

LASSO-ing potential pregnane X receptor agonists

**Sean Ekins¹²³ , Antony J. Williams⁴, Zsolt Zsoldos⁵,
Aniko Simon⁵, Orr Ravitz⁵, Valery Tkachenko⁴**

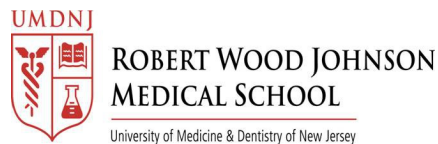
¹ Collaborations in Chemistry, Jenkintown, PA.

² School of Pharmacy, Department of Pharmaceutical Sciences, University of Maryland, Baltimore, MD.

³ Department of Pharmacology, University of Medicine & Dentistry of New Jersey-Robert Wood Johnson Medical School, Piscataway, NJ.

⁴ Royal Society of Chemistry, Wakeforest, NC.

⁵ SimBioSys, Toronto, Canada



PXR agonists

Endogenous

- Bile salts, Cholesterol metabolites

Drugs

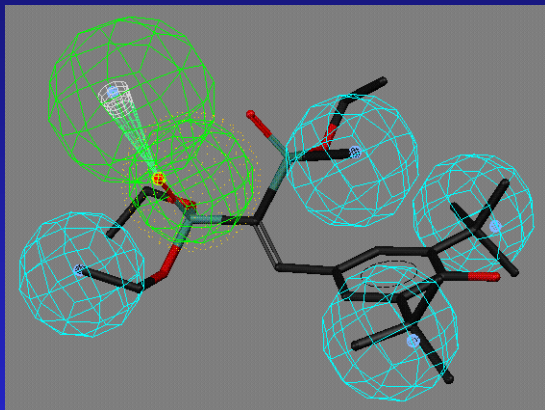
- Statins, PPAR antagonists, Calcium channel modulators
- Synthetic peptide mimetics, Anticancer compounds
- HIV protease inhibitors

Exogenous

- Herbal components, Carotenoids, Vitamins

Environmental Contaminants

- Endocrine disrupters, Pesticides, Plasticizers



Pharmacophore Features

Hydrophobic
Hydrogen bond acceptors
Hydrogen bond donor
(occasionally)

How to avoid interaction with the protein

Attaching hydrogen bonding groups on one of the hydrophobic features, adding larger more rigid groups as well as removing central H-bond acceptors

Growing role for PXR agonists

Interaction between hyperforin in St Johns Wort and irinotecan
= reduces efficacy

Ablating the inflammatory response mediated by exogenous toxins e.g.
inflammatory diseases of the bowel

Cholesterol metabolism pathway control - a negative effect

Mediating blood-brain barrier efflux of drugs modulation of efflux
transporters e.g. mdr1 and mrp2.

decrease retention of CNS drugs e.g. anti-epileptics and pain killers,
decreasing efficacy

PXR induces cell growth and is pro-carcinogenic

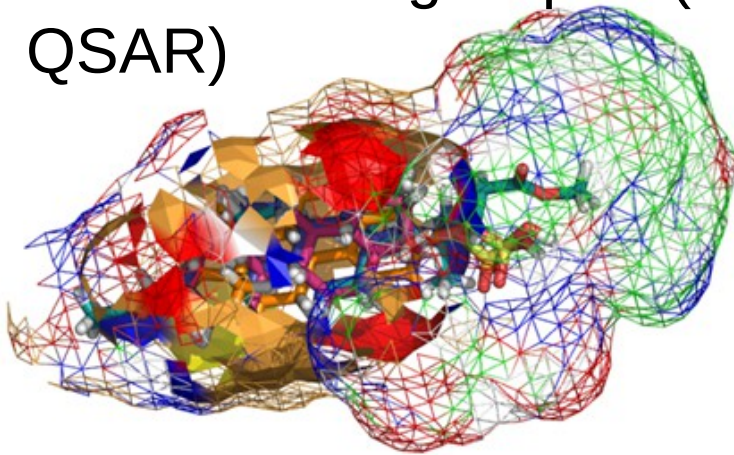
PXR Agonist Machine Learning and Docking Comparison

- Compared different methods for predicting agonist binding
- Training set 98 hPXR activators and 79 hPXR non-activators (Ung et al., *Mol Pharmacol* **71**:158-168 2007)
- Recursive partitioning (RP)
- Random forest (RF)
- Support Vector Machines (SVM)
- Using VolSurf 3D descriptors
- Docking (FlexX)
- Large external test set N = 145 molecules (82 active, 63 inactive)
- Machine learning (63-67% Accuracy) better than docking – did not look at enrichment

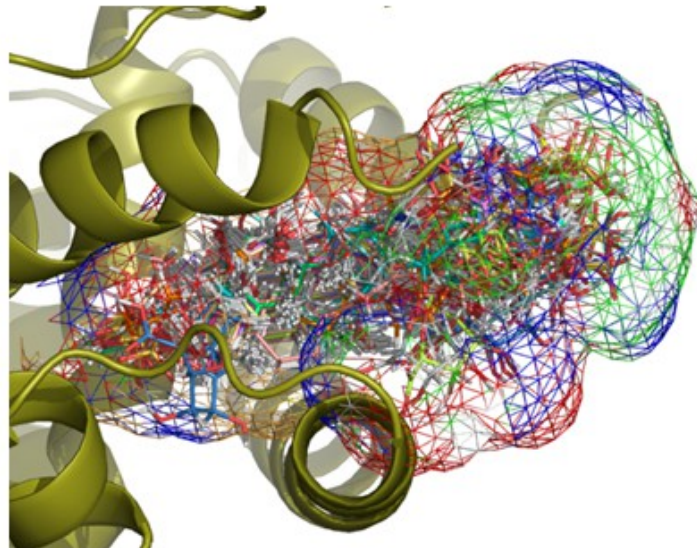
Steroidal compounds

Steroidal Bayesian model

A Receptor model for PXR obtained using Raptor (5D-QSAR)



B



A

 G1: -2126785789 4 out of 5 good Bayesian Score: 0.830	 G2: 2091601364 4 out of 5 good Bayesian Score: 0.830	 G3: -1073865179 4 out of 5 good Bayesian Score: 0.830	 G4: -629400856 4 out of 5 good Bayesian Score: 0.830	 G5: 1821873555 4 out of 5 good Bayesian Score: 0.830
 G6: -525615358 4 out of 5 good Bayesian Score: 0.830	 G7: 635792517 4 out of 5 good Bayesian Score: 0.830	 G8: -528241481 4 out of 5 good Bayesian Score: 0.830	 G9: 1741488496 4 out of 5 good Bayesian Score: 0.830	 G10: 363007390 4 out of 5 good Bayesian Score: 0.830
 G11: 1104405562 7 out of 11 good Bayesian Score: 0.800	 G12: 1618154665 7 out of 11 good Bayesian Score: 0.800	 G13: 1479455918 7 out of 11 good Bayesian Score: 0.800	 G14: 203677720 7 out of 11 good Bayesian Score: 0.800	 G15: 907007053 7 out of 11 good Bayesian Score: 0.800
 G16: 1186303932 7 out of 11 good Bayesian Score: 0.800	 G17: -1441604540 7 out of 11 good Bayesian Score: 0.800	 G18: -2002183168 7 out of 11 good Bayesian Score: 0.800	 G19: -453677277 7 out of 11 good Bayesian Score: 0.800	 G20: 16 7 out of 11 good Bayesian Score: 0.800

B

 B1: 1251059317 0 out of 27 good Bayesian Score: -1.997	 B2: -376420834 0 out of 22 good Bayesian Score: -1.823	 B3: 977703876 0 out of 22 good Bayesian Score: -1.823	 B4: 712326478 0 out of 22 good Bayesian Score: -1.823	 B5: -247917990 0 out of 22 good Bayesian Score: -1.823
 B6: -1980251006 0 out of 22 good Bayesian Score: -1.823	 B7: -2109497546 0 out of 22 good Bayesian Score: -1.823	 B8: -1946897685 0 out of 22 good Bayesian Score: -1.823	 B9: -1447052923 0 out of 22 good Bayesian Score: -1.823	 B10: 1089446752 0 out of 19 good Bayesian Score: -1.701
 B1: 1251059317 0 out of 27 good Bayesian Score: -1.997	 B2: -376420834 0 out of 22 good Bayesian Score: -1.823	 B3: 977703876 0 out of 22 good Bayesian Score: -1.823	 B4: 712326478 0 out of 22 good Bayesian Score: -1.823	 B5: -247917990 0 out of 22 good Bayesian Score: -1.823
 B6: -1980251006 0 out of 22 good Bayesian Score: -1.823	 B7: -2109497546 0 out of 22 good Bayesian Score: -1.823	 B8: -1946897685 0 out of 22 good Bayesian Score: -1.823	 B9: -1447052923 0 out of 22 good Bayesian Score: -1.823	 B10: 1089446752 0 out of 19 good Bayesian Score: -1.701

Introduction to LASSO



- What's in a name ?
 - ◆ Spawned from the eHiTS docking / scoring
 - ◆ Ligand Activity in Surface Similarity Order
- Lasso is a ligand-based vHTS tool that is focused on similarity of biomolecular activity rather than structural similarity
- Key component: Interacting Surface Point Type (ISPT) molecular descriptor

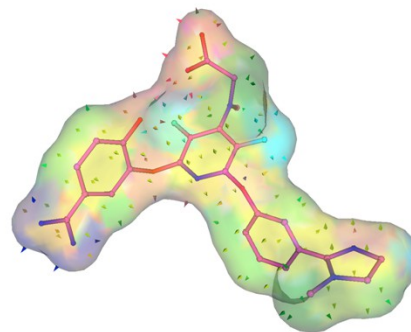
LASSO Ligand Activity by Surface Similarity Order

LASSO uses 23 kinds of Interactive Surface Point Descriptors and

- ◆ is conformation independent
- ◆ screens at 1 million structures/min
- ◆ provides scaffold hopping

ISPT descriptor for 1FJS ligand:

0	4	0	0	1	0	4	6	1	0	0	0	8	8	0	0	23	5	2	2	0	6	0
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---

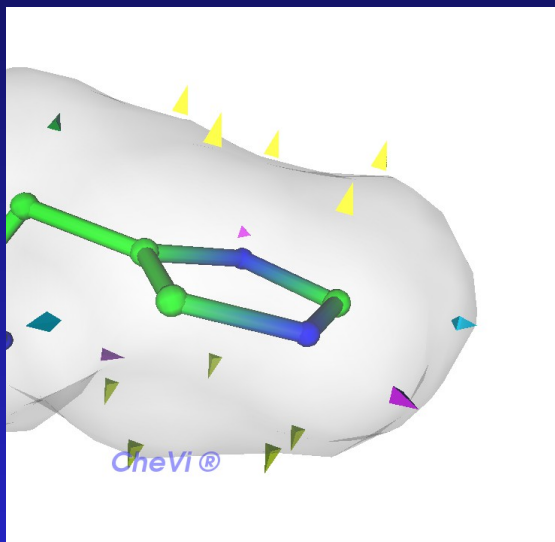


The surface point types (SPT), indicated by triangles are shown on the histidine ring.

Each SPT has associated chemical properties (indicated by their color), such as H-bond donor, H-bond acceptor, hydrophobic, π -stacking, etc.

The count for each of the 23 surface point types creates the feature vector for that ligand.

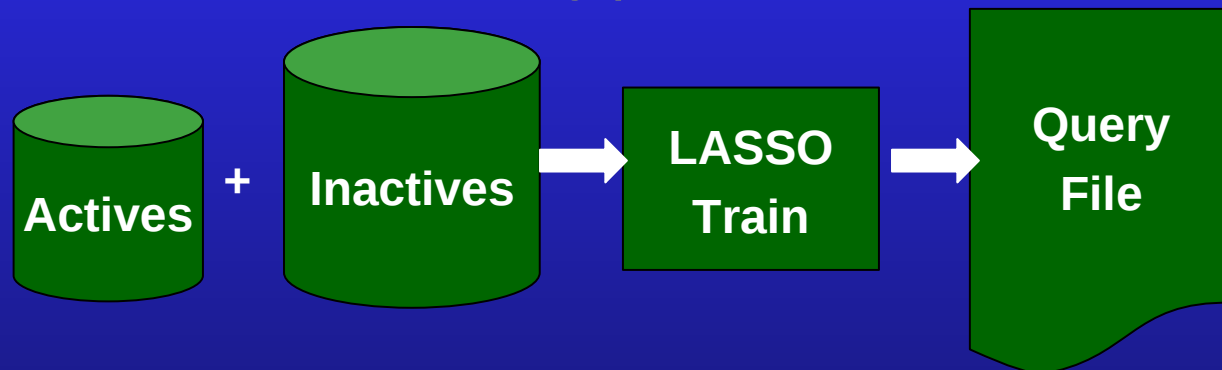
Assumption that ligands with similar feature vectors have similar activity.



LASSO Model building process

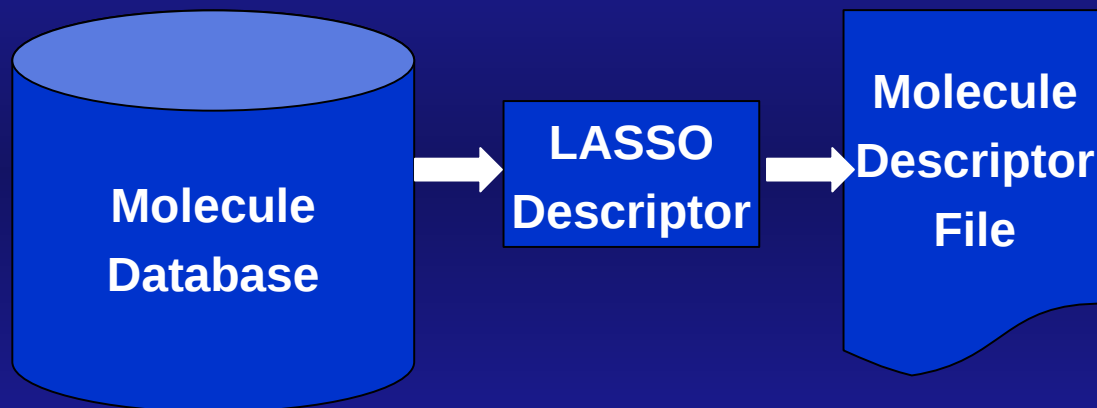
Step 1:

Create a query file using known active ligands and inactive molecules



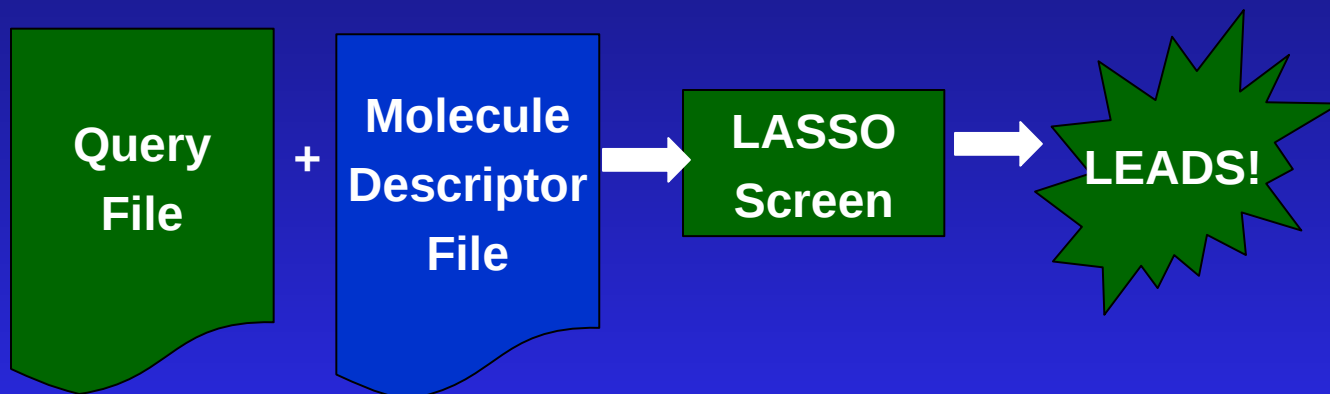
Step 2:

Create a LASSO descriptor file of the database you wish to screen



Step 3:

Screen your database (descriptor file) using your knowledge base to find new leads



PXR datasets

- 1 – hpxr_test 80 actives and 64 inactives (drug-like)
 - 2 – hpxr_train 93 actives and 75 inactives (drug-like)
 - 3 – PXR119 30 actives and 89 decoys (steroids)
-
- Used actives to train models and test on remaining sets combined with decoys
 - Then combined all actives (N = 203) and took increasing numbers and combined with decoys as well
 - Actives = $EC_{50} < 10 \mu\text{M}$

Ekins S, et al PLoS Comput Biol 5(12): e1000594, (2009)
Khandelwal et al., Chem Res Toxicol, 21:1457-67 (2008) .

Model building Phase I

7 models were built using JUST the ACTIVES:

Model 1:

trained on 1st dataset (hpxr_test); tested with the other two.

Model 2:

trained on 2nd dataset (hpxr_train_with-activity); tested with the other two.

Model 3:

trained on 3rd dataset (PXR119-class); tested with the other two.

Model 4-6:

trained on (1+2, 1+3, or 2+3); tested with the remaining one.

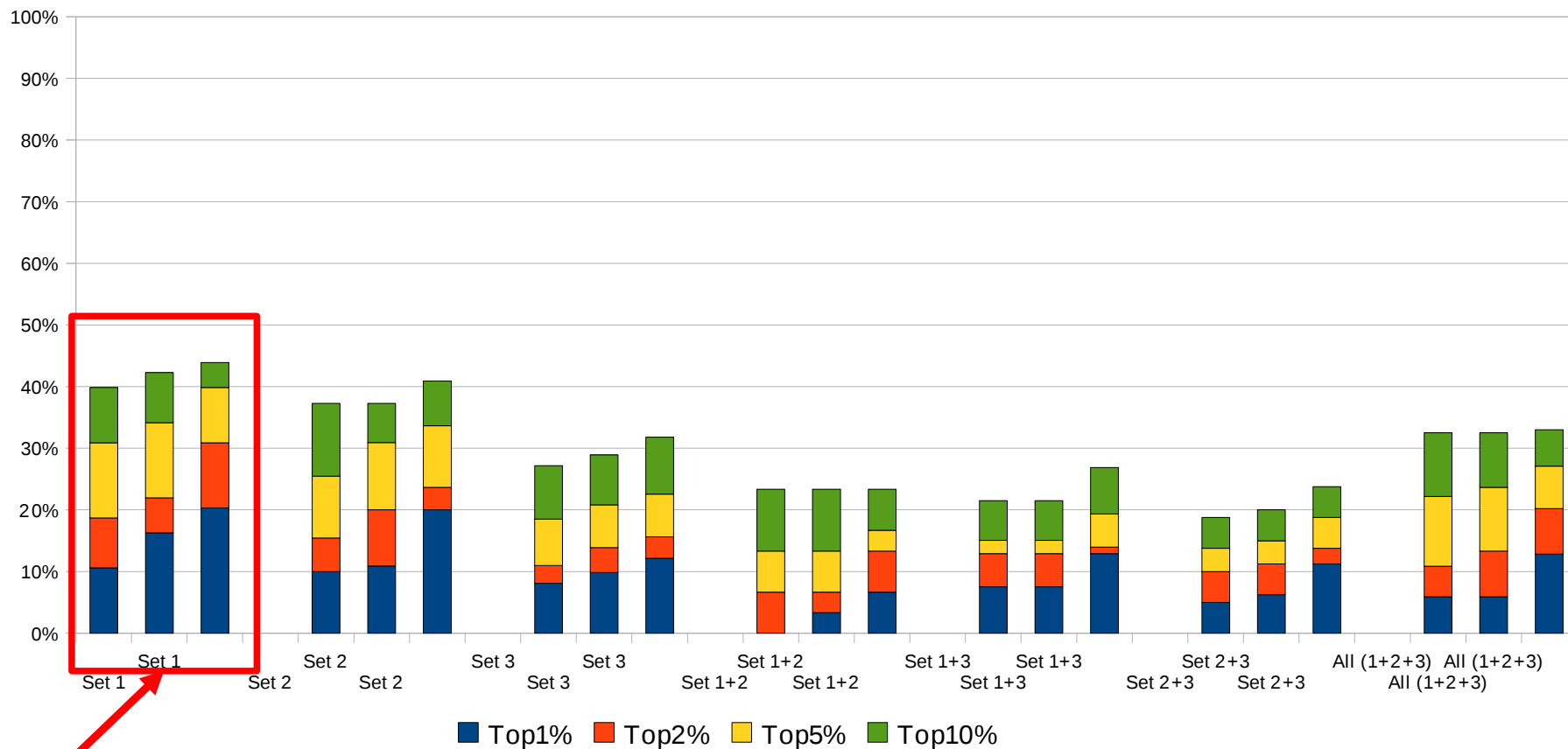
Model 7:

trained on ALL (1+2+3); tested on the SAME

PXR LASSO models - phase I

Lasso 7 models

tested with 3k, 8k and 24k drug-like decoys



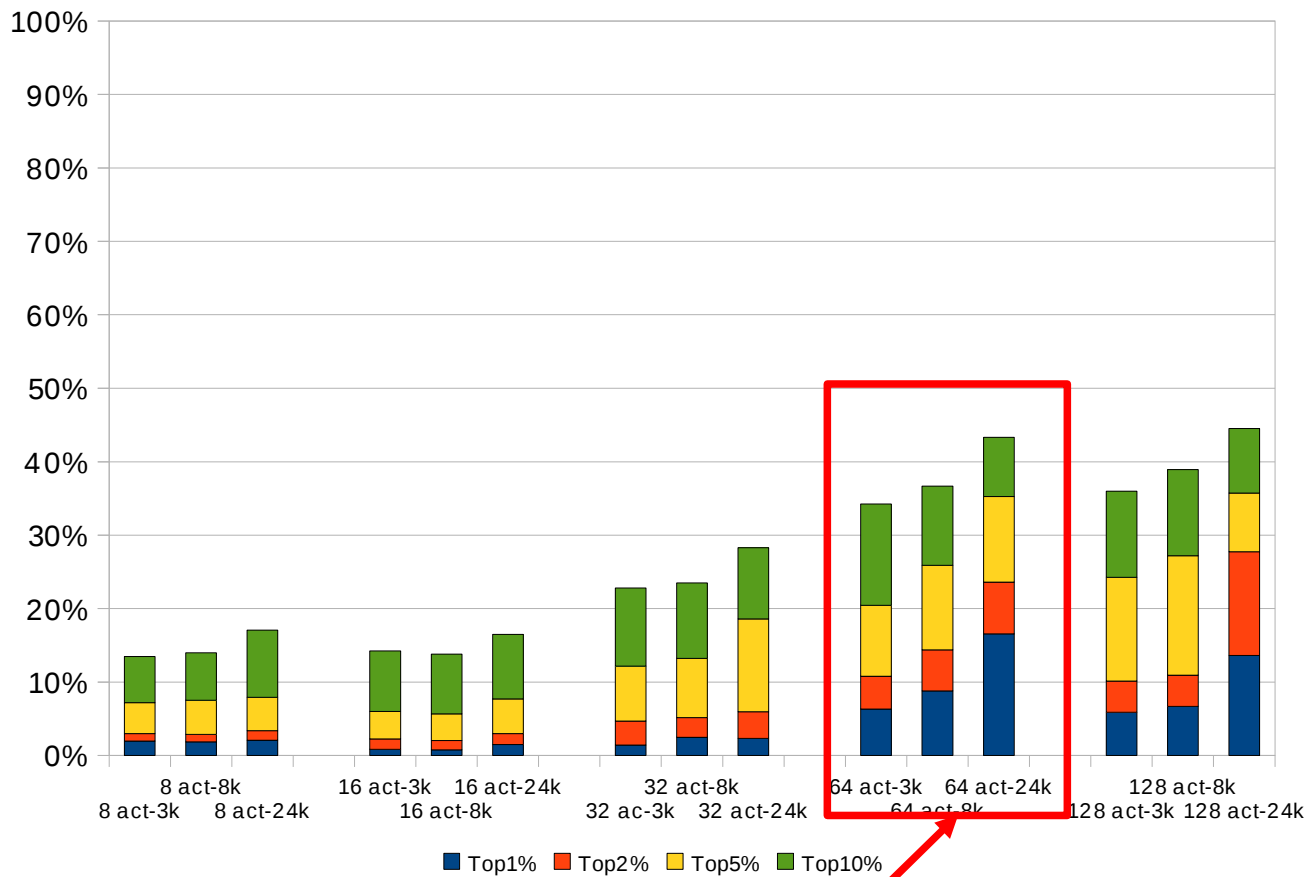
Model building Phase II

- The Actives from the 3 datasets were all merged, resulting in an SDF file with 203 ligands.
- 25 models were built by:
 - Selecting 8, 16, 32, 64, and 128 actives
 - Starting from positions: 1st, 9th, 17th, 33rd, 64th position in the merged ACTIVES file
- The models were tested for enrichment factor using:
 - Actives: the total active set minus the ones used for training
 - Decoys: 3k, 8k and 24k drug-like decoys, in each case the DECOYS (228 structures) from the provided set was merged into the decoy set.

PXR LASSO Phase II

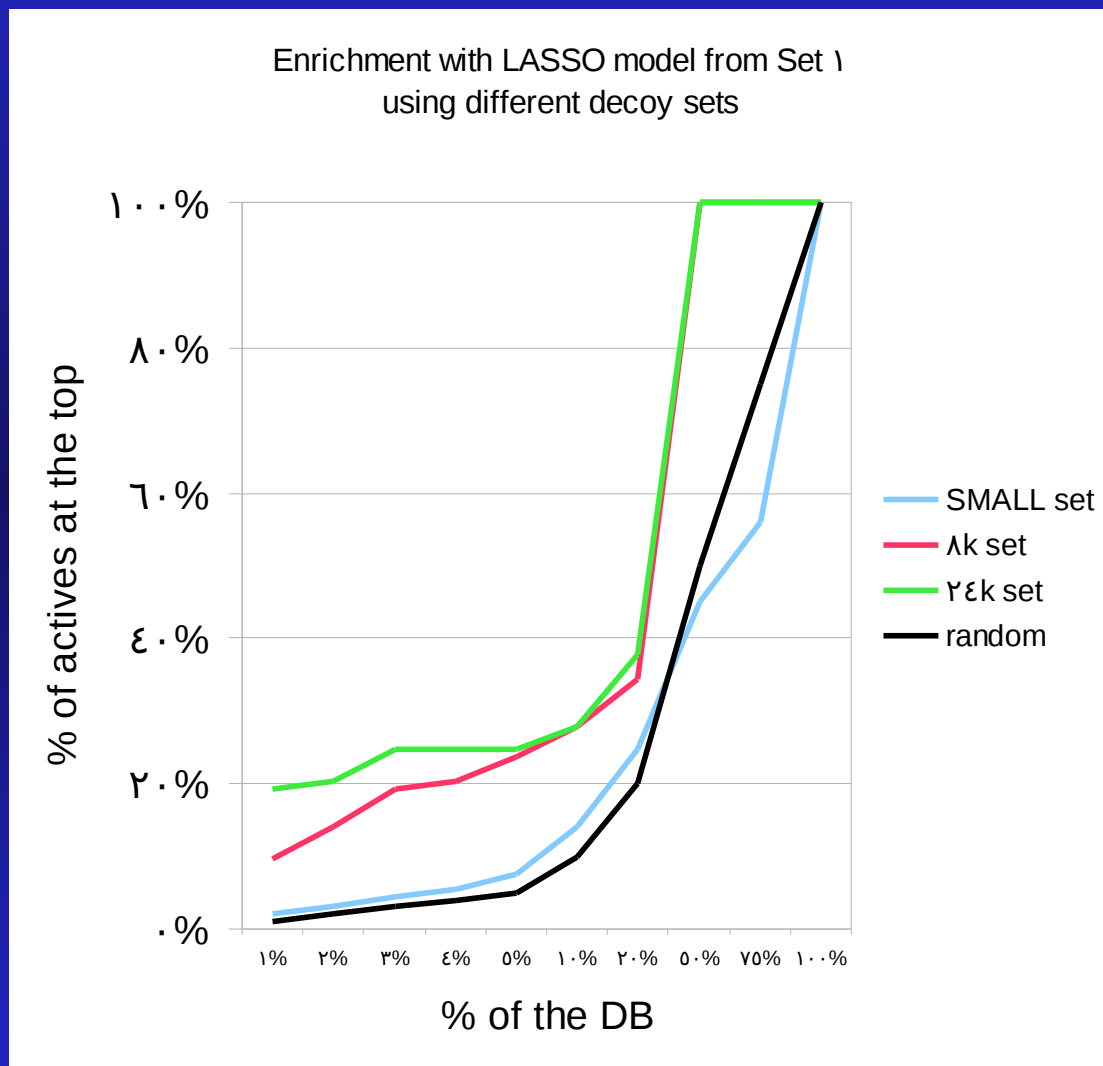
Lasso AVG's of 5 models (over 5 runs)

with 3k, 8k and 24k decoys



PXR dataset 1 vs 2 - results

- Used LASSO model for dataset 1 to predict dataset 2 (N = 168)
- Sensitivity 12%
- Specificity 99%
- Accuracy 51%
- Matthew's correlation 0.2
- Using dataset 2 to predict dataset 1 gives similar results



PXR datasets – results and Chemspider

- 10% in the top 1% of the screened DB, i.e EF at 1%=10 x better than random
- 40% of the ligands can be pushed into the top 10% of the screened DB i.e EF at 10%=4 x better than random

Future work – assess enrichment data with other target proteins

Already 40 models for targets on www.chemspider.com

Can add PXR next...

SIMBIOSys LASSO			
Descriptors: 0, 0, 0, 0, 0, 0, 0, 3, 0, 0, 0, 0, 3, 8, 2, 0, 17, 0, 1, 1, 0, 3, 0, 0			
Category	Target	PDB Code	LASSO Score
Other Enzymes	PARP, poly(ADP-ribose) polymerase	1efy	0.74
Kinases	PDGFRb, platelet derived growth factor receptor kinase	N/A	0.70
Other Enzymes	COX-2, cyclooxygenase-2	1cx2	0.21
Kinases	EGFR, epidermal growth factor receptor	1m17	0.09
Nuclear Hormone Receptors	PPARg, peroxisome proliferator activated receptor	1fm9	0.08
Other Enzymes	HIVRT, HIV reverse transcriptase	1rt1	0.02
Nuclear Hormone Receptors	RXRa, retinoic X receptor R	1mvc	0.02
Kinases	P38 MAP, P38 mitogen activated protein	1kv2	0.02
Kinases	VEGFR2, vascular endothelial growth factor receptor	1vr2	0.02
Serine Proteases	FXa, factor Xa	1f0r	0.02
Metalloenzymes	PDE5, phosphodiesterase 5	1xp0	0.02
Kinases	GSK-3, glycogen synthase kinase-3	1gsk	0.01

ToxCast: docking pesticides in human PXR

- 10 Groups have contracts with EPA to test ~300 conazoles & pesticides, etc with various biological assays (cell based, receptor etc)
- We have docked all the molecules into the PXR agonist site of 5 structures
- GOLD (ver 4) -genetic algorithm explores conformations of ligands and flexible receptor side
- 20 independent docking runs
- Used the regular goldscore to classify compounds
- Comparing their respective scores to the corresponding goldscores of the co-crystallized ligands.
- Majority vote across the five structures.
- Verified predictions in vitro ourselves

ToxCast: docking pesticides in PXR

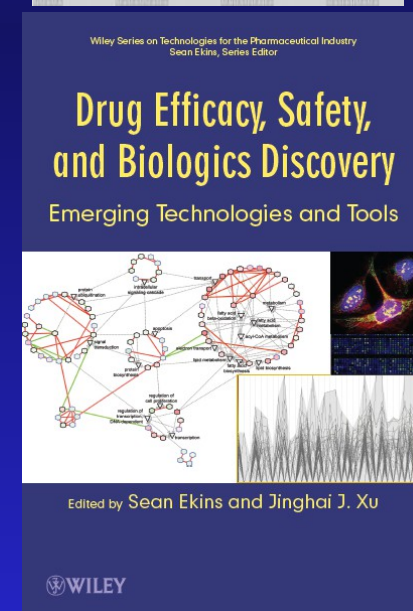
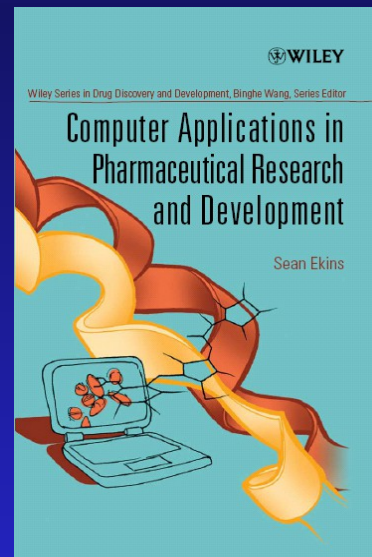
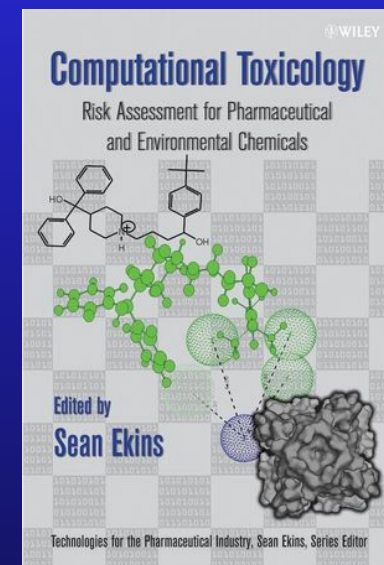
- Activities of most activators more potent vs NCGC data
- We correctly predict ~70% of compounds and 75% of activators
- Including other predicted pesticides from Lemaire, G et al., Toxicol Sci. 2006; 91:501-9, (2006).

Kortagere et al Environ Health Perspect ASAP 2010

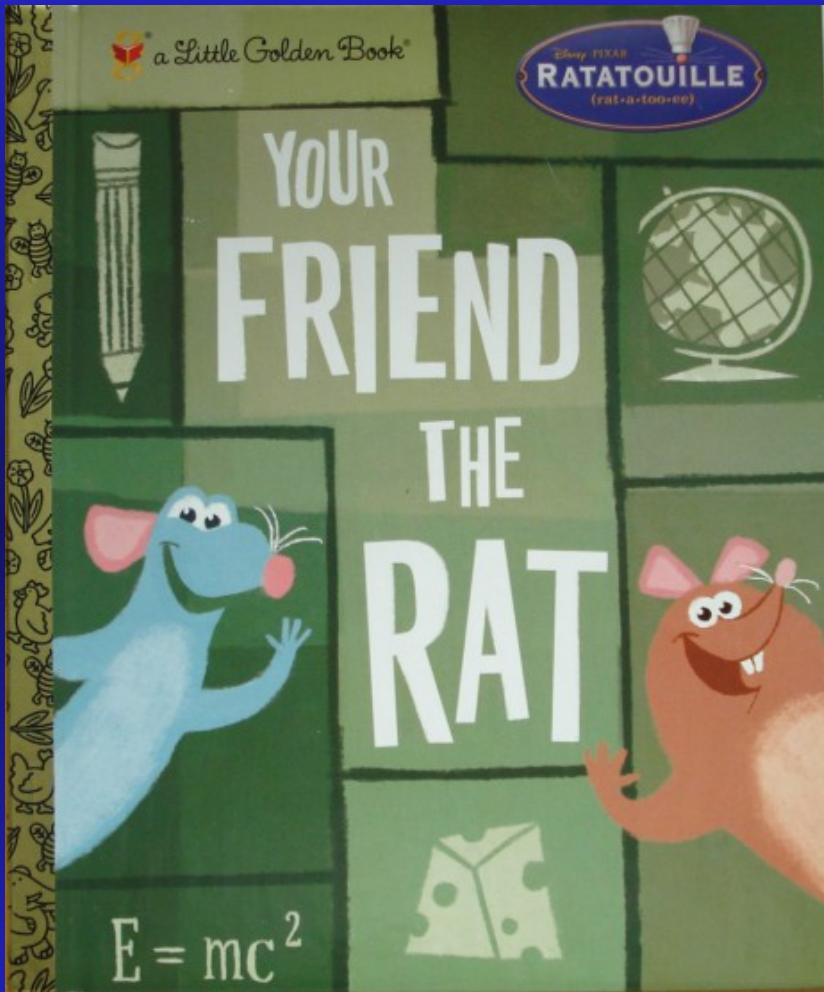
Compound	Docking classification	Actual PXR activity	EPA ToxCast phase I data (NCGC) PXR human activity (μM)	Human PXR EC ₅₀ (μM)	Efficacy Relative to 10 μM rifampicin
Mancozeb	N	N	N	No effect	
Mesosulfuron-methyl	A	N	N	No effect	
Diethylhexyl phthalate	A	A	A (20.75)	1.8	0.63
Methylhexyl phthalate	N	-	N		
Bensulide	A	-	A (1.57)		
Foramsulfuron	A	N	N	>50	
Bensulfuron methyl	A	N	N	89.4	0.18
Esfenvalerate	A	A	A (26.98)	1.5	0.64
Z,E-Fenpyroximate	A	N	N	No effect	
Butafenacil	A	A	N	6	0.53
α -Cypermethrin	A	A	A (18.3)	1.6	0.54
Triflusulfuron methyl	A	-	N	-	
β -Cyfluthrin	A	A	A (19.7)	2.5	0.54
Permethrin	A	A	A (20.26)	5.4	0.53
Oxasulfuron	A	-	N	-	
Fenarimol ⁺	N	A			
Propiconazole ⁺	A	A	A (36.81)		
Fenbuconazole ⁺	A	A	N		
Prochloraz ⁺	A	A	N		
Imazalil ⁺	N	N	A (36.54)		
Oxadiazon ⁺	A	A	A (5.49)		
Alachlor ⁺	N	A	A (15.35)		
2,4-D ⁺	N	N	N		
Diuron ⁺	N	N	N		
Atrazine ⁺	N	N	N		
Fipronil ⁺	N	A (weak)	A (12.55)		
Thiabendazole ⁺	N	N	N		
Carbaryl ⁺	N	N	N		

Acknowledgments

- Matthew D. Krasowski, Erica J. Reschly (University of Iowa)
- Akash Khandelwal, Peter W. Swaan, (University of Maryland)
- Sandhya Kortagere (Drexel University)
- Markus Lill (Purdue University)



Species differences in PXR – can we LASSO these?



- Species dependent effects on transporter and enzyme induction is due to activation of PXR and other NHRs

Species differences in Rifampin agonism

Human, monkey, chicken, dog & Rabbit but not rat or mouse

PCN - rat but not human

and mouse, rabbit, zebrafish, chicken...