Cheminformatics and Toxicogenomics for Toxicity Prediction and Mechanistic Insight

George Daston

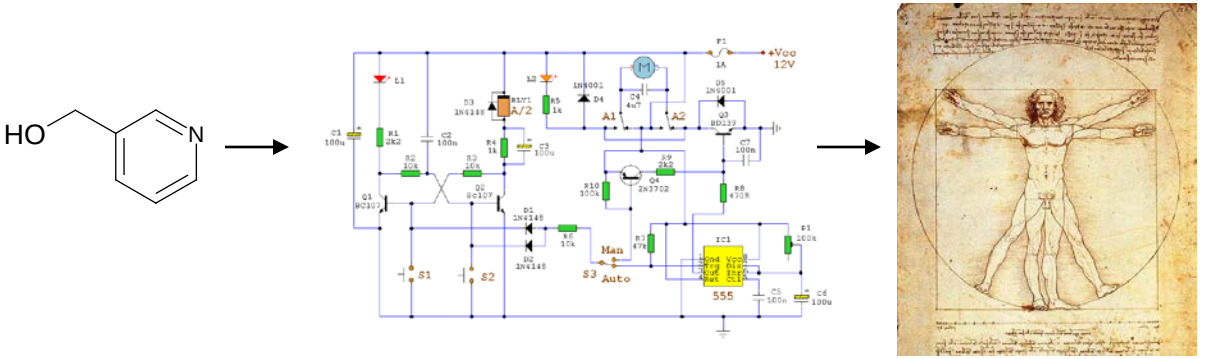
Overview

- Tiered approach to predicting toxicity of new chemicals
  - Cheminformatics- supported SAR
  - High-content methods to assess SAR solutions
- Identifying MOA using cheminformatics and toxicogenomics
  - MOA ontology
  - Connectivity mapping

Toxicology: From an Empirical to a Predictive Science

Traditional Approach (Black box): Use a model that we have (some) confidence in, but incomplete understanding of how it works

Desired Approach: Predictions based on deep, fundamental understanding

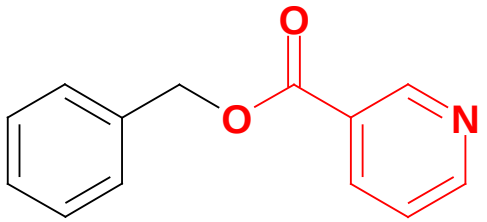


Taking Advantage of the Existing Literature

- Considerable outcome data in DART (almost 12,000 entries in publicly available databases)
- Pressing need is to identify initial molecular events
- Effort needed to connect initial events with tissue/organ level effects

Initial Screening for Human Hazards

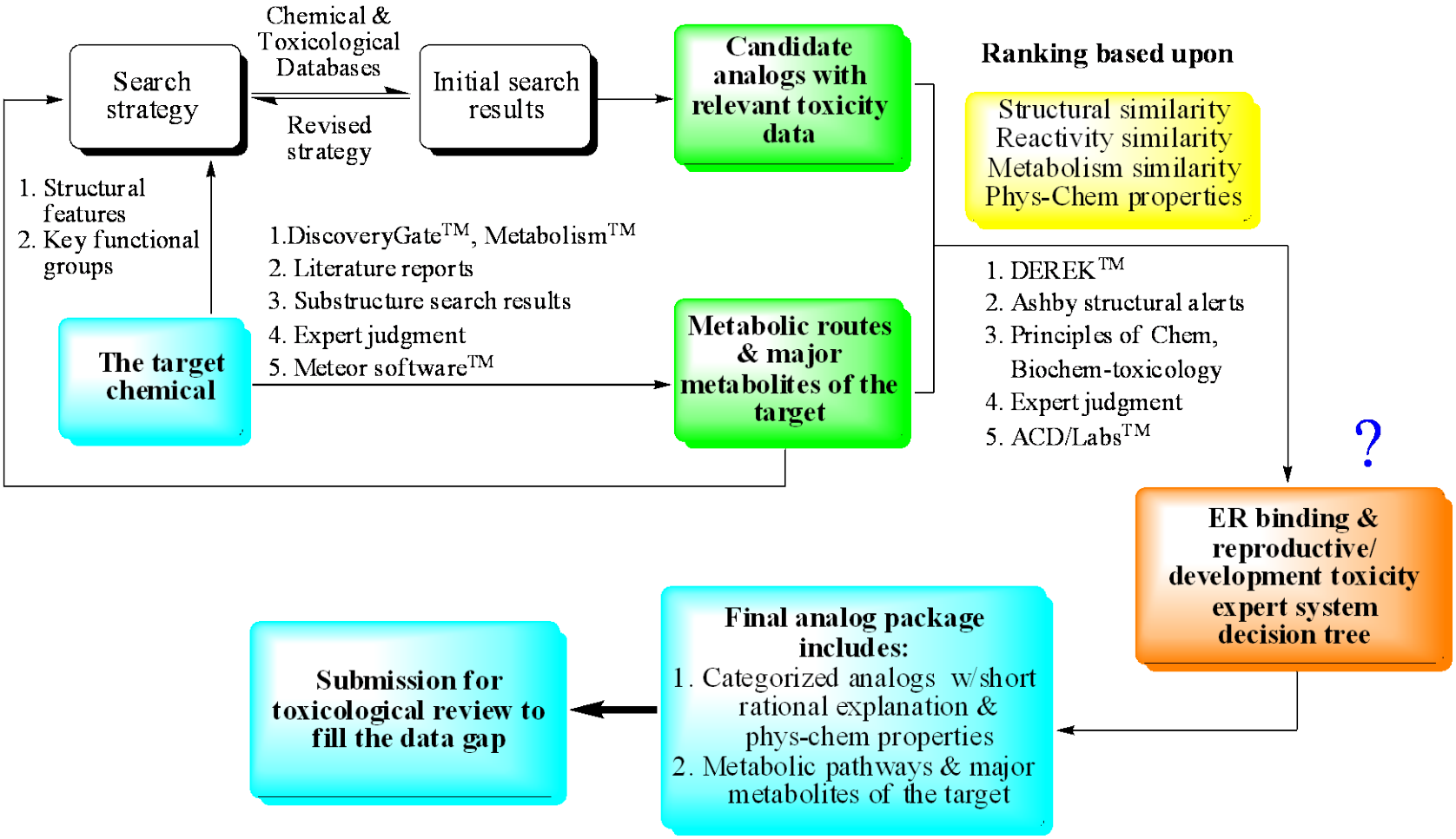
- Substructure searching
  - Genotoxicity (19,300)
  - Carcinogenicity (15,800)
  - Skin Sensitization (9,400)
  - Skin Irritation (10,400)
  - Reproductive/Developmental Toxicity (11,300)
  - Subchronic/Chronic Toxicity (15,100)
  - Acute Toxicity (68,500)



- All assessment captured in CHS

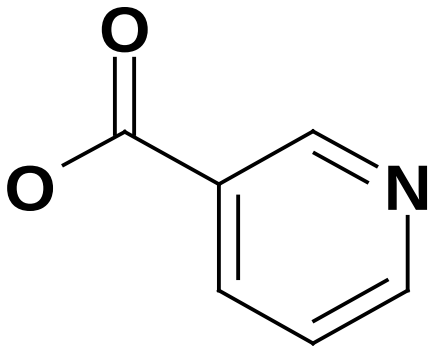
- External Data Sources: BIBRA*, Cal Prop 65*, CTFA*, HERA*, HPV*, OECD*, IPCS*, NICNAS*, RIFM/FEMA*, SCCP*, WHO/JECFA*, SciFinder, ToxNet, ATSDR, CPDB, ECETOC, ECB, IARC , Thompson/MicroMedix, NTP, RTECS/NIOSH, Scopus, TSCATS, others

Flow chart of new analog identification & evaluation process

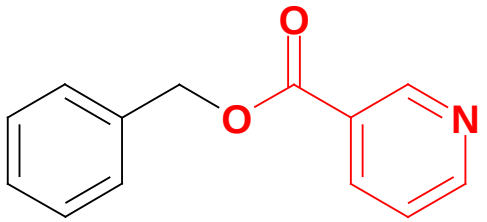


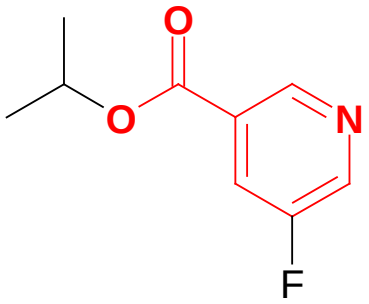
Wu et al., RTP, 2010

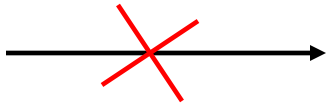
Searching GRASP- Substructure Searching

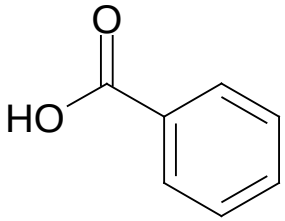


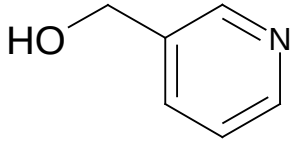
Search Structure







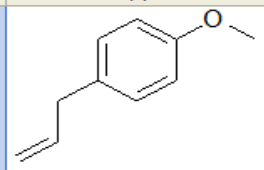
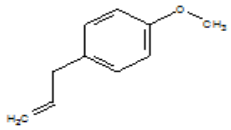
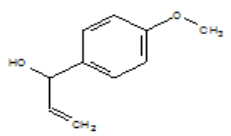




Output – Substructure Searching

RS3 Excel [rpdp v2.3.0 :: TL5734] - Carcinogencity Example 2.xls

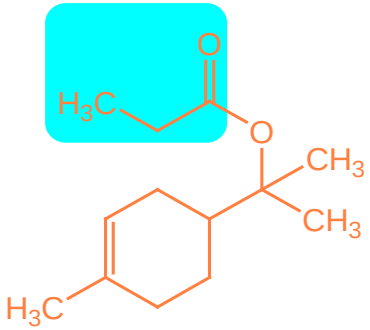
File Edit View Insert Format Tools Data Window Help RS3 Discovery

	A	B	C	D	E	F	G	H
1		SEARCH STRUCTURE						
2	Substances							
3		*A. Alias*	*B. CAS No.*	*C. Study Type*	*D. Species*	*E. Route of Admin.*	*F. Result*	*N. Ref. Journal*
4	 103652	ANISOLE, p-ALLYL-	140-67-0	BIOASSAY	MOUSE	DIET		RIFM
5	103652	ANISOLE, p-ALLYL-	140-67-0	BIOASSAY	MOUSE	GAVAGE		RIFM
6	103652	ANISOLE, p-ALLYL-	140-67-0	BIOASSAY	MOUSE		+	RTECS
7	103652	ANISOLE, p-ALLYL-	140-67-0	BIOASSAY	MOUSE		F +	http://potency.berkeley.e
8	103652	ANISOLE, p-ALLYL-	140-67-0	CA PROP 65				http://www.oehha.org/p
9	103652	ANISOLE, p-ALLYL-	140-67-0	OTHER	MOUSE	INTRAGASTRIC	M +	PHS149
10	103652	ANISOLE, p-ALLYL-	140-67-0	VSD	MOUSE	DIET		VSD
11	 104455	BENZYL ALCOHOL, p-METHOXY-alpha-VINYL-	51410-44-7	BIOASSAY	MOUSE		+	RTECS
12	104455	BENZYL ALCOHOL, p-METHOXY-alpha-VINYL-	51410-44-7	BIOASSAY	MOUSE		F +	http://potency.berkeley.e

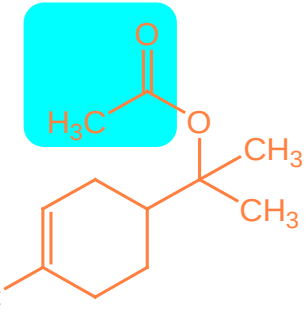
Carcinogenicity Example 2 | Sheet2 | Sheet3 | Sheet4 | Sheet5 | Sheet6 | Sheet7 | She

Ready

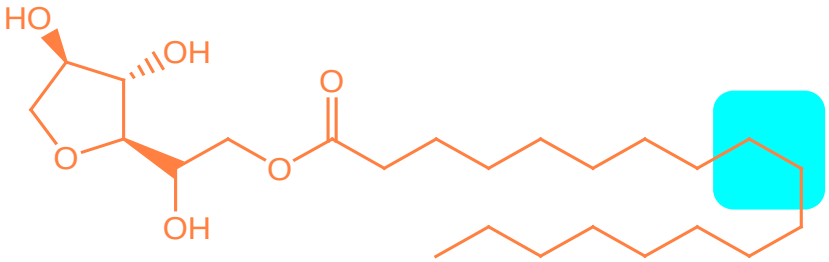
Suitable Analogs



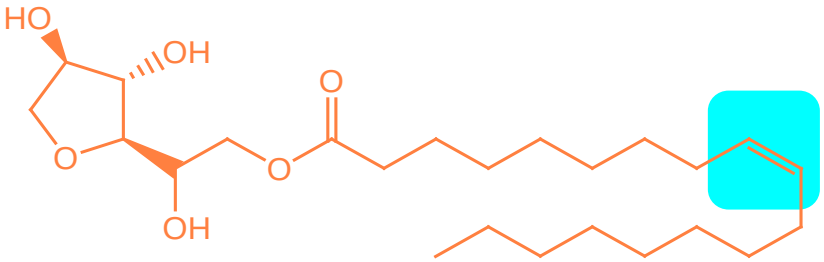
CAS# 80-27-3



CAS# 8007-35-0

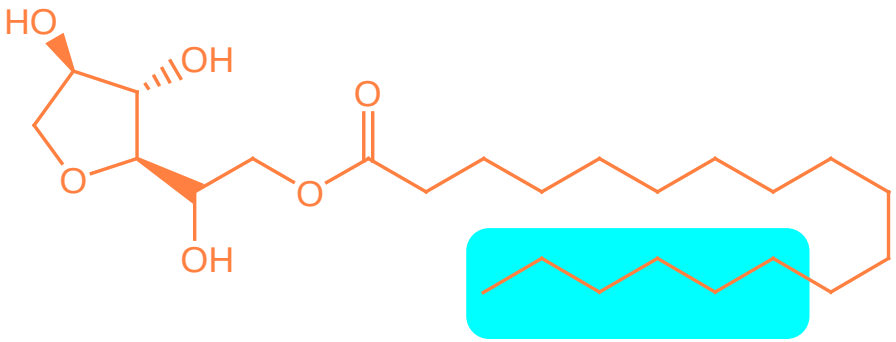


CAS# 1338-43-8

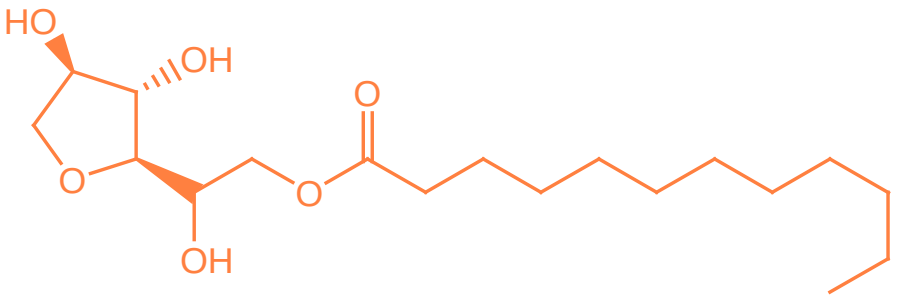


CAS# 8007-43-0

Possibly Suitable

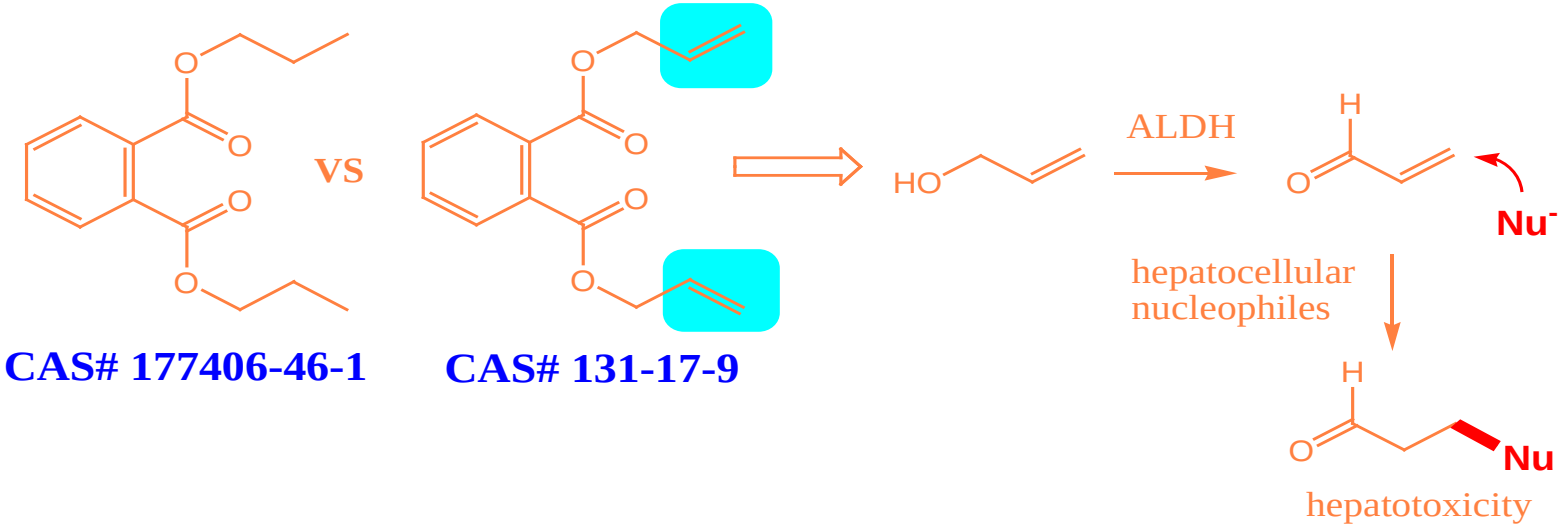


CAS# 1338-43-8



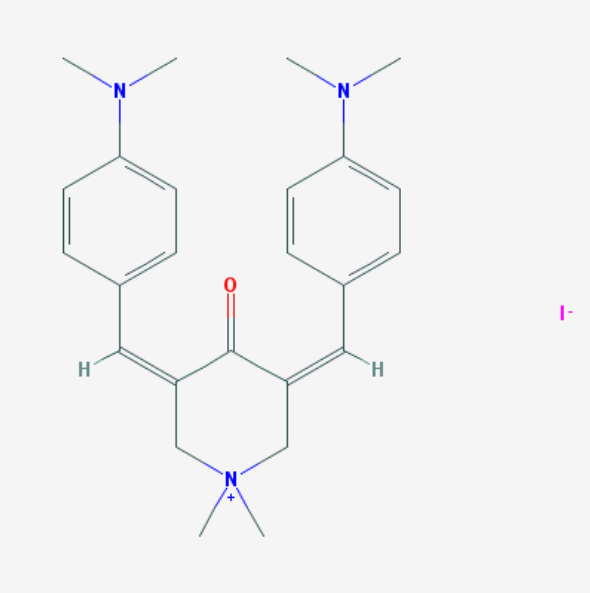
CAS# 1338-92-3

Unsuitable Analogs



CAS# 177406-46-1 **CAS# 131-17-9**

PubChem Compound



- Nrf2 qHTS screen for inhibitors: counterscreen for cytotoxicity
- qHTS Assay for Inhibitors of RanGTP induced Rango (Ran-regulated importin-beta cargo) – Importin beta complex dissociation
- qHTS Assay for Inhibitors of JMJD2A-Tudor Domain

Chemical Probe Active Inactive Inconclusive Unspecified

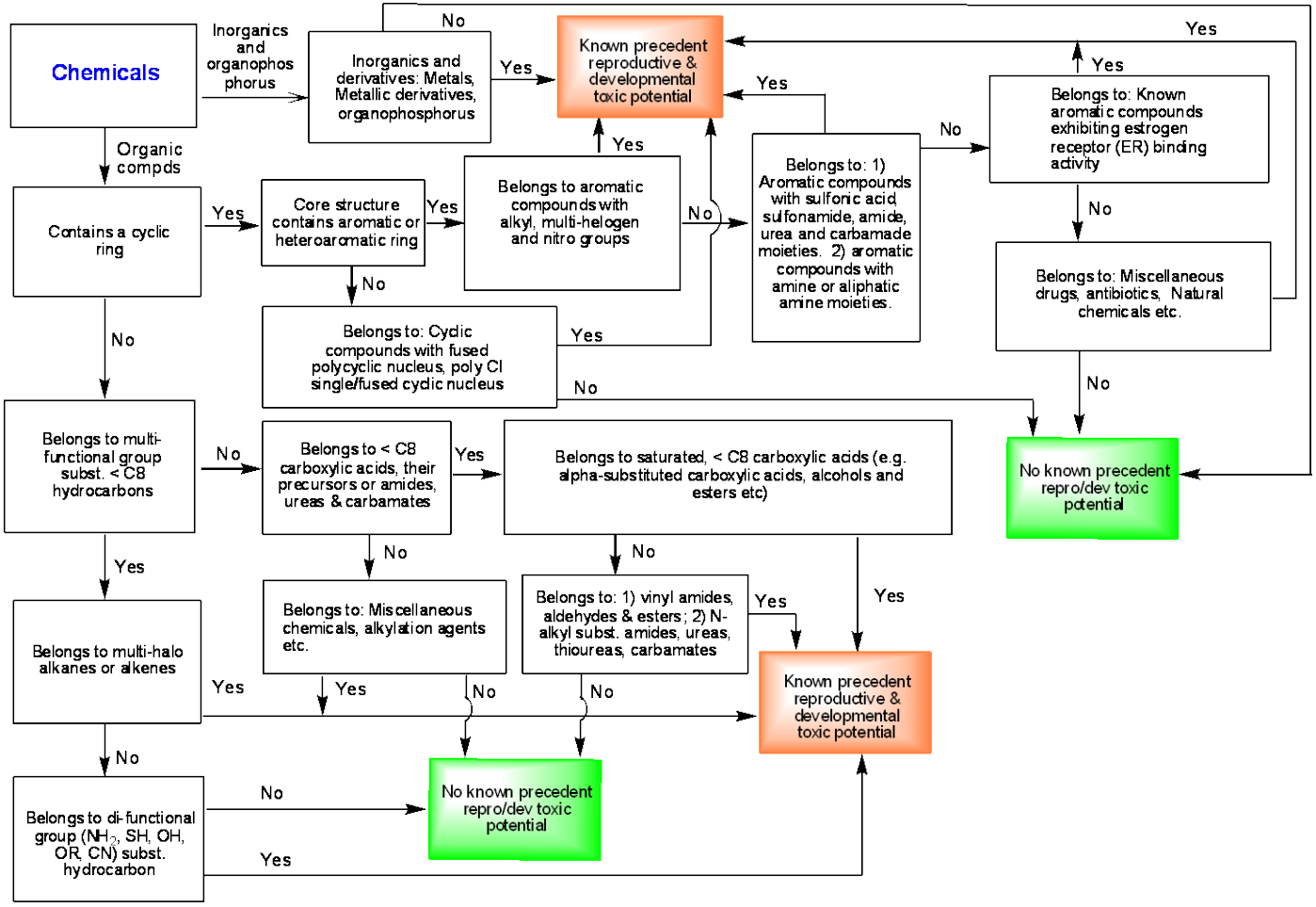
Cheminformatics: Ontology

- Use large database to organize chemicals into mode of action groupings
- Start to estimate the extent of “the universe of toxicity mechanisms”
- This will allow us to design a suite of model systems that is comprehensive

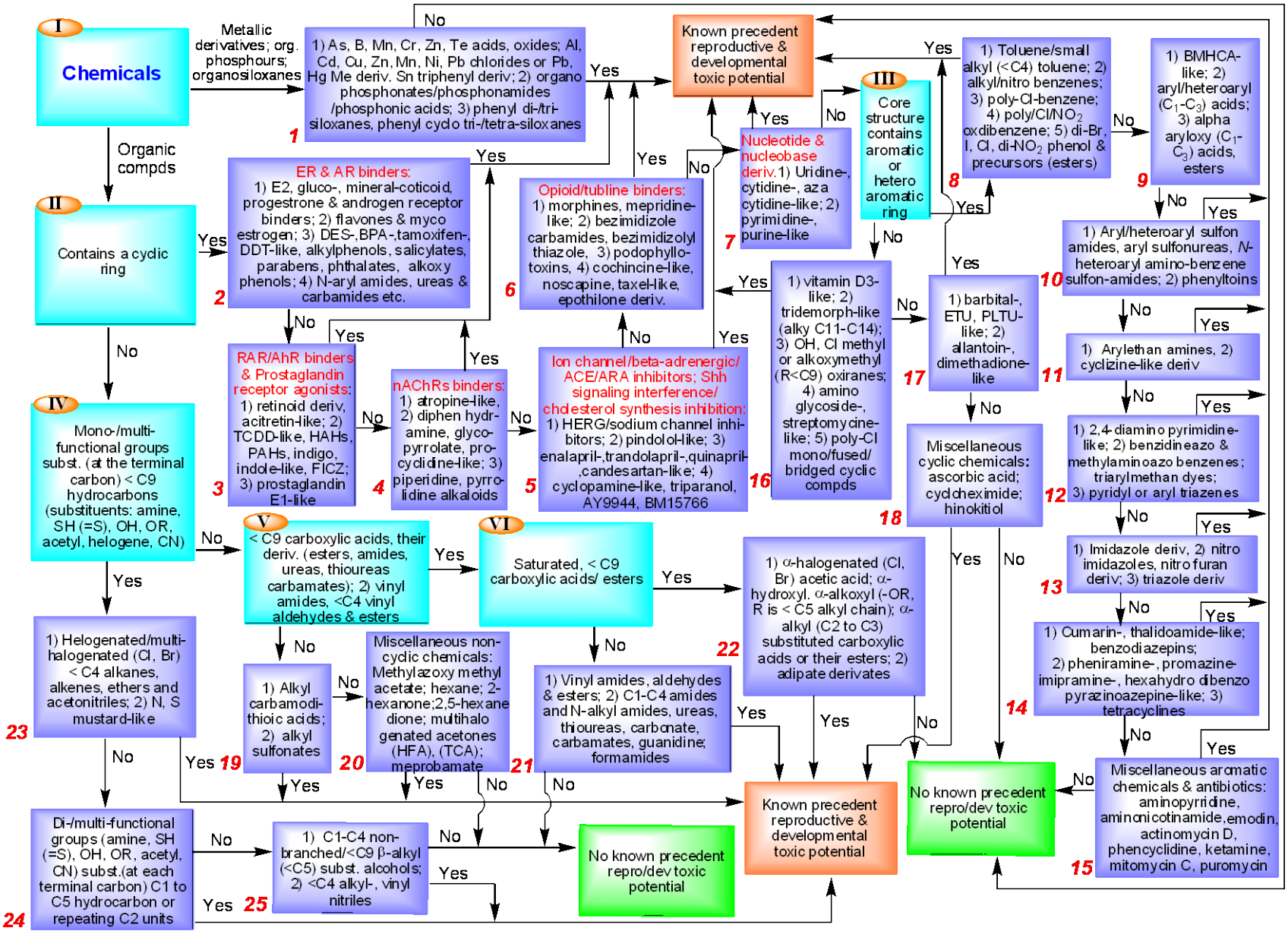
Initial Concept

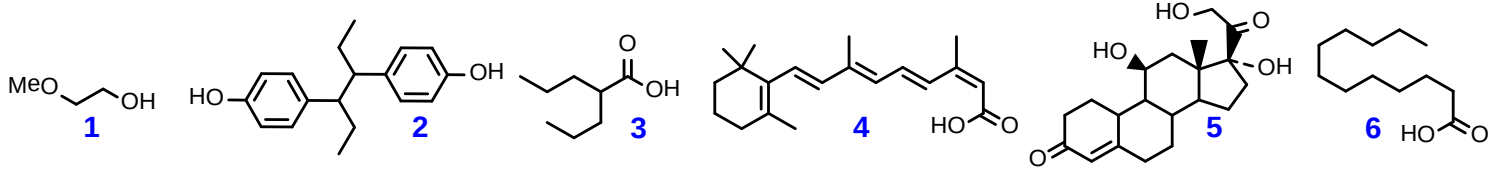
- An initial list of ~ 260 chemicals with DART data was originally developed as part of an evaluation of Threshold of Toxicologic Concern (TTC) (Laufersweiler et al., 2012)
- These chemicals were grouped based on their chemical characteristics and this tree was published in concept in Blackburn et al. (2011)

Original Tree



Expert system decision tree for repro/dev toxicity





I
Chemicals
1, 2, 3, 4, 5, 6

1, 2, 3
4, 5, 6
Organic compds

II
Contains a cyclic ring

Yes
2, 4
5
2

ER & AR binders:
1) E2, gluco-, mineral-corticoid, progestrone & androgen receptor binders; 2) flavones & myco estrogen; 3) DES-,BPA-,tamoxifen-, DDT-like, alkylphenols, salicylates, parabens, phthalates, alkoxy phenols; 4) N-aryl amides, ureas & carbamides etc.

4
No
Yes
4

RAR/AhR binders & Prostaglandin receptor agonists:
1) retinoid deriv, acitretin-like; 2) TCDD-like, HAHs, PAHs, indole-like; 3) prostaglandin E1-like
3

1, 3, 6
No

IV
1) Mono-/multi-functional groups subst. < C9 hydrocarbons (substituents at the termial carbons must be amine, SH (=S), OH, OR, OAc, halogene, CN) 2) Metal chealators

No
3, 6

V
< C9 carboxylic acids, their deriv. (esters, amides, ureas, thioureas carbamates); 2) vinyl amides, <C4 vinyl aldehydes & esters

Yes
3

VI
Saturated, < C9 carboxylic acids/ esters

Yes
3

1) α-halogenated (Cl, Br) acetic acid; α-hydroxyl. α-alkoxyl (-OR, R is < C5 alkyl chain); α-alkyl (C2 to C3) substituted carboxylic acids or their esters; 2) adipate derivates
22

1
Yes

1) Halogenated/multi-halogenated (Cl, Br) < C4 alkanes, alkenes, ethers and acetonitriles; 2) N, S mustard-like
23

6
No

1)Alkyl carbamodi-thioic acids; 2) alkyl sulfonates; 3)perfluorooctane sulfonic acid derivatives
19

No
6

Miscellaneous non-cyclic chemicals: Methylazoxy methyl acetate; hexane; 2-hexanone;2,5-hexanedione; multihalogenated acetones (HFA), (TCA); meprobamate
20

1) Vinyl amides, aldehydes & esters; 2) C1-C4 amides; thioamides and N-alkyl amides, ureas, thioureas, carbonates, carbamates, guanidines; formamides; 3) O-(methylamino)carbonyl derived carbamates
21

1
No

1) Di-/multi-functional groups (amine, SH (=S), OH, OR, OAc, CN) subst.(at each terminal carbon) C1 to C5 hydrocarbon or repeating C2 units; 2) Metal chelators
24

No
Yes
25

1) C1-C4 non-branched/<C9 β-alkyl (<C5) subst. alcohols; 2) <C4 alkyl-, vinyl nitriles

No known precedent repro/dev toxic potential

Known precedent reproductive & developmental toxic potential

Known precedent repro/dev toxic potential

No known precedent repro/dev toxic potential

Metallic derivatives; org. phosphorus; organosiloxanes

1) As, B, Mn, Cr, Zn, Te acids, oxides; Al, Cd, Cu, Zn, Mn, Ni, Pb chlorides or Pb, Hg Me deriv. Sn triphenyl deriv; 2) organo phosphonates/phosphonamides /phosphonic acids; 3) phenyl di-/tri-siloxanes, phenyl cyclo tri-/tetra-siloxanes
1

Yes
2, 5

Opioid/tubuline binders: 1) morphines, meperidine-like; 2) bezimidizole carbamides, bezimidizolyl thiazole, 3) podophyllotoxins, 4) colchicine-like, noscapine, taxol-like, epothilone deriv.
6

nAChRs binders & AChE inhibitors 1) atropine-like, 2) diphenyl hydramine, glycopyrrolate, procyclidine-like; 3) piperidine, pyrrolidine alkaloids; 4) O-aryl carbamates
4

Ion channel/beta-adrenergic/ACE/ARA inhibitors; Shh signaling interference/ cholesterol synthesis inhibition: 1) HERG/sodium channel inhibitors; 2) pindolol-like; 3) enalapril-,trandolapril-,quinapril-,candesartan-like; 4) cyclopamine-like, tripabanol, AY9944, BM15766
5

Nucleotide & nucleobase deriv 1) Uridine-, cytidine-, aza cytidine-like; 2) pyrimidine-, purine-like
7

Core structure contains aromatic or hetero aromatic ring
8

1) Toluene/small alkyl (<C4) toluene; 2) nitro benzenes; 3) poly-Cl-benzene; 4) poly-Cl/NO2 oxydibenzene; 5) di-halo, di-NO2 phenol & precursors (esters)
9

1) BMHCA-like; 2) aryl/heteroaryl (C_1-C_3) acids; 3) alpha aryloxy (C_1-C_3) acids, esters

1) Vitamin D3-like; 2) tridemorph-like (alky C11-C14); 3) OH, Cl methyl or alkoxymethyl (R<C9) oxiranes; 4) amino glycoside, streptomycin-like; 5) avermectin & spiramycin-like; 6) poly-Cl mono/fused/ bridged cyclic compds
16

1) barbital-, ETU, PLTU-like; 2) allantoin-, dimethadione-like 3) cyclic/hetero cyclic 6 member ring (e.g.piperazine -like)
17

1) Aryl/heteroaryl sulfon amides, aryl sulfonureas, *N*-heteroaryl amino-benzene sulfon-amides; 2) phenytoins
10

1) Arylethane amines, 2) cyclizine-like deriv
11

Miscellaneous cyclic chemicals: hydantoin; cycloheximide; hinokitol, tepraloxydin-like
18

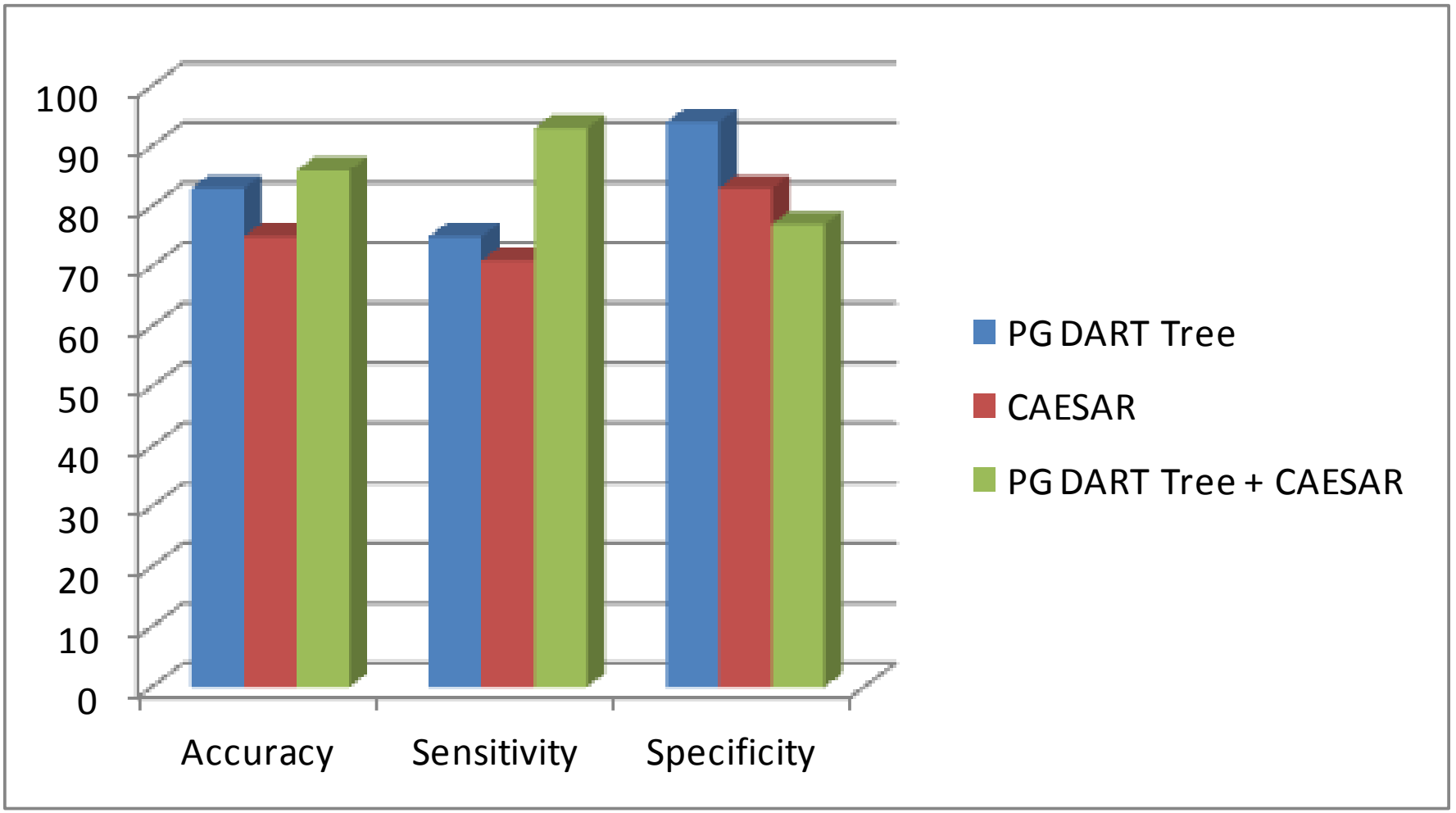
1) 2,4-diamino pyrimidine-like; 2) benzidineazo & methylaminoazo benzenes; triarylmethane dyes; 3) pyridyl or aryl triazenes
12

1) Imidazole deriv, 2) nitro imidazoles, nitro furan deriv; 3) triazole deriv.
13

1) Coumarin-, thalidomide-like; quinolones; benzodiazepines; 2) pheniramine-, promazine-imipramine-, hexahydro dibenzo pyrazinoazepine-like; 3) tetracyclines
14

Miscellaneous aromatic chemicals: Isoniazid, aminonicotinamide, aminopicoline, procarbazine, phencyclidine, aminoglutethimide, ketamine, phenidione, haldol, cypermar-like
15

P&G DART tree + CAESAR for test set (106 active, 73 non-active)



Accuracy: ~86%, Sensitivity: 93% and Specificity: 77%

Putative MOA Grouping by Chemical Structure

- 25 major categories, multiple sub-categories
- Highest level of confidence has
  - Similar structures
  - Identified molecular target
  - Similar DART outcome (e.g., common syndrome or highly specific effect)
- Along with toxicogenomics, has the potential to accelerate assigning MOA to DART compounds

Hierarchy Examples

- Nuclear hormone receptor ligands
- Prostaglandin receptor ligands
- Nicotinic ACh receptor ligands and AChesterase inhibitors
- Shh signaling interference/ cholesterol synthesis inhibitors
- Nucleotide derivatives

Nuclear Hormone Receptor Ligands

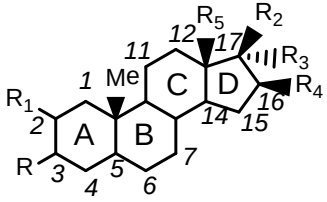
- Estrogen and androgen receptor ligands
- Glucocorticoid receptor ligands
- Retinoic acid receptor ligands
- Thyroid hormone receptor ligands
- Ah receptor ligands

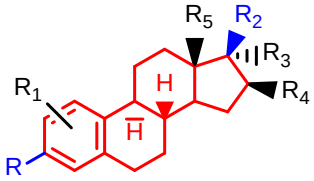
Nuclear hormone receptor ligands

- Estrogen and androgen receptor ligands
  - steroid nucleus-derived compounds
    - Estradiol-like
    - Progesterone, androgens, steroidal anti-androgens
  - Non-steroidal compounds
    - Flavones and mycoestrogens
    - Alkylphenols
    - N-aryl-substituted ureas, carbamides, amides
    - other

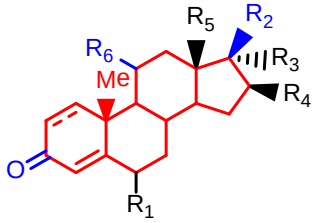
Nuclear hormone receptor ligands

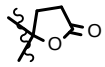
- Estrogen and androgen receptor ligands
  - steroid nucleus-derived compounds
    - Estradiol-like
    - Progesterone, androgens, steroidal anti-androgens
  - Non-steroidal compounds
    - Flavones and mycoestrogens
    - Alkylphenols
    - N-aryl-substituted ureas, carbamides, amides
    - other





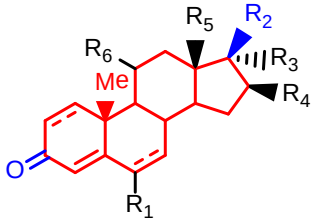
R=OH, OMe
R_1=OH @ C-2 or C-4
R_2=OH
R_3=H, alkyne
R_4=H, OH
R_5=Me, H
17-OH (R_2,R_3) also can be C=O, H



R_1=H, F
R_2= -$COCH_2OH$, -$COCH_2Cl$
R_3=OH, CO, H
R_4=H, Me
R_2,R_3 can form 

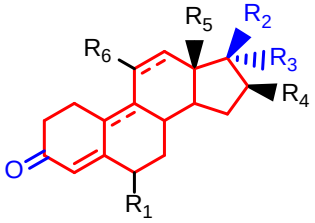
R_3,R_4 can form an acetal/ketal
R_5=Me, -CHO
R_6=H, OH, -CO
C-1, C-2 can be C=C/C-C bond
C-3 contains H or OH - in very few cases.

Key functional groups:
C=O at C-3; -$COCH_2OH$ and -$COCH_2Cl$ at C-17; OH, C=O at C-11



R_1=H, Cl, Me
R_2= -$COCH_3$
R_3=OH, H, $OCOCH_3$, $OCOCH_2CH_3$
R_4=H
R_5=Me
R_6=H
C-1, C-2 and C-6, C-7 can be saturated/unsaturated
C-1, C-2 can form cyclopropane

Key functional groups:
C=O at C-3; -$COCH_3$ and OH, OAc at C-17

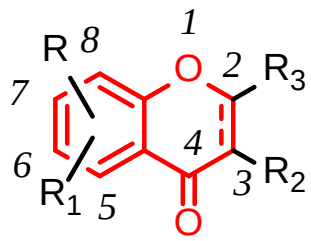


R_1=H
R_2=OH
R_3=H, Me, Et, ethyn or allyl, actonitrile etc.
R_4=H
R_5=Me, Et
R_6=H
C-9, C-10 and C-11, C-12 can be C=C/C-C bond

Key functional groups:
C=O at C-3; -OH and alkyl (C1-C3 carbons), ethyn at C-17

Nuclear hormone receptor ligands

- Estrogen and androgen receptor ligands
  - steroid nucleus-derived compounds
    - Estradiol-like
    - Progesterone, androgens, steroidal anti-androgens
  - Non-steroidal compounds
    - Flavones and mycoestrogens
    - Alkylphenols
    - N-aryl-substituted ureas, carbamides, amides
    - other

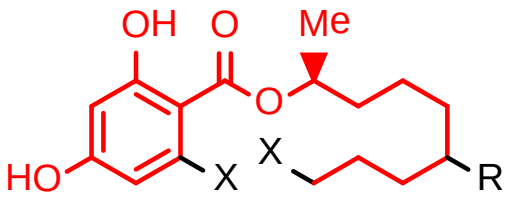


R=OH, H; R_1=OH

R_2=mono-, di-, tri-, OH-Ph, MeO-Ph

R_3=mono-, di-, tri-, OH-Ph

R_2 and R_3 can not be present at C-2 and C-3 simutaneously

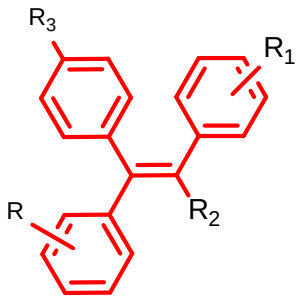


R=OH, =O

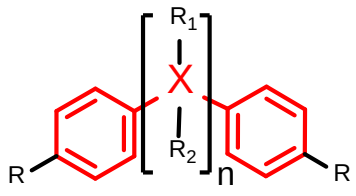
X-X=C-C, C=C

Nuclear hormone receptor ligands

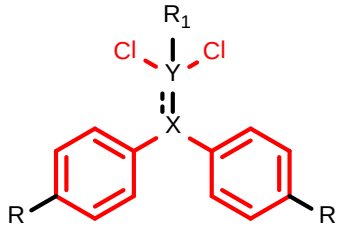
- Estrogen and androgen receptor ligands
  - steroid nucleus-derived compounds
    - Estradiol-like
    - Progesterone, androgens, steroidal anti-androgens
  - Non-steroidal compounds
    - Flavones and mycoestrogens
    - Alkylphenols
    - N-aryl-substituted ureas, carbamides, amides
    - other



R=H, 4-OH
R_1=H, OMe
R_2=Cl, Me, Et
R_3=H, $(Me)_2CH_2CH_2O$-



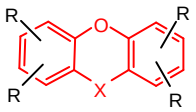
R=OH, NH_2
n=1, X=C, R_1=Alkyl (C1-C4)
R_2=Me
R_1,R_2=isobenzofuranone
n=2, R_1 and R_2 are on different C's
n=2, X=C-C, R_1, R_2=H, Me, Et
n=2, X=C=C, R_1, R_2=H, Me, Et
n=1, X=O, S, SO_2, R_1=R_2=none



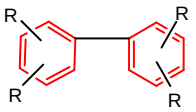
X-Y= C-C
R=OH, Cl, OMe
R_1=H, Cl
X-Y= C=C
R=OH, Cl, OMe
R_1=none

Ah Receptor Ligands

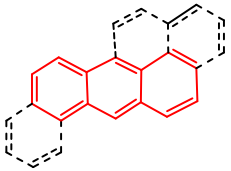
- TCDD-like chemicals
  - cleft palate, hydronephrosis and reproductive system defects
- Indole-related compounds: repro system
- Polycyclic aromatics
- Halogenated aromatics (e.g., PCBs)
  - Liver cyp induction leads to DART effects?

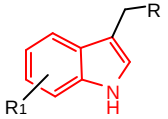


R=Cl, Br
X=O, none



R=Cl, Br





R=OH
R_1=H

Problems with the chemical approach

- Promiscuous chemicals that have more than one molecular target
- Seemingly similar compounds that have different developmental outcome
  - PK differences?
  - More than one target?
  - Insufficient potency against target?

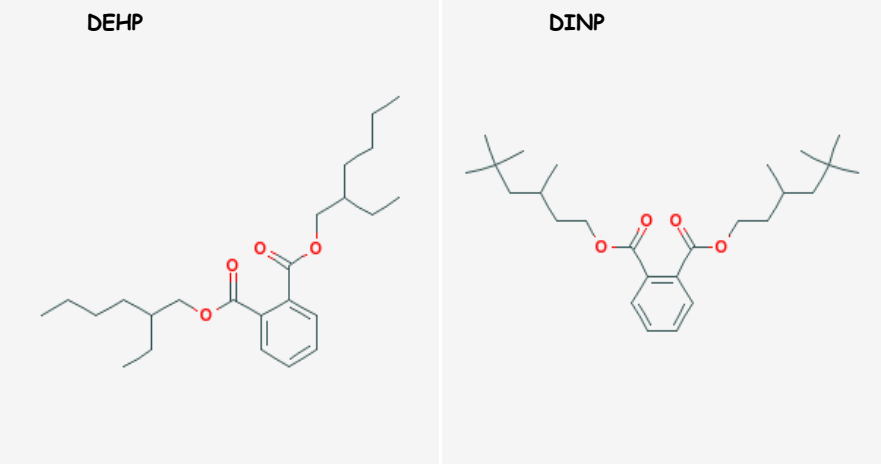
Outcome of chemistry assessment is hypothesis generation

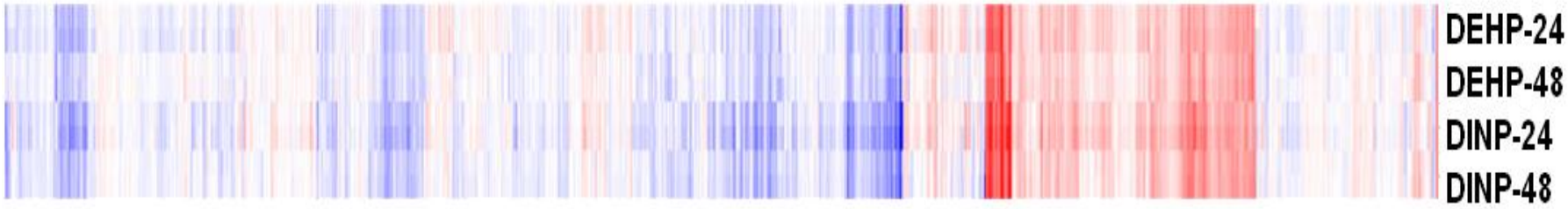
- Chemical is metabolized to a tested chemical, or to a known active metabolite
  - Currently, assessment is done by wet lab metabolism
- Chemical is sufficiently similar in structure to analogs of known toxicity that similar biological activity is inferred
  - Currently, assessment is done by MOA-specific evaluation
  - Add ToxCast and other PubChem data to our databases and our expert considerations about mechanism
  - Global analysis of gene expression

Using Gene Expression Analysis to Inform MOA and AOP

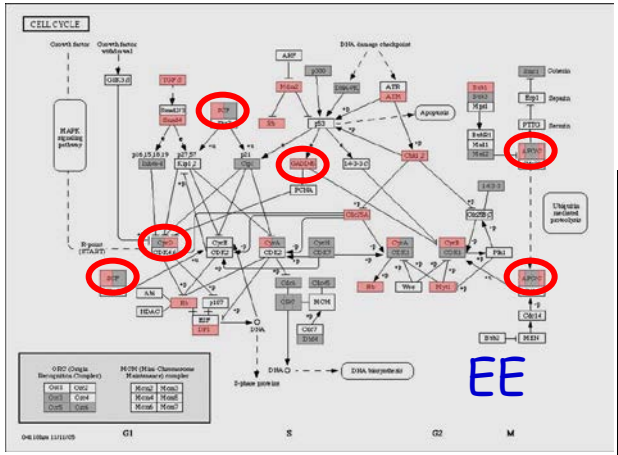
- Gene expression is specific for MOA
- In vitro models may have great potential to identify MOA via gene expression

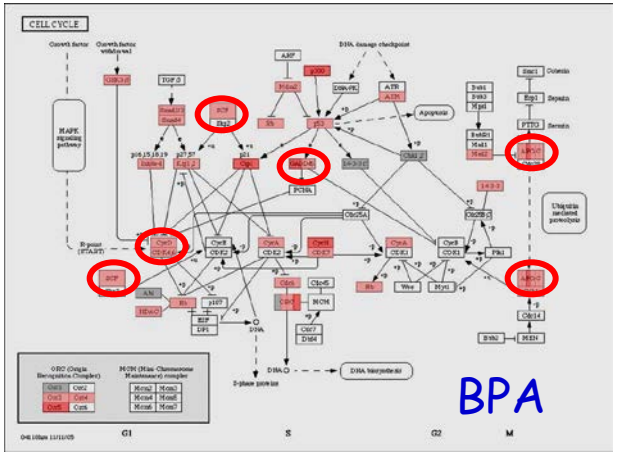
Two Close Structural Analogs

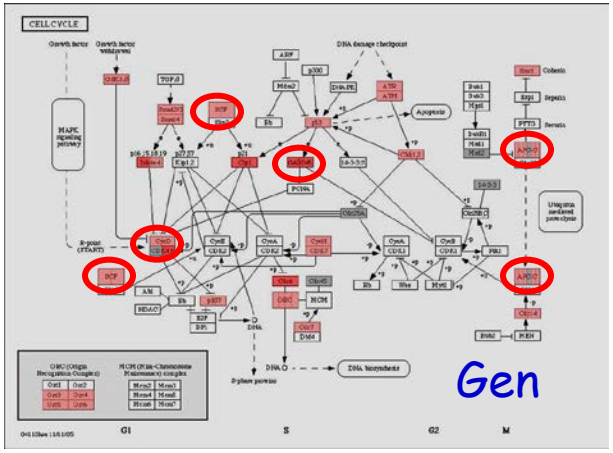


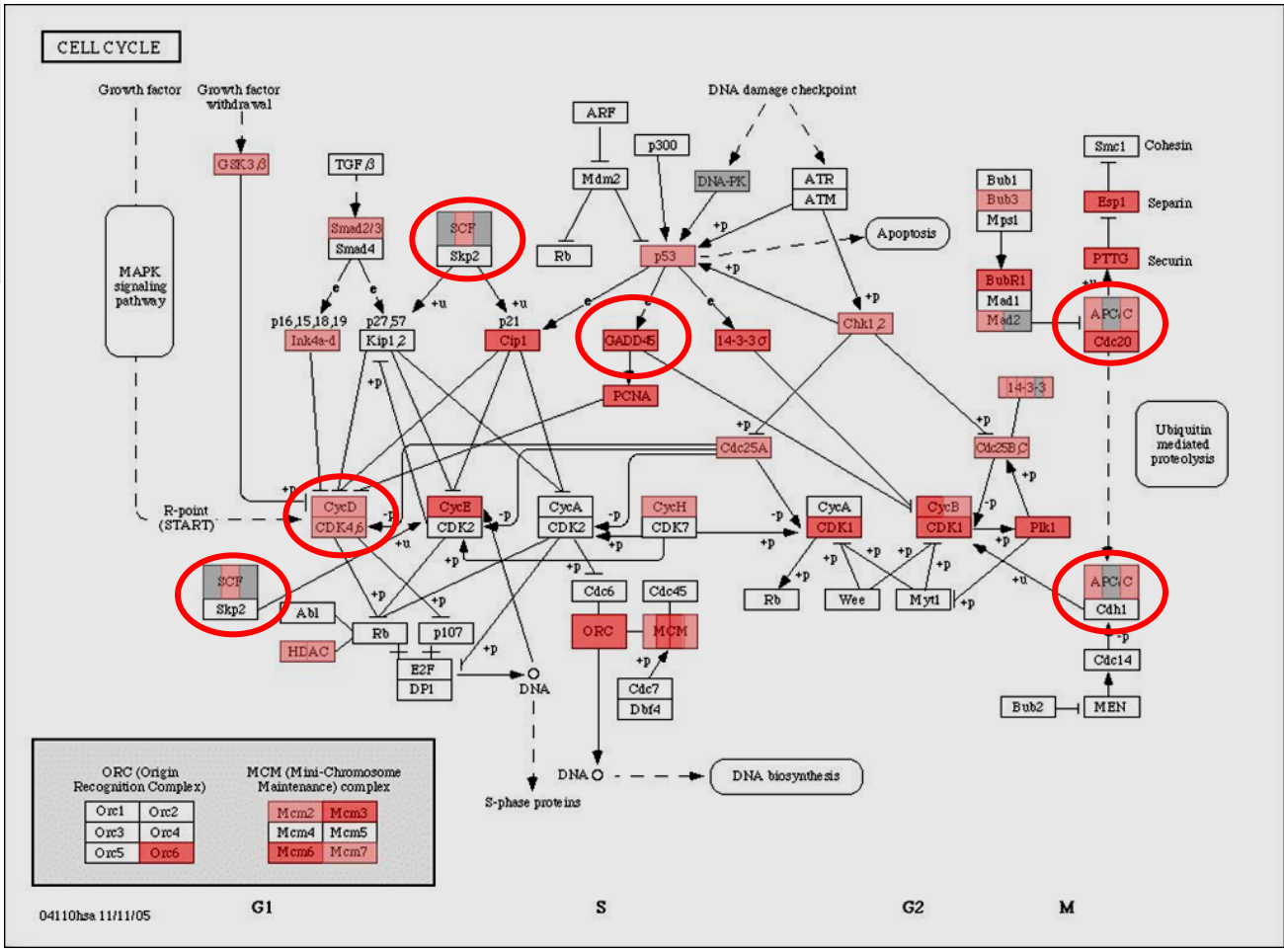


KEGG Cell cycle Example









EE-RAT-24hr
(Up-regulated)

fc <= -2.0
-2.0 < fc <= -1.2
-1.2 < fc < 1.2
1.2 <= fc < 2.0
fc >= 2.0

Connectivity Mapping: High-throughput toxicogenomics

- Concept developed by Lamb in 2006
  - A relatively small number of carefully selected cell types contained all of the pathways necessary to define gene expression profiles for all therapeutic agents in current use
- Can we do the same for toxicants?
  - Cell types: rich in either small molecule receptors or metabolizing enzymes

MOAs to Interrogate with CMAP

- Estrogens, environmental estrogens
- Anti-estrogens
- PPAR agonists
- Anti AndrogenAndrogens
- CAR/PXR agonists
- RAR agonists
- TR agonists
- AhR agonists
- Vitamin D agonist
- Glucocorticoid receptor agonists
- EGFR receptor agonists
- FXR receptor agonists
- Progesterone receptor agonists
- EGFR antagonist
- Steroid synthesis inhibitors
- HDAC inhibitors
- Folate/one-carbon metabolism inhibitor
- Glycolytic inhibitors
- Oxidative phos/mitochondral inhibitors
- Iron chelators
- Microtubule inhibitors
- Liver cholestasis inducerss

Genes Significantly Changed

Chemical	MCF7	Ishikawa	HepG2
Bisphenol A	**76**	**5262**	**9247**
Trenbolone	**188**	**18**	**3**
methotrexate	**3296**	**16**	**5376**
vorinostat	**17342**	**19432**	**21798**
RU486	**106**	**4**	**22**
Vitamin D3	**519**	**93**	**2**
Amoxicillin	**6**	**29**	**810**

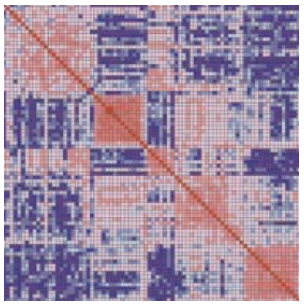
Connectivity Mapping: Example

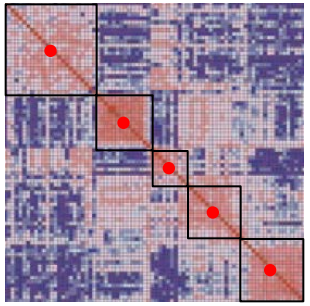
Bisphenol A comparisons

DES
resveratrol
epitiostanol
equilin
genistein
genistein
estrone
genistein
estradiol
levonorgestrel
resveratrol
equilin

landmark genes

- **expression of 978 *landmark* genes measured**
  - **selected from large, diverse, high-quality microarray dataset**
  - **orthogonal expression and validated predictive power**
- **inputs for genome-wide inference model**
  - **compute expression of transcripts not explicitly measured**
  - **flagged as *LM* (rather than *INF*) in output data file**





>100,000 Affy U133 scans

gene

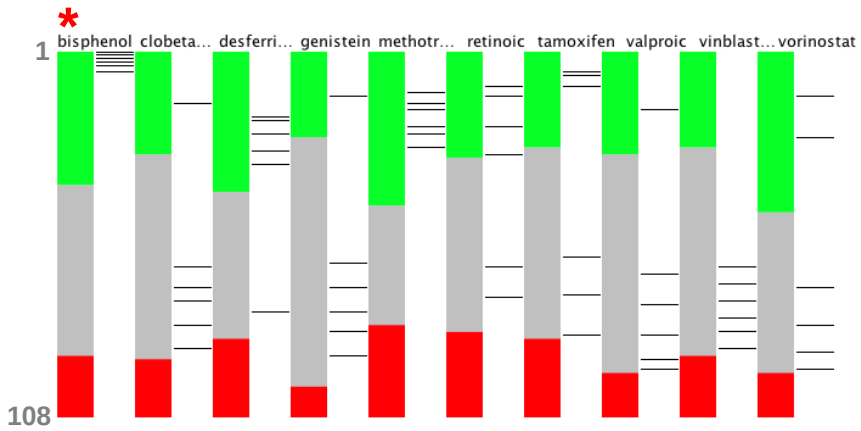
gene correlation

landmarks

AFFX *versus* L1000

- create *signatures* for each treatment from AFFX data
  - treatment (n=1) *versus* corresponding vehicle control (n=1)
  - 50 up- and 50 down- regulated genes by signal-to-noise
- create *instances* for each treatment from L1000 data
  - rank all features by extent of differential expression
  - treatment compared with matched control sample
- compute enrichment of each signature in each instance
  - rank instances based on these connectivity scores
  - AFFX signature finds expected L1000 instances in 9 of 10 tests

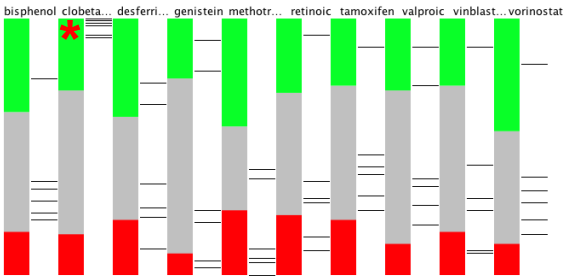
signature: bisphenol A



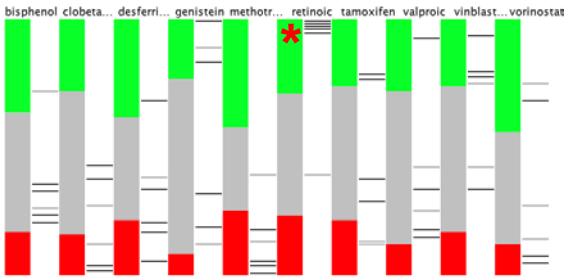
AFFX *versus* L1000

ranks of L1000 *instances* of each treatment with specified AFFX *signature*

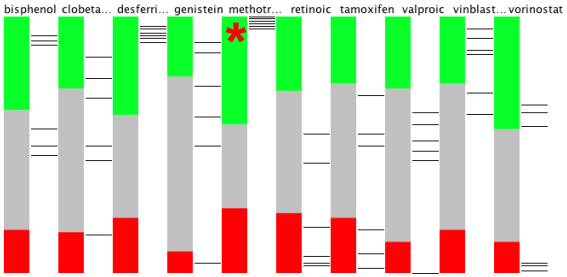
clobetasol



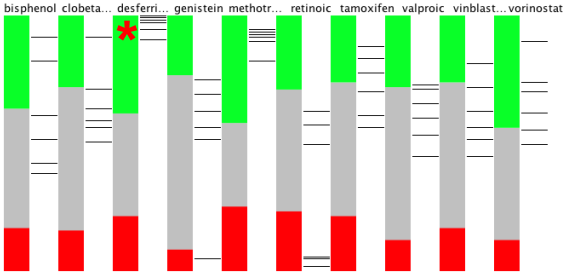
retinoic acid



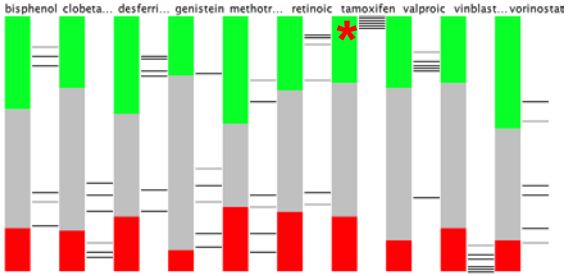
methotrexate



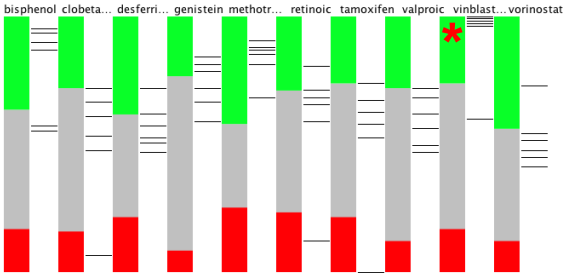
desferrioxamine



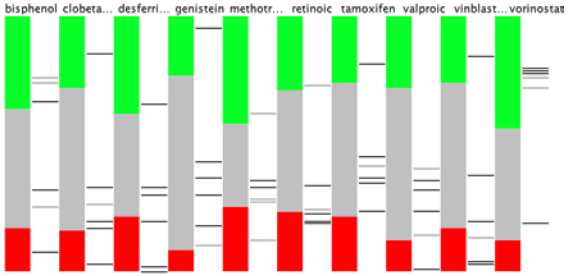
tamoxifen



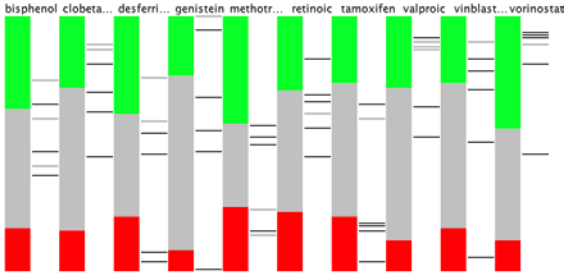
vinblastine



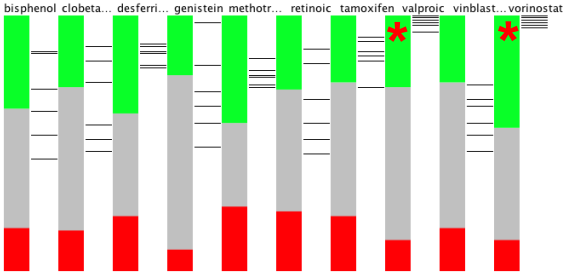
genistein



valproic acid



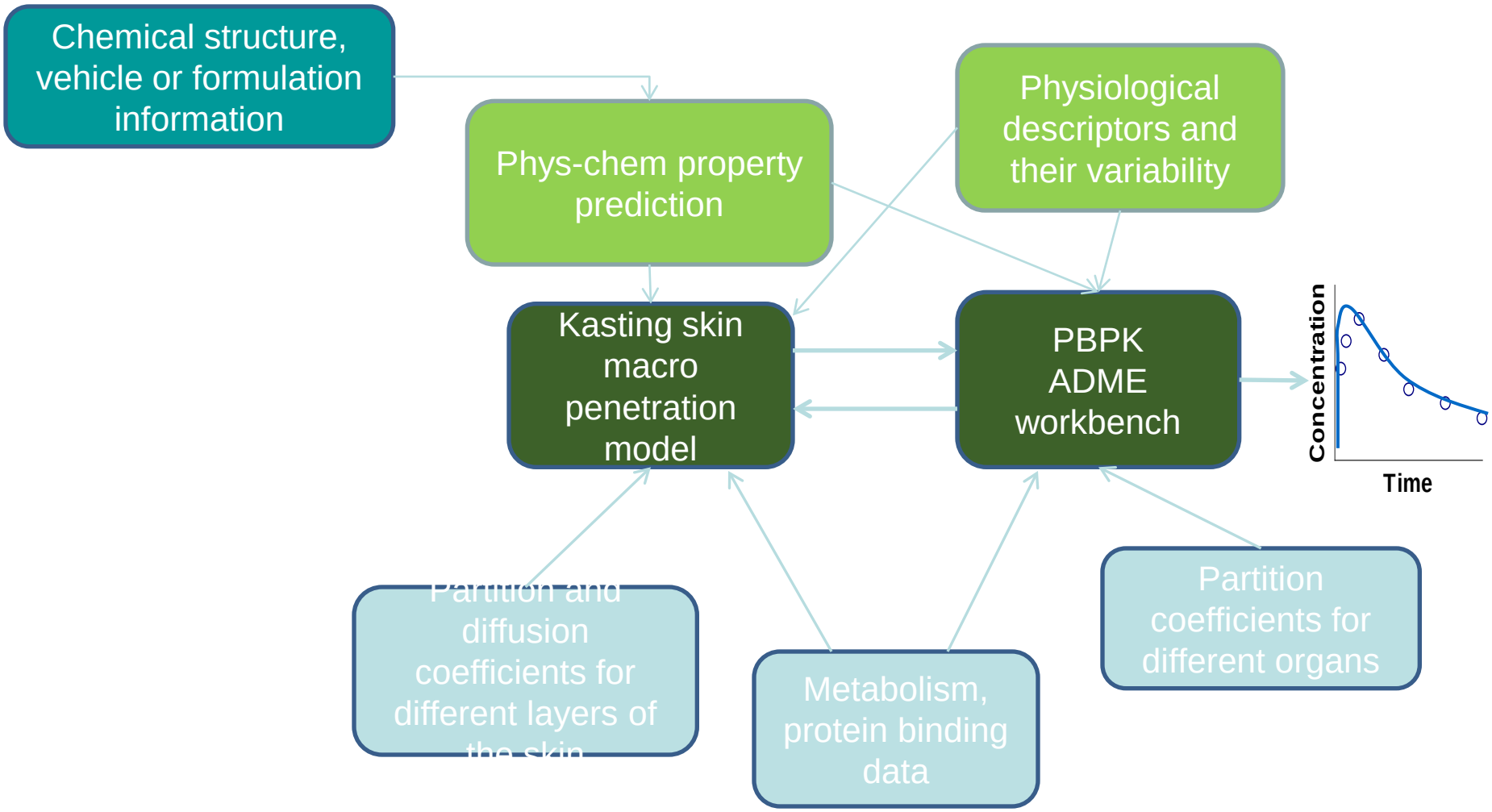
vorinostat



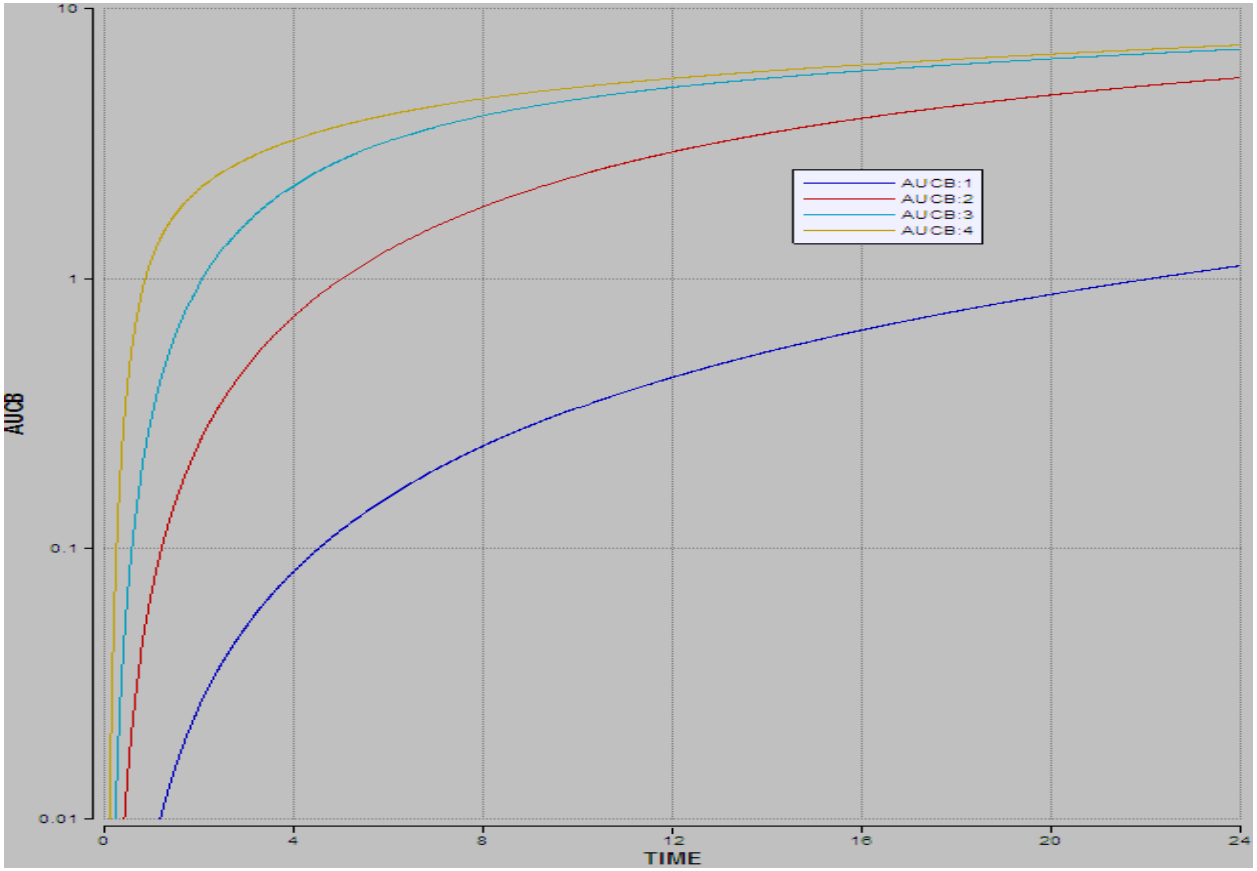
Modeling PK to Ensure the Right In Vitro Concentration

- Dose matters: data obtained in vitro at irrelevant concentrations is also irrelevant in predicting risk
- Active concentrations at the target tissue in vivo are predictable

PK workflow


Chemical structure, vehicle or formulation information
Phys-chem property prediction
Physiological descriptors and their variability
Kasting skin macro penetration model
PBPK ADME workbench
Concentration
Time
Partition and diffusion coefficients for different layers of the skin
Metabolism, protein binding data
Partition coefficients for different organs

Modeling AUC for a Range of Absorption Values



Conclusions

- Chemical ontology can aid in assigning chemicals to groups with same putative MOA
- It is possible to estimate the size of the MOA universe
- Linking initial molecular event with outcome will require considerable hypothesis testing, aided by gene expression data, modeling and simulation
- It is already possible to estimate tissue dose using computation, phys chem parameter estimation, and judicious data generation

Acknowledgements

- Cheminformatics
  - Shengde Wu
  - Karen Blackburn
  - Jorge Naciff
  - Joan Fisher
- Toxicogenomics
  - Jorge Naciff
  - Yuching Shan
  - Xiaohong Wang
  - Jay Tiesman
  - Greg Carr
  - Nadira deAbrew
- PK Modeling
  - Joanna Jaworska
  - Russell Devane
  - John Troutman
- CMAP (external)
  - Justin Lamb
  - Rusty Thomas
  - Ed Carney