Experiment-Assisted Computational Drug Discovery

Alexander Tropsha

Laboratory for Molecular Modeling

UNC Eshelman School of Pharmacy

UNC-Chapel Hill

Experiment-Assisted Computational Drug Discovery? Shouldn't it be the other way around?

'The problems of how enzymes are induced, or how proteins are synthesized, or how antibodies are formed, are closer to solution than is generally believed... If you stop doing experiments for a little while and *think* how proteins can possibly be synthesized, there are only 5 different ways, not 50! And it will take only a few experiments to distinguish these'

L. Szilard

OUTLINE

- Methodology
  - Predictive QSAR Modeling Workflow
  - Examples of the Workflow applications : virtual screening and hit/lead identification
- Emerging Areas
  - Integration of QSAR modeling with other knowledge mining approaches
  - QSAR modeling using hybrid chemical/biological descriptors
- Conclusions
  - models are tools for testable hypothesis generation → focus on accurate, experimentally confirmed predictions

The chief utility of QSAR models: identification of novel hits in external libraries

MML
UNC.EDU

CHEMICAL STRUCTURES

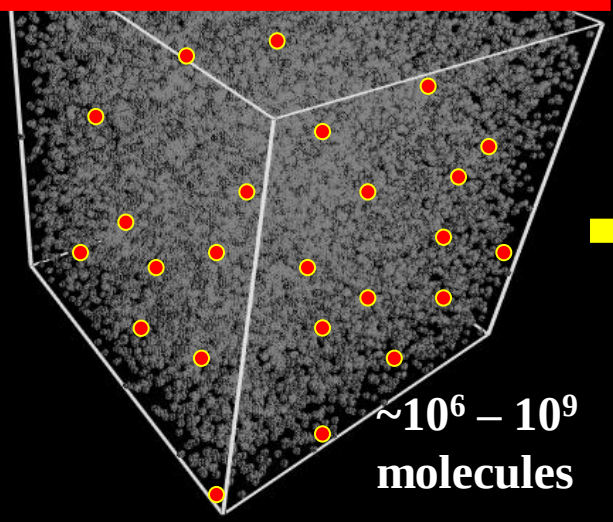
CHEMICAL DESCRIPTORS

PREDICTIVE QSAR MODELS

PROPERTY/ ACTIVITY

QSAR MAGIC

CHEMICAL DATABASE



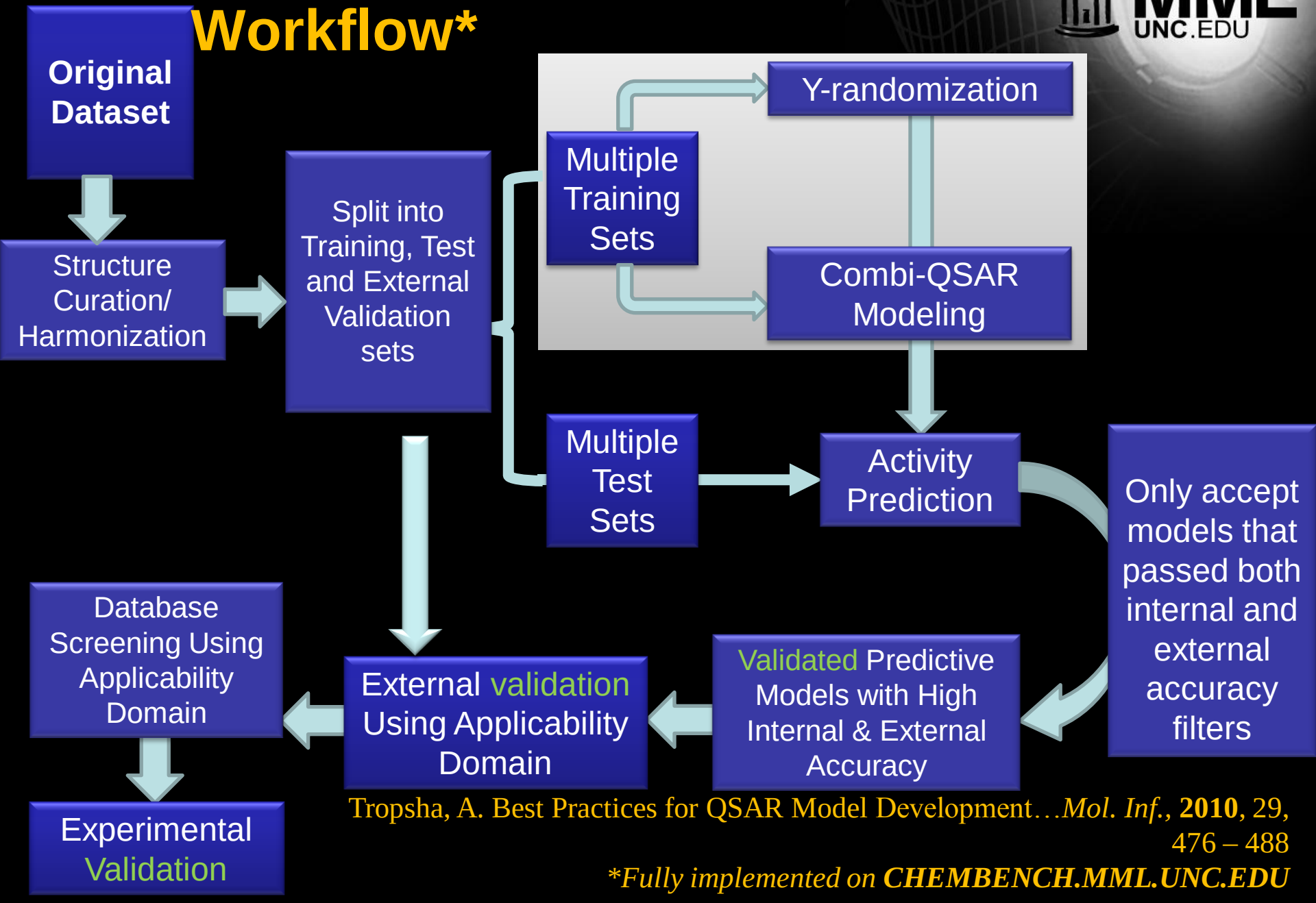
~10^6 – 10^9 molecules

VIRTUAL SCREENING

HITS (confirmed actives)

INACTIVES (inactives)

Predictive QSAR Modeling Workflow*



Tropsha, A. Best Practices for QSAR Model Development…*Mol. Inf.*, **2010**, 29, 476 – 488

Fully implemented on **CHEMBENCH.MML.UNC.EDU*

How not to develop QSAR* (examples of errors)

1. Failure to take account of data heterogeneity
2. Use of inappropriate endpoint data
3. Use of collinear descriptors
4. Use of incomprehensible descriptors
5. Error in descriptor values
6. Poor transferability of QSAR/QSPR
7. Inadequate/undefined applicability domain
8. Unacknowledged omission of data points
9. Use of inadequate data
10. Replication of compounds in dataset
11. Too narrow a range of endpoint values
12. Over-fitting of data
13. Use of excessive numbers of descriptors in a QSAR/QSPR
14. Lack of/inadequate statistics
15. Incorrect calculation
16. Lack of descriptor auto-scaling
17. Misuse/misinterpretation of statistics
18. No consideration of distribution of residuals
19. Inadequate training/test set selection
20. Failure to validate a QSAR/QSPR correctly
21. Lack of mechanistic interpretation

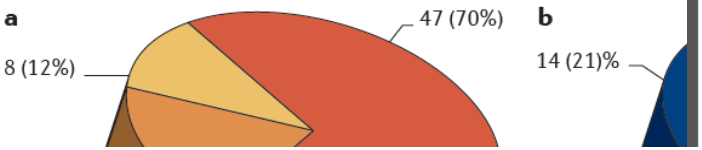
*Dearden et al., SAR QSAR Environ Res. 2009;20(3-4):241-66

Data dependency and data quality are critical issues in QSAR modeling

MML UNC.EDU

- CORRESPONDENCE
  Florian Prinz, Thomas Schlange
  Disc. Sep 2011

Believe it or not: how m
rely on published data
drug targets?



Drug Discovery Today • Volume 16, Numbers 17/18 • September 2011 EDITORIAL

ELSEVIER

editorial

Antony J. Williams

medicine and now drug repositioning or repurposing efforts. Their utility depends on the quality of the underlying molecular structures used. Unfortunately, the quality of much of the chemical structure-based data introduced to the public domain is poor. As an example we describe some of the errors found in the recently released NIH Chemical Genomics Center 'NPC browser' e as an example. There is an urgent need ernment funded data curation to improve lity of internet chemistry and to limit the ation of errors and wasted efforts.

Full Papers

QSAR & Combinatorial Science

QCS

Are the Chemical Structures in Your QSAR Correct?

Douglas Young[a]*, Todd Martin[a], Raghuraman Venkatapathy[b], and Paul Harten[a]

[a] US Environmental Protection Agency, 26 West Martin Luther King Drive, Cincinnati, OH 45268, USA; E-mail: young.douglas@epa.gov
[b] Pegasus Technical Services, 26 West Martin Luther King Drive, Cincinnati, OH 45268, USA

Keywords: Databases, *N*-octanol/water partition coefficient, Quantitative structure-activity relationships, SMILES

Received: June 26, 2008; Revised: August 13, 2008; Accepted: August 21, 2008

DOI: 10.1002/qsar.200810084

ling agencies have been investing in the development of main chemistry platforms with the primary attention n to the informatics platform itself rather the quality of content. This is clearly exemplified by the recently PC browser from the NIH Chemical Genomics Center 1]. Public online databases such as PubChem, ChemID- d the EPA's ACToR [3], to name just a few, have rapidly usted valuable resources which researchers rely on for able chemical structures and associated data. While emistry databases can certainly be of value, we feel the ould be immediately alerted to consider issues of data en using these resources and we call into question both s and the trust we place in them. To our knowledge the raise, using the example of a recently released database, been described elsewhere and the user community, and

QSAR Comb. Sci. 27, 2008, No. 11-12, 1337–1345

In the last ten years, public online databases have

In the last decade numerous attempts have been made to

Data dependency and data quality are critical issues in QSAR modeling

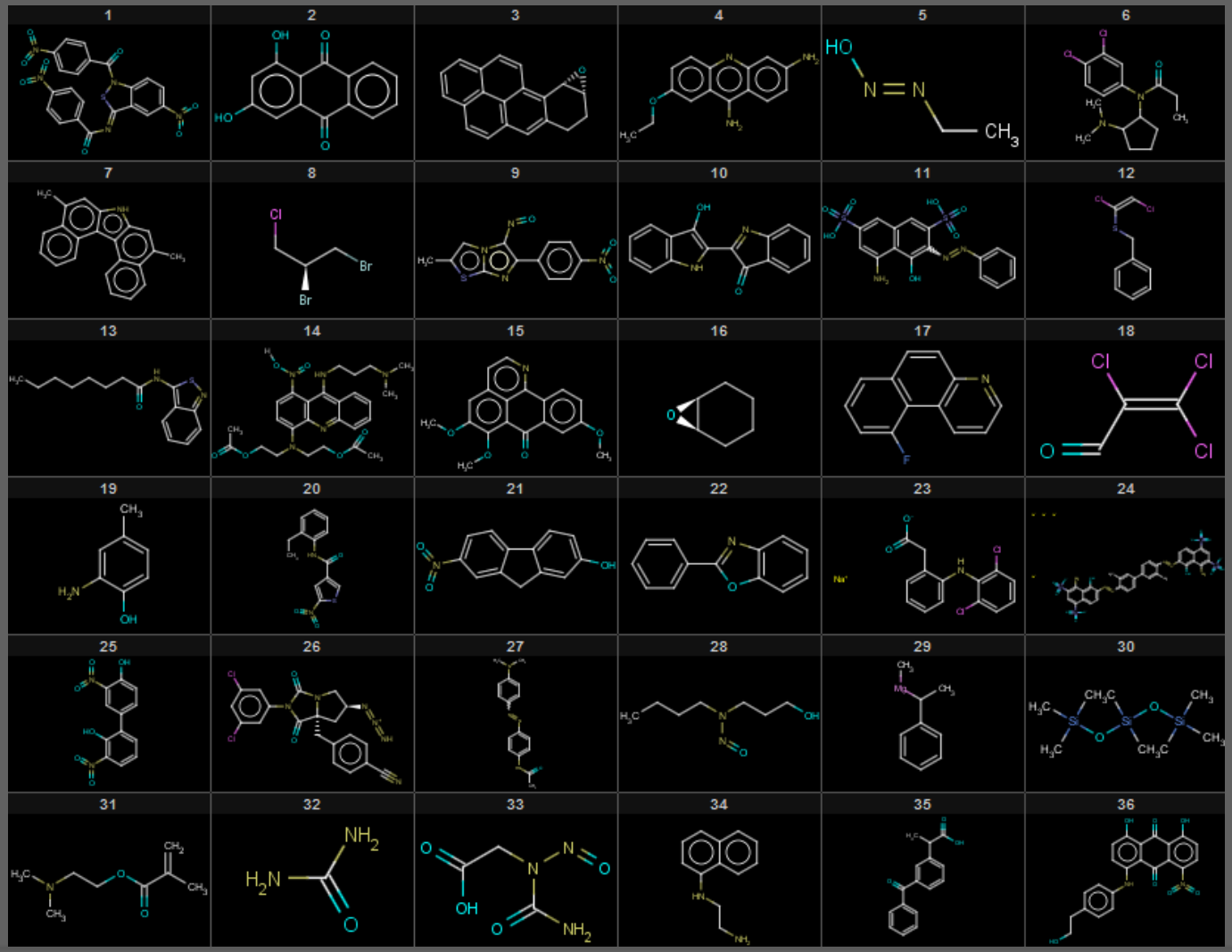
- Cheminformaticians are at the mercy of data providers with respect to data quality.

- Both chemical and biological data in a dataset may be inaccurate and in need of thorough curation

- The number of published QSAR models that were poor or not too successful due to data quality issue is unknown but possibly large
  - error rates range from 0.1 to 10 %
  - small structural errors could lead to significant loss of predictive power

- Often considered trivial, the **basic steps to curate a dataset** of compounds are not so obvious especially for beginners.

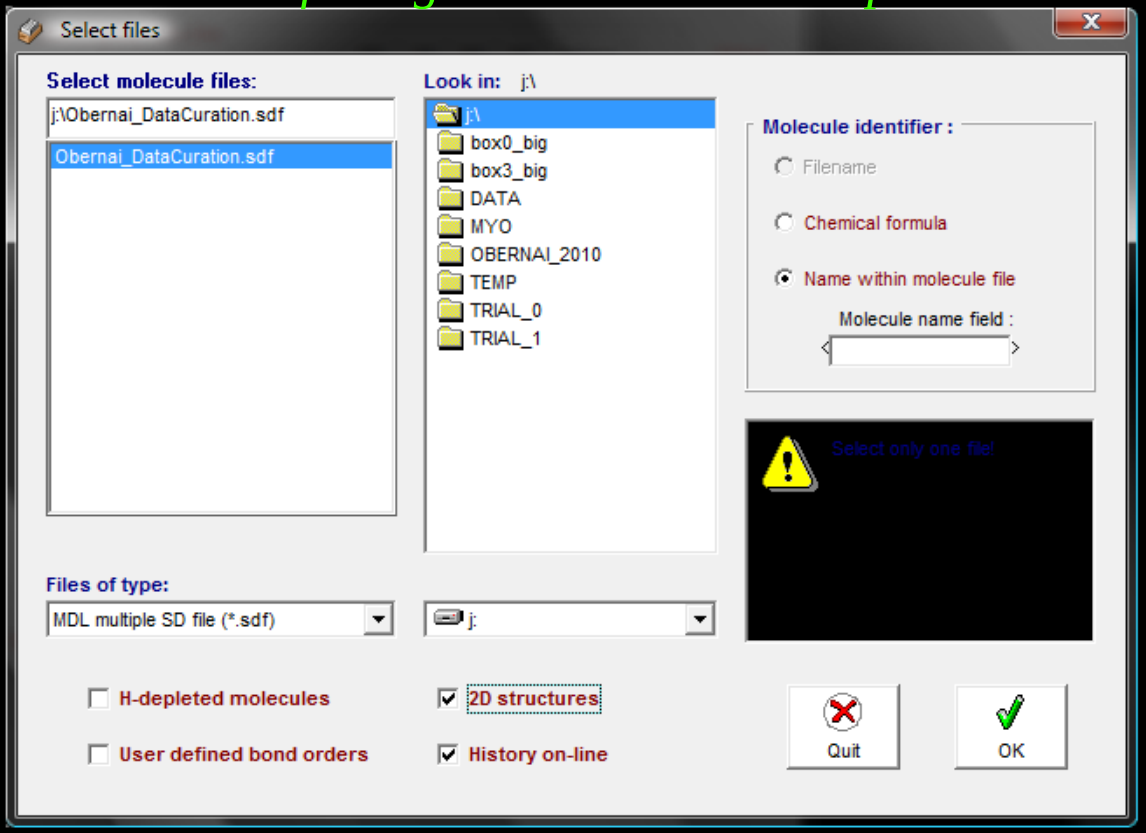
242 chemical records / one binary activity

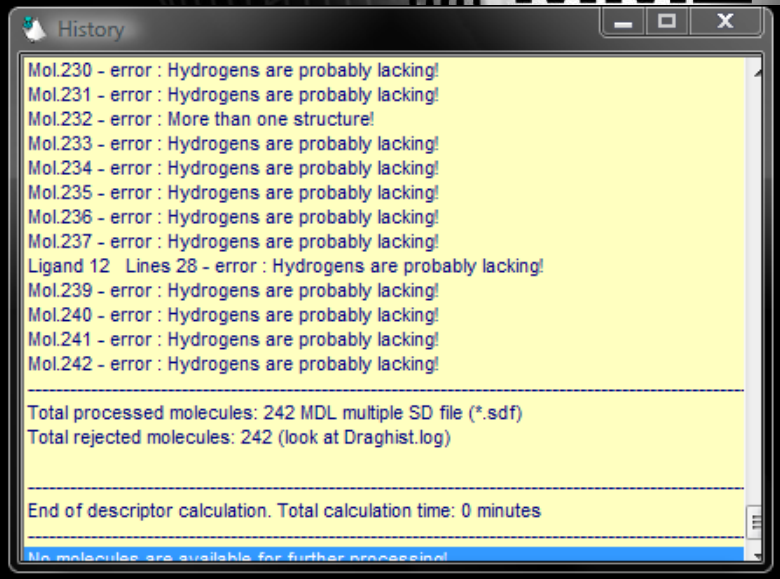


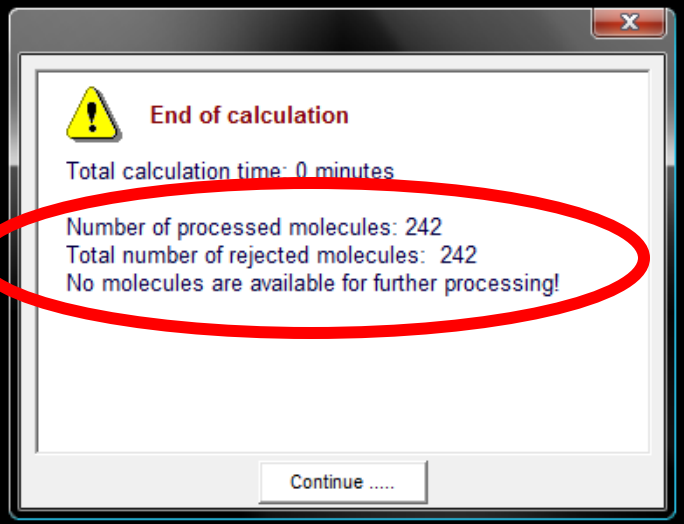
Looks clean …

Looks clean … but …

Calculation of Dragon molecular descriptors



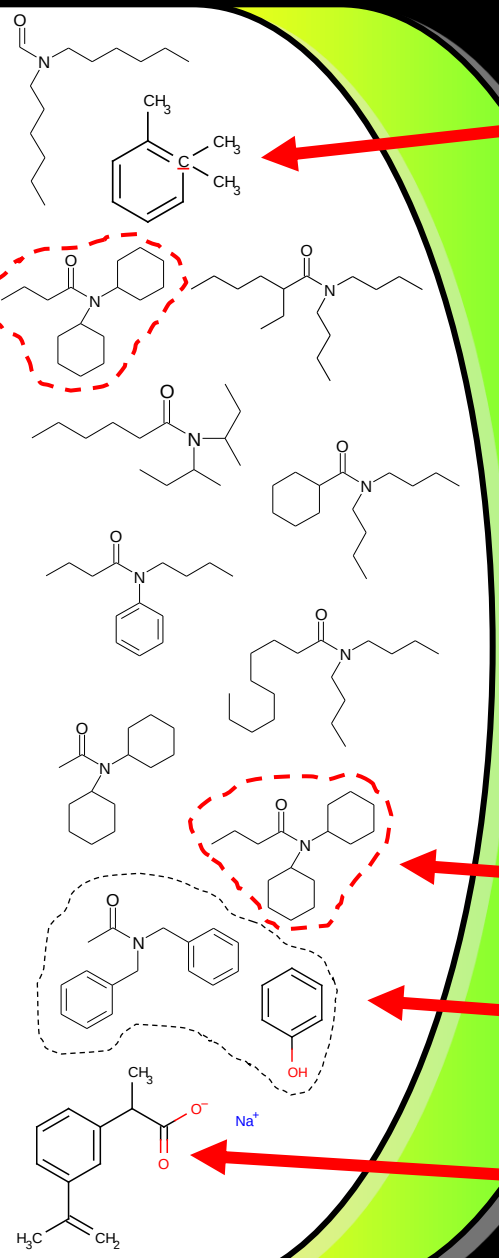




All compounds are in fact incorrect (presence of inorganics, salts, organometallics, duplicates; certain hydrogens are lacking; wrong standardization; etc.)

http://chembench.mml.unc.edu

QSAR modeling with non-curated datasets



DESCRIPTORS

Presence of **ERRONEOUS AND/OR WRONG STRUCTURES**

Presence of **MISPRINTS AND WRONG NAMES**

ERRORS in the calculation of DESCRIPTORS

QSAR MODELS ???

Presence of **DUPLICATES**

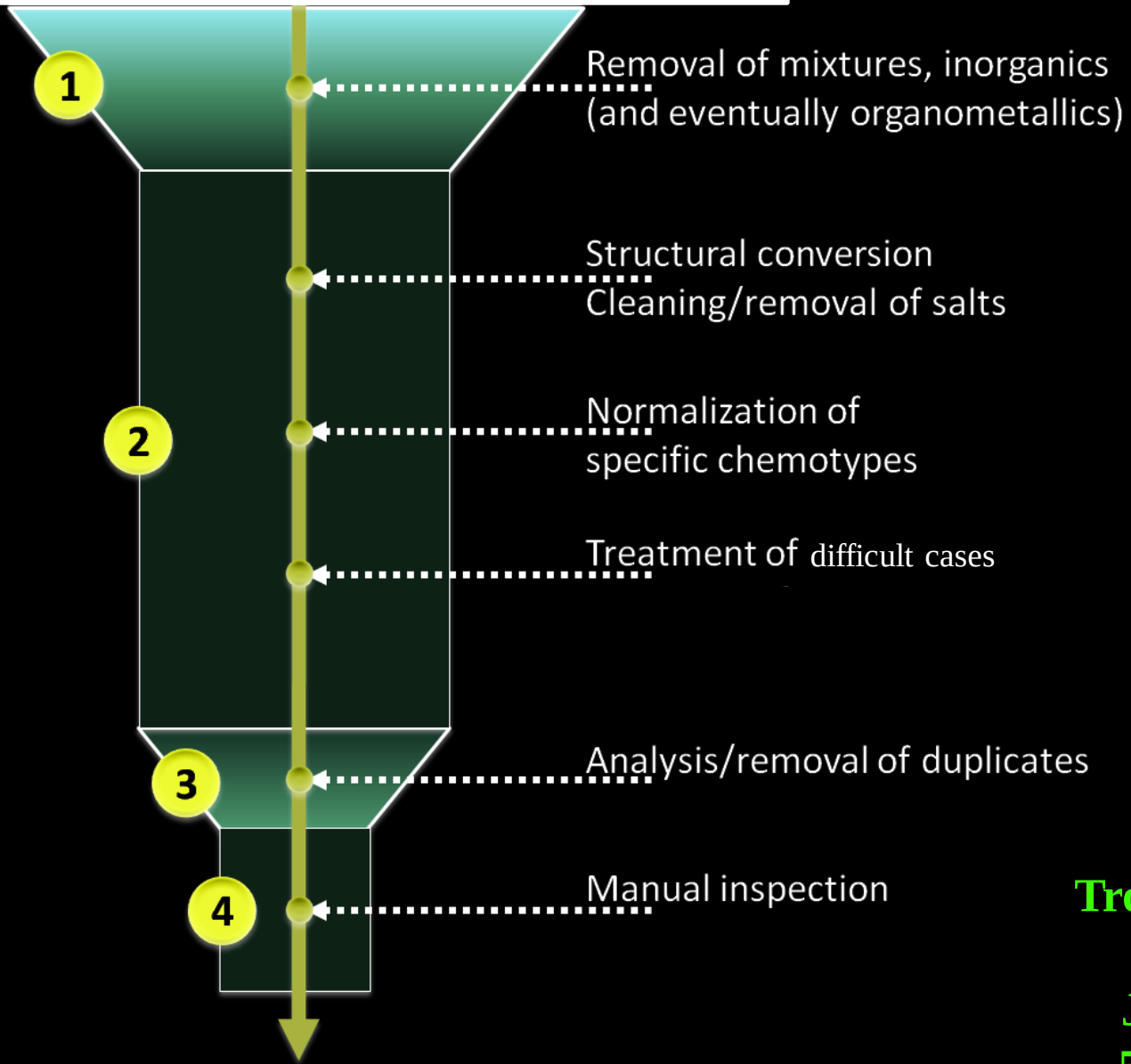
Presence of **MIXTURES**

Presence of **SALTS**

Etc.

0.613
0.380
-0.222
7.08
1.146
0.491
0.301
0.141
0.956
0.256
0.799
1.195
1.005

INITIAL LIST OF SMILES/STRUC

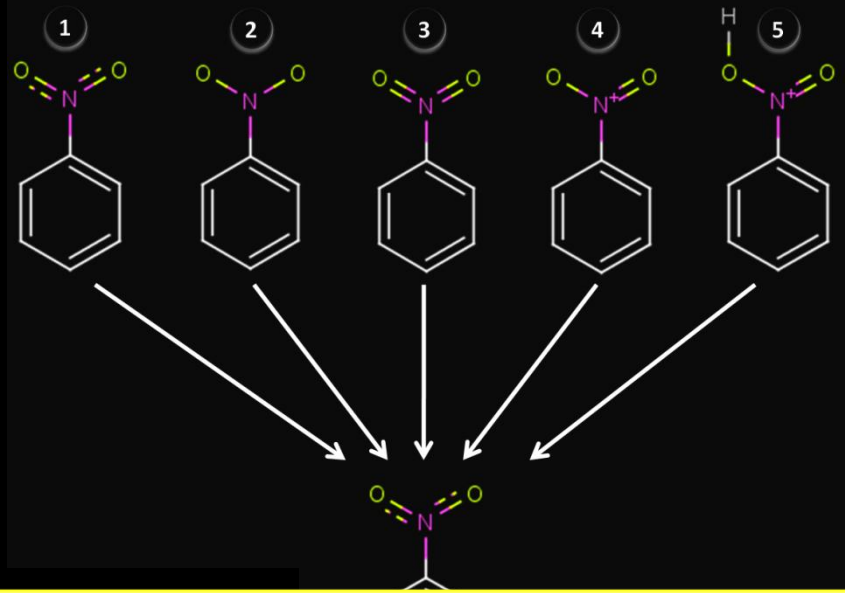


MML
UNC.EDU

Fourches, Muratov, Tropsha. Trust but verify. JCIM, 2010, 50:1189-204.

QSAR modeling of nitro-aromatic toxicants

-Case Study 1: 28 compounds tested in rats, log(LD50), mmol/kg.
-Case Study 2: 95 compounds tested against *Tetrahymena pyriformis,* log(IGC50), mmol/ml.

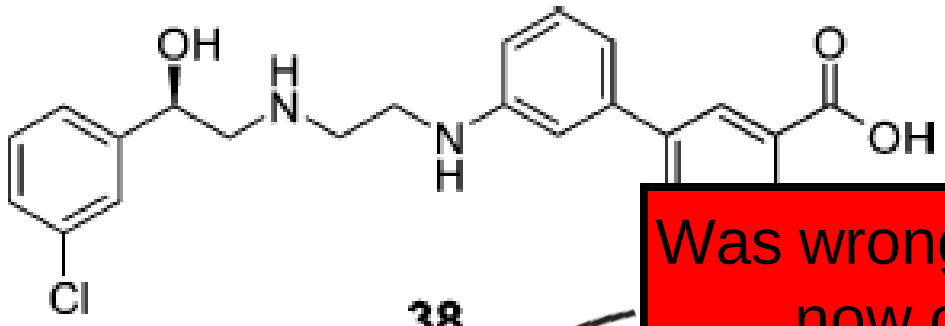


Data curation affects the accuracy (up or down!) of QSAR models

Even small differences in structure representation can lead to significant errors in prediction accuracy of models

Artemenko, Muratov et al. J. SAR QSAR 2011, 22 (5-6), 1-27.

Possible Source of Errors: inaccurate extraction from literature



38

Correct

Was wrong In ChEMBL, now corrected

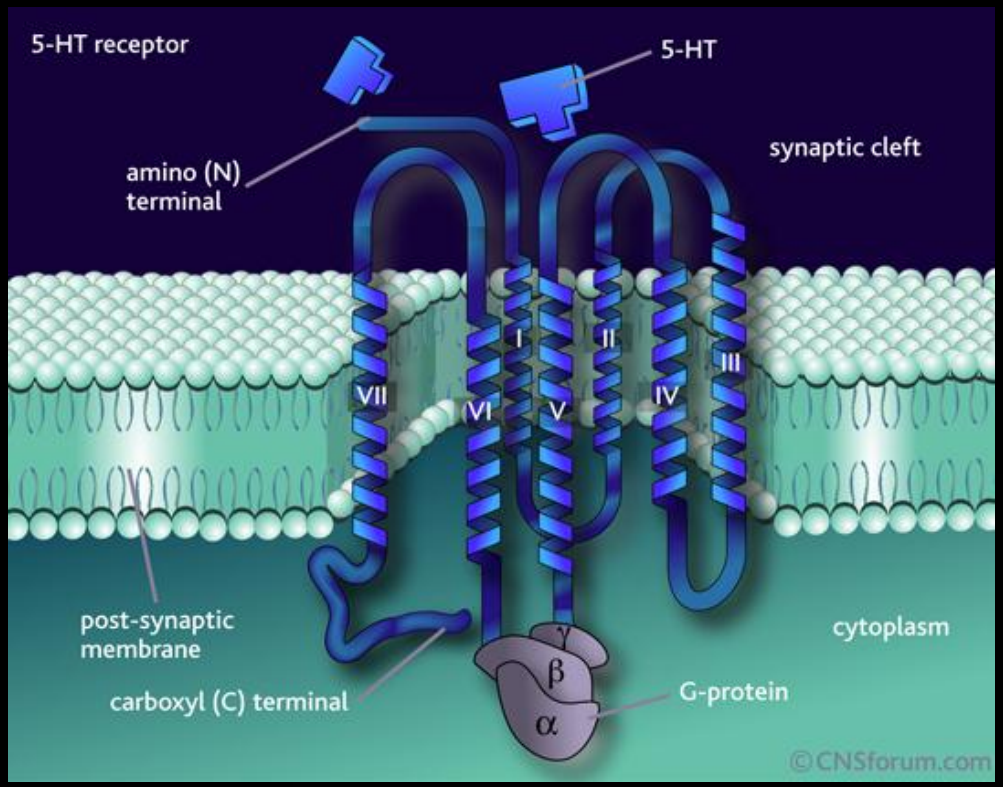
β_3 AR activity	β_2 AR[a] binding	β_3 functional/β_2 binding[b]	β_1 AR binding[a]	β_3 functional/β_2 binding[b]
8.4 ± 0.2	5.8 ± 0.5	398	6.4 ± 0.5	100

[a] The binding constant pK_i of compound **38** ($n = 3$) against β_2 or β_1 ARs; see Experimental Section. [b] The ratio of the pIC_{50} of the compound for β_3 AR relative to the binding constant for β_2 or β_1 ARs.

J. Med. Chem. 2006, 49: 2758-2771.

Case study 1: 5-HT_7 Receptor binders

- A member of the GPCR superfamily of cell surface receptors.
- Involved in various cognitive and behavioral functions.
- A potential drug target for psychotic disorders such as schizophrenia and major depression.



** Basic and clinical pharmacology, 8th edition.2001:265–291*

Study Design

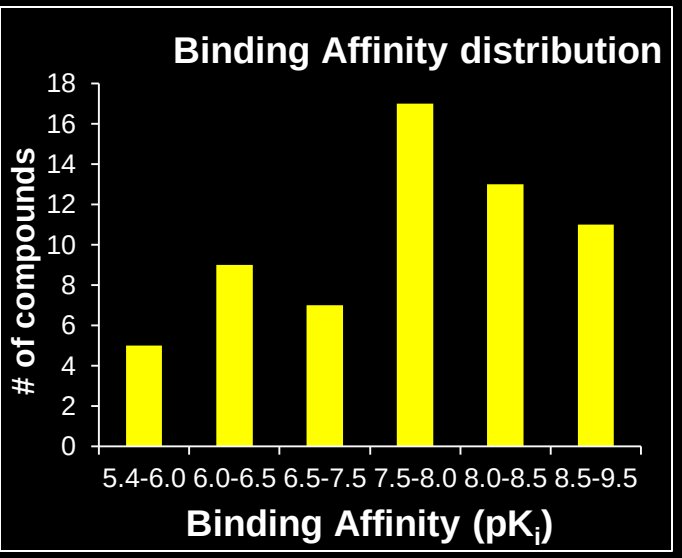
Curated 5-HT$_7$ Data*

k Nearest Neighbors

Continuous kNN Model (62 cpds)

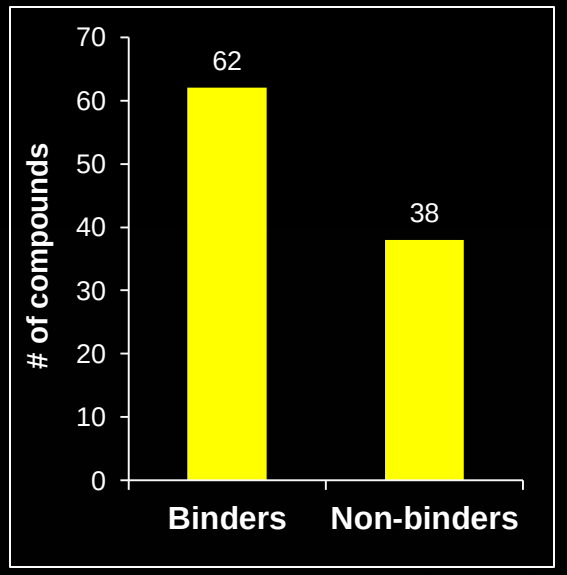
DWD Classification Model (100 cpds)

Distance Weighted Discrimination



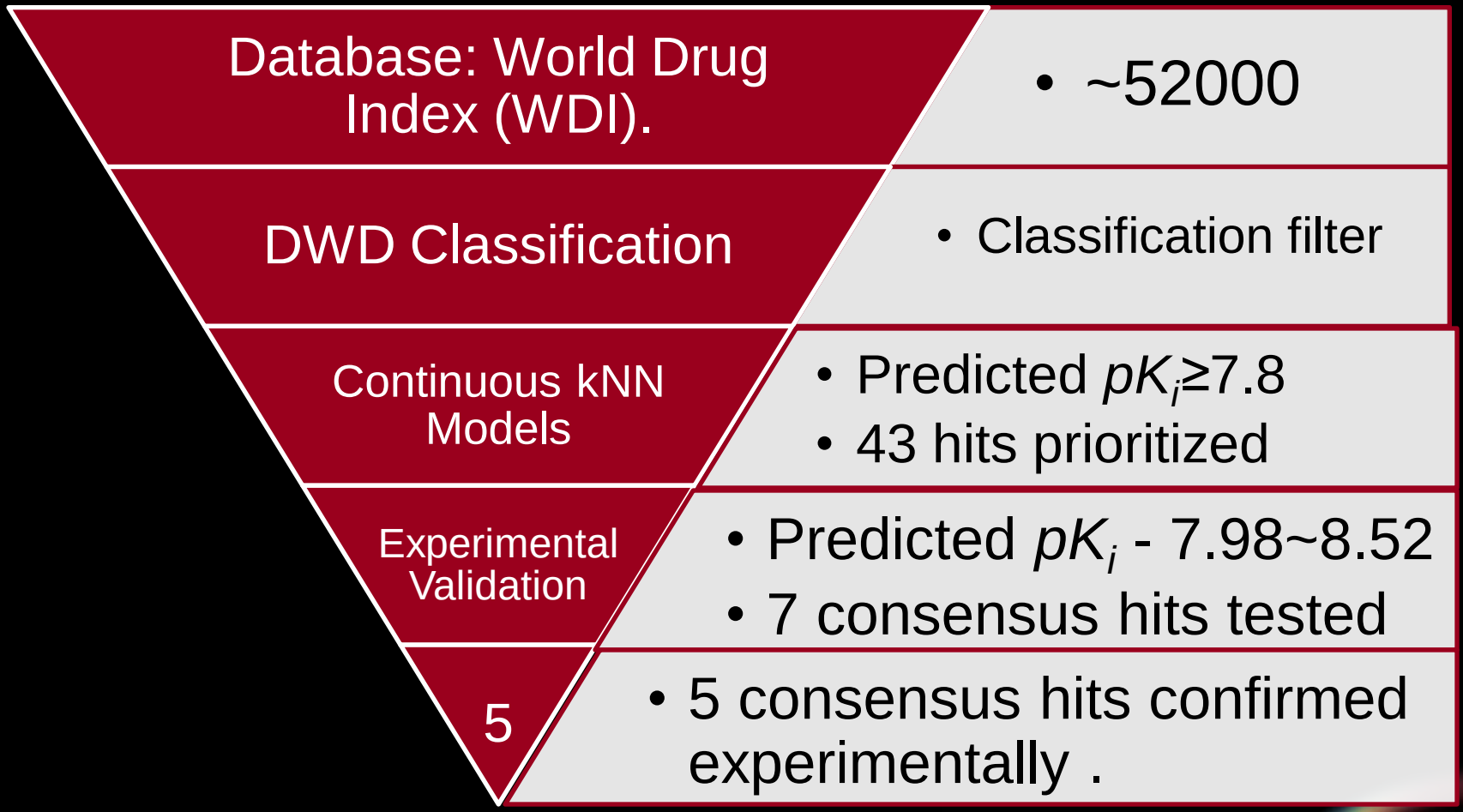
World Drug Index (~52,000 cpds)

VS Hits



* Data were collected from PDSP database provided by Prof. Roth's lab.

Virtual Screening Workflow to identify and confirm $5HT_7$ binders



Experimental Validation*: 5 out of 7 Tested Hits Are confirmed 5-HT_7 Binders

Name	Predict K_i(nM)	K_i (nM)	Function	Therapeutic Category	Mechanism of action
Droperidol	**3.24**	**3.5**	**Antagonist**	Butyrophenone antiemetic and antipsychotic agent	Ligand of postsynaptic GABA and dopaminergic receptors; selectively blocks α-adrenergic receptors.
Perospirone	**7.08**	**8.6**	**Antagonist**	Atypical antipsychotic agent	Antagonist of 5-HT2A and dopamine D_2 receptors
Altanserin	3.39	143.0	N/A	Used in Human neuroimaging study	Strong 5-HT_{2A} ligand
Pravadoline	9.55	3184.0	N/A	Cannabinoid analgesic agent	Inhibit cyclooxygenase (COX)
Clomipramine	13.80	46.0	N/A	Tricyclic antidepressant; antiobsessional agent	Presynaptic receptors are affected: α_1 and β_1 are sensitized, α_2 are desensitized
Clazolam	6.46	>10000	N/A	N/A	N/A
Sulazepam	14.13	>10000	N/A	Sedative and anxiolytic agent	N/A

*data from B. Roth's lab.

Case study 2: 5-HT_{2B}-receptor binders

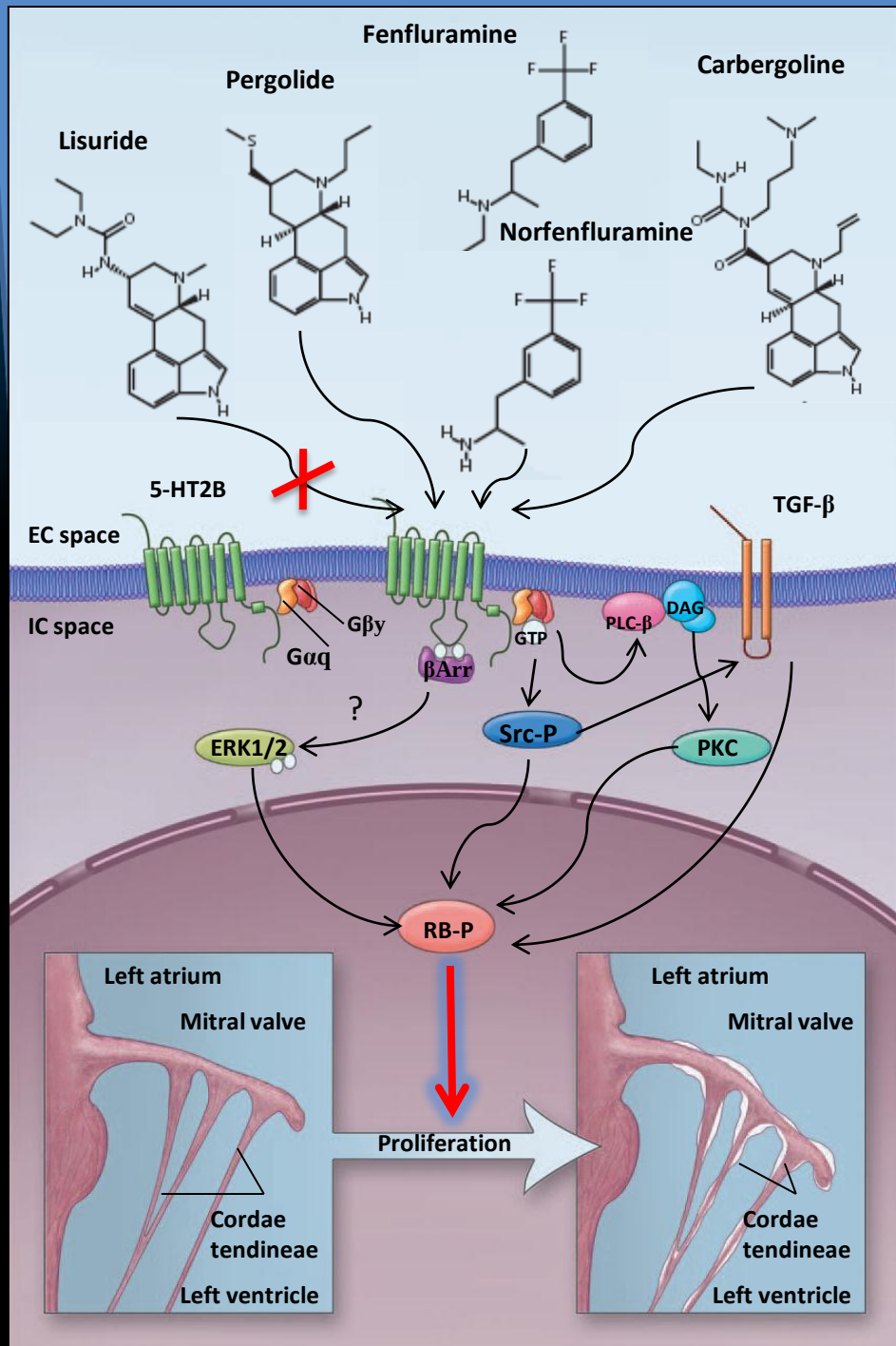
Possible Explanation of cardiotoxicity:

- Activation of 5-HT_{2B} receptors leads to the dissociation of the G protein
- Activation of phospho lipase C-β (PLC-β)
- Activation of Src
- Activation of ERK1/ERK2
- Phosphorylation of retinoblastoma protein

↓ mitogenesis

Overgrowth valvulopathy and subsequent valvular dysfunction.

Roth, B.L. *N ENGL J MED,* 356;1 (2007)



5-HT_{2B} models and VS results

Dataset

800 cps.

Dataset curation

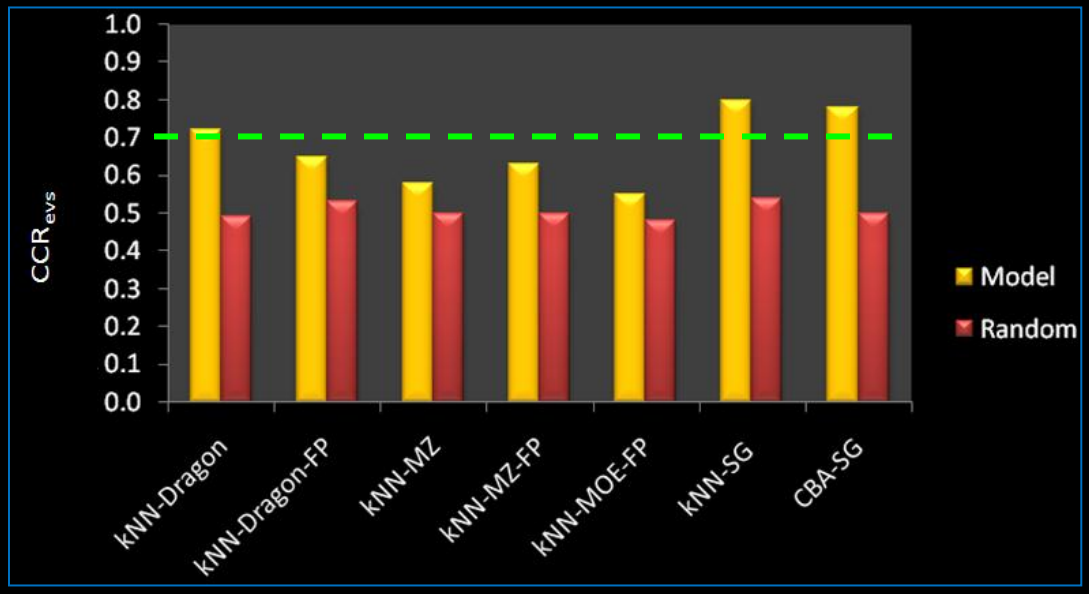
608 Inactives (0)

146 Actives (1)

Huang, X., *et al*. ***Molecular Pharmacology*** (2009)

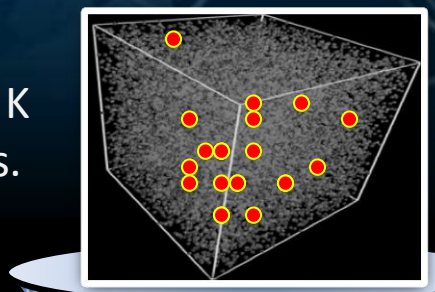
Source: Roth lab, UNC

Model statistics



Virtual screening

59 K cps.



5-HT_{2B} predictor

122 VS Hits

Select for Testing

10 VS Hits

Experimental Testing

9 Validated Actives

Hajjo R. et al, *J Med Chem.* 2010 11;53(21):7573-86

Results of VS and radioligand binding assays

Compound	Experimental K_i (nM)
Methylergometrine	0.8
6-Fluoromelatonin	2495
Adrenoglumerulotropin	491
CGP-13698	>10000
PIM-35	1617
Fendiline	3217
Fluspirilene	151.4
PNU-96415E	69.6
Prestwick-559	33.1
Raloxifene	69

Success rate for active *vs.* inactive models = **90 %**

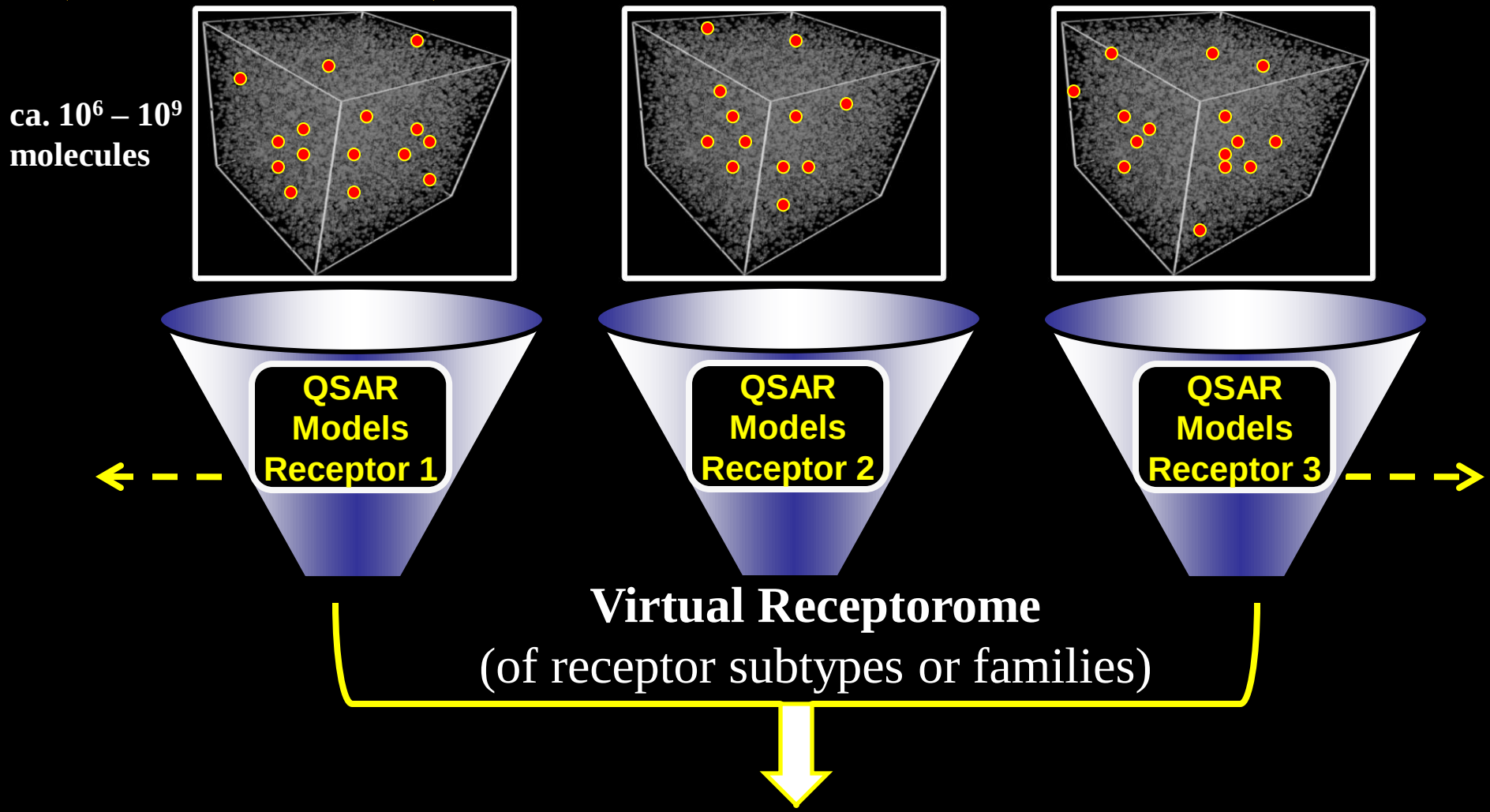
Tested by collaborators at PDSP.

Can we identify these same hits with simple similarity searches??

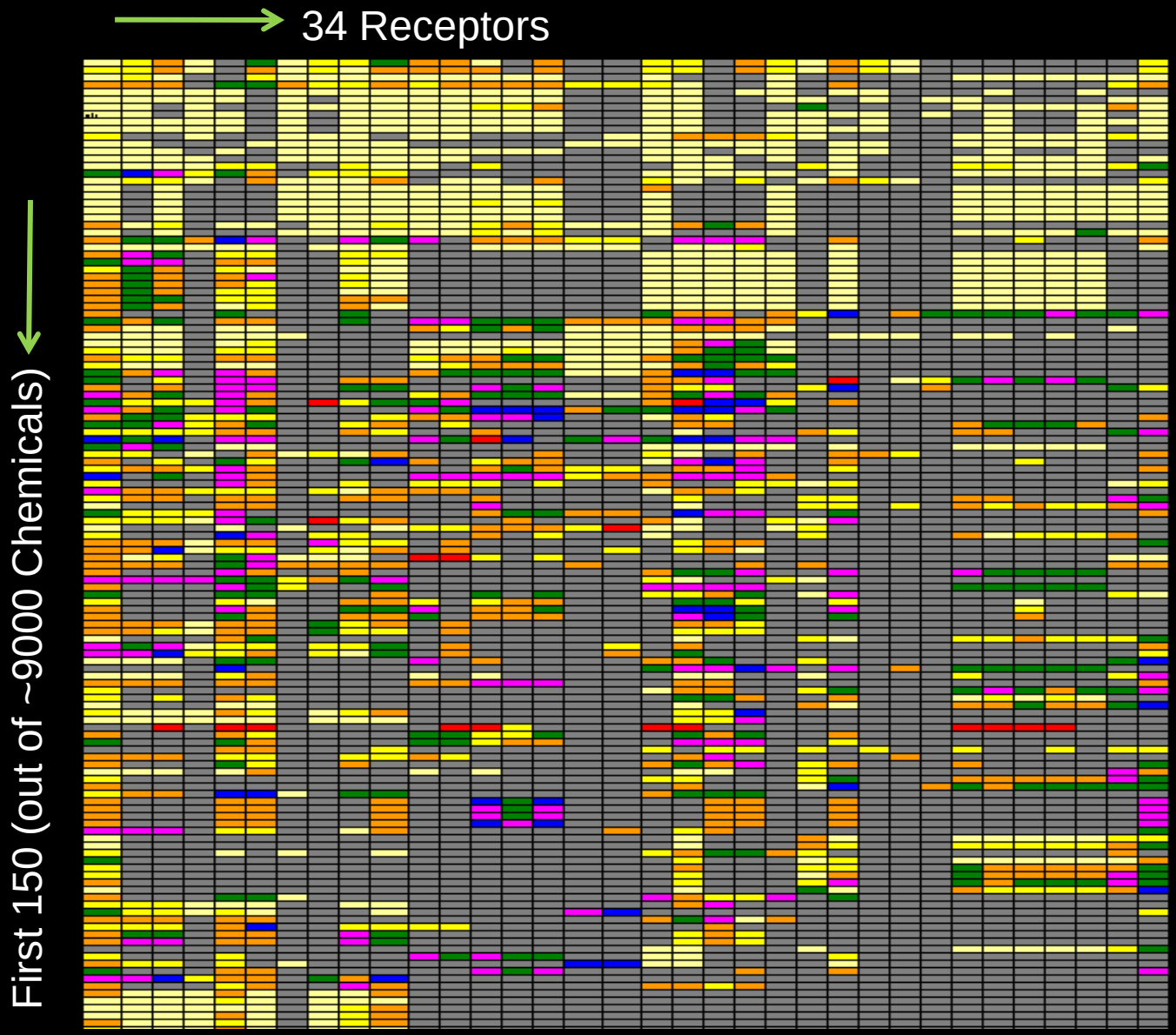
Tc	WDI Compounds	122 VS Hits	10 Tested Hits
≥ 0.9	286	2	2
≥ 0.8	1341	4	3
≥ 0.7	7048	13	8
≥ 0.6	21431	38	9
≥ 0.5	36719	81	9
≥ 0.4	44208	115	10
≥ 0.3	45860	122	10
≥ 0.2	46220	122	10
≥ 0.1	46301	122	10
≥ 0.0	46406	122	10

Tanimoto coefficients (Tc) & 166 MACCS structural keys were used for similarity calculations

Case study 3: QSAR-based virtual receptoromics (QSAR-omics)



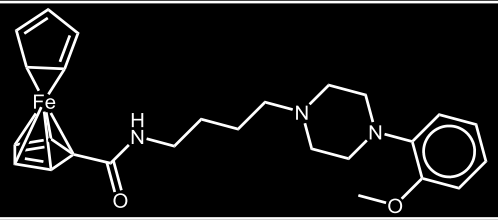
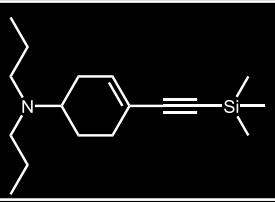
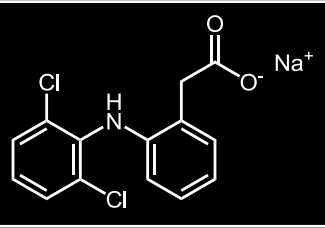
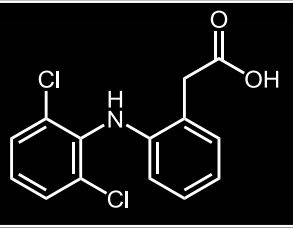
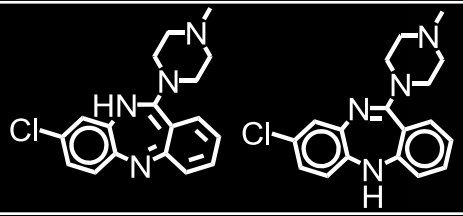
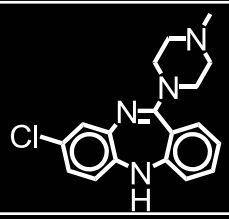
GPCRome Data Matrix: filling the gaps



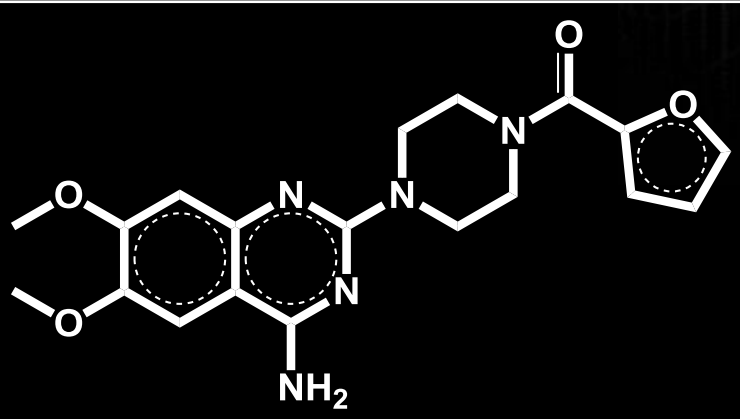
pK_i
Not tested
<5
5-6
6-7
7-8
8-9
9-10
>10

Degree of sparsity = 93.25%
# of tested ligands per receptor: >100

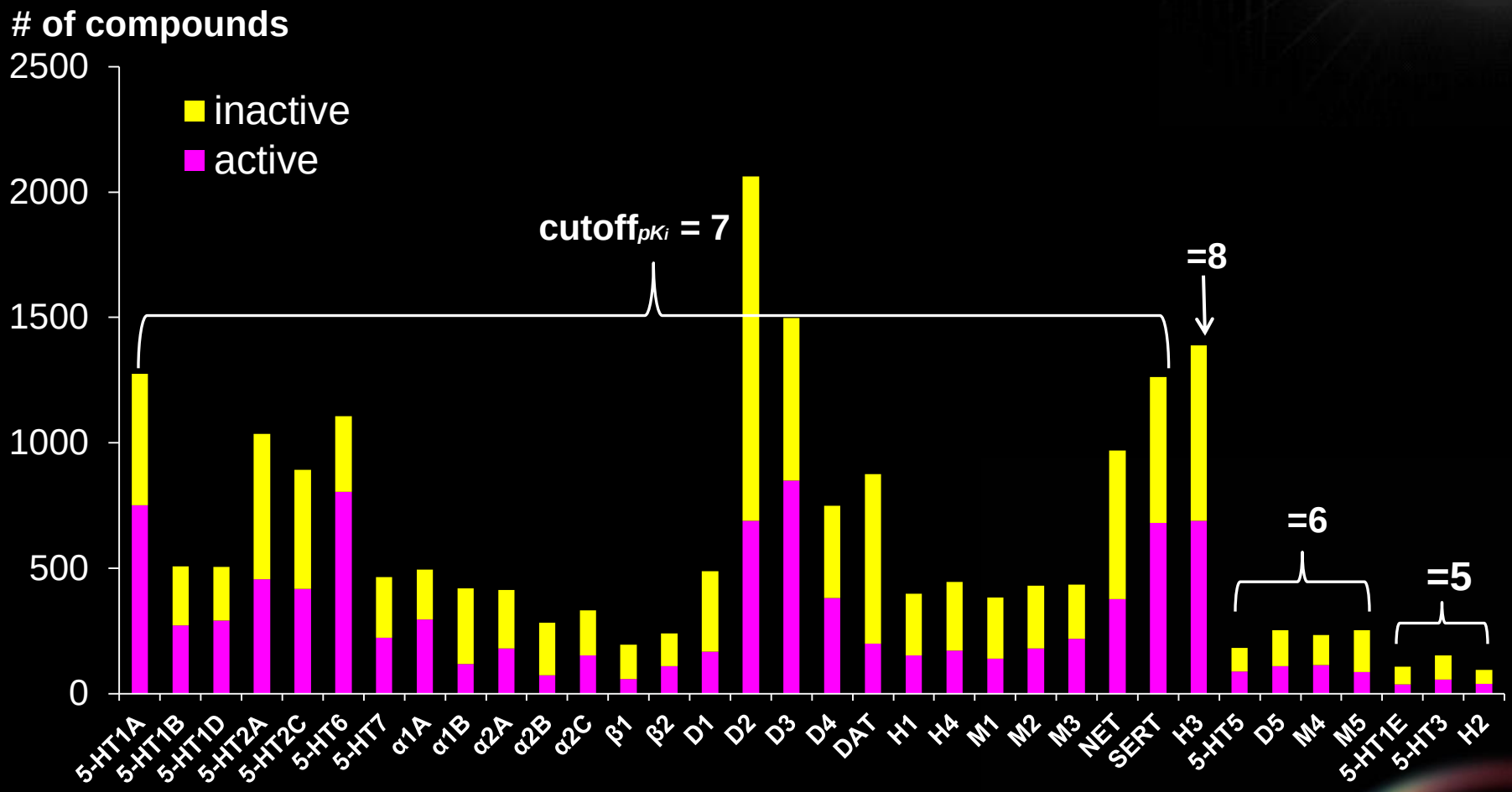
Examples of structure curation

Issues	Source	Before curation	After curation
Organometallics	ChEMBL		Deleted
Organosilicon	PDSP		Deleted
Salts	PDSP		
Tautomers	ChEMBL PDSP		

Biological Data Curation

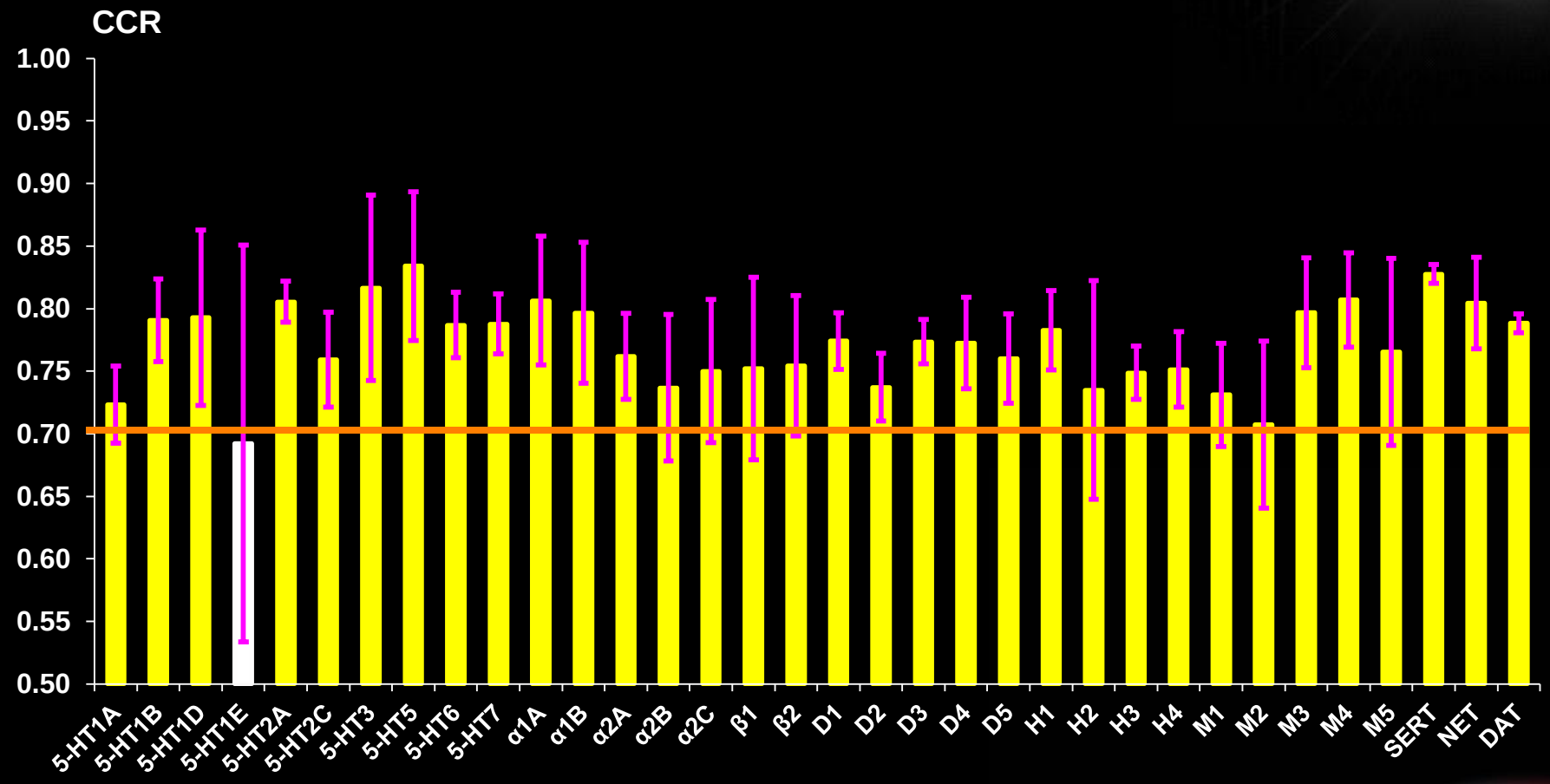
Prazosin			
Targets	5-HT_{2A}	α-1A	D2
Standard Deviation	2.80	0.63	0.4
Assay records (pK_i)	5.15 5.45 [illegible]	9.16 10.22 8.74 8.14 9.29 9.23 9.23	7.24 7.51 7.84 7.97 7.02

34 Datasets: Distribution of Actives and Inactives



Different cutoff values were used to balance the ratio of actives and inactives.

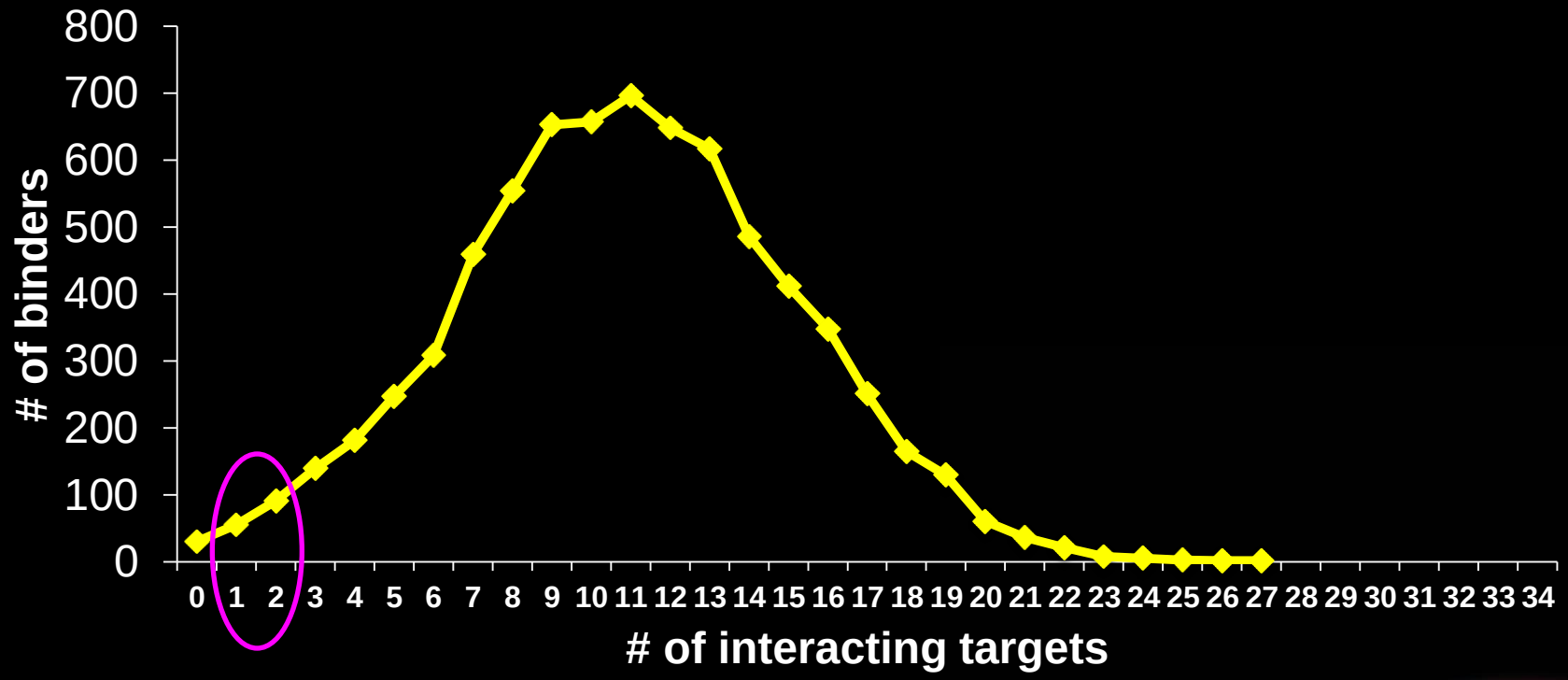
External Prediction Accuracy



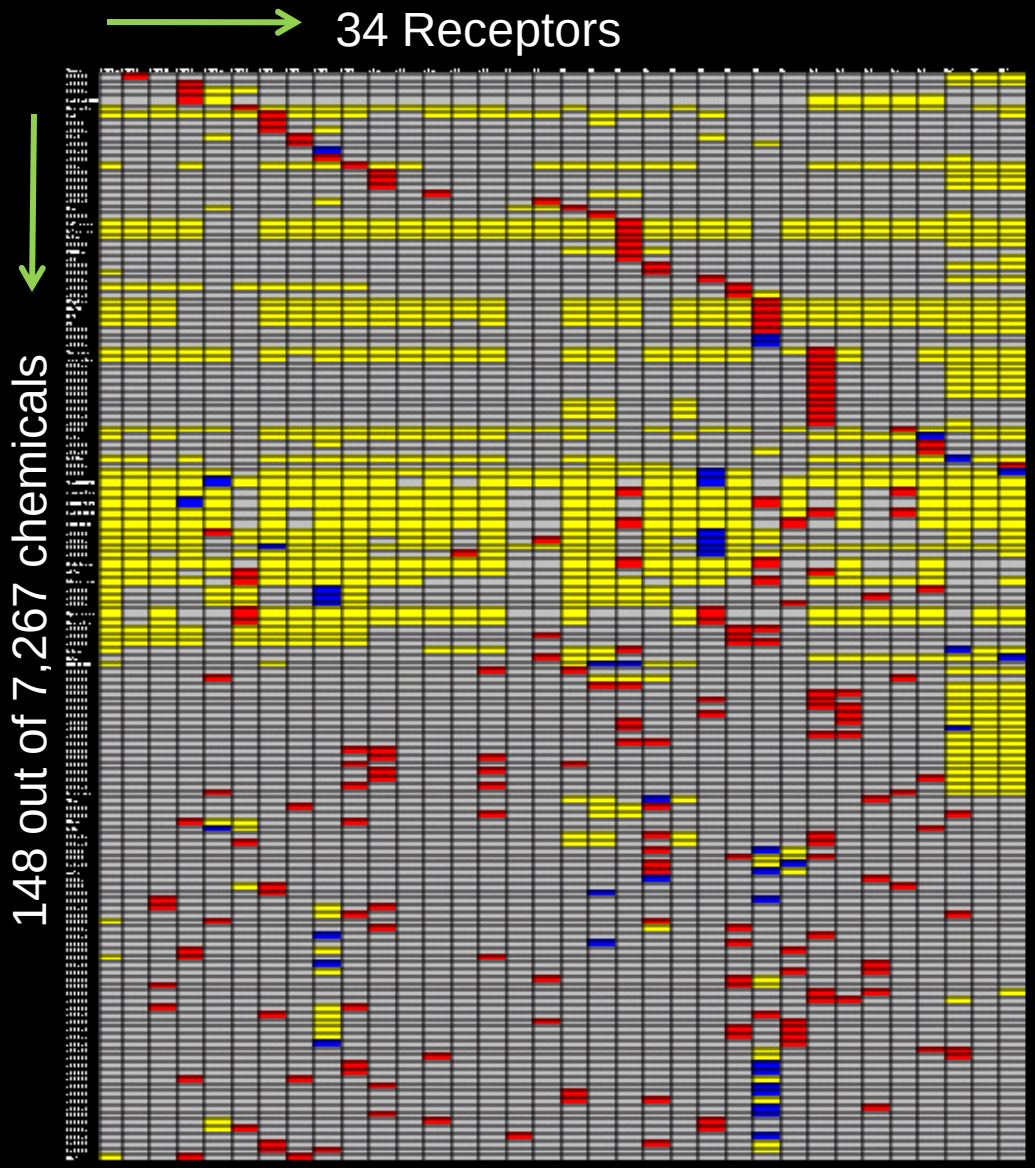
33 out of 34 models have 5-fold external CV cumulative balanced accuracy > 0.7

Binding Promiscuity

- Most compounds are predicted to bind several GPCRs.



Selective Ligands

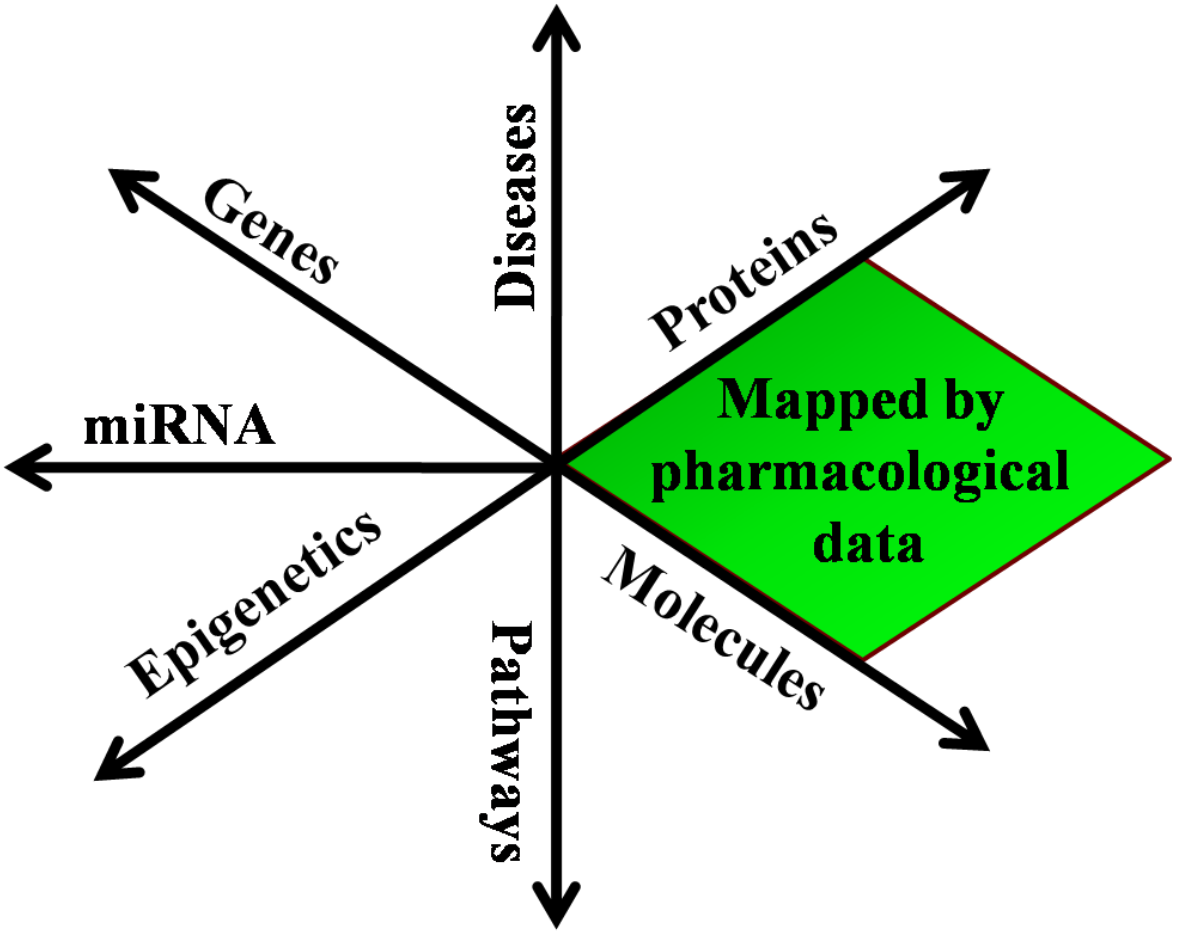


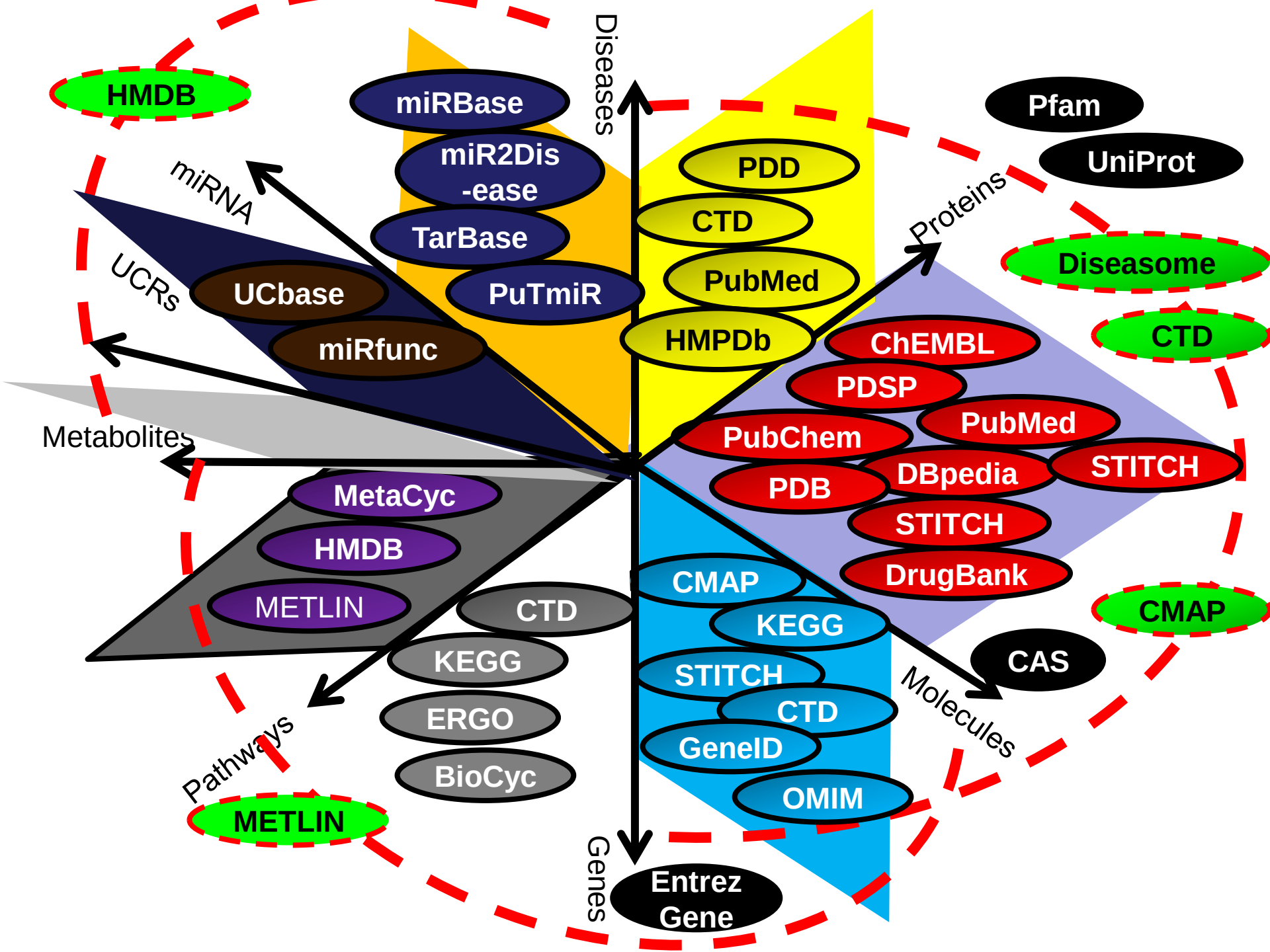
predicted non-binder
experimental non-binder
predicted binder
experimental binder

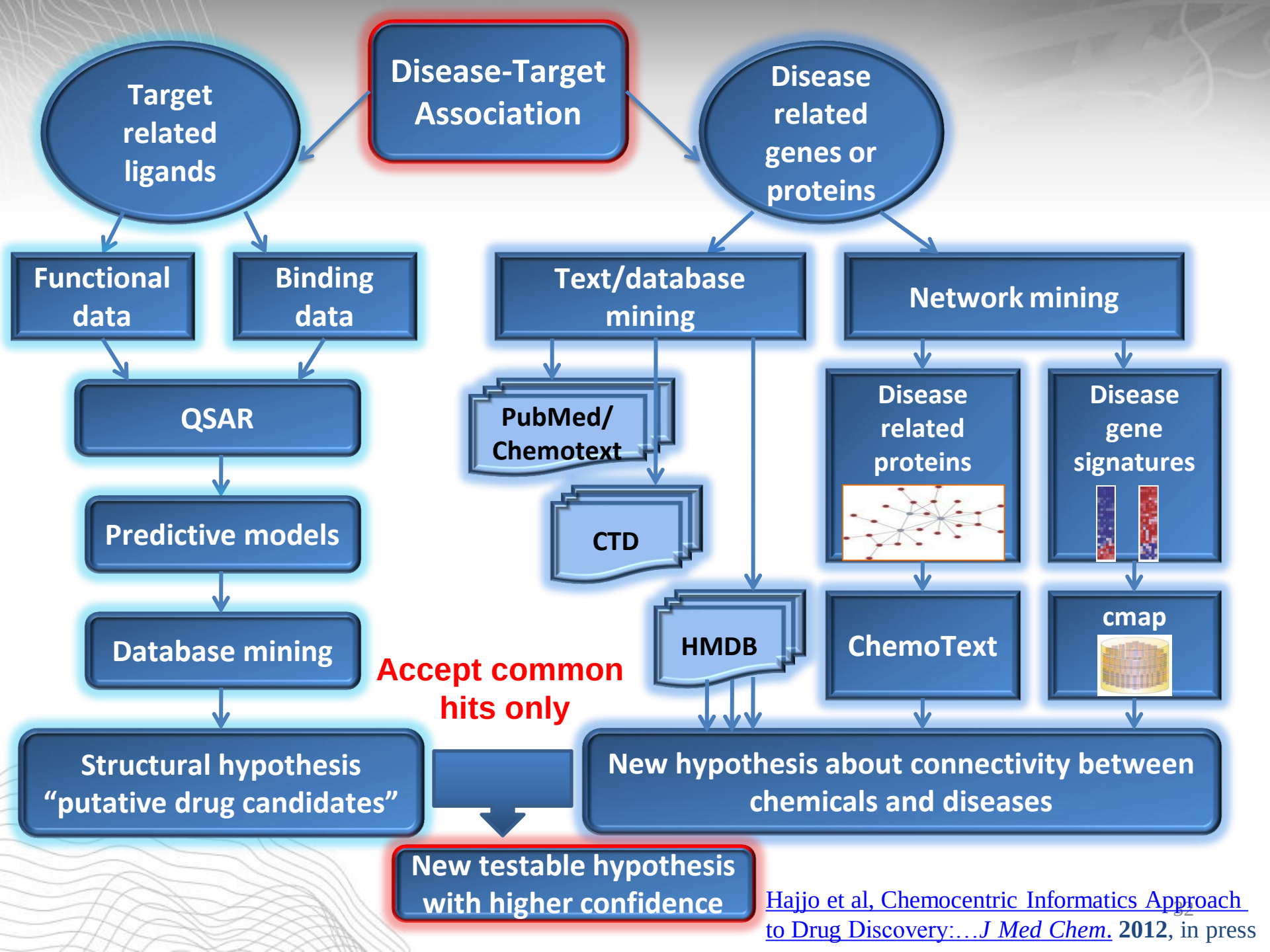
- 148 compounds were identified to bind one or two GPCRs.
  - ✓ 55 selective
  - ✓ 93 dual selective

- These compounds are selected for further experimental investigation in B. Roth lab.

Case study 4: Chemocentric Integrative Informatics? Application to 5HT6 lihgands




HMDB
miRBase
miR2Dis
-ease
TarBase
PuTmiR
miRNA
UCRs
UCbase
miRfunc
Diseases
PDD
CTD
PubMed
HMPDb
Pfam
UniProt
Proteins
Diseasome
CTD
ChEMBL
PDSP
PubMed
PubChem
DBpedia
STITCH
PDB
STITCH
DrugBank
Metabolites
MetaCyc
HMDB
METLIN
CTD
KEGG
ERGO
BioCyc
Pathways
METLIN
CMAP
KEGG
STITCH
CTD
GeneID
OMIM
CAS
CMAP
Molecules
Genes
Entrez
Gene



Hajjo et al, Chemocentric Informatics Approach to Drug Discovery:…*J Med Chem.* **2012**, in press

5-HT$_6$ receptor QSAR models & QSAR-based VS

UNC ESHELMAN SCHOOL OF PHARMACY

Dataset

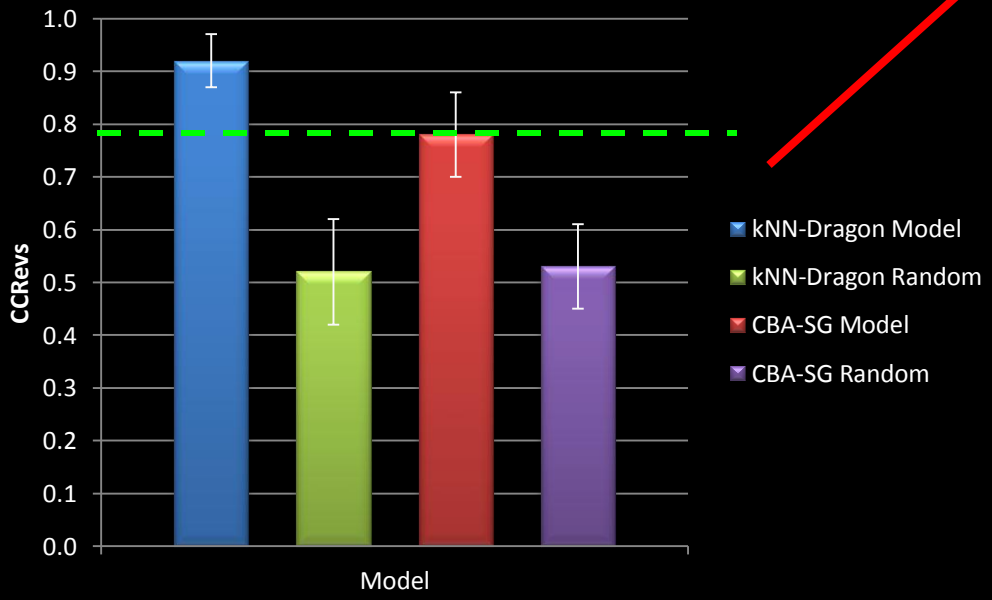
196 cps.

94 Inactives
$K_i \geq 10$ µM

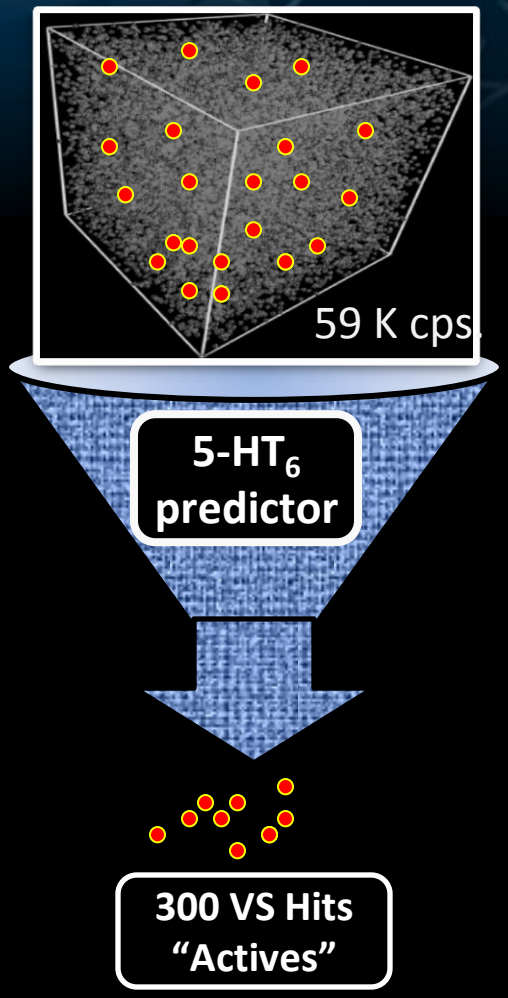
102 Actives
$K_i < 10$ µM

Source: PDSP K_i-DB

Model statistics



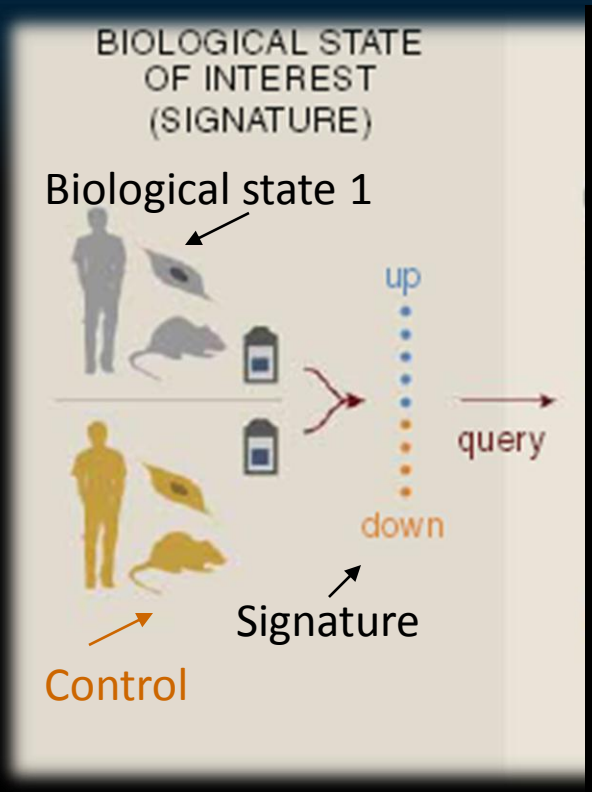
Virtual screening



The connectivity map

UNC ESHELMAN SCHOOL OF PHARMACY

Input **Database** **Output**



Step1: upload signature

Step2: query the cmap

Step3 : list of correlated compounds

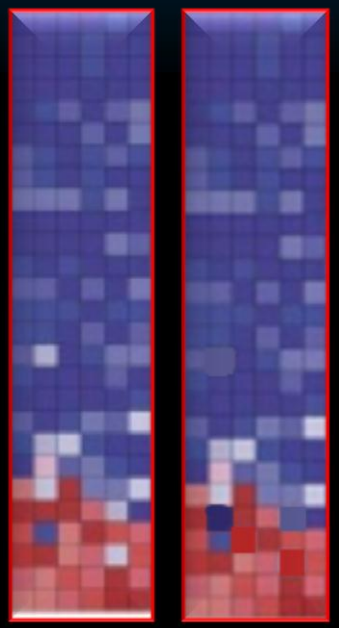
Lamb, J. *et al*. **Science**, 313, 1929-1935 (2006)
Lamb, J. Nature 7, 54-60 (2007)

Querying the cmap

UNC ESHELMAN SCHOOL OF PHARMACY

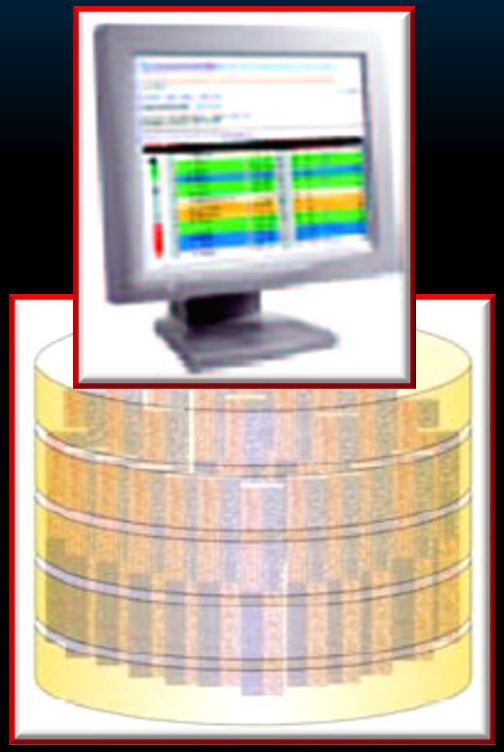
Upload signature

(S1) (S2)



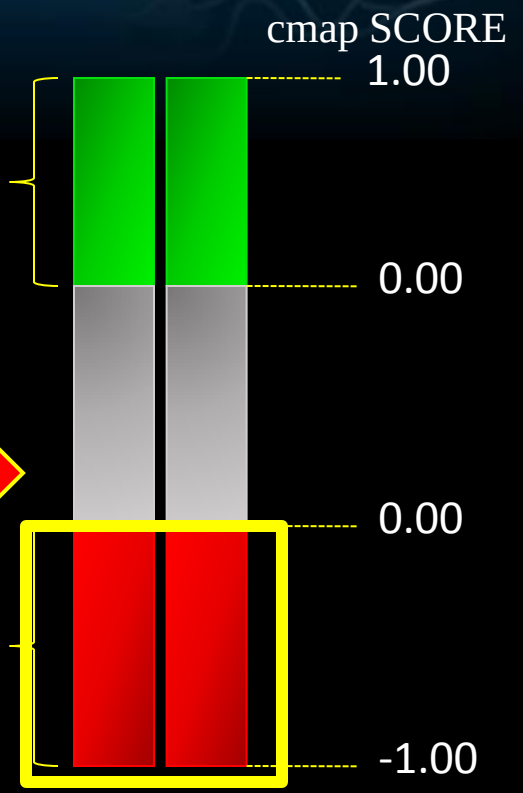
Alzheimer's disease gene signatures

Query the cmap

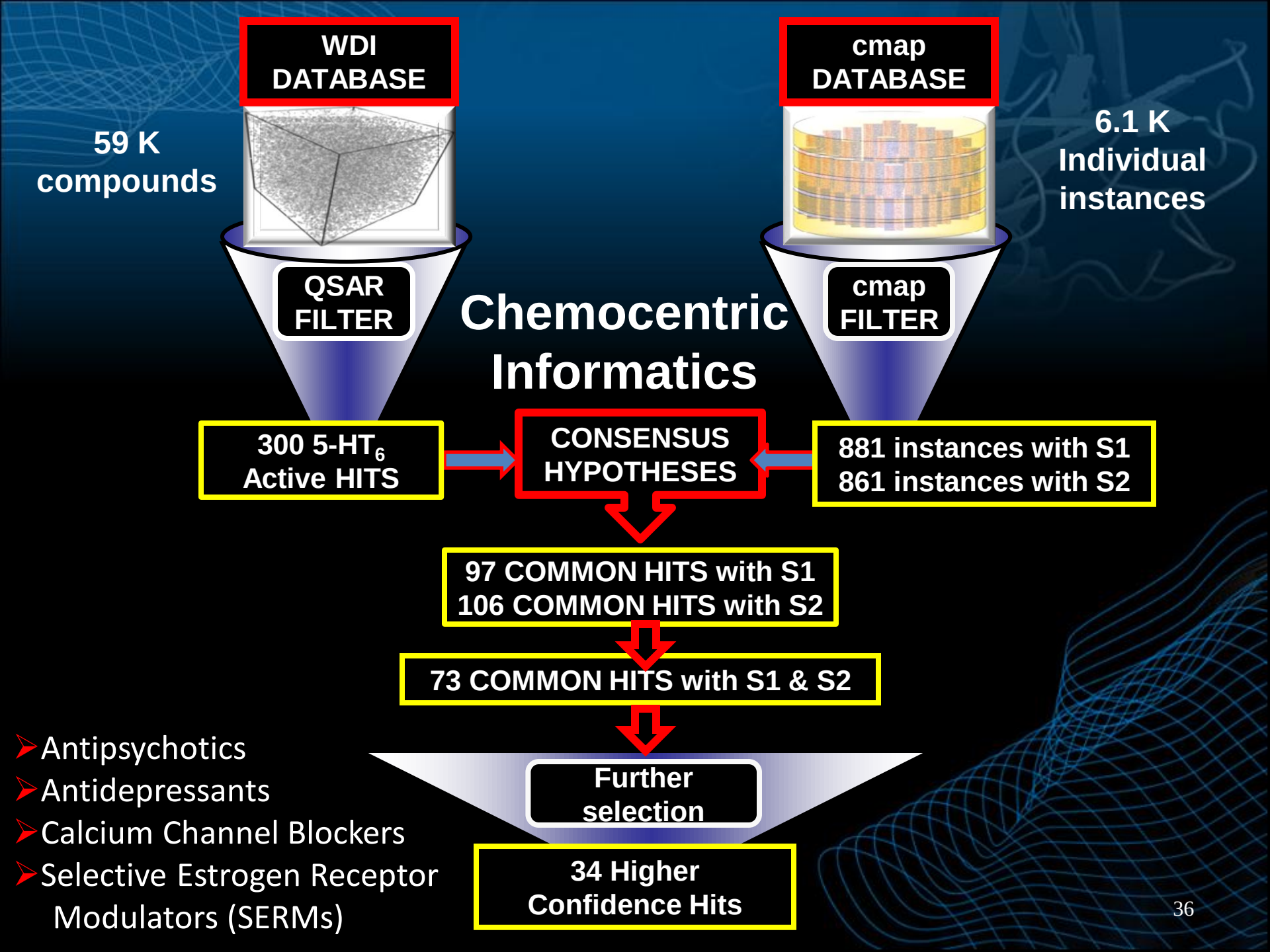


cmap

List of compounds



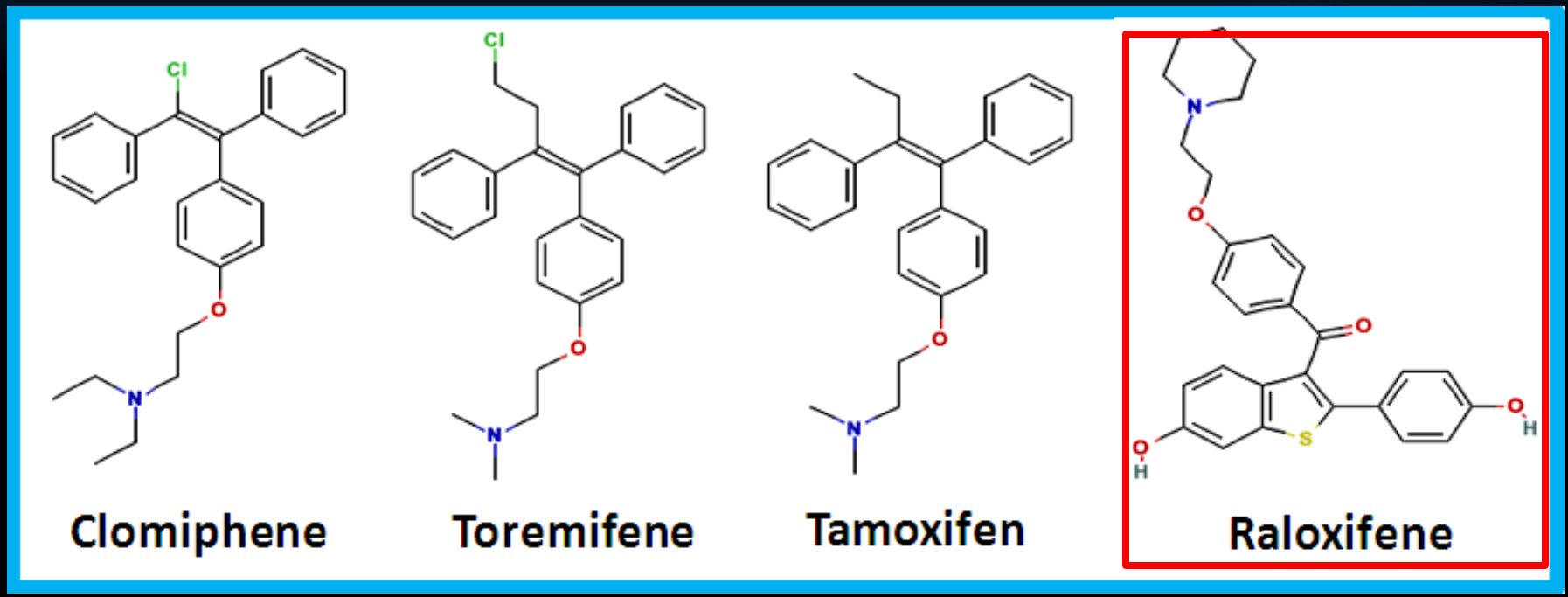
S1: Hata, R. et al., ***Biochem. Biophys. Res. Commun*** 284, 310 (2001).
S2: Ricciarelli, R. et al., ***IUBMB Life*** 56, 349 (2004).



- Antipsychotics
- Antidepressants
- Calcium Channel Blockers
- Selective Estrogen Receptor Modulators (SERMs)

SERMs predicted as 5-HT_6 receptor ligands

UNC ESHELMAN SCHOOL OF PHARMACY



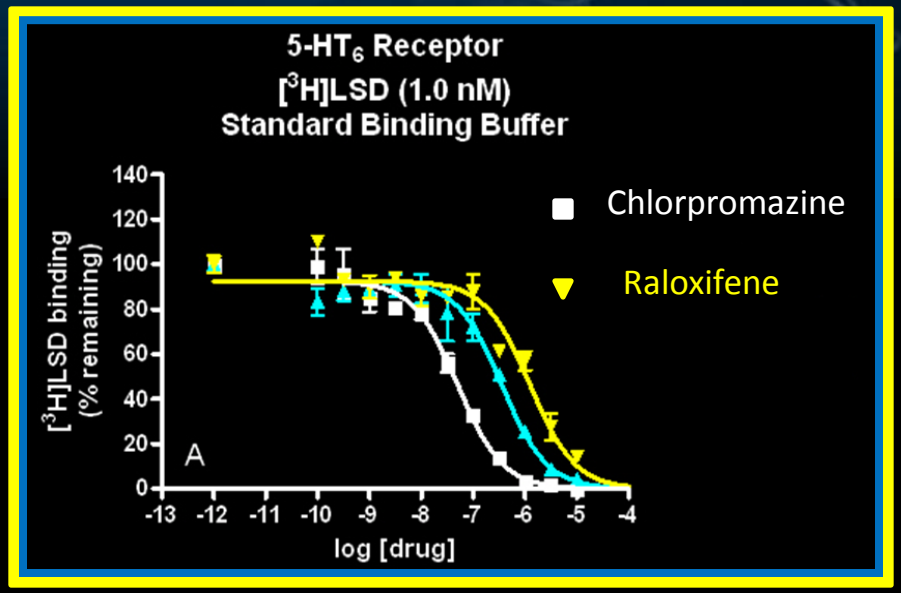
Raloxifene identified as a 5-HT_6 receptor ligand and potential preventative for Alzheimer's disease

UNC ESHELMAN SCHOOL OF PHARMACY

- Raloxifene binds to 5-HT_6 receptor with a K_i**= 750 nM.**
- Raloxifene given at a dose of 120 mg/day led to reduced risk of cognitive impairment in post-menopausal women.

  Yaffe, K. *et al.*, ***Am J Psychiatry,*** 162, 683–690 (2005).
- A newly funded study by NIH is ongoing to evaluate its effects in AD patients.

  http://www.nia.nih.gov/alzheimers/publications/adprevented/



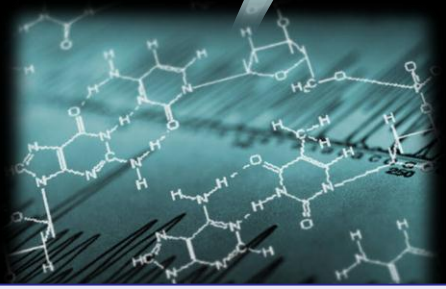
Competition binding at 5-HT_6 receptors for raloxifene (yellow triangle) and chlorpromazine (square) versus [3H] LSD. Tested by our collaborators at PDSP.

Exploration and exploitation of diverse data streams

MML
UNC.EDU

Cheminformatics

Inherent chemical properties



Bioinformatics

Multiple biological assays

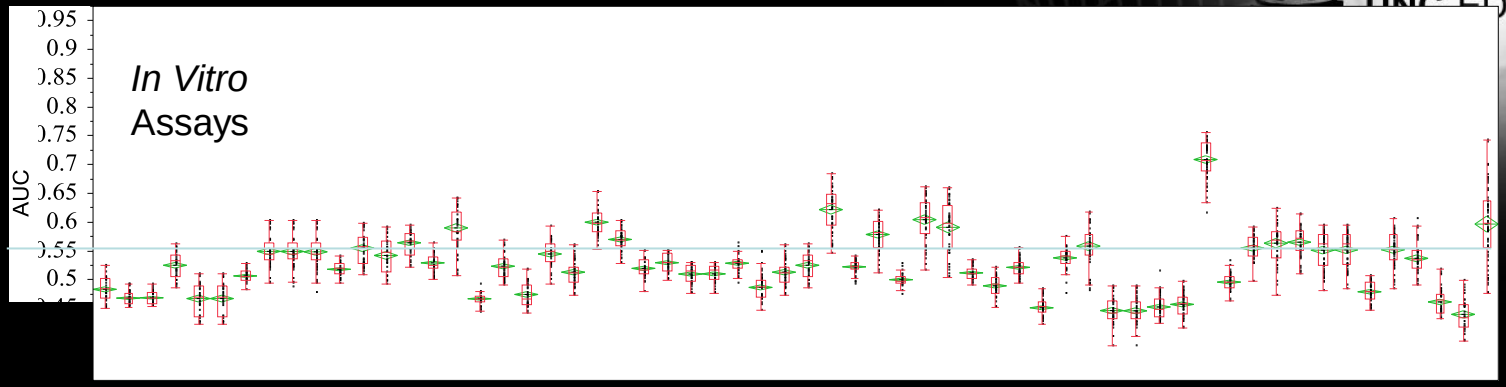
Integrate cheminformatics and short term assay data to improve predictive power and interpretability

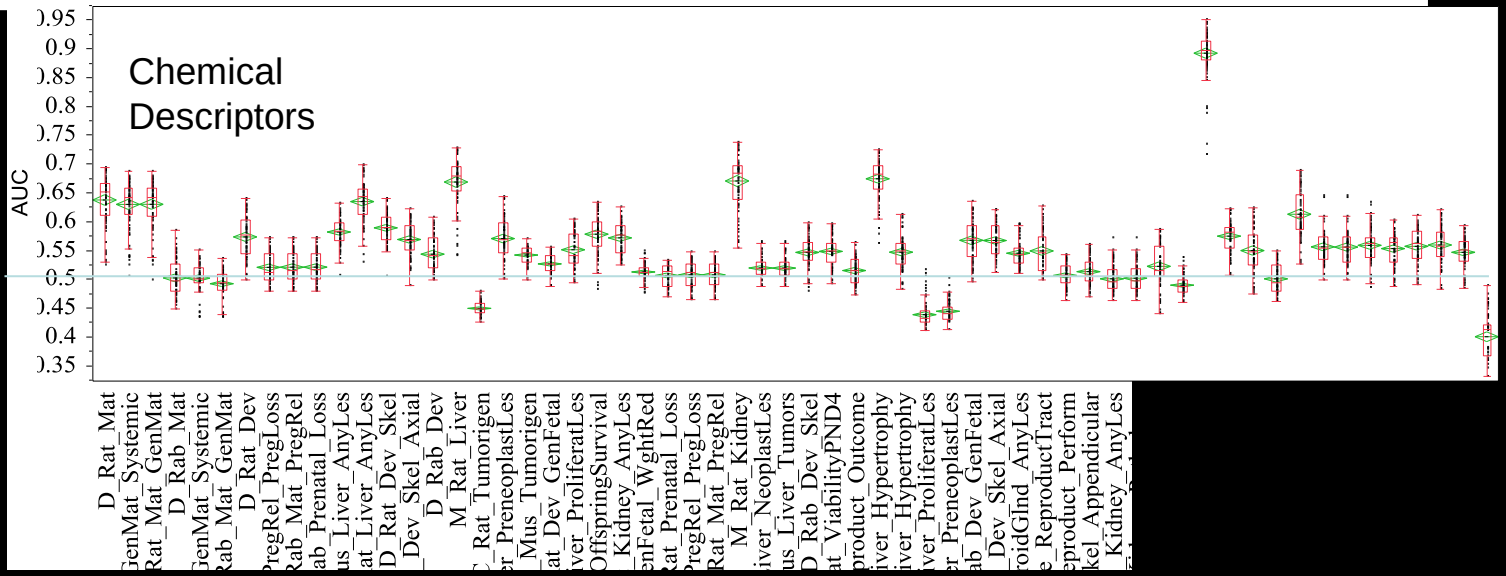
Tox21

NexGen

Human Effects

In vitro data alone cannot explain in vitro effects





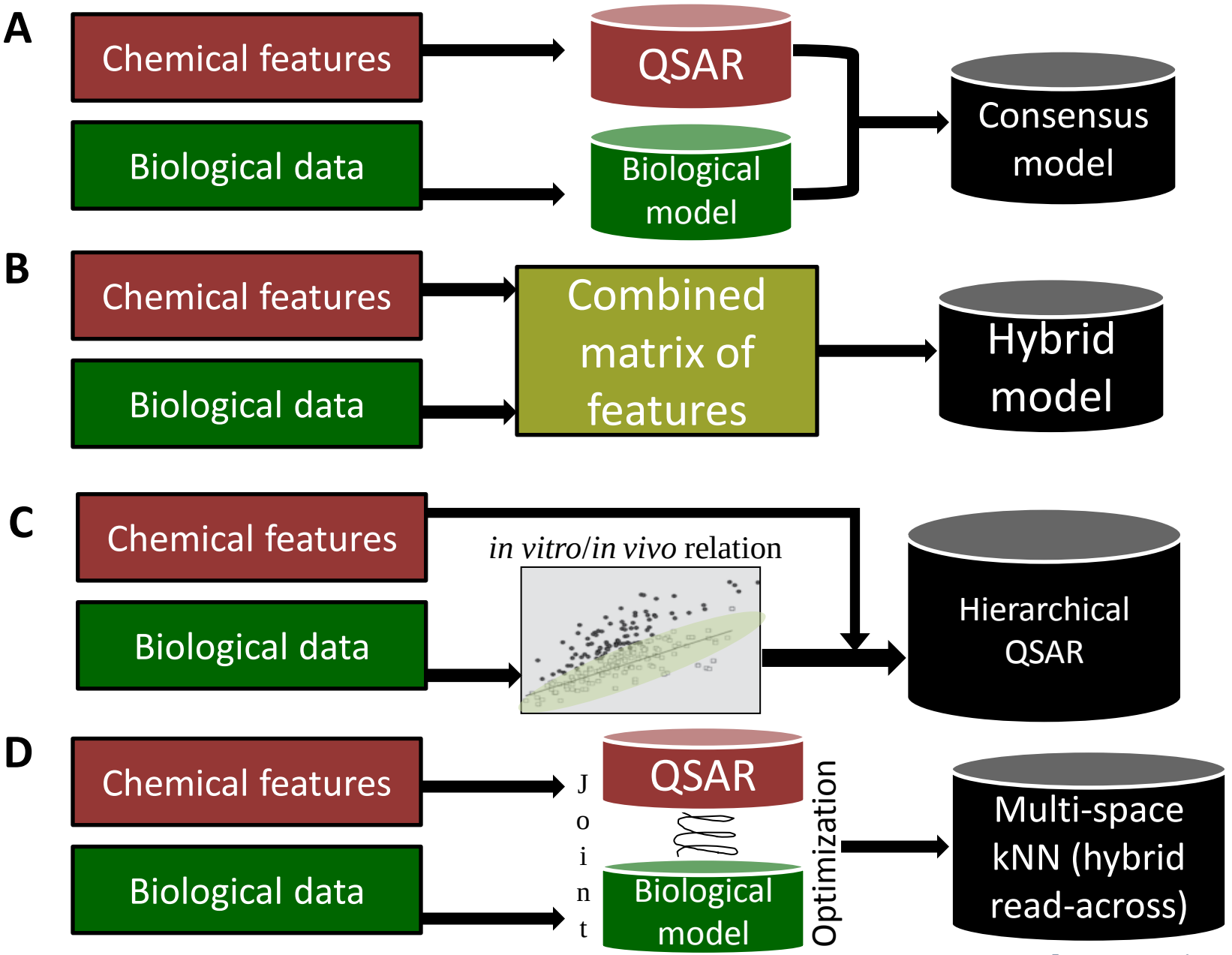
Thomas R., et al. A Comprehensive Statistical Analysis of Predicting In Vivo Hazard Using High-Throughput In Vitro Screening. *Toxicol Sci*. **2012** May 31. [Epub ahead of print]

Emerging approaches combining cheminformatics and short-term assays:

The Use of Biological Screening Data as Additional Biological Descriptors Improves the Prediction Accuracy of Conventional QSAR Models of Chemical Toxicity

- Zhu, H., Rusyn I, Richard A, Tropsha A. Use of cell viability assay data improves the prediction accuracy of conventional quantitative structure-activity relationship models of animal carcinogenicity. *EHP*, 2008, (116): 506-513
- Sedykh A, Zhu H, Tang H, Zhang L, Richard A, Rusyn I, Tropsha A. Use of in vitro HTS-derived concentration-response data as biological descriptors improves the accuracy of QSAR models of in vivo toxicity. *EHP*, 2011, 119(3):364-70.
- Low et al., Predicting drug-induced hepatotoxicity using QSAR and toxicogenomics approaches. *Chem Res Toxicol*. 2011 Aug 15;24(8):1251-62
- Rusyn et al, Predictive modeling of chemical hazard by integrating numerical descriptors of chemical structures and short-term toxicity assay data. *Tox. Sci*., 2012, 127(1):1-9

Approaches to Hybrid QSAR Modeling



Rusyn et al,. *Tox. Sci.*, 2012 127(1):1-9

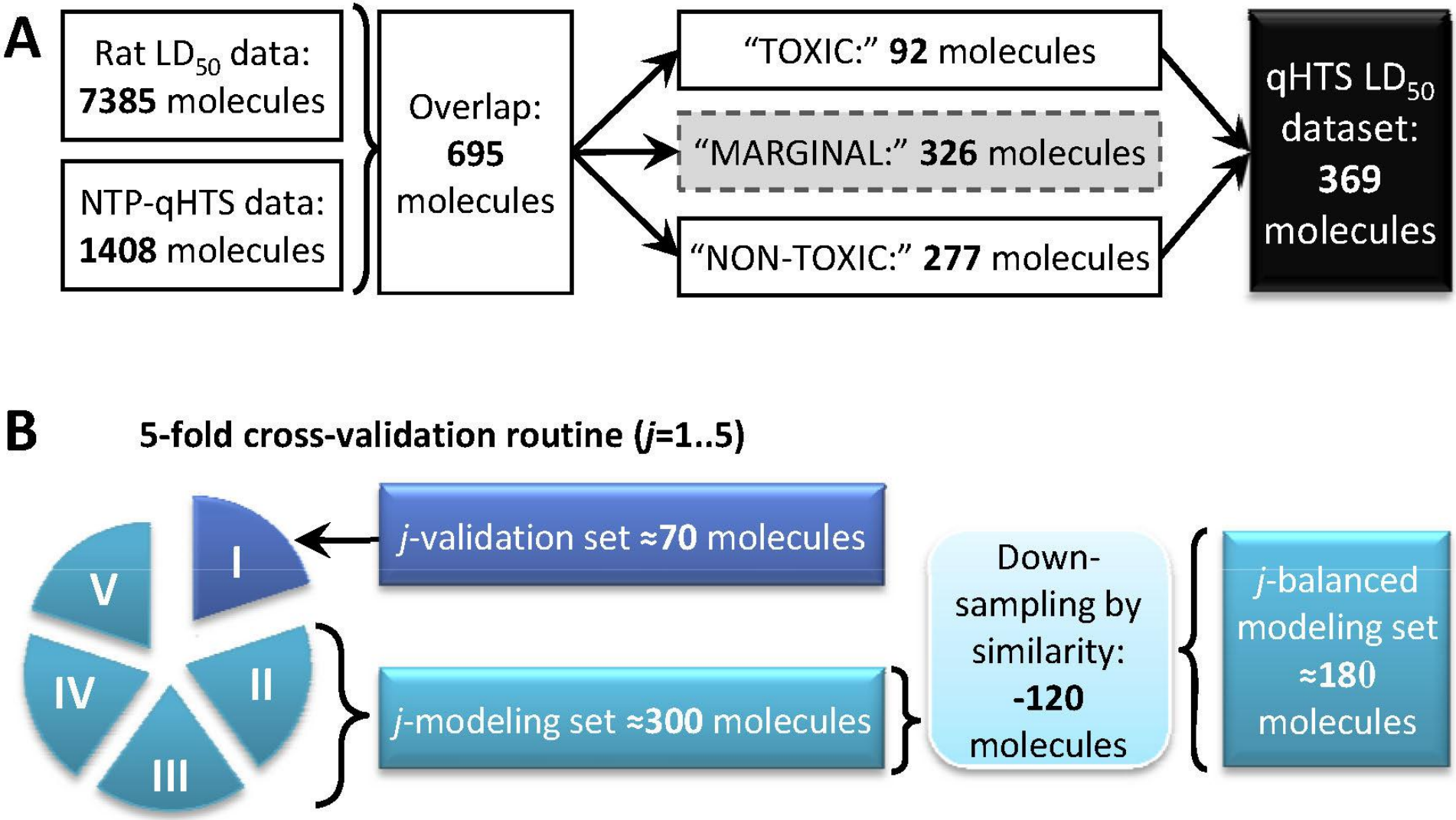
Case study 5. In vitro dose-response data improve the predictive power of QSAR models of in vivo toxicity (rat LD_{50})

- 1408 substances
- 382 chemical structure descriptors (Dragon v5.5)
- 13 *in vitro* NCGC cell viability assays * :
  - qHTS (quantitative HTS) data
  - 14 test concentrations: 0.6nm .. 92.2µm

May yield up to 13x14 = 182 *in vitro* qHTS descriptors, but the issue of data noise becomes important.

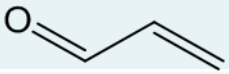
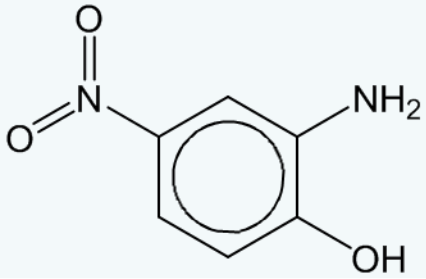
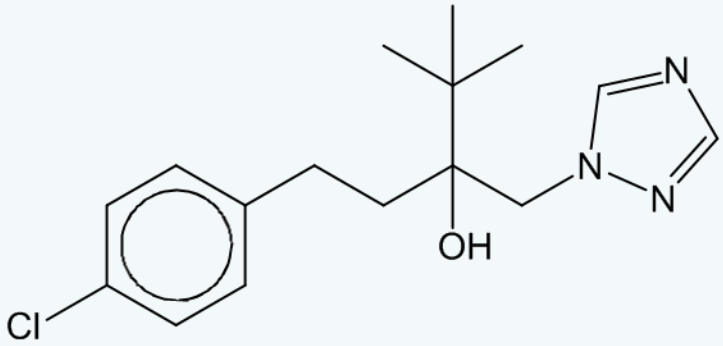
*Inglese J., Douglas S. A. et al. *PNAS,* **2006**, v103(31), p11473

Modeling Workflow

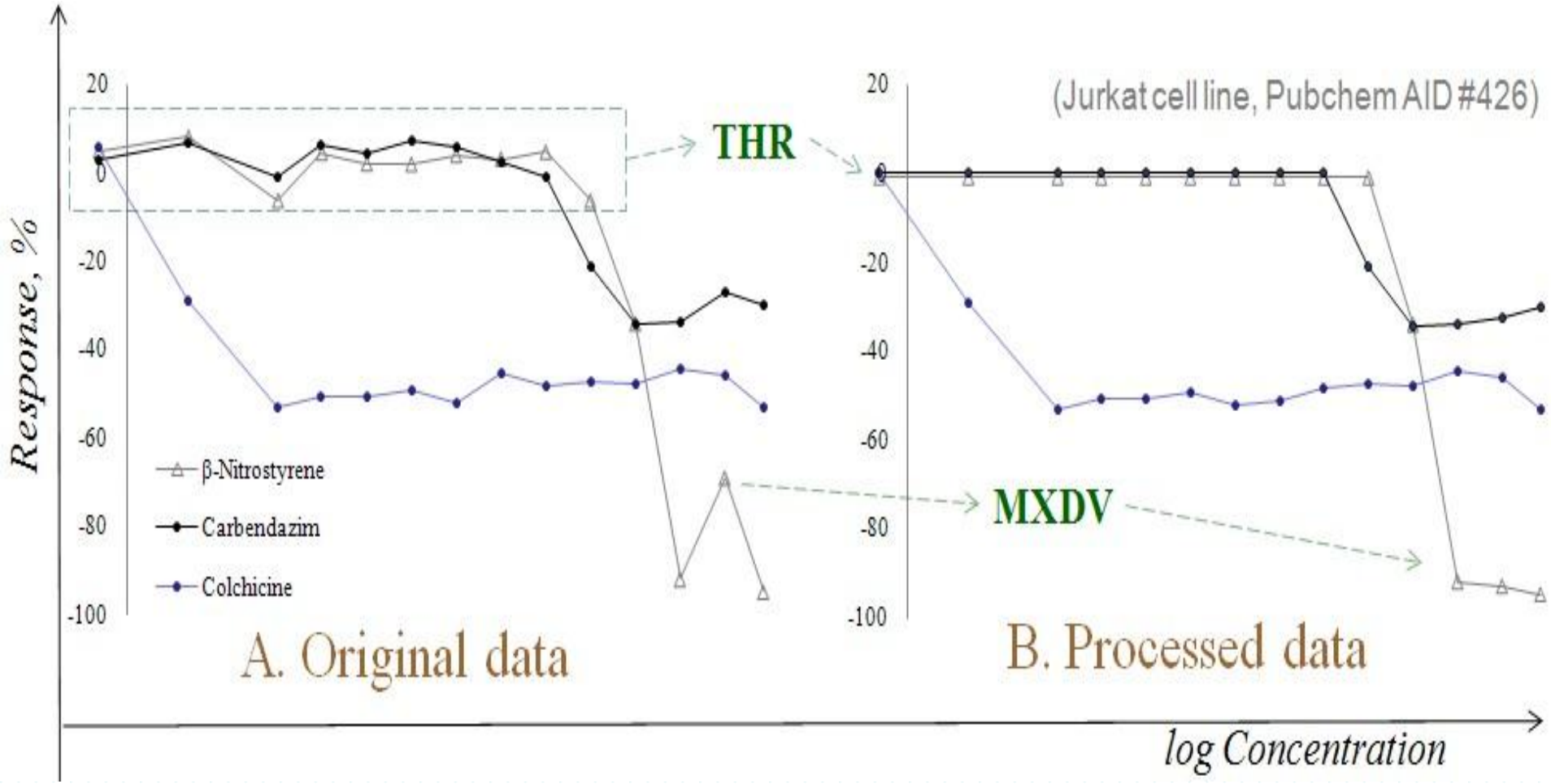


QSAR Table – qHTS descriptors

Descriptor #: 1 2 … 182

ID	Name	Structure	3T3 9.2mkM	3T3 21mkM	…	SHSY 92mkM
1	Acrolein		0	0	…	-92
2	2-Amino-4-nitrophenol		0	-22	…	0
…	…	…	…	…	…	…
369	Tebuco-nazole		-21	-24	…	-18

SMOOTHING CONCENTRATION-RESPONSE CURVES.



Sedykh A, Zhu H, Tang H, Zhang L, Richard A, Rusyn I, Tropsha A. EHP, 2011, 119(3):364-70

Smoothing the concentration-response data improves the prediction accuracy of hybrid models.

	%	Chemical descriptors only	Hybrid descriptors (Original)	Hybrid descriptors (THR=15%)
*k*NN models	*Sensitivity*	68±8	63±9	76±5
	Specificity	85±4	86±4	87±2
	CCR	**76 ±5 ***	**74 ±5**	**82 ±3**
Random Forest (RF) models	*Sensitivity*	74±9	66±8	77±10
	Specificity	82±7	87±4	86±3
	CCR	**78 ±4 ***	**77 ±5**	**82 ±5**

Shown are averaged results of five-fold external validation. *Chemical descriptors only models were significantly different ($p < 0.05$) from all other models of the corresponding group by the permutation test (10,000 times).

Hybrid QSAR models have higher predictive power than commercial software TOPKAT

%	TOPKAT	Chemical descriptors only		Hybrid descriptors (Original)		Hybrid descriptors (THR=15%)	
		kNN	*RF*	*kNN*	*RF*	*kNN*	*RF*
Sensitivity	0.45	0.73	0.73	0.55	0.82	0.91	0.91
Specificity	0.93	0.78	0.80	0.85	0.78	0.85	0.83
CCR	**0.69 ***	**0.75**	**0.77**	**0.70**	**0.80**	**0.88**	**0.87**

Results are shown for 52 compounds in our external validation sets, which were also absent in the TOPKAT training set.

*TOPKAT model was significantly different ($p < 0.05$) from all other models by the permutation test (10,000 times).

Conclusions and Outlook

- Methodology
  - data curation is critical (NB: QSAR models could be used to spot and correct erroneous data!)
  - Rigorous external model validation is mandatory and should precede any mechanistic interpretation
  - Consensus (collaborative!) prediction using all acceptable models affords the highest accuracy and chemical space coverage
  - Novel chemical descriptors for (so far) uncommon substances (mixtures, materials, nanomaterials)
  - outcome: decision support tools for prioritizing compounds for experimental screening and/or regulatory decision making

Conclusions and Outlook

- Emerging trends in QSAR modeling
  - Rapid accumulation of large biomolecular datasets (especially, in public domain)
  - Non-traditional sources of datasets (text mining of biomedical literature, patents, EMRs, …)
  - Extension of QSAR modeling beyond organic molecules (mixtures, materials, nanomaterials, …)
  - Integration of inherent chemical properties with short term biological profiles (biodescriptors) in the context of *structure – in vitro – in vivo* extrapolation
  - Interpretation of significant chemical and biological descriptors emerging from externally validated models to inform the selection or design of effective and safe chemicals

QSAR Modeling: Where have you been, where are you going?

Where have you been?
Where are you going to?
I want to know what is new
I want to go with you
What have you seen?
What do you know that is new?
Where are you going to?
Because I want to go with you

Chris Rea, “The Blue Café” song

Experiment-Assisted Computational Drug Discovery?

Recent examples of experimentally validated QSAR-based predictions

- Anticonvulsants: Shen, M. *et al*, *J. Med. Chem.* **2004**, 47, 2356-2364.
- HIV-1 reverse transcriptase inhibitors: Medina-Franco, J., *et al*, *J. Comput. Aided. Mol. Des*., **2005,** 19, 229–242
- D1 receptor antagonists: Oloff et al, *J. Med. Chem*., **2005,** 48, 7322-32
- Anticancer agents: Zhang *et al*, *J. Comp. Aid. Molec. Des*., **2007**, 21, 97-112.
- AmpC inhibitors: Hsieh, J.-H.. et al, *J. Comp. Aid. Molec. Des*., **2008,** 22(9):593-609
- HDAC inhibitors: Wang, S. *et al*, (JCIM, **2009**, 49, 461-76)
- GGT-I inhibitors: Wang, Peterson, et al (JMC, **2009**, 52(14):4210-20; provisional patent)
- 5Ht2B binders: Hajjo et al, JMC, **2010**, 11;53(21):7573-86
- 5HT6 binders: Hajjo et al, JMC, 2012 (in press)
- 5HT7 binders; 5HT1A ligands, etc...(in preparation)

CHEM BENCH

http://chembench.mml.unc.edu

HOME | MY BENCH | DATASET | MODELING | PREDICTION | CECCR BASE

Toxicity Predictors

These are public predictors useful for toxicity prediction.

Select	Name	Date Created	Modeling Method	Descriptor Type	Description
☐	5HT2B_Binder_DragonkNN	2010-09-16 03:57	KNN	DRAGONH	This predictor contains models generated using Dragon and kNN by R Hajjo; etal in http://dx.doi.org/10.1021/jm100600y. These models built and validated using 304 compounds with binder/non-binder classification defined based on functional assays.
☐	Ames_Genotoxicity_kNN	2011-06-14 15:28	KNN	DRAGONH	
☐	Ames_Genotoxicity_SVM	2011-06-14 15:28	SVM	DRAGONH	
☐	cb101--ld50_369_cdk_RF	2011-08-28 20:46	RANDOMFOREST	UPLOADED CDK	
☐	cb101--ld50_369_hts_RF	2011-09-09 23:03	RANDOMFOREST	UPLOADED HTS	
☐	cb101--ld50_369_hybrid_RF	2011-08-28 20:46	RANDOMFOREST	UPLOADED HYBRID	
☐	cb101--ld50_369_sdf_RF	2011-08-30 11:22	RANDOMFOREST	CDK	
☐	ER_binding_affinity	2011-09-12 14:07	SVM	UPLOADED	
☐	RAT-ACUTE-LD50_DragonkNN	2010-09-23 03:57	KNN	DRAGONH	This predictor contains models generated using Dragon and kNN by H Zhu; etal in http://dx.doi.org/10.1021/tx900189p. These models built and validated using 3472 compounds predict Acute Toxicity (pLD50(mol/kg)) in Rats.
☐	T.Pyriformis	2009-10-09 16:46	KNN	MOLCONNZ	This predictor contains the kNN-MolconnZ models generated by H Zhu; et al in http://dx.doi.org/10.1021/ci700443v. These models built using 983 compounds (644 training/339 external test) predict aquatic toxicity (pIGC50) against Tetrahymena Pyriformis.

The Laboratory for Molecular Modeling

Principal Investigator
Alexander Tropsha

Research Professors
Alexander Golbraikh, **Denis Fourches**, **Eugene Muratov**

Graduate Research Assistants
Andrew Fant,
Stephen Bush,
Yen Low

Postdoctoral Fellows
Aleck Sedykh,
Ashutosh Tripathy
Regina Politi

Former members:
Guiyu Zhao
Rima Hajjo

Adjunct Members
Weifan Zheng, Shubin Liu

MAJOR FUNDING

NIH
- R01-GM66940
- R01-GM068665

EPA (STAR awards)
- RD832720
- RD833825
- RD834999