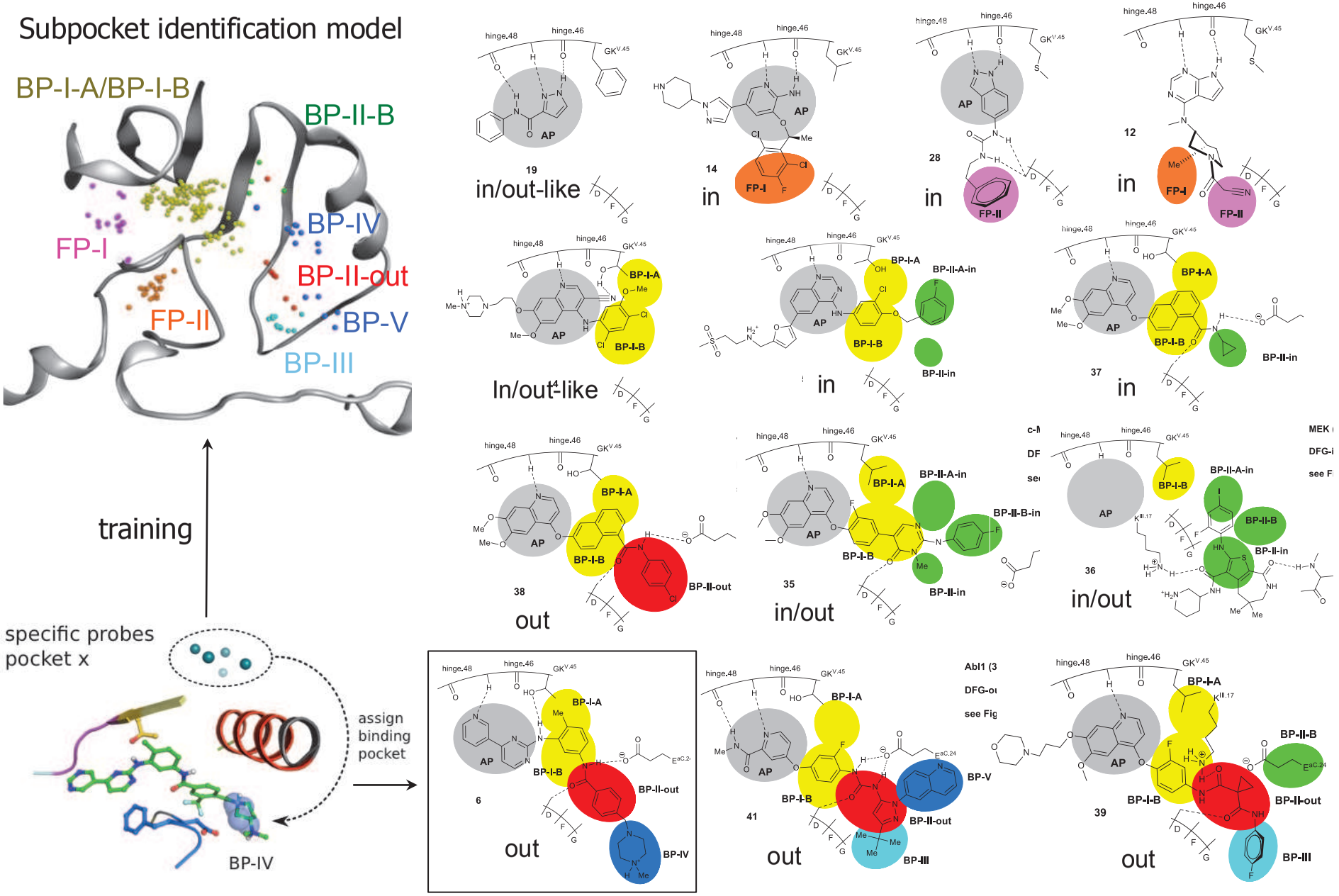
Van Linden & Kooistra et al. (2014) *J Med Chem* 57: 249

Binding mode classification and subpocket identification



Van Linden & Kooistra et al. (2014) *J Med Chem* 57: 249

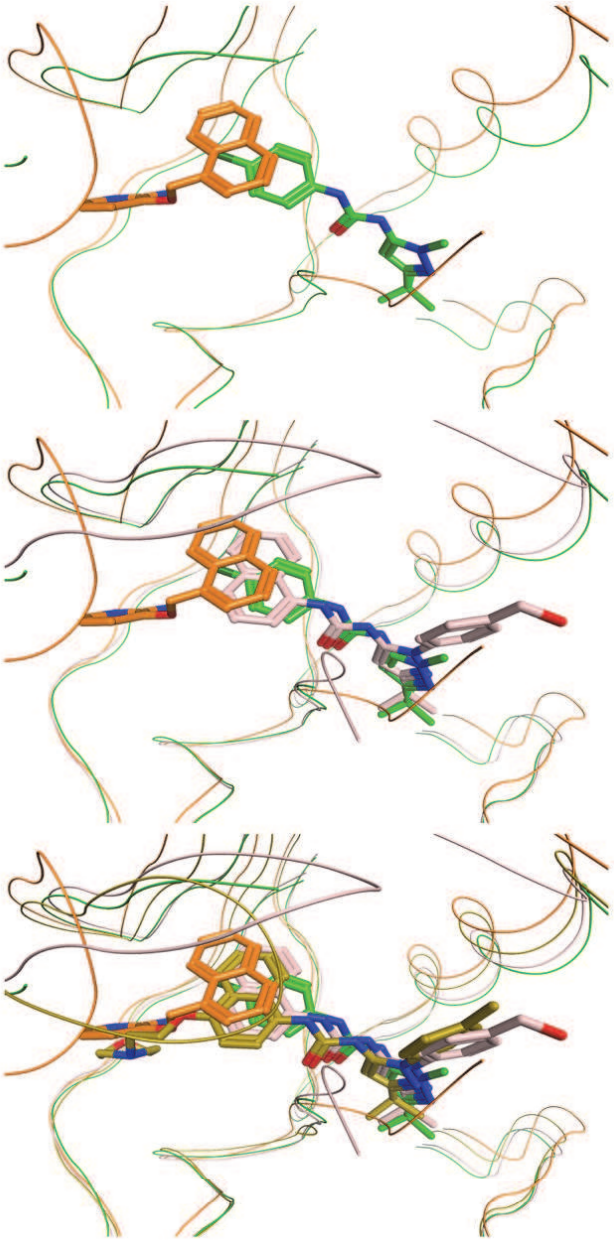
Binding mode classification and subpocket identification

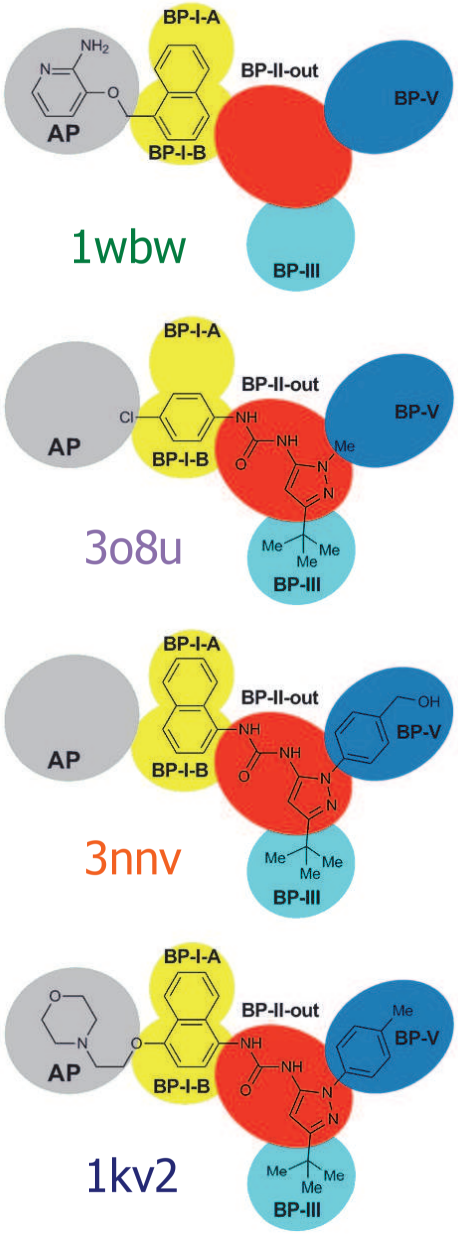


Kooistra et al. (2016) Nucl. Acids Res. 44: D365

Van Linden & Kooistra et al. (2014) *J Med Chem* 57: 249

KLIFS: Structure-based medicinal chemistry data mining

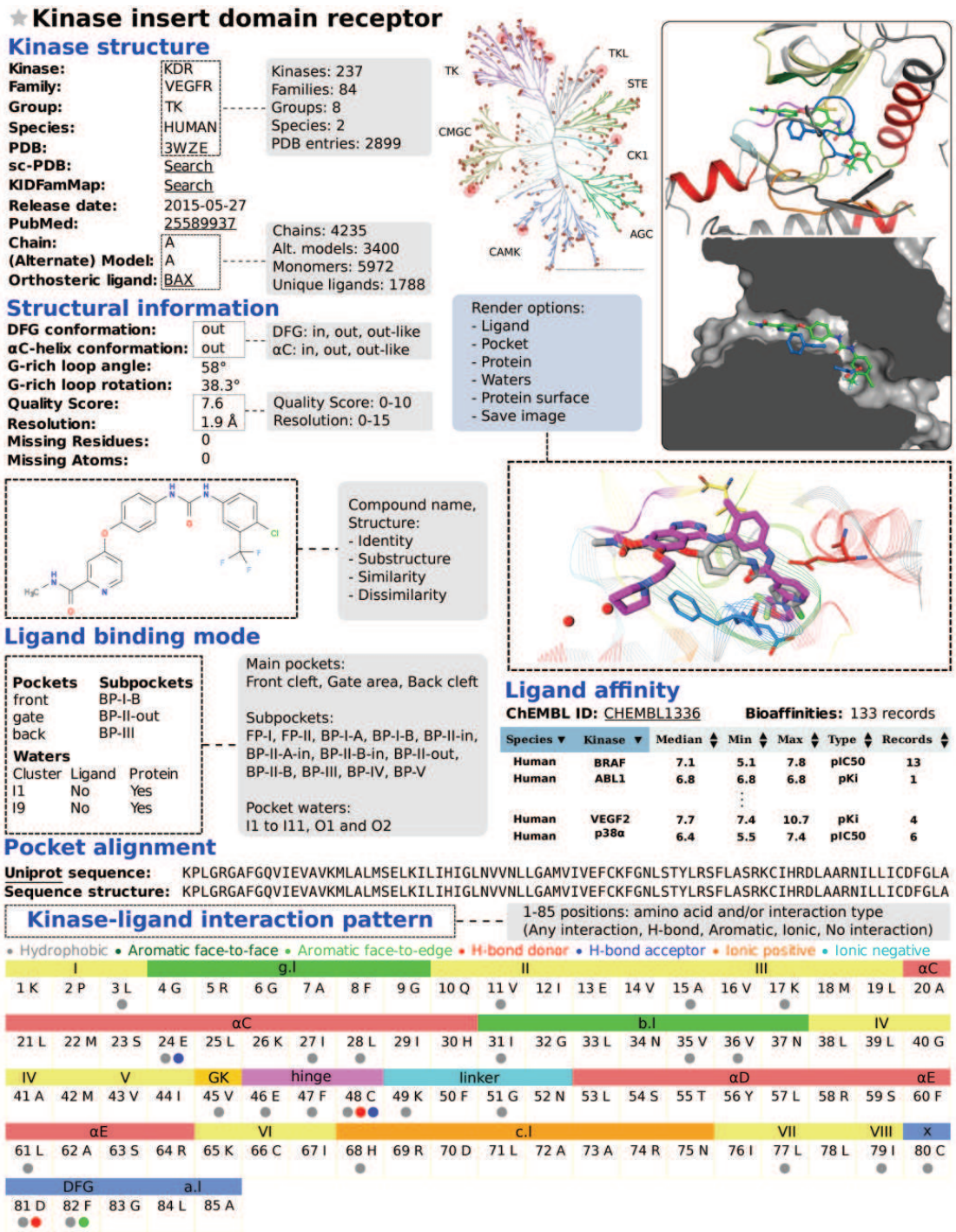




Combining:
- ligand chemical similarity
- kinase-ligand IFP similarity analyses

Structure-affinity/selectivity/ kinetics relationships in 3D

Recipees for **structure-based ligand design**



KLIFS website:

klifs.vu-compmedchem.nl

Weekly updated, in sync with PDB

Kooistra et al. **2016** Nucl. Acids Res. 44: D365

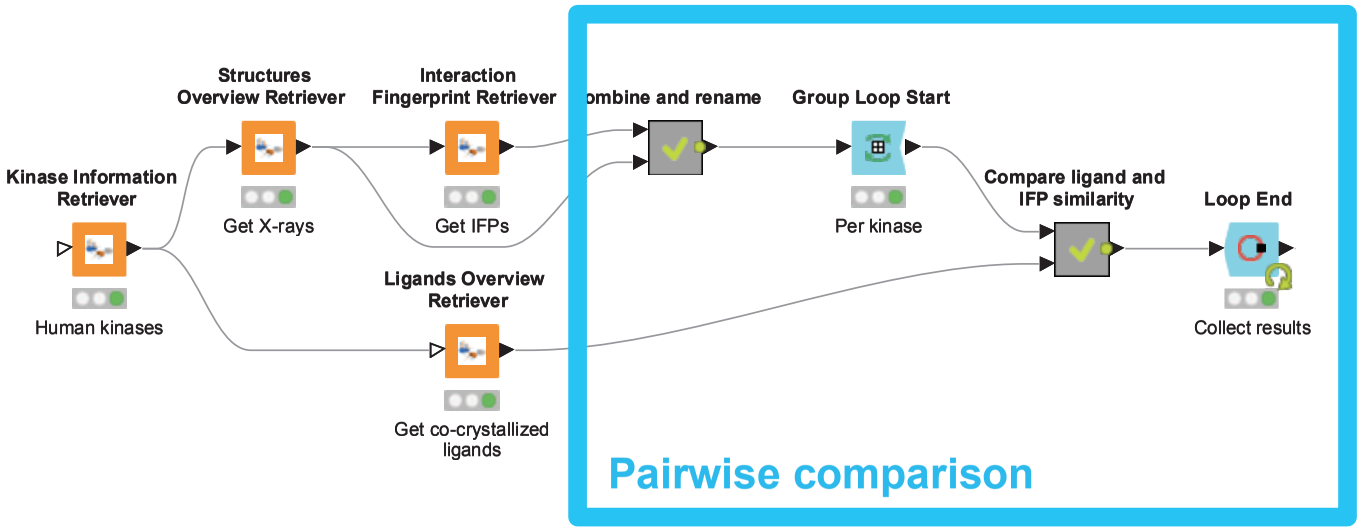
KLIFS KNIME nodes:

3d-e-chem.github.io

KNIME
Data Analytics Made Easy

McGuire, Verhoeven. Vass, et al. **2017** *JCIM* 57: 115
Kooistra, Vass et al. **2018** *ChemMedChem* 13, 614

Structural interaction based ligand scaffold hopping



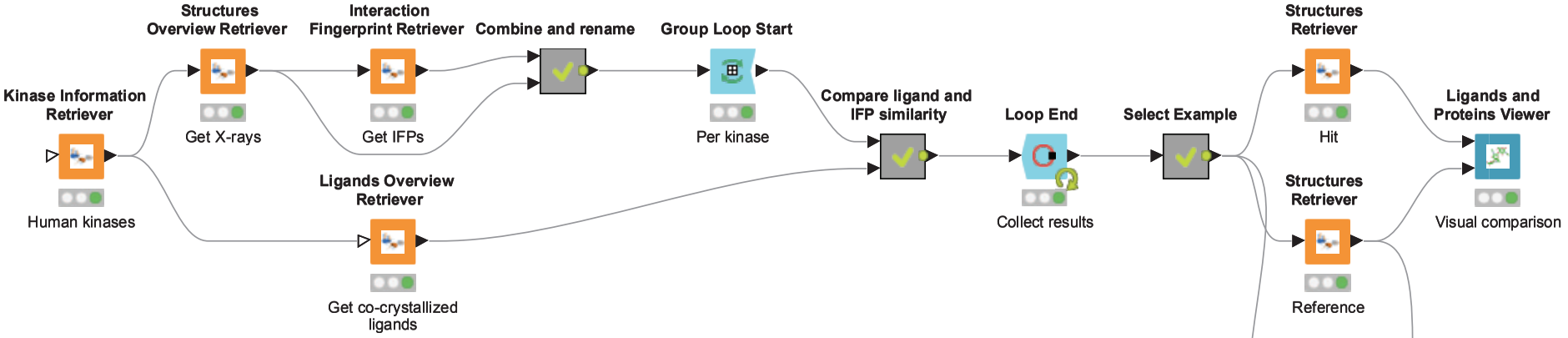
Goal:
- Dissimilar ligands
- Similar interactions

Method:
- High IFP (≥0.8)
- Low ECFP-4 (<0.4)

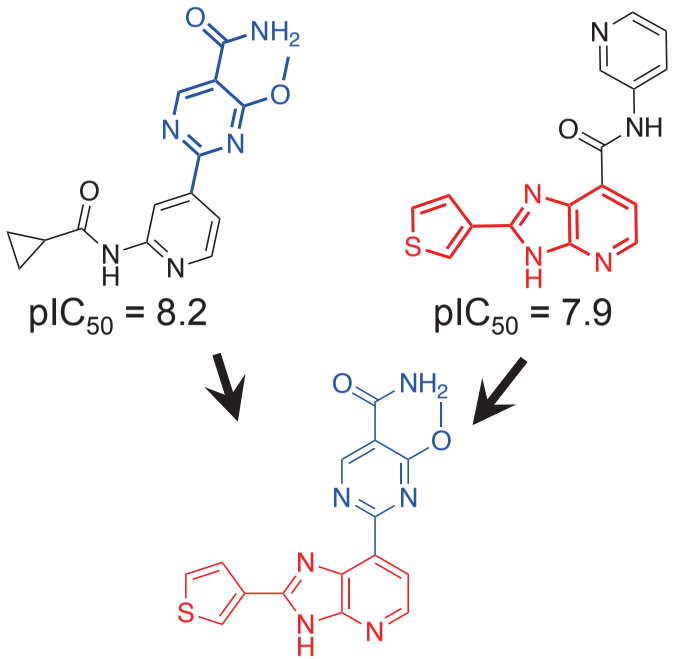
Table "default" – Rows: 21868 | Spec – Columns: 10 | Properties | Flow Variables

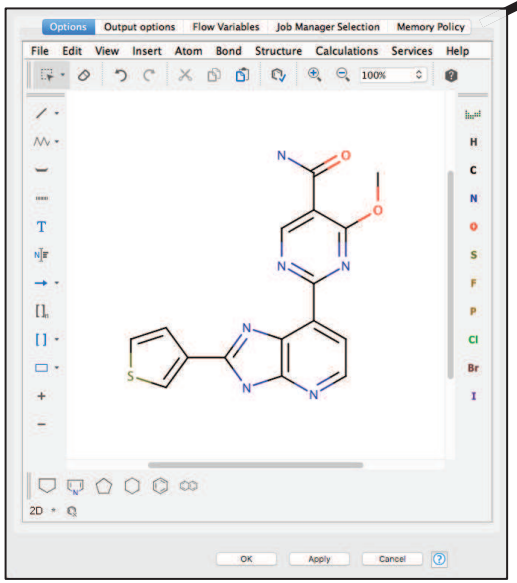
Row ID	Structure ID	PDB	Chain	IFP Similarity	Reference ...	ECFP-4 similarity	Hit str...	Hit SMILES
1982_Row...	1982	4dhf	B	0.8		0.129	7221	
689_Row0...	689	4dit	A	0.773		0.154	632	
689_Row0...	689	4dit	A	1		0.188	634	

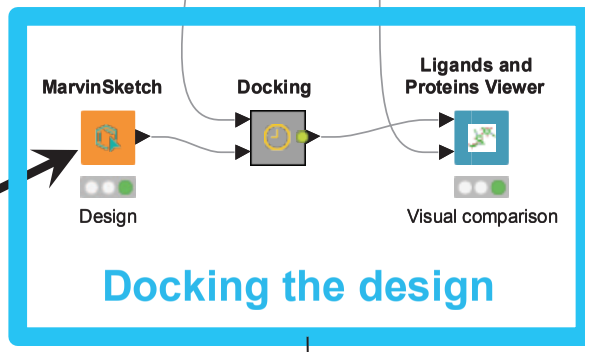
Structural interaction based ligand design



Glycogen synthase kinase 3-beta







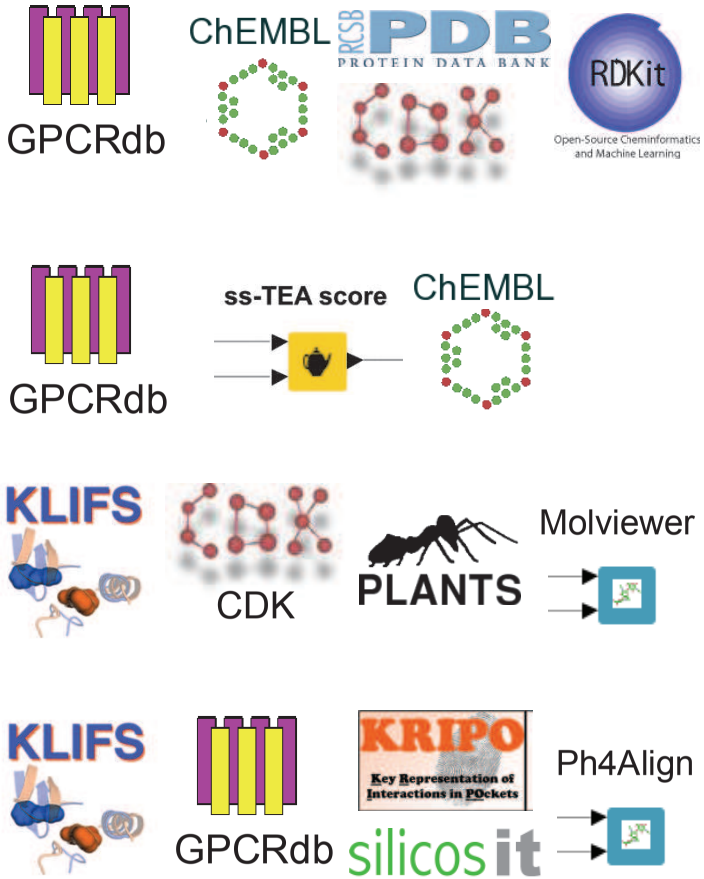
- Genetic algorithm
- Ant colony optimization
- ChemPLP scoring
- Free academic license

PLANTS

Korb/Exner et al. (2009) *J Chem Model Inf* 49: 84

Structural Cheminformatics Workflows

- Structural coverage of chemical GPCR ligand space
- Structure-based GPCR mutation data mapping
- Sequence-based GPCR binding site prediction
  ligand repurposing
- Structural interaction-based ligand scaffold hopping
- Structure-based kinase-GPCR ligand repurposing



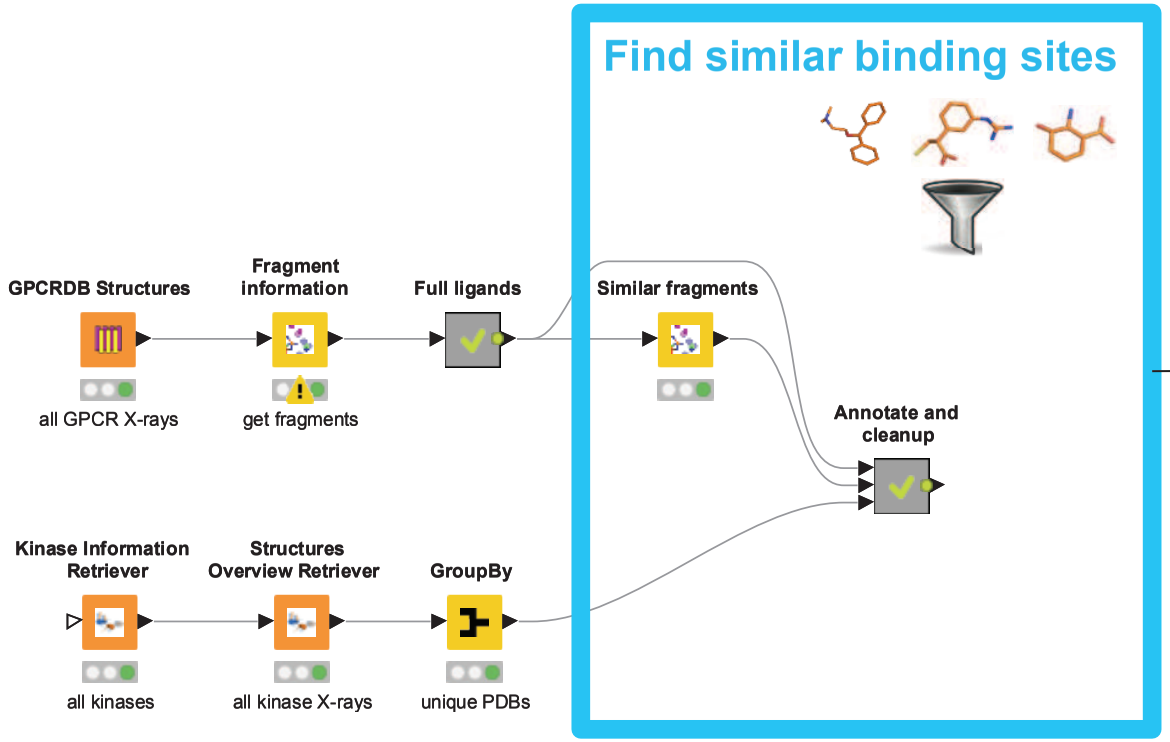
Vass et al. *Tr Pharm Sci* 2018, doi:10.1016/j.tips.2018.02.004
Kooistra et al. *ChemMedChem*, 2018, 13: 614

Vass et al. *In preparation* 2018
McGuire et al. *J Chem Info Model* 2017

- Structure-based virtual screening - SAR - SKR

docking - IFP - RAMD

From protein similarity to ligand repurposing

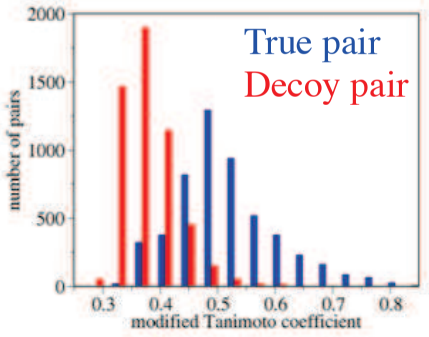


KRIPO

- Ph4 features
  - Hydrophobic
  - Donor/acceptor
  - Aromatic
  - Ionic
- Fingerprint
  - Fast comparison

KRIPOdb web service

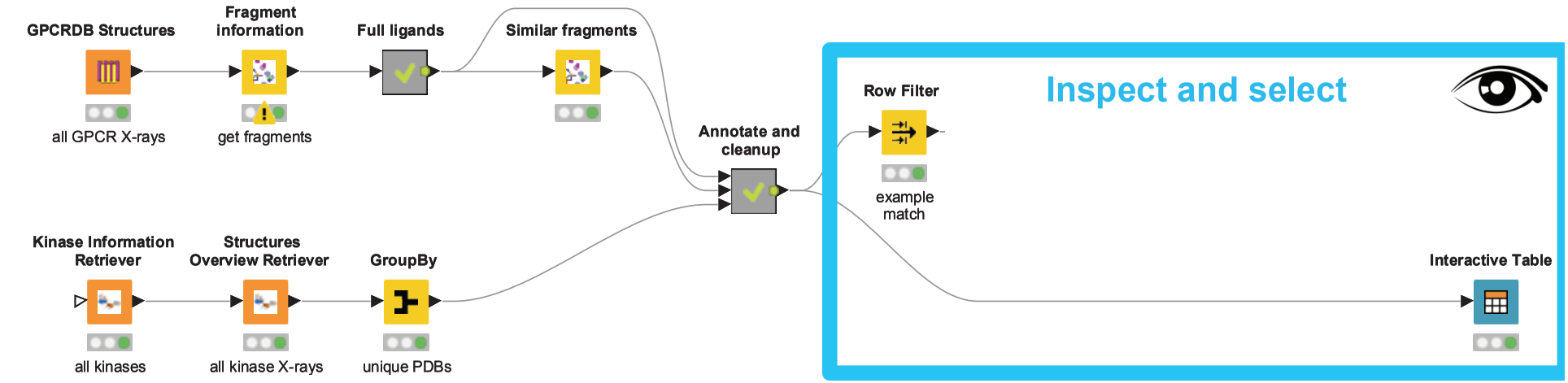
- 59 876 PDBs
- 569 907 Fingerprints
- 569 907 Ph4s
- 162 billion scores



- Cutoff 0.5

McGuire & Verhoeven & Vass, et al. (2017) *J Chem Info Model* 57: 115 Wood (2012) *J Chem Info Model* 52: 2031

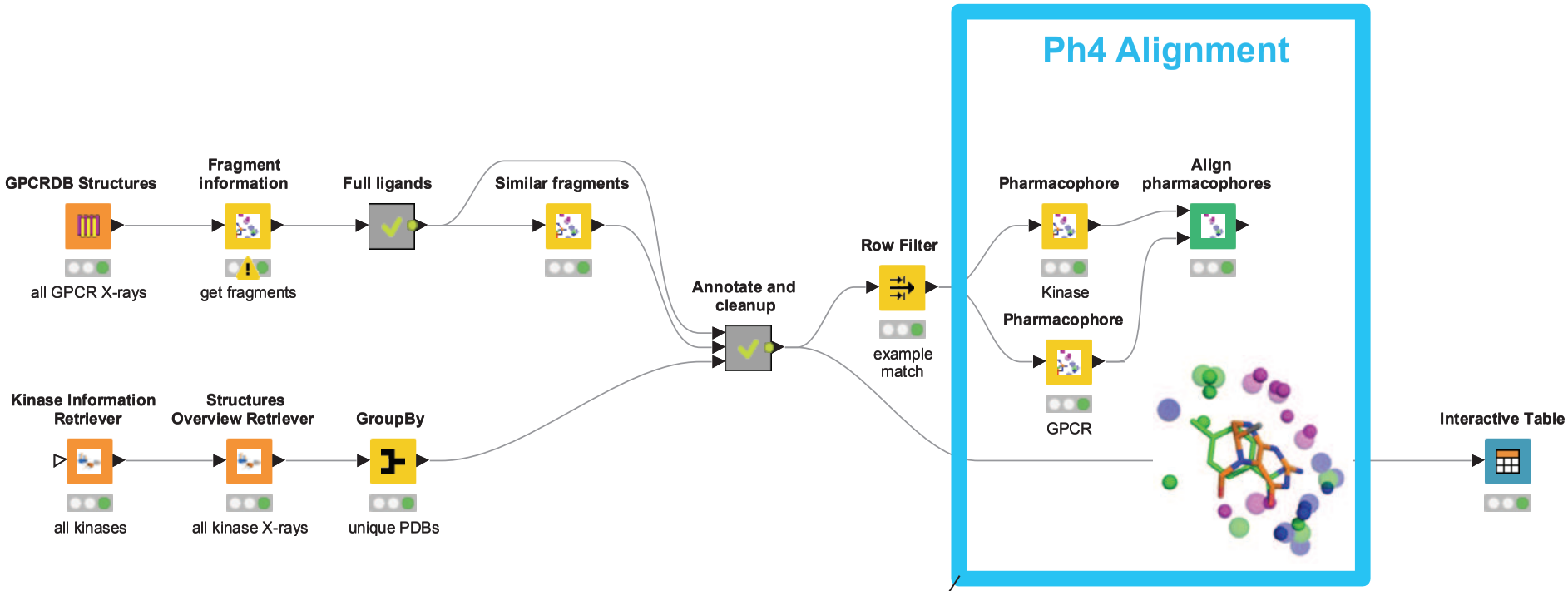
From protein similarity to ligand repurposing



Row ID	GPCR_frag_id	kinase_frag_id	score	GPCR_ligand	GPCR_name	kinase_lig...	kinase_name
1656	5uig_8D1_frag1	4eyj_N61_frag1	0.537		Adenosine receptor A2a,Soluble c...		Mitogen-activated protein kinase 13
1657	5uig_8D1_frag1	5ljj_AD5_frag1	0.535		Adenosine receptor A2a,Soluble c...		Dual specificity protein kinase TTK
1658	5uig_8D1_frag1	5j9y_6HL_frag1	0.534		Adenosine receptor A2a,Soluble c...		Epidermal growth factor receptor

Kooistra, Vass, de Graaf et al. ***ChemMedChem***, 2018, doi: 10.1002/cmdc.201700754

From protein similarity to ligand repurposing

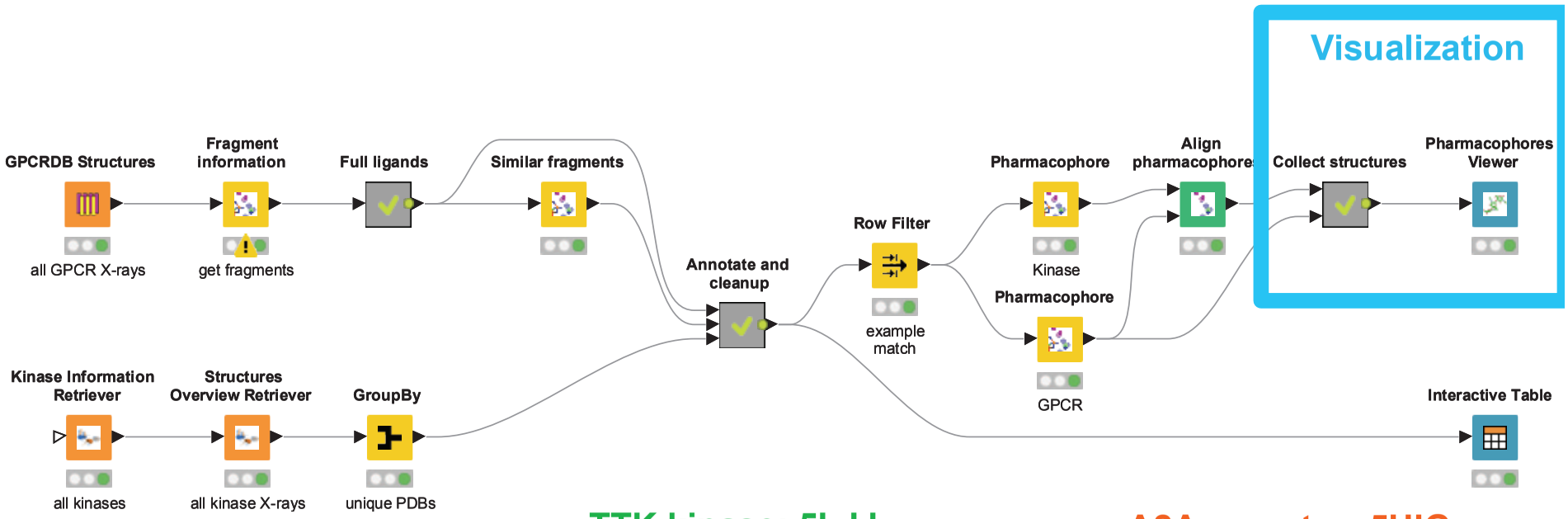


Steps:

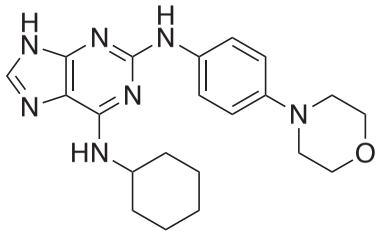
- Obtain fragments Ph4s from web service
- Independent alignment
  - Clique detection algorithm
- Apply rotational matrix to PDB

Kooistra, Vass, de Graaf et al. ***ChemMedChem***, 2018, 13: 614

From protein similarity to ligand repurposing



TTK inhibitor reversine



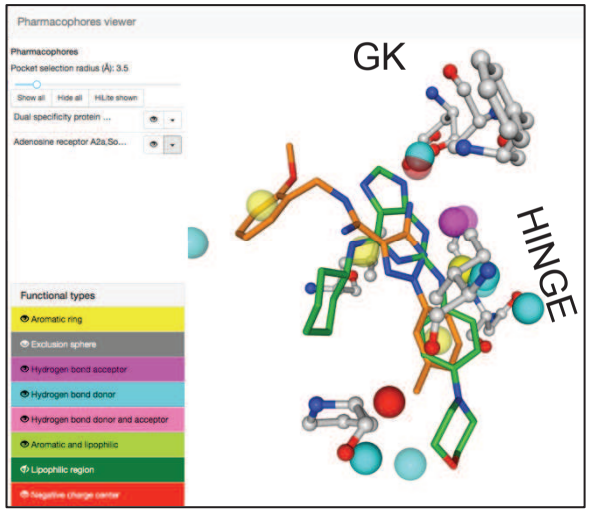
TTK pIC_{50} = 8.6

$A_{2A}R$ inhibition 30% (10μM)

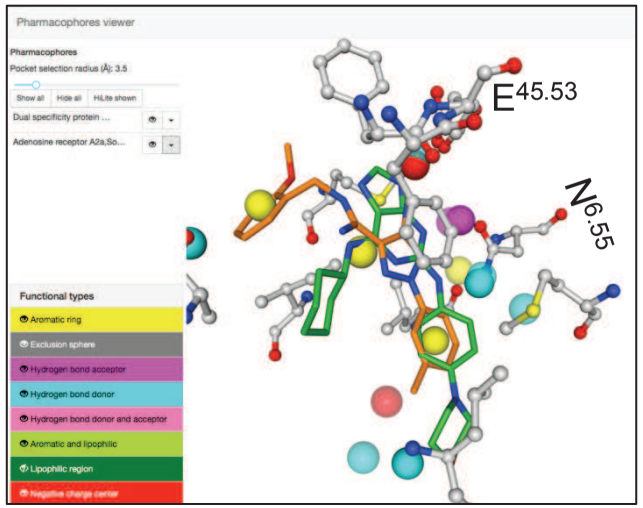
A_3R pK_i = 6.2

Perreira et al. (2005) *J Med Chem* 48: 4910

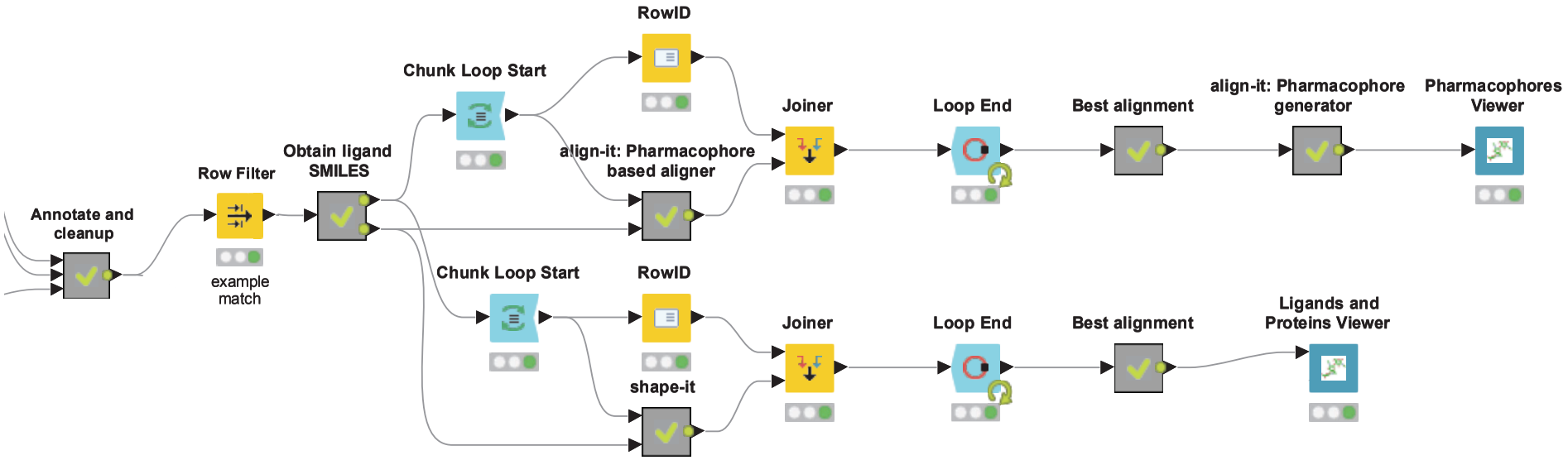
TTK kinase: 5LJJ

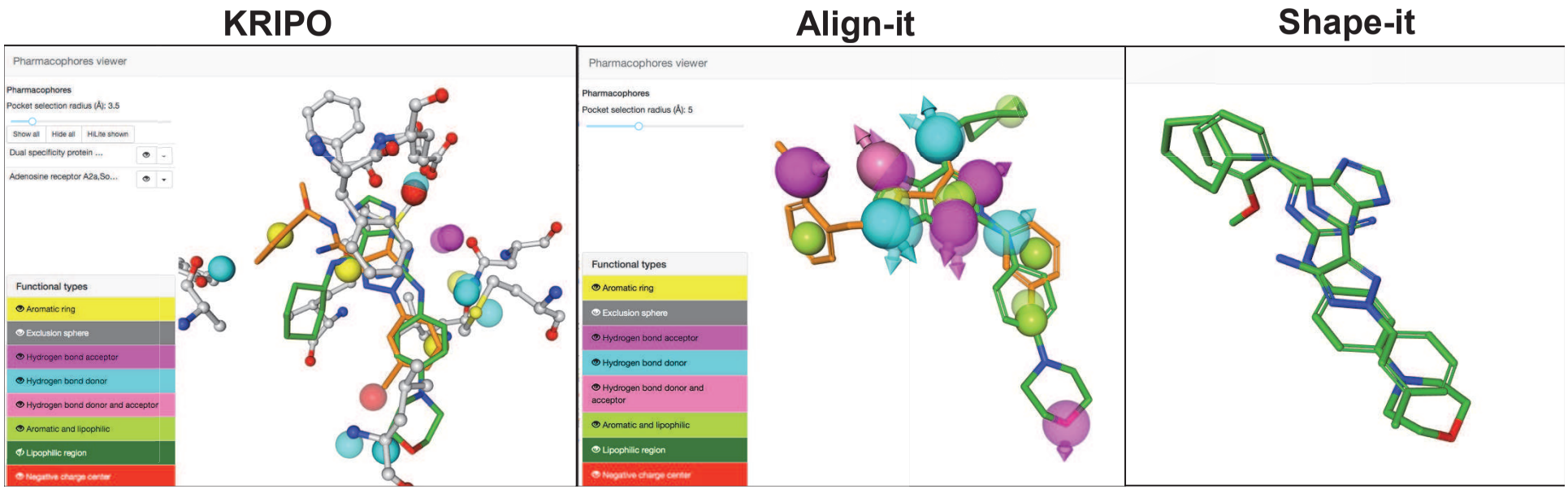


A2A receptor: 5UIG



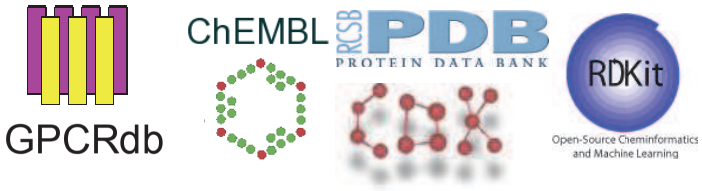
Structural chemogenomics: protein-ligand match





From structural interactions to structure-kinetics relationships

- Structural coverage of chemical GPCR ligand space

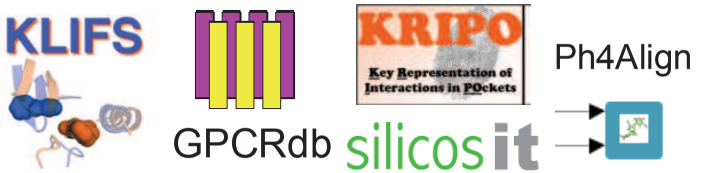


- Structure-based GPCR mutation data mapping

- Sequence-based GPCR binding site prediction
  ligand repurposing

- Structural interaction-based ligand scaffold hopping

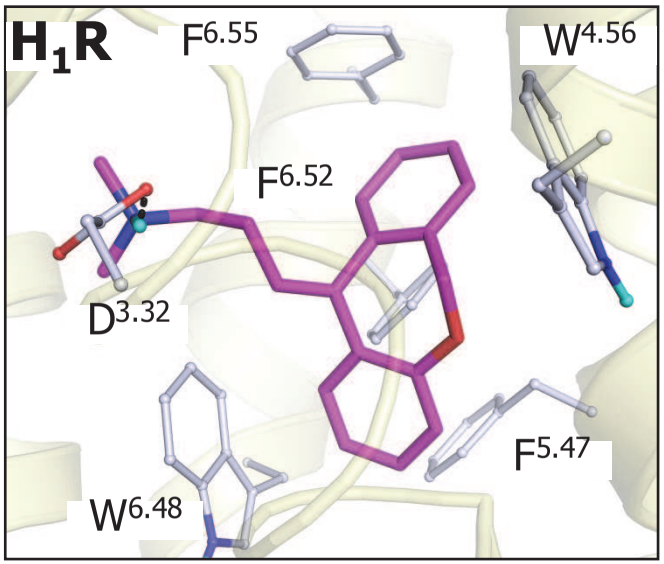
- Structure-based kinase-GPCR ligand repurposing



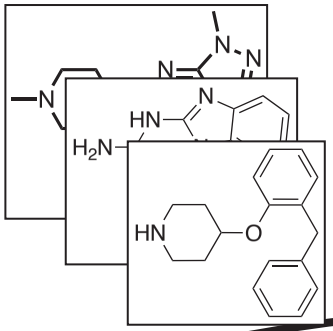
- Structure-based virtual screening - SAR - SKR

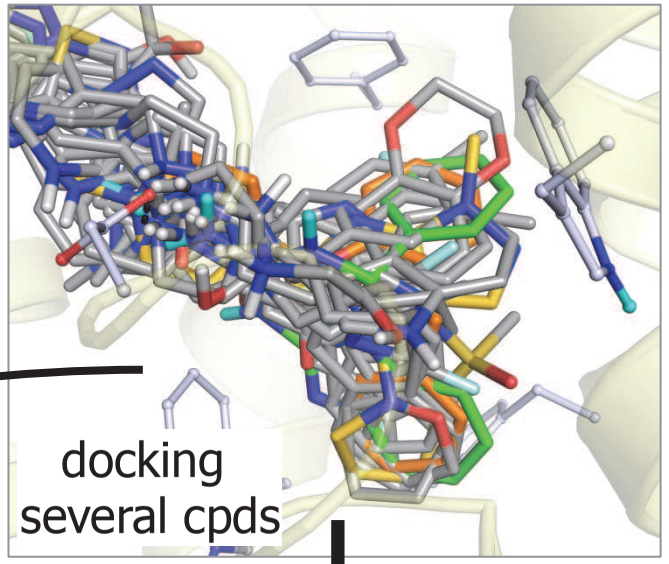
docking - IFP - RAMD

IFPs: GPCR structure-based virtual screening

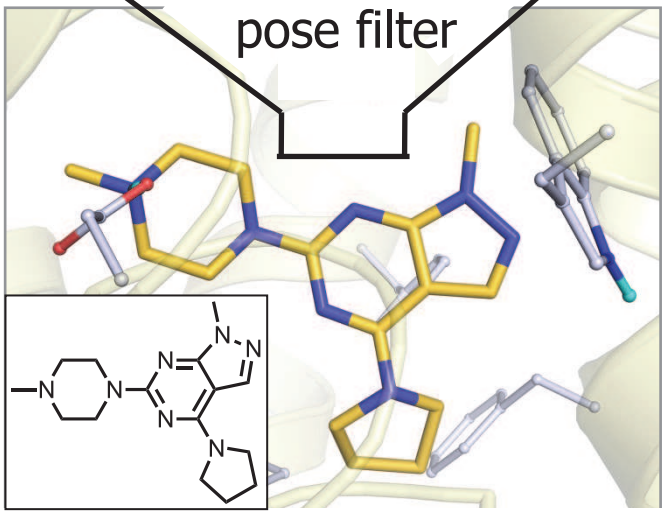


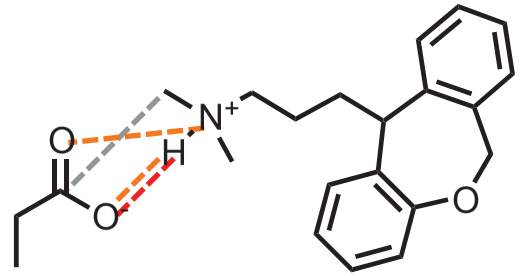
database





pose filter





D3.32 W6.48 F6.52 F6.55 F5.47 W4.56

1 0 0 0 1 0 1 1 0 1 0 0 0 0 1 0 0 1 0 0 0 1 0 1 0 0 0 0 1 0 0 0 0 0 0 1 0 1 0 0 0 0

1. hydrophobic
2. aromatic face to face
3. aromatic face to edge
4. H-bond (donor-acceptor)
5. H-bond (acceptor-donor)
6. ionic (positive-negative)
7. ionic (negative-positive)

interaction bits

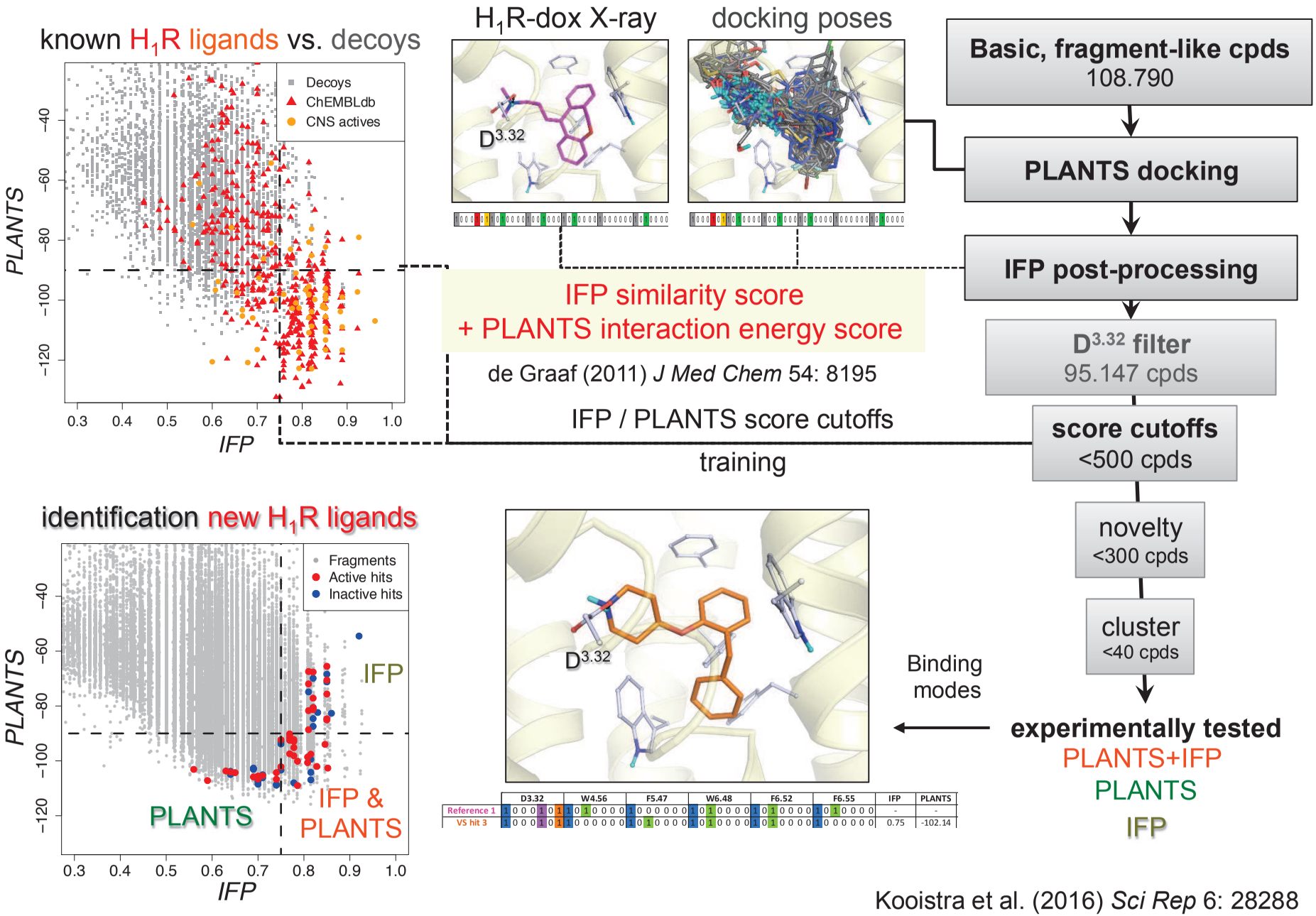
similarity score

D3.32 W6.48 F6.52 F6.55 F5.47 W4.56

1 0 0 0 1 0 1 1 0 0 0 0 0 0 1 0 0 1 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 1 0 0 0 0

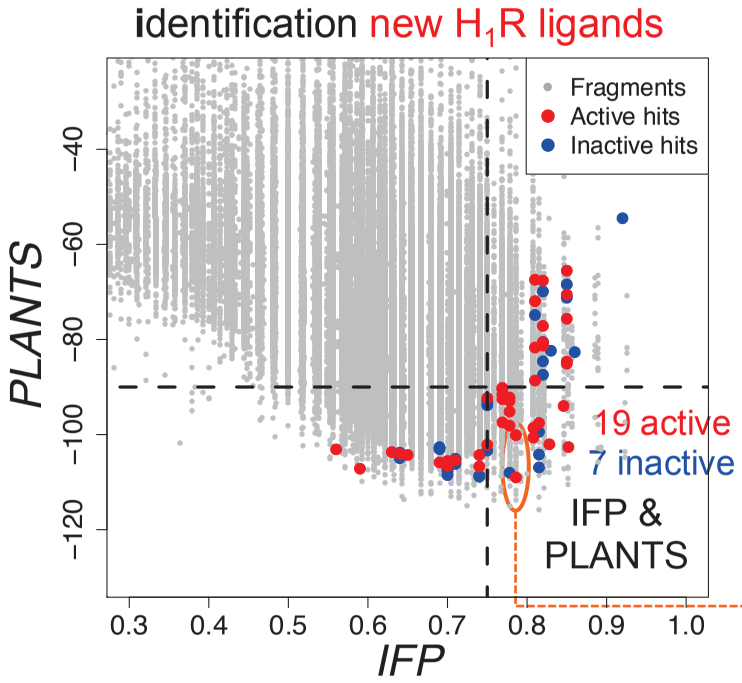
Marcou (2007) *JCIM* 47:195; de Graaf (2008) *JMC* 51:4978

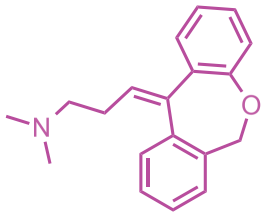
H_1R structure-based virtual fragment screening



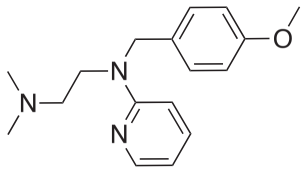
Kooistra et al. (2016) *Sci Rep* 6: 28288

H_1R structure-based virtual fragment screening

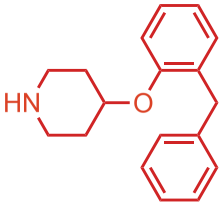




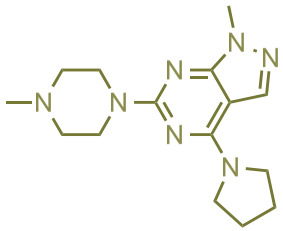
doxepin (ref. **1**)
K_i = 0.2 nM



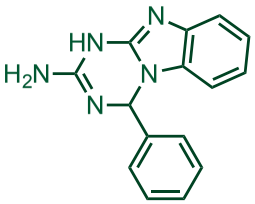
mepyramine (ref. **2**)
K_i = 2.1 nM



VUF14544 (hit **3**)
K_i = 6.4 nM

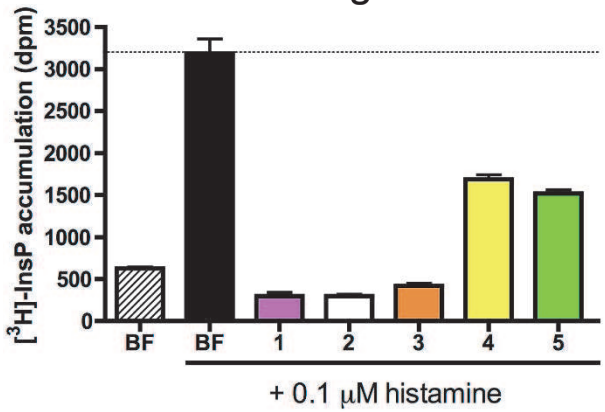


VUF13810 (hit **4**)
K_i = 62 nM

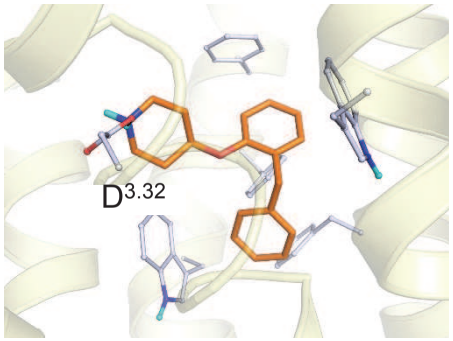


VUF13806 (hit **4**)
K_i = 427 nM

Inverse agonists

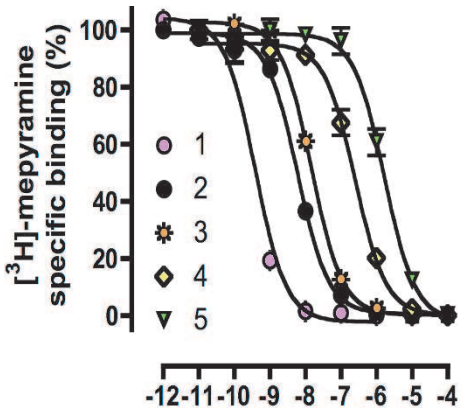


73% hit rate



VUF14544 (hit **3**)

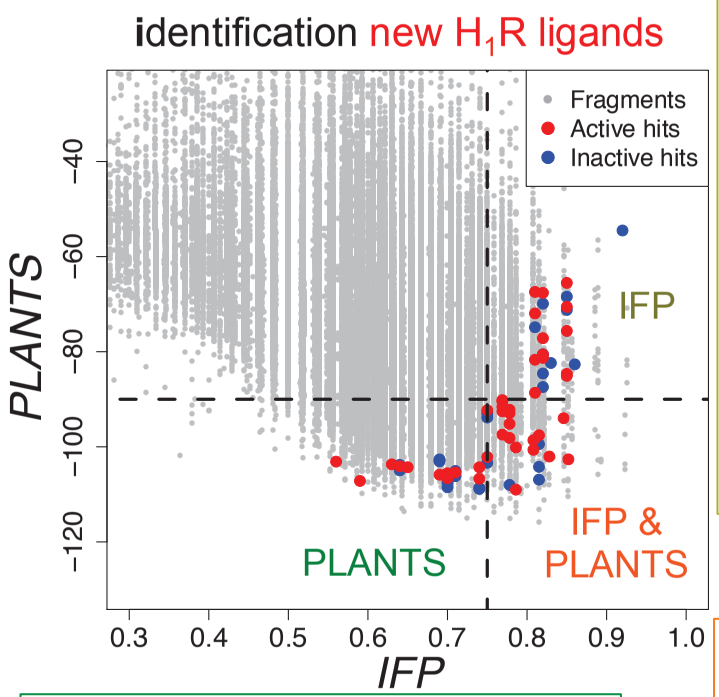
high affinity hits

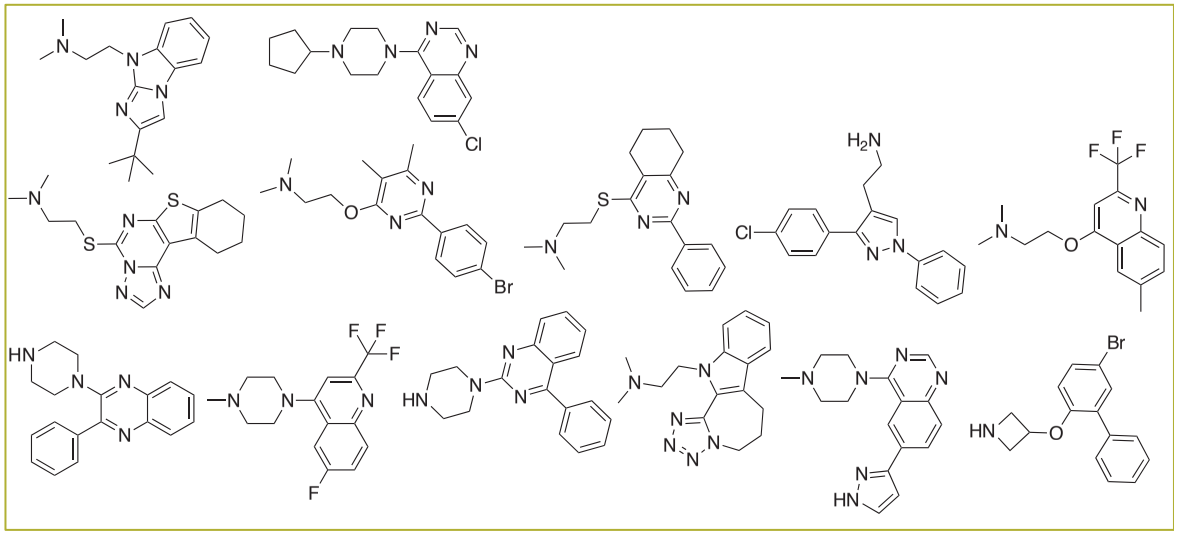


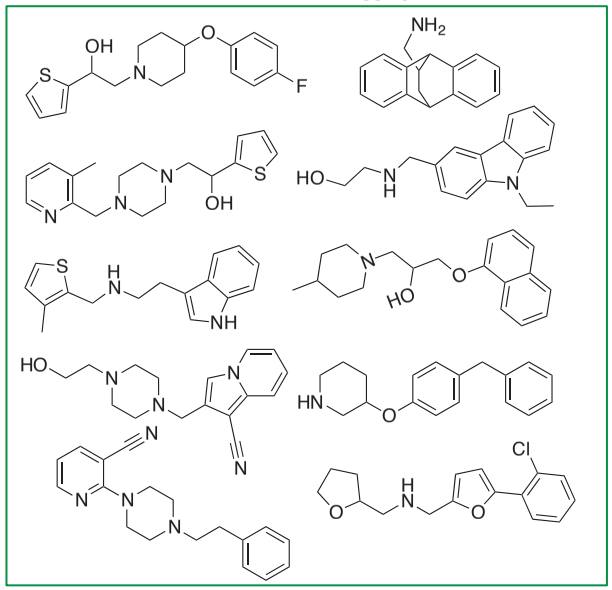
H.F. Vischer

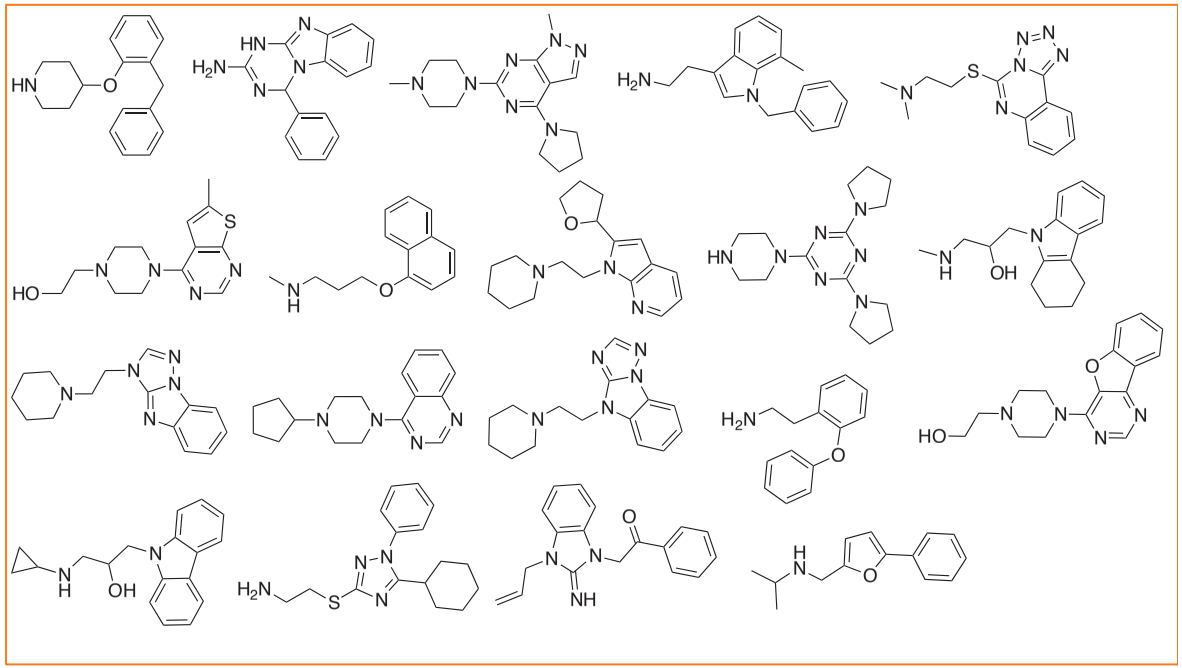
de Graaf (2011) *J Med Chem* 54: 8195

Improved structure-based VS by *combined* scoring?



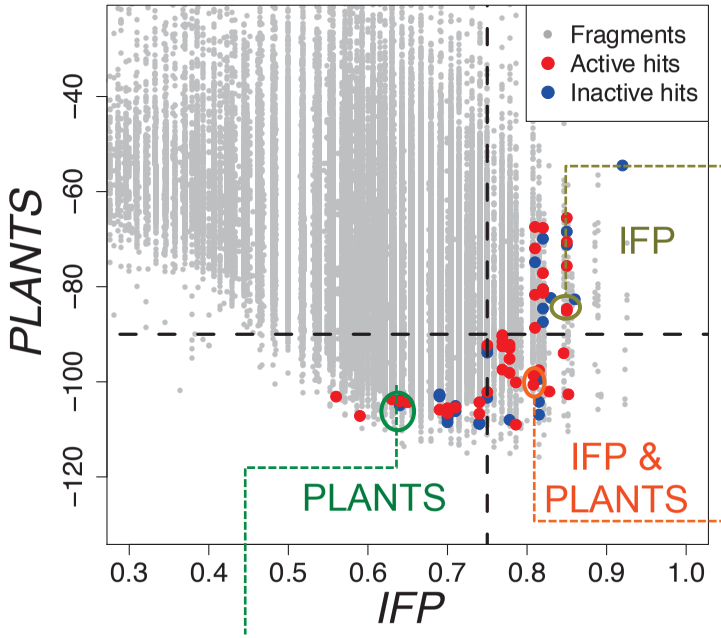




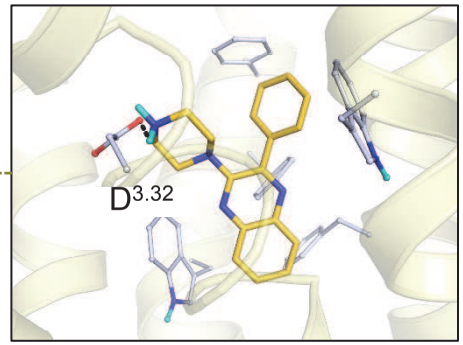


Improved structure-based VS by combined scoring

identification new H_1R ligands

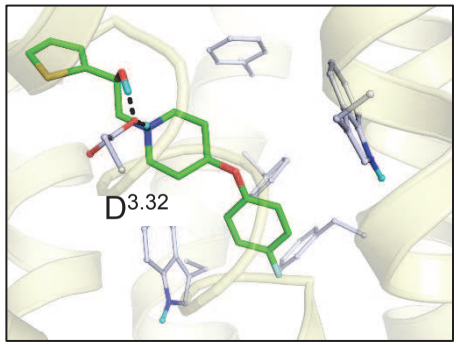


IFP: **61% hit rate**



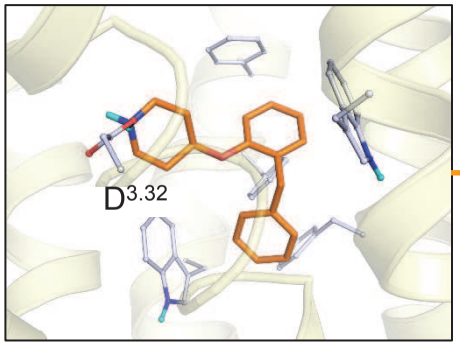
VUF14040
pK_i=7.1

PLANTS: **45% hit rate**



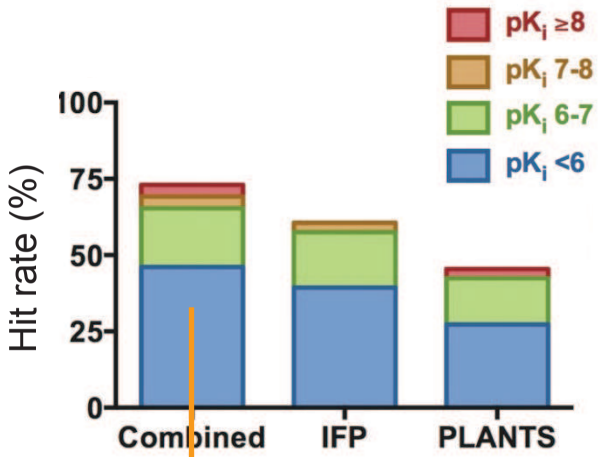
VUF hit 35
pK_i=7.1

PLANTS+IFP: **73% hit rate**



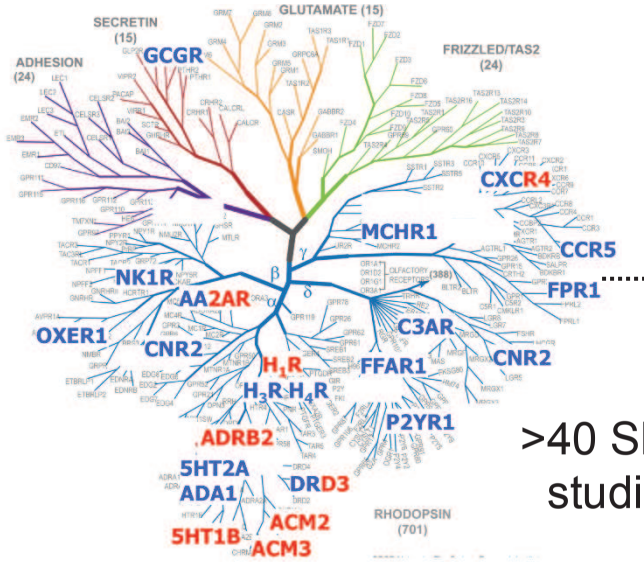
VUF hit 3
pK_i=8.2

IFP-PLANTS:
Improved hit rate & quality



Kooistra et al. (2016) *Sci Rep* 6: 28288

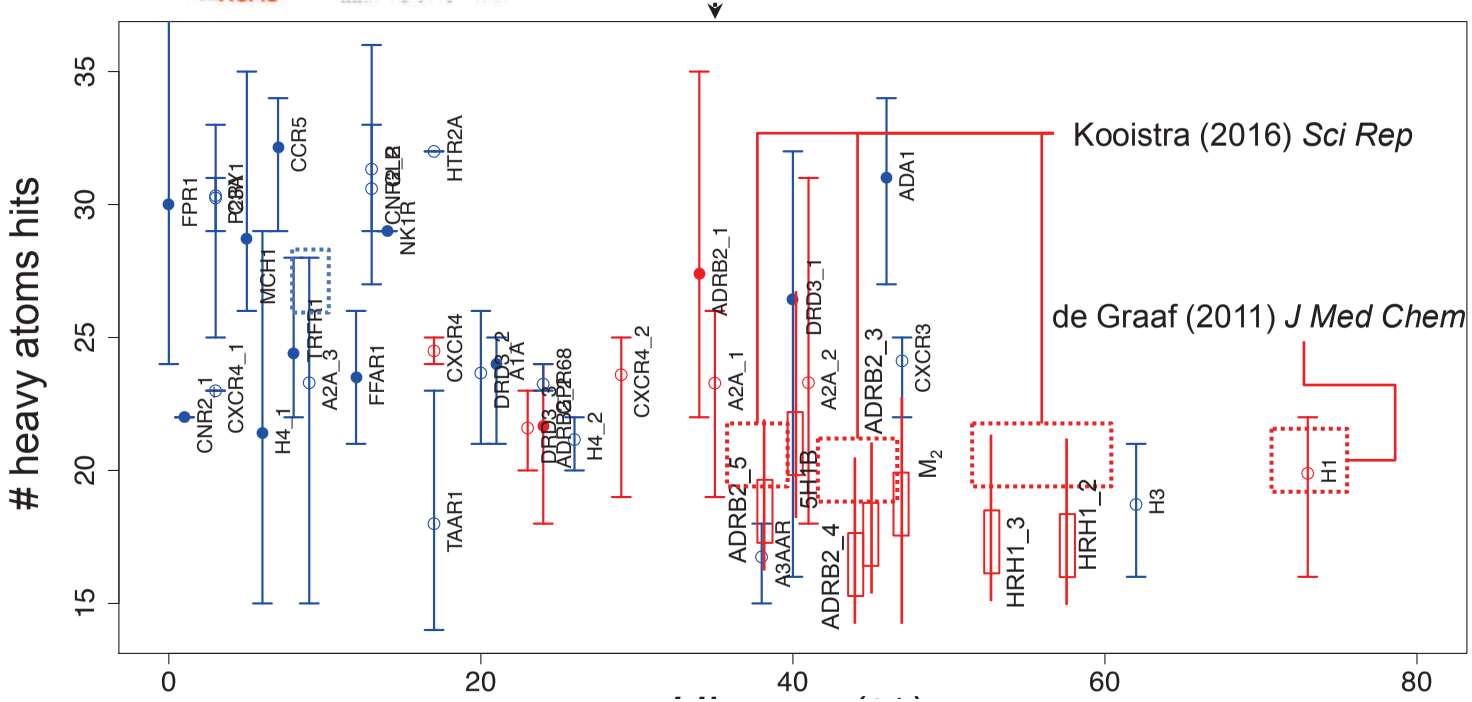
X-ray structure/homology model-based discovery GPCR ligands



>40 SBVS studies

Fragment-like hit molecules: structure-based basis for:

- Ligand optimization
- Binding site exploration



Kooistra (2014) *Adv Exp Med Biol* 796:129

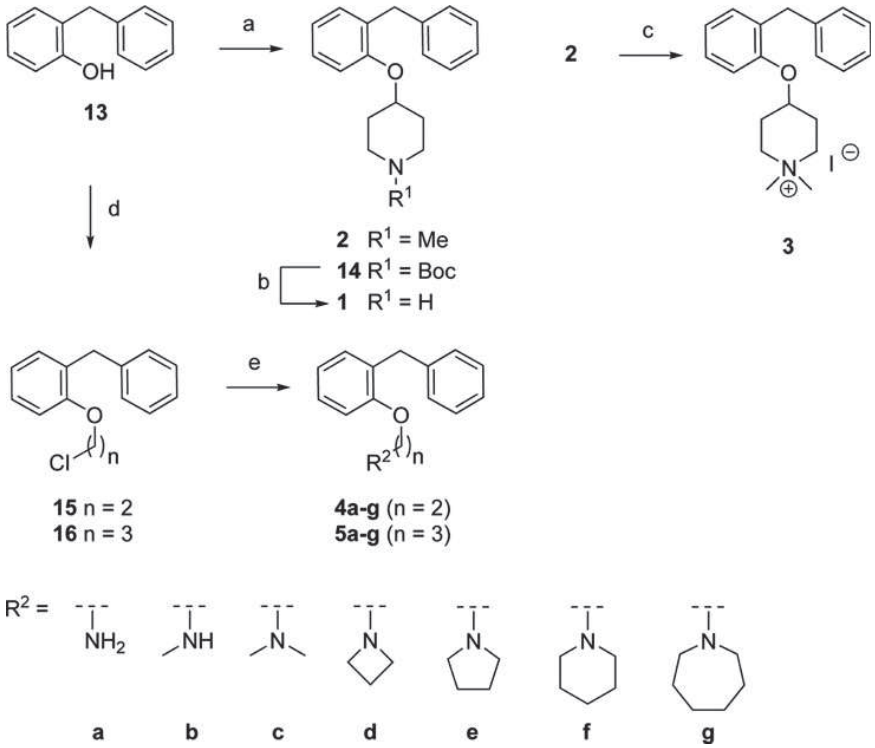
Probing H_1R binding affinity hot spots

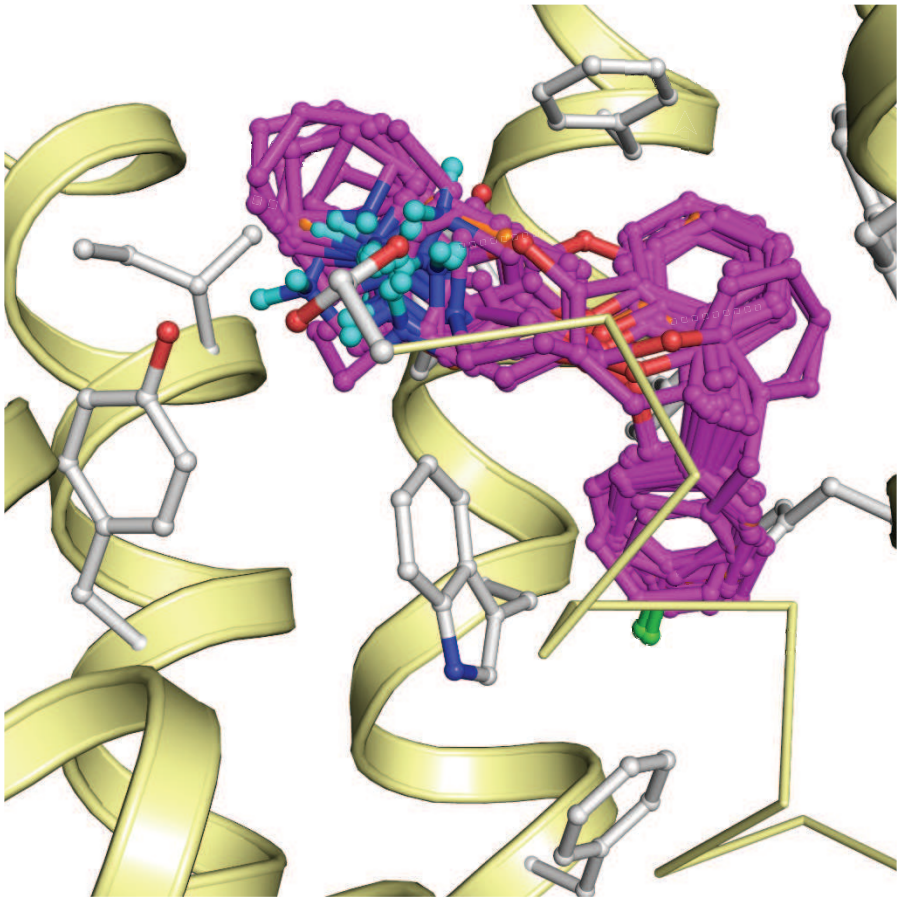
S. Kuhne

Starting from fragment-like SBVS hit:

Design & Synthesis 24 tool cpds:

- ligand deconstruction
- ring systems, ring opening
- substitution





Kuhne (2016) *J Med Chem.* 59: 9047

Probing H_1R binding affinity hot spots

S. Kuhne

Starting from fragment-like SBVS hit:

Design & Synthesis 24 tool cpds:

- ligand deconstruction
- ring systems
- ring opening
- substitution

Group Efficiency (**GE**) hotspots:

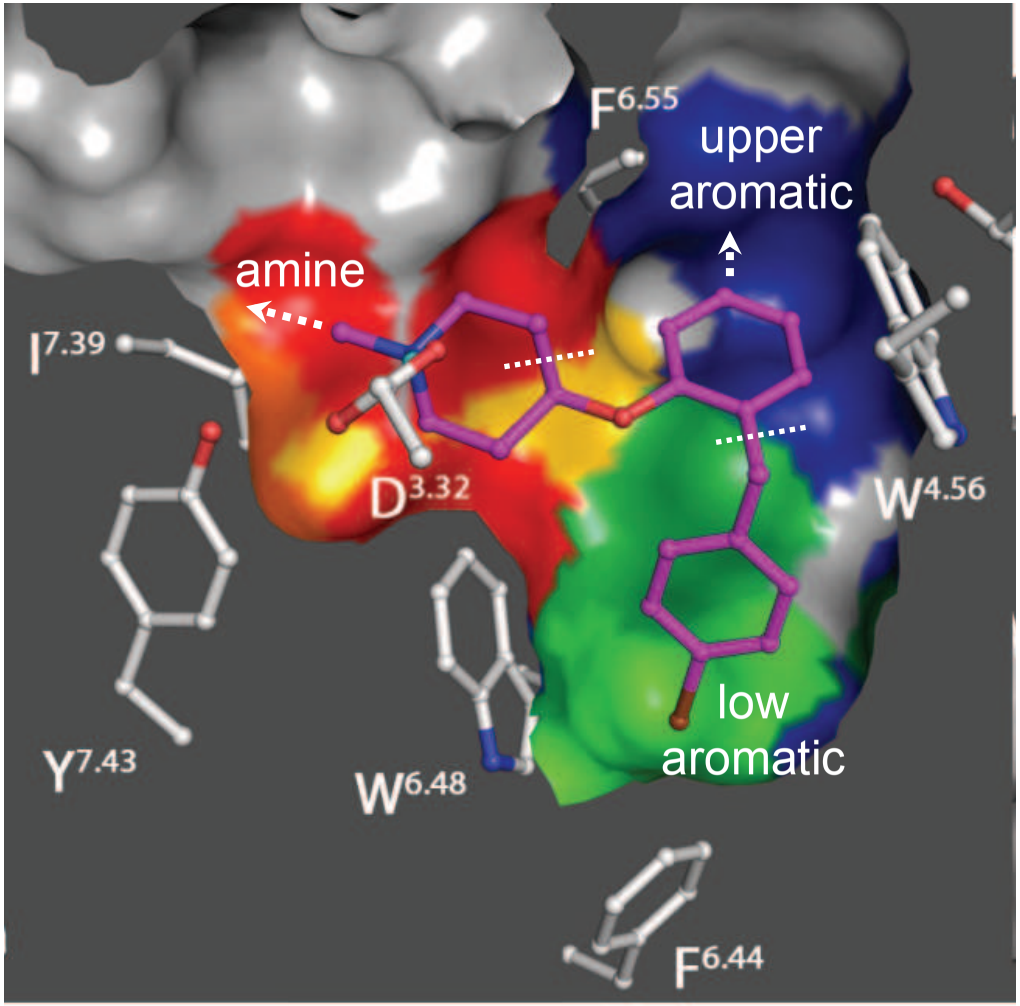
- Amine pocket
- Lower aromatic pocket

GE= $-\Delta\Delta G_{ligand}/\Delta N$

$\Delta G = -RT \ln(K_i)$

$\Delta\Delta G_{lig} = \Delta G_{lig} - \Delta G_{reference}$

$\Delta N = HA_{lig} - HA_{reference}$

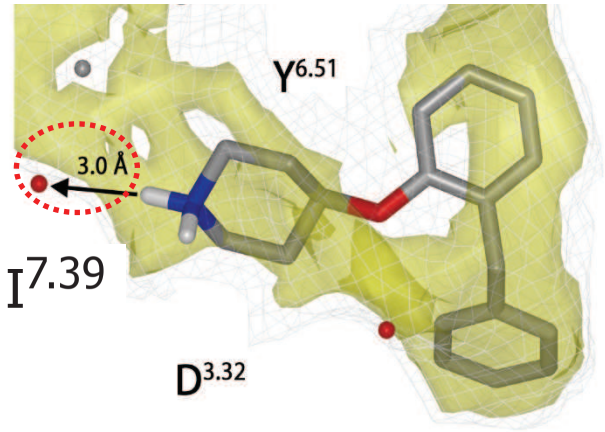


Kuhne (2016) *J Med Chem.* 59: 9047

N-methyl effect: unhappy water displacement?

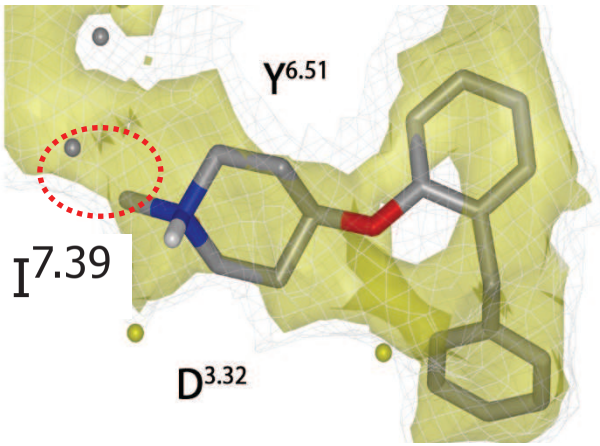
H_1R wild-type

secundary amine

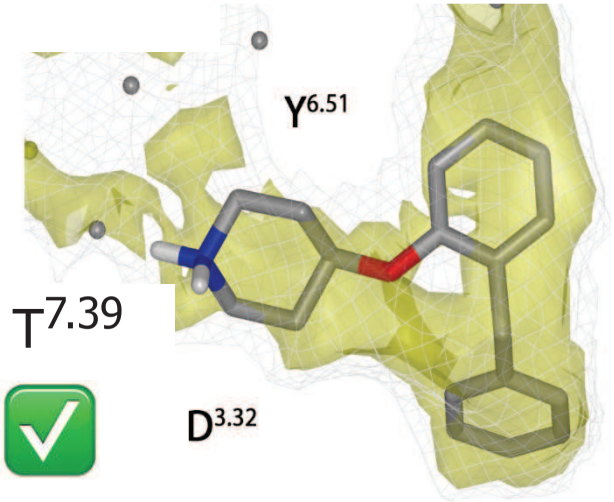


N-methyl effect: $\Delta pK_i = 0.9$

tertiary amine



$I454^{7.39}T$: **minimize** effect?

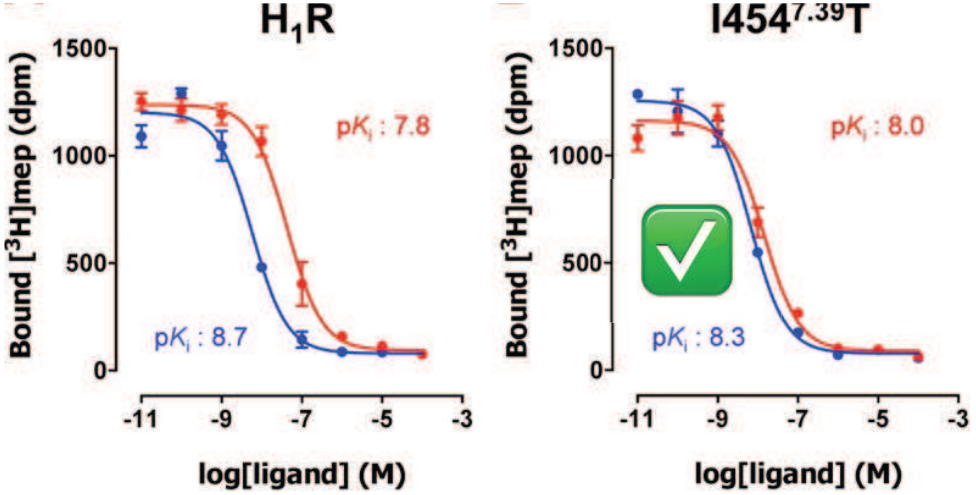


No N-methyl effect: $\Delta pK_i = 0.3$

A.J. Kooistra

R. Bosma

unhappy water not so happy water bulk water



Collaboration Bortolato, Mason

HEPTARES therapeutics

https://3d-e-chem.github.io

Medicinal Chemistry, VU University Amsterdam:

A.J. Kooistra **M. Vass** G.K. Kanev
M. Arimont S. Sun S. Kuhne
R. Bosma S. Nijmeijer H.F. Vischer
M. Wijtmans M.J. Smit I.J.P. de Esch
R. Leurs

ONCORNET, K4DD,GLISTEN

Veni, TOP PUNT
LIFT, 3D-e-Chem

CMBI, Radboudumc:
R. McGuire G. Vriend

NLeSC:
S. Verhoeven

HITS Heidelberg:
R.C. Wade D. Kokh

Heptares Therapeutics:
B. Tehan J. Mason
G. Bottegoni F. Deflorian
J.C. Mobarec C. Scully
R. Smit *A. Bortolato* (Schrodinger)

Kyoto University:
T. Shimamura S. Iwata

Imperial College London
G. Verdon K. Beis

GPCR Consortium:
M. Hanson

USC / ShanghaiTech
R.C. Stevens

Shanghai Institute Materia Medica
M.-W. Wang

University Copenhagen
D. Gloriam and GPCRdb team

Polish Academy Sciences:
S. Podlewska A. Bojarski