Human Health Hazard Indicators Workshop
March 15-16, 2010
California EPA
Sacramento, CA

Computational toxicology screening methods

Dale Johnson, PharmD, PhD
UC Berkeley and Emiliem, Inc

Presentation Outline

- CompTox at UC Berkeley
- Processes and methods: 20+ years of validation
- >100K chemicals and the data gap issue
- QSARs – and the applicability domain
- Using information from related chemicals
- Structural motifs correlating with biology
- Carcinogenicity is too complex for single method
- Biological – signaling pathways critical
- Cannot stop with single chemical information or we miss products

Computational Toxicology at UC Berkeley

- Undergraduate Molecular Toxicology major
  - Computational toxicology is required
- Chemicals in the environment & therapeutics
- ~150 students – over 5 years
  - 4 hr credit course – individual projects; independent study & honors research
  - ~10% of students gain internships at FDA
  - Unbiased use and critique of software and databases
  - Profound innovation by undergraduates
  - 9 students with publications, 5 papers presented at national meetings, 1 book chapter

UCB Computational Toxicology Definition

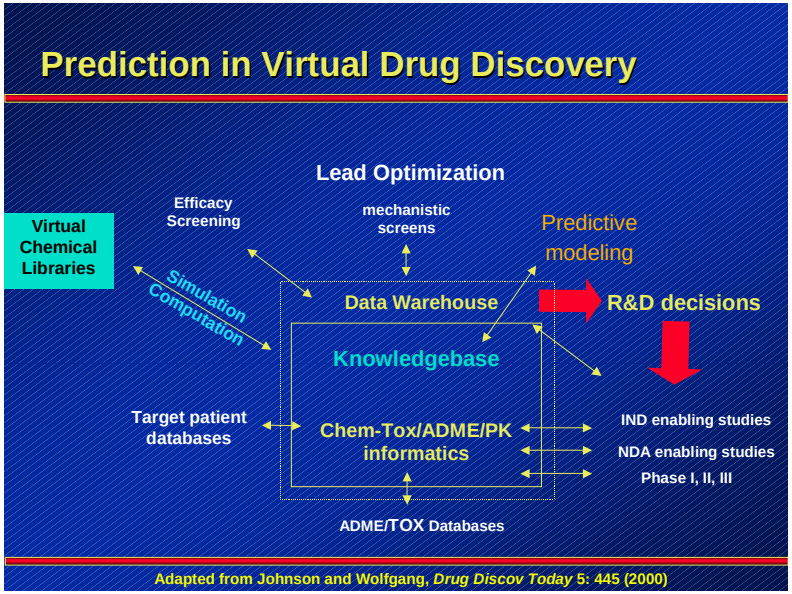
- The application of computer technology and mathematical / computational models to analyze, model and/or predict potential toxicological effects from:
  - Chemical structure (parent compound or metabolites)
  - Inference from similar compounds
  - Exposure, bioaccumulation, persistence
    - Biomonitoring data
    - Plasma or tissue concentrations
  - Differential indicators or patterns related to exposure
  - Networks of biological pathways affected by the chemical
- To further understand mechanisms of toxicity
  - Organism specific
  - Organ specific
  - Disease specific
- To explain why certain individuals are more susceptible
- Key methods
  - Chemical fragment or structural similarities (structural alerts)
  - Categorization or grouping
    - Analogs, categories based on mechanism, mode of action
  - QSARs
  - Biological pathway perturbations

Computational Toxicology had early roots in combinatorial chemistry

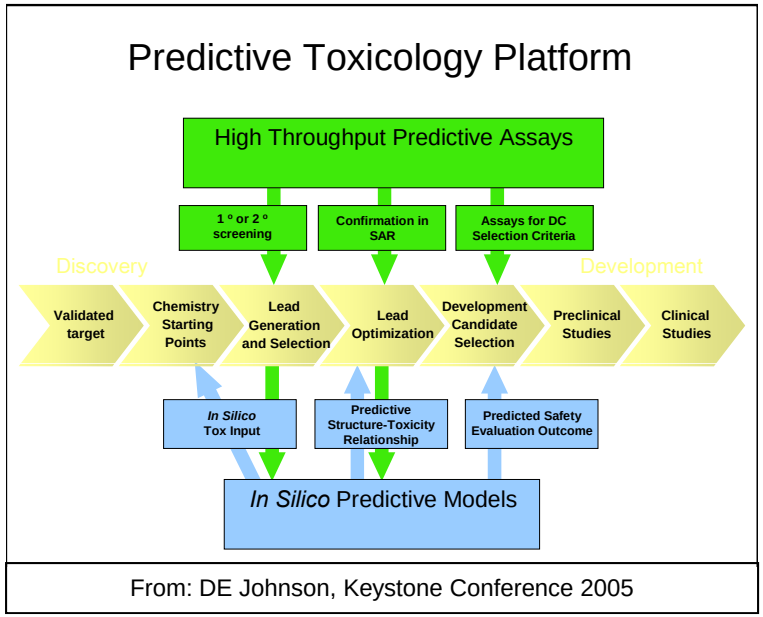
- Rapid synthesis or computer simulation of a large number of different but structurally related molecules or materials – (by building blocks)
- Highly parallel or split-pool chemical synthesis, resulting in thousands to millions of compounds
- 1000's of compounds in mixture (liquid state or solid state)
- De-convolution by:
  - structural similarity categories
  - rank order elimination algorithms based on targeted screening

- **The key lessons:**
  - Analog identification and categorization crucial for unknowns
  - Structural features are related to chemical-biological effects
  - SAR & QSAR could be used to fill data gaps with caution
  - Huge difference in rank ordering and predicting endpoints
  - Proper weighting of endpoint criteria essential

This lead to Data warehouse model



And to the screening based *in silico* platform

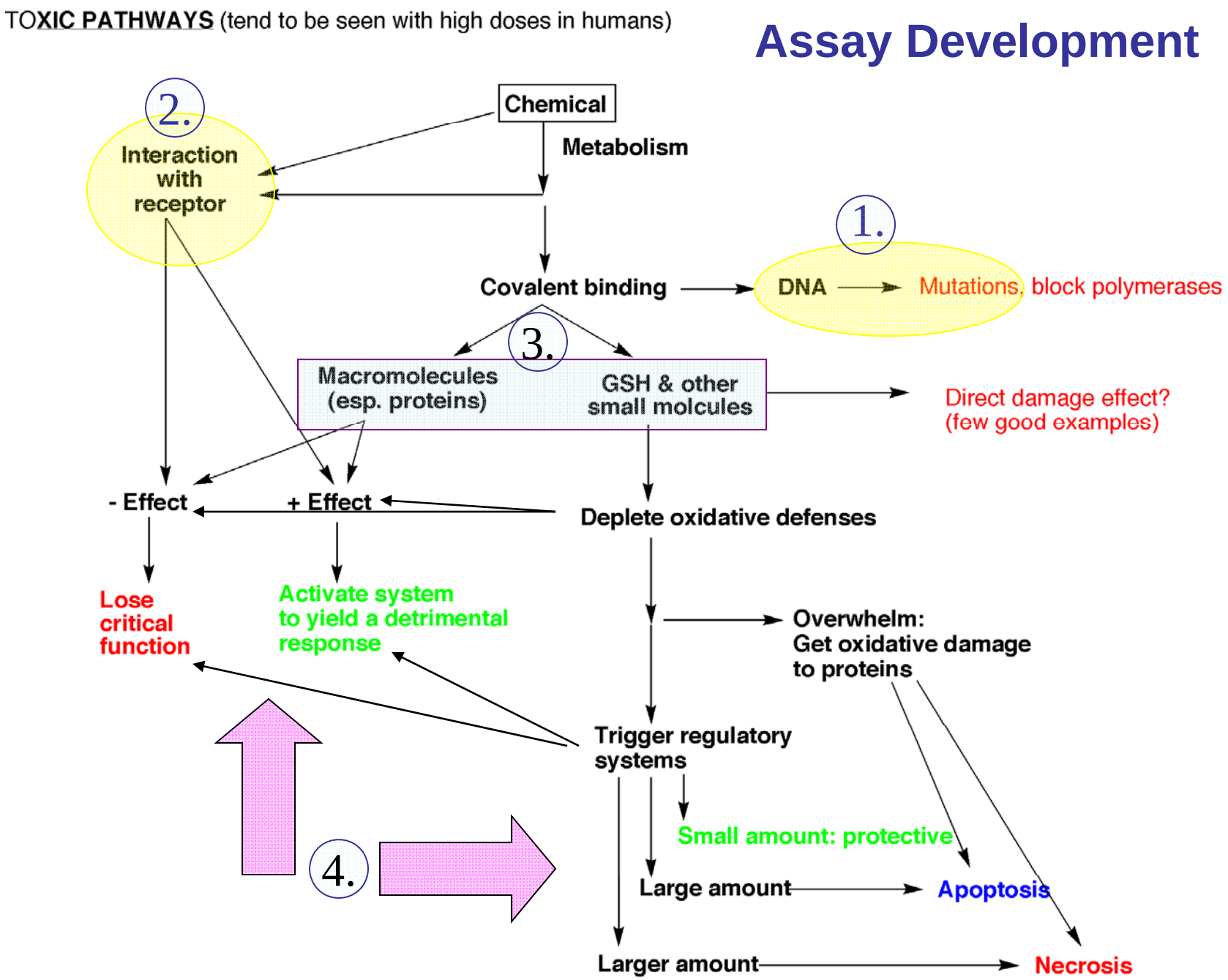


Millions of chemical compounds have been screened

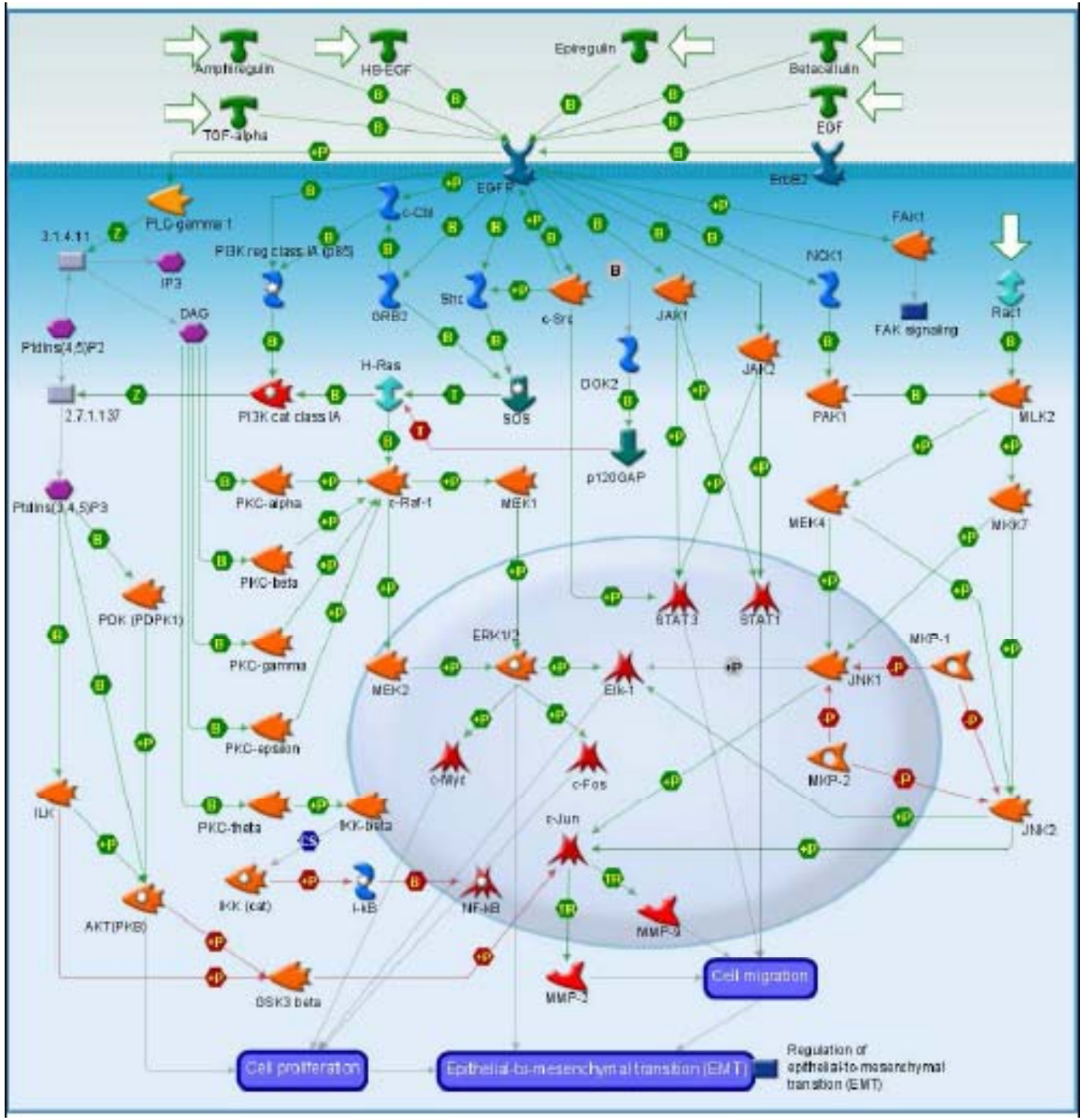
SARs have been established for multiple targets

Positive and negative “filters” are applied computationally

Assay Development



Liebler & Guengerich (2005) *Nature Rev. Drug Disc.* **4**, 410-420


Amphiregulin
HB-EGF
Epiregulin
Betacellulin
TGF-alpha
EGF
EGFR
ErbB2
PLC-gamma 1
c-Cbl
3.1.4.11
PI3K reg class IA (p85)
IP3
DAG
GRB2
Shc
c-Src
JAK1
NCK1
FAK1
FAK signaling
Rac1
Ptdins(4,5)P2
H-Ras
DOK2
JAK2
2.7.1.137
PI3K cat class IA
SOS
PAK1
MLK2
p120GAP
Ptdins(3,4,5)P3
PKC-alpha
c-Raf-1
MEK1
MEK4
MKK7
PKC-beta
PDK (PDPK1)
STAT3
STAT1
ERK1/2
MKP-1
PKC-gamma
MEK2
Elk-1
JNK1
PKC-epsilon
c-Myc
c-Fos
MKP-2
ILK
PKC-theta
IKK-beta
c-Jun
JNK2
IKK (cat)
I-kB
NF-kB
MMP-9
AKT(PKB)
GSK3 beta
MMP-2
Cell migration
Cell proliferation
Epithelial-to-mesenchymal transition (EMT)
Regulation of epithelial-to-mesenchymal transition (EMT)

The ~100K Chemical Challenge

- Data gap filling– Specific experimental data is preferred but often scarce
  - Modeled data is sometimes unreliable (e.g. outside domain of applicability)
- Use available “read-across” physical or chemical data from an analogous chemical or chemicals (e.g. water solubility)
  - Make predictions for missing toxicological and fate data
  - Quantitative or qualitative
- Enables grouping of chemicals – Separate similar assessments or one category assessment
  - Results partly based on common properties and modes of action
  - Increase consistency between assessments– Interpretation of data,
  - Areas of similarity and uncertainty

Methods to fill “data gaps”

- **SAR and QSAR**
  - Global models
  - Local models
  - Category specific QSAR models
  - Transparency of chemistry – no “black box” approach
  - OECD criteria
    - a defined end point
    - an unambiguous algorithm,
    - a defined domain of applicability,
    - appropriate measures of goodness-of-fit, robustness, and predictivity,
    - a mechanistic interpretation, if possible

What is a QSAR model?

$$Y_{i\,(est)} = \beta_0 + \beta_1 X_i + \varepsilon$$

$$\varepsilon_i = y_i - y_{i\,(est)}\,,\ i = index\ of\ obs.$$

$\mathbf{Y_i}$ is the dependent variable. the relevant biological activity

$\mathbf{X_i}$ is the independent variable. A descriptive property of the structure of the molecule.

"i" is an index for our observations. So Y_1 would be compound 1, Y_2 compound 2, etc.

ε_i represents the error term, or the deviation of our observed values of Y compared to predicted values of Y

β_0 is a constant term; it is the intercept of the line - to what degree, given a change in the independent variable (X), will we see a corresponding change in our dependent variable Y. In this example, this means that for a one unit change in X, we will see a β_1 change in Y.

Multiple Regression

$$Log(IC_{50}) = \beta_0 + \beta_1 log(X_1) + \beta_2 log(X_2) + \ldots + \beta_n log(X_n) + \varepsilon$$

Exactly similar to the previous example, only the X's represent different molecular descriptors.

Statistic model "Model 1"

	Molecules	pMED	pMED(calc)	Residuals
1	C6H5 CH3	4.3300	4.4557	-0.1257
2	CH3 (CH2)3C6H5	4.4700	4.4167	0.0533
3	C5H11 C5H11	4.6300	4.5658	0.0642
4	CH(CH3)2 CH(CH3)2	4.7700	5.0403	-0.2703
5	CH3 (CH2)3C6H5 CH2C	4.9500	4.9631	-0.0131
6	C3H7 C3H7	4.9200	5.1204	-0.2004
7	CH3 CH2)C6H5	5.1500	5.2078	-0.0578
8	CH2OCON(CH3)2 CH2OI	5.1600	4.8564	0.3036
9	C2H5 C2H5	5.4600	5.5353	-0.0753
10	CH3 (CH2)2OCH3	5.5700	5.7036	-0.1336
11	OCH3 OCH3	5.5900	5.5786	0.0114
12	CH3 CH(CH3)2	5.6000	5.5098	0.0902
13	C3H7 C3H7	5.6300	5.9594	-0.3294
14	CH3 CH3	5.6600	5.9810	-0.3210
15	H CH(CH3)2	5.6800	5.6697	0.0103
16	CH3 CH(OCH3)C2H5	5.6800	5.4860	0.1940
17	C3H7 CH2CH2OCONH2	5.6800	5.8560	-0.1760
18	CH3OC2H4 CH3OC2H4	5.6900	5.4395	0.2505
19	C2H5 CH(C2H5O)CH2OC	5.7600	5.8463	-0.0863
20	CH3 (CH2)2OCONH2	5.7800	5.9810	-0.2010
21	CH3 (CH2)3 dimer	5.8200	5.5472	0.2728
22	CH3 C2H5	5.8600	5.7569	0.1031
23	CH3 CH(OC2H4OCOCH3	6.0300	6.2264	-0.1964
24	CH3 CH2CH(CH3)OCON	6.1400	6.1414	-0.0013
25	C2H5 CH(CH2OCONH2)(	6.1600	6.1718	-0.0118
26	CH3 CH(CH2CH3)CH2OI	6.1800	6.2112	-0.0312
27	CH3 CH(OC2H5)CH2OC	6.1800	6.0730	0.1070
28	CH3 (CH2)3OCONH2	6.1800	6.1590	0.0210
29	CH3 (CH2)2OCONH2	6.2100	6.2878	-0.0778
30	C2H5 (CH2)2OCONH2	6.2500	6.0649	0.1851

Algorithm < Ordinary Multiple Regression >. Regression equation :

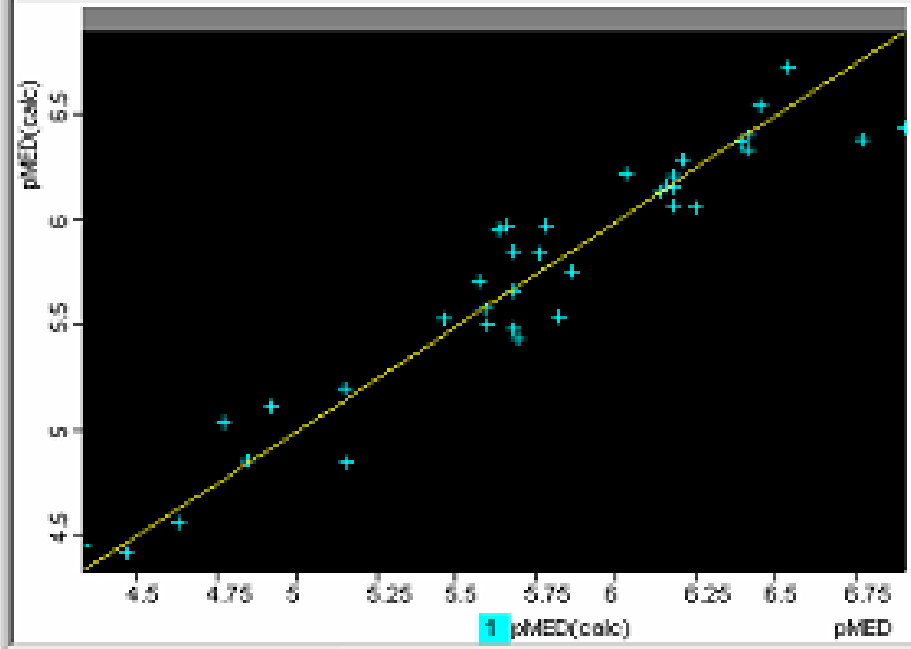
pMED = -1.389*SaasC - 0.241*xv1 + 1.975*xvpc4 - 0.3464*SsCH3 - 0.4404*SsssCH

Regression quality :

Multiple R-Squared = 0.9143 Standard error of estimation = 0.2
F-statistic = 66.17 P-value = 0
Multiple Q-Squared = 0.8688 Cross validation RSS = 1.838

Conclusion :

The training set is very well described by the regression equation, statistically very significant. Cross-validation shows that the constructed model can be used to predict



Selected 15 descriptors

In model used 5 descriptors and 37 molecules

Chemical descriptors must be relevant to the endpoint

Physicochemical interpretation and possible mechanistic involvement: examples of two widely used:

- Octanol/water partition coefficient: log *K*ow
  - Hydrophobicity / lipophilicity may influence absorption
  - Transport and location in tissues
  - Binding to active sites and/or receptors
  - Bioaccumulation
- Hammett electronic substituent constant (σ)
  - Electron-donating or -accepting properties of an aromatic substituent in the ortho, meta, or para position
  - Model electrophilic or nucleophilic reactivity

Methods to fill “data gaps” (cont.)

- **Structural alerts**
  - ToxTree, and combinations of models
- **Analog identification**
  - AIM (analog identification methodology – EPA)
  - OECD Toolbox
  - CAESAR and Lazy QSAR
  - ToxMatch
- **Categorization**

ToxTree Structural alerts: sensitive predictor for genotoxicity & <u>related</u> carcinogenicity

204 chemicals

with + mammary tumors in mice and/or rats (multiple sites)

[Rudel R, et al. 2007 Cancer 109(12 Supp): 2635]

315 chemicals:

107 with + mammary tumors in rats (multiple sites)

30 with + mammary tumors (only site)

104 with + carcinogenicity (but not mammary)

104 with no carcinogenicity; 22 with + Ames

[Cunningham A, et al. 2008 Chem Res Toxicol 21:1970]

167 of 204 chemical dataset had structural alerts for genetoxicity and carcinogenicity

Nitro-aromatic (41); primary aromatic amine, hydroxyl amine and esters (37); polycyclic aromatic hydrocarbons (21); hydrazine (20); alkyl and N-nitroso groups (19); aliphatic halogens (15); α,β unsaturated carbonyls (14); epoxides and aziridines (10)

Decision trees for the OECD Toolbox, a primary local QSAR model builder for green chemistry initiatives

S. Maxey, G. Mao, HW Kim, J. Chan, and D. Johnson

Department of Nutritional Science and Toxicology, University of California, Berkeley

#205—421

Abstract

This project was designed to evaluate the OECD Toolbox as a primary QSAR model builder for multiple green chemistry initiatives. The ability to identify and use relevant sets of compounds to form local QSAR models for multiple endpoints is essential for a wide range of potential green chemistry users.

In this study we used 96 hr LC50 data from fathead minnow ecotoxicity studies and compared results from the Toolbox to other freeware programs. Key compounds used in the analysis were butylamine, 4-fluoraniline, pentachlorophenol, 4-phenoxybenzaldehyde, furaldehyde, acetaldehyde oxime, triclosan, and ethanol. Experimentally generated LC50 data for each compound were used as the comparator controls in the analysis.

The Toolbox uses a variety of chemical, biochemical, and biological criteria, such as a structural similarity search, protein binding categorization, or US EPA and ECOSAR schemes to cluster compounds into relevant subsets. In our evaluation, the Organic Functional Groups, Structural Similarity, and Fragment Analysis categories tended to bias the user toward a global rather than a local comparison, hindering the later development of a local model—the strength of the Toolbox. The mechanistic ECOSAR and US EPA categories were found to be the most useful for the fathead minnow analysis as they focused only on chemical motifs documented to affect ecological toxicity. The Toolbox performed superiorly to other software in guiding the user toward more relevant local models.

Although initial use is far from straight forward, by carefully exploiting OECD tools like categorization and sub-categorization, the user can readily create a database of suitable compounds to produce relevant local QSAR models that can be further tuned with built-in non-linear regression capabilities. To aid other green chemistry users we created a simplified decision tree for practical application of the software.

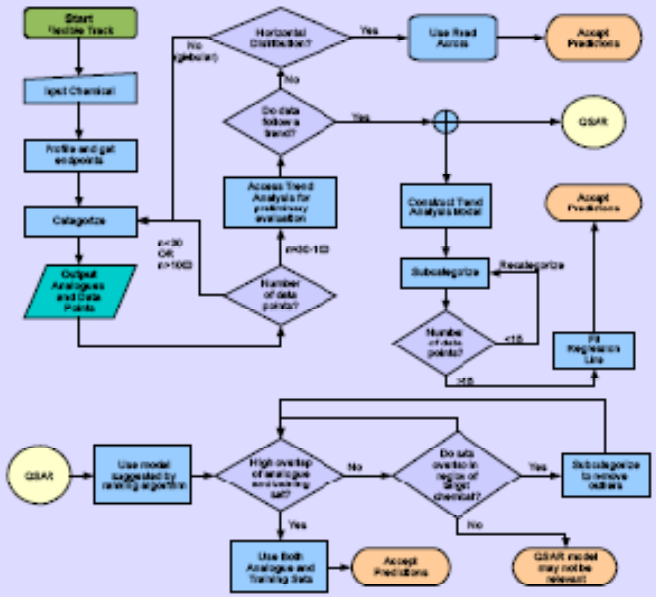
Conclusions

- The OECD toolbox can guide users towards a more local prediction model
- Care must be taken to select an appropriately sized analogue set
- Some chemicals respond better to a simple Read-Across average rather than sophisticated QSAR analysis.
- Compared with other free programs, the Toolbox gives superior results

However:

- The Toolbox still lacks many useful features, including data transformation, more complex regression techniques, and ease of data point manipulation
- Accurate results still require an understanding of the underlying toxicology

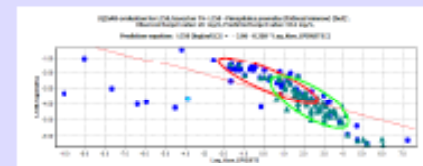
Flowchart



Example Chemical

Butylamine
CAS: 109-73-9

- Profiled by 'US EPA Categorization': Aliphatic Amines
- Received adequate data: 63 unique analogues, 141 data points
- Sinusoid trend, used QSAR module to predict toxicity, not Read-Across
- Moderate overlap of training set data and analogues:



- Subcategorize to improve results: Include only ECOSAR Classification 'Aliphatic Amines':

- Results:
  Tested Toxicity: 287 mg/L
  Predicted (Initial): 33.8 mg/L
  Predicted (Improved): 70.7 mg/L

Comparative Performance

- OECD Toolbox compared to two other free, online tools and to experimental toxicity.
- Toolbox, when used correctly, outperforms these other tools.

Compound	MW (g/mol)	Experimental Toxicity (mM)	LazyQSAR (mM)	Danish Database (mM)	Best Toolbox Estimate(s) (mM)	Method used
Butylamine	73.14	3.63	3.714	2.654	3.72 3.54	Read-Across QSAR
4-fluoro-aniline	111.12	0.170	0.239	0.486	0.170	Read Across, Trend Analysis
Furaldehyde	96.09	0.224	0.545	10.4	0.275 0.234	Trend Analysis, Read-Across
Acetaldehyde Oxime	59.06	1.29	N/A	16.9	1.30	Read-Across
Pentachloro-phenol	266.35	0.000977	0.00604	0.00139	0.000955	Read-Across, Trend-Analysis
4-phenoxybenzaldehyde	198	0.0234	0.0228	0.0226	0.0234	Trend-Analysis
Triclosan	289.55	0.00105	0.00770	0.00131	0.00195	Trend-Analysis
Ethanol	46.07	89.1	N/A	21.7	95.5	Trend-Analysis

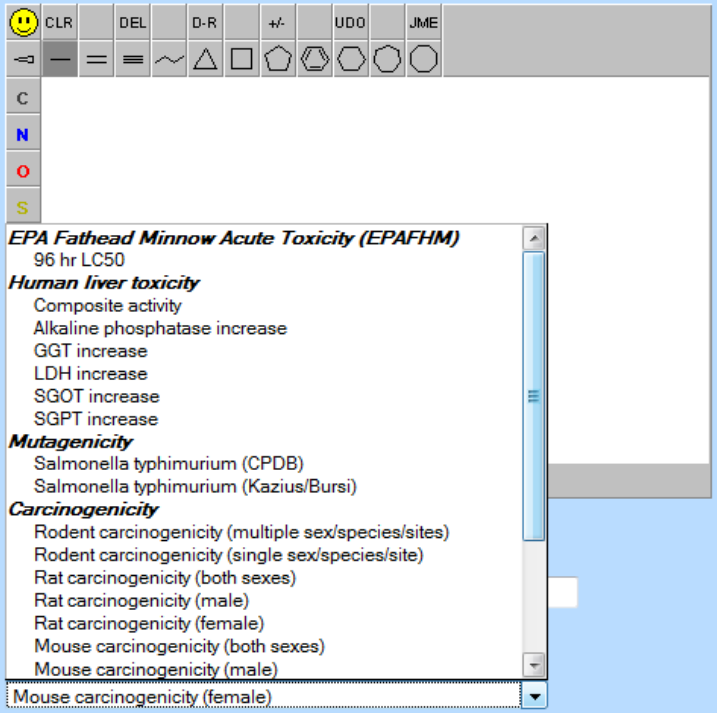
Student Affiliations

Sunny Maxey
Undergraduate, UC Berkeley
smaxey@gmail.com
Guandong Mao
University of Texas, Southwestern Medical Center, Dallas
thermac@gmail.com
Heul won Kim
UC Berkeley Alumnus
hyerik87@gmail.com
Jason Chen
Denny Lab, UCLA
jason.chen1187@gmail.com

Input | Validation | Documentation

Lazar Toxicity Predictions

1. Draw a chemical structure (help)



2.

3. Predict

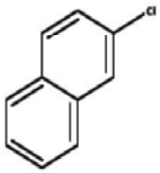
© in silico toxicology 2004-2008
Built with: lazar, OpenBabel, CDK, Ruby on Rails
JME Editor courtesy of Peter Ertl, Novartis

Several models have been developed under the CAESAR project for the five chosen endpoints relevant for REACH legislation:

In order to make these models accessible and easy to use, a **java-based application package** has been developed in collaboration with *developers of Dragon, AGDesigner, Mr. Davide Bigoni* and *Dr. Todd Martin* (US EPA).

CAESAR QSAR model for bioconcentration factor (BCF) in fish - version 1.0

Prediction for the compound no. 1: [Cl]/C2=C/C1=CC=CC=C1C=C2

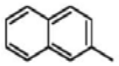


BCF value: 1112 (L/Kg) whole body weight

Log BCF value: 3,05

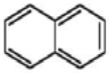
Remarks for the prediction:

The following chemicals similar to the query compound have been identified in the CAESAR database:



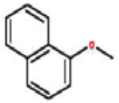
Dataset id: 20
SMILES: Cc1ccc2c(c1)cccc2
Similarity: 0,988

Experimental Log BCF: 2,41
Predicted Log BCF: 2,72



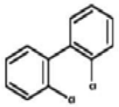
Dataset id: 19
SMILES: c1c2c(ccc1)cccc2
Similarity: 0,982

Experimental Log BCF: 2,50
Predicted Log BCF: 2,48



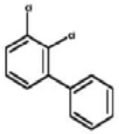
Dataset id: 280
SMILES: COc1cccc2c1cccc2
Similarity: 0,981

Experimental Log BCF: 2,21
Predicted Log BCF: 1,58



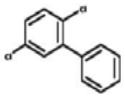
Dataset id: 177
SMILES: Clc1c(cccc1)c1c(cccc1)Cl
Similarity: 0,964

Experimental Log BCF: 3,53
Predicted Log BCF: 3,88



Dataset id: 178
SMILES: Clc1c(cccc1Cl)c1ccccc1
Similarity: 0,964

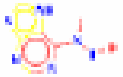
Experimental Log BCF: 3,72
Predicted Log BCF: 3,88



Dataset id: 179
SMILES: Clc1c(cc(cc1)Cl)c1ccccc1
Similarity: 0,964

Experimental Log BCF: 3,86
Predicted Log BCF: 3,88

Carcinogenicity - Mouse carcinogenicity (female) [Validation and endpoint definition]

Predicted Activity (Confidence)	Structure	Measured Activity	Additional Information
active (0.00919008) **unreliable:** unknown/infrequent features low confidence (<0.025)		not available	Relevant Fragments DSSTox database PubChem database

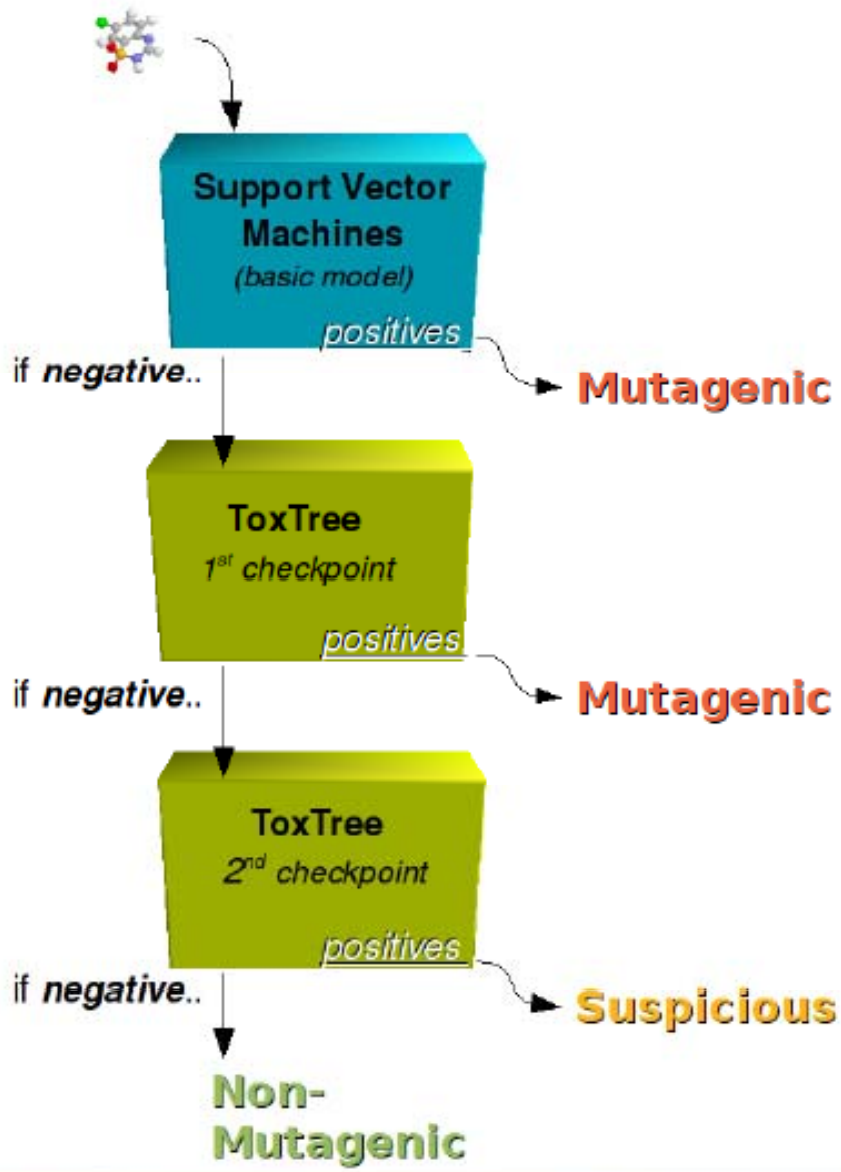
Similar Structures (10 from 28 Neighbors)

Similarity	Structure	Measured Activity	Additional Information
1.00		inactive	Original data (DSSTox) PubChem database
0.67		active	Original data (DSSTox) PubChem database
0.56		active	Original data (DSSTox) PubChem database
0.46		active	Original data (DSSTox) PubChem database

Architecture of the integrated system

3 steps in cascade:

- **statistical model** (based on *chemical descriptors*)
- **knowledge-based filter** (based on *structural alerts*)



Integrated model statistics

CAESAR test set	**Suspicious** taken as **non-mutagenic**	**Suspicious** taken as **mutagenic**
accuracy:	*83.3%*	*82.1%*
sensitivity:	*88.3%*	*90.9%*
specificity:	*77.1%*	*71.2%*

CONFIDENT CHOICE

✓ *Accuracy* close to the reliability of the experimental test (85%)

PRUDENT CHOICE

✓ *Sensitivity* boosted over 90%

Non-mutagenic carcinogens more challenging – proper categorization essential

- Co-initiation
  - Facilitating original mutagenic changes in stem or progenitor cells that start cancer process
  - Induction of enzymes activating other chemicals
- Promotion
  - Enhancing growth vs. differentiation/death of initiated clones
- Progression
  - Enhancing growth, malignancy, or spread of already developed tumors
  - Suppression of immune surveillance
  - Hormonally mediated growth stimulation for tumors with approapriate receptors (eg. estrogen)
- Multi-phase
  - Epigenetic silencing of tumor suppressor genes

From: Hattis, et al. Crit Rev Toxicol (2009) 39 (2) 97-138

Methods to fill “data gaps” (cont.)

- **ID targets and biological pathways affected**
  - Genego Metadrug and Metacore

CHEMICAL EXPOSURE AND BREAST CANCER: IDENTIFYING COMPOUNDS OF CONCERN

#1999 - 726

Chan, E.[1]; Dairkee,S.[2]; Janssen, S.[3]; Latimer, J.[4]; Rudel, R.A.[5]; Schwarzman, M.[1]; Zeise, L.[6]; JeBaily, L.[7]; Johnson, D.E.[1]

[1] University of California, Berkeley, Berkeley, CA,; [2] California Pacific Medical Center, San Francisco, CA,; [3] Natural Resources Defense Council, San Francisco, CA,; [4] University of Pittsburgh Cancer Institute, Pittsburgh, PA,; [5] Silent Spring Institute, Newton, MA,; [6] Office of Environmental Health Hazard Assessment, California Environmental Protection Agency, Oakland CA,; [7] Genego Inc., St. Joseph, MI.

Abstract

This research was conducted to evaluate in silico methods for identifying and prioritizing chemicals of concern that may increase the risk of human breast cancer. Candidates identified would reflect possible mechanistic causality. Hence they could enter a pipeline for subsequent rigorous testing through an experimental methods-based procedure proposed by the California Breast Cancer and Chemicals Policy Project. Criteria for selection were NIAMBS biomonitoring data, presence in breast tissue/milk, hormonal activity, compound reactivity, modulation of critical metabolic enzymes or transporters in breast tissue, and interaction with known human breast cancer targets. The primary dataset included 216 compounds associated with mammary tumors in animals (Rudel R, et al. Cancer 2007). A systematic process for filling datagaps was established using Lazar, CAESAR, OECD toolbox, Toxtree, Pubmed, Toxnet, -among others; and Metabolog and MetaCore from GeneGo. Key molecular targets and activating pathways were identified for each compound using MetaLog and pertinent published information (Weigert R, et al. J Pathol 2006; Yamoah C and Rahman N. Annu Rev Genom Human Genet 2008). The potential for each compound to activate key metabolic enzymes expressed in breast tissue was assessed as suggested by the literature (Williams J and Phillips D. Cancer Ler 2000) and QSAR modeling. A majority of compounds testing positive for mutagenicity and carcinogenicity contained structural alerts associated with such activity. Moreover, 11 compounds predicted modulation of human breast cancer related targets/pathways, and 35 predicted activation of key metabolic enzymes known to play an important role in carcinogen metabolism. This approach provides valuable guidance in the development of a roadmap to guide chemical screening for agents likely to be involved in breast cancer development and progression.

Datasets for Analysis

- 204 chemicals with + mammary tumors in mice and/or rats (multiple sites) (Rudel R, et al. 2007 Cancer 109(12 Suppl): 2635)
- 315 chemicals: (Cunningham A, et al. 2008 Chem-Biol Toxicol 21:290)
  - 107 with + mammary tumors in rats (multiple sites)
  - 30 with + mammary tumors (only site)
  - 104 with + carcinogenicity (but not mammary)
  - 104 with no carcinogenicity; 32 with + Ames

Data gaps identified

In 315 'well characterized' chemicals:

- 96 (30%) had data gaps for genotoxicity

In 104 chemicals carcinogenic but not mammary

- 22 (21%) had data gaps for female mammary tissue

Hazard ID – Data Gap Evaluation

In this on-going project, several in silico methods are being evaluated, with the goal of selecting freely available tools combined with a multi-purpose package that includes systems biology functionality. We have established an evaluation procedure focused on breast cancer, but consistent with any disease-specific health hazard. This involves (1) establishing disease-specific endpoints that can be evaluated by data inquiry, (2) conducting hazard identification / data gap analysis on each chemical, (3) filling data gaps via in silico evaluation (4) defining a testing battery to generate endpoint data, (5) prioritizing chemicals for further testing (6) defining hazard outcome by chemical – documentation in disease-specific Db (7) identifying chemicals of concern, highest level move to action required status, (8) exposure assessments by chemical including targeted, susceptible populations (9) disease-specific risk assessment, combining exposure with hazard outcome- documentation in disease-specific Db (10) if risk not adequately controlled, chemical(s) move to action required status. The above is not a linear step-wise approach but a dynamic set of criteria subject to advances in science and technology and new data.

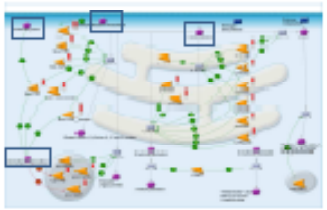
Since human breast cancer has not been adequately represented in past hazard data generation, chemicals could move to action required status expeditiously.

This poster presents preliminary results and highlights work focused on identifying biological target and pathway interactions with one set of chemicals. 204 compounds were analyzed against Phase I & II metabolizing enzymes, relevant transporters, and multiple genes and networks known to be associated with breast cancer. Hypotheses have been proposed for markers for future testing and MIS.

The legend below defines objects in network diagrams.

Chemicals modulating Breast Cancer targets / pathways

Estradiol Metabolism with examples of compound interactions

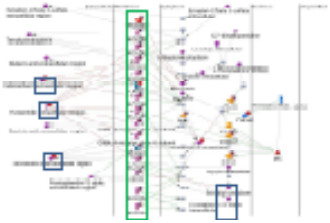


Trans-Activation ligand-dependent activation of RXR/VDR pathway

Lipid Metabolism With nuclear pathway

Breast cancer targets interacting with chemicals marked in red or indicated by red circle or sphere

Example of cellular function of targets



Antiapoptosis signaling

Results and Conclusions

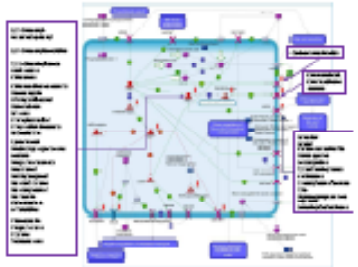
Structural alerts from Toxtree most sensitive predictor for genotoxicity and mechanistically related carcinogenicity:

167 of 204 chemical dataset had structural alerts for genotoxicity and carcinogenicity

Top tier: Nitro-aromatic (41); primary aromatic amine, hydroxyl amine and esters (37); polycyclic aromatic hydrocarbons (21); hydrazine (20); alkyl and N-nitroso groups (19); aliphatic halogens (16); α,β-unsaturated carbonyls (14); epoxides and aziridines (10)

Hypothesis

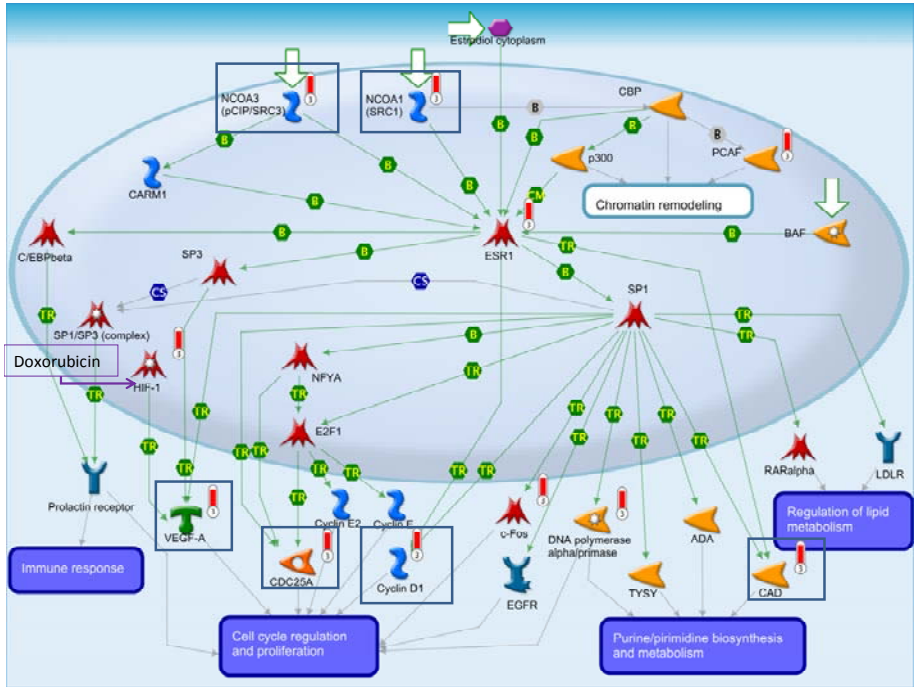
Several compounds activate NR1I2/PXR (Pregnane X receptor): binds as a response element in the CYP3A4 and ABCB1/MDR1 gene promoter; activates expression in response to a wide variety of endobiotics and xenobiotics (by structural similarity). Activation and downstream effects suggest it may be a xenosensor of endocrine disruption (Kretschmer & Baldwin 2006 Chem-Biol Interact 160:171)



PXR-related transport; and network of PXR upstream- and downstream effectors, breast cancer target interactions

In silico interrogation of chemical interactions with breast cancer related genes and biological pathways can help to find common mechanistic pathways for chemicals and lead to defining biomarkers for future testing schemes.

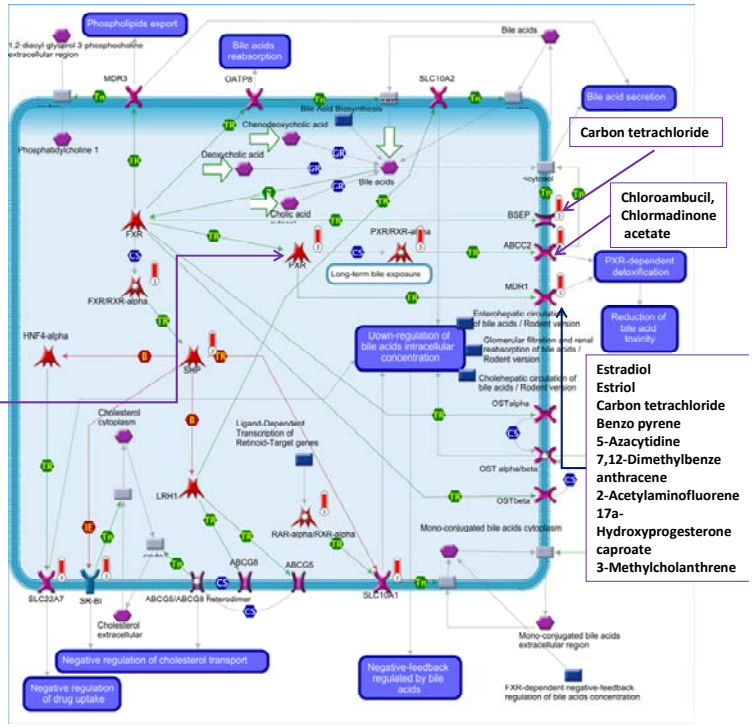
Biological pathway analysis: testing large sets of compounds to understand molecular targets



3,2'-Dimethyl-4-aminobiphenyl

3,3'-Dimethylbenzidine

7,12-Dimethylbenze Anthracene
Chlordane
Chlormadinone acetate
Daunomycin
Diethylstilbestrol
Doxorubicin
Estrone
Ethinylestradiol
Ethynodial diacetate
Griseofulvin
Lynestrenol
Medroxyprogesterone acetate
Megestrol acetate
Mestranol
Methyleugenol
Norethisterone
Norethynodrel
Norlestrin
Ochratoxin A
o-Toluidine
Phenesterin
Progesterone
Styrene
Testosterone



204 compounds were analyzed against Phase I & II metabolizing enzymes, relevant transporters, and multiple genes and networks known to be associated with breast cancer (~120 models)

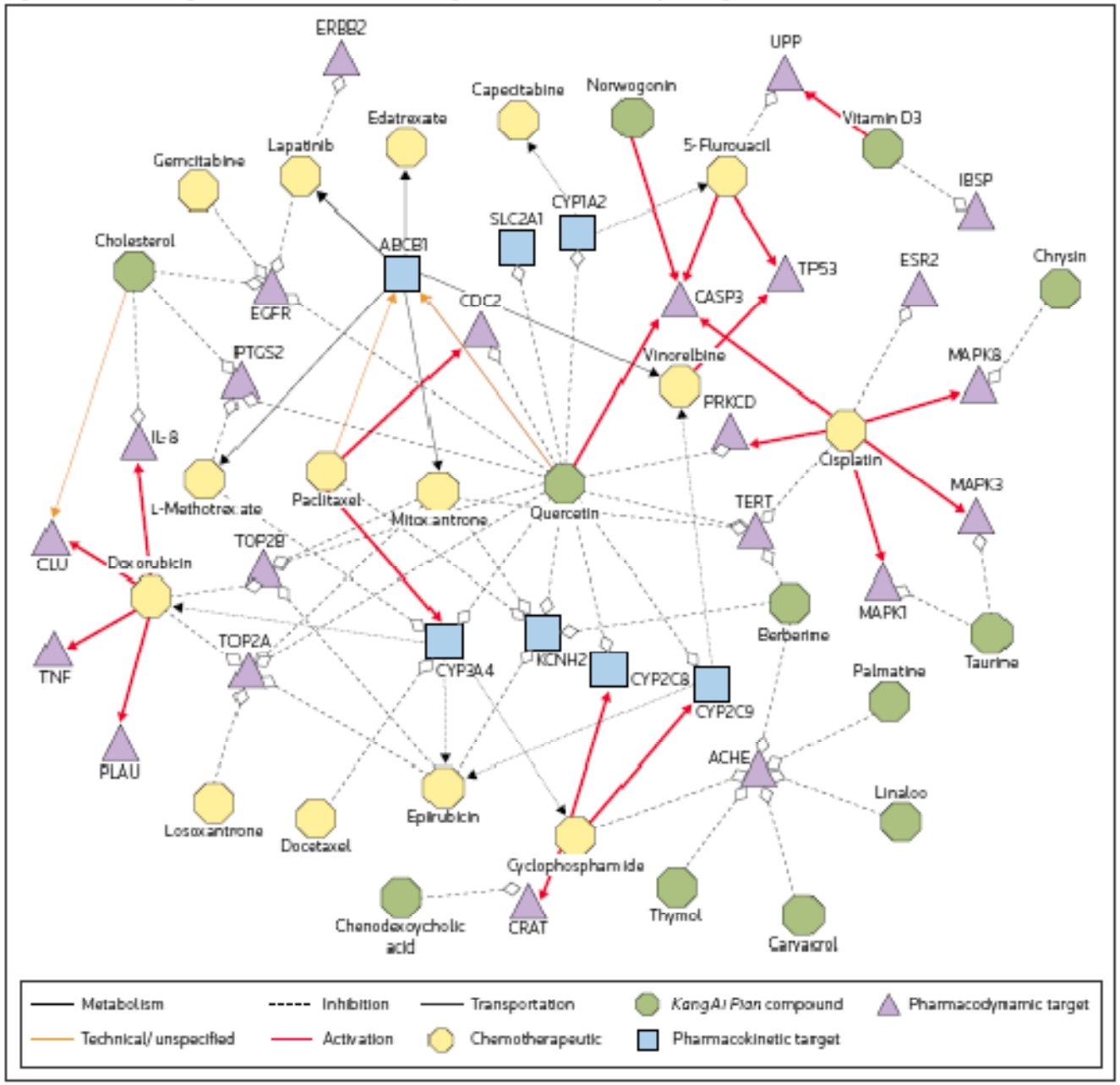
25 Breast Carcinogens w/Molecular Targets involved in breast cancer

	Breast Carcinogen	# of affected breast cancer targets	Action	Breast cancer gene/ target affected by carcinogen	Target (gene) description	Source
1	17a-Hydroxyprogesterone caproate	1	transported by	ABCB1	ATP-binding cassette, sub-family B (MDR/TAP), member 1	QSAR
2	Chlormadinone acetate	2	transported by	ABCB1	ATP-binding cassette, sub-family B (MDR/TAP), member 1	QSAR
			inhibition	AR	androgen receptor	PUBMED
3	Daunomycin	1	inhibition	ABCB1	ATP-binding cassette, sub-family B (MDR/TAP), member 1	pubmed & QSAR
4	Diethylstilbestrol	1	inhibition	SLC2A1	solute carrier family 2 (facilitated glucose transporter), member 1	PUBMED
5	Doxorubicin	3	inhibition	ABCB1	ATP-binding cassette, sub-family B (MDR/TAP), member 1	QSAR
			inhibition	HIF1A	hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)	PUBMED
			activation	PLAU	plasminogen activator, urokinase	PUBMED
6	Estradiol-17b	13	inhbition/ unspecificied	AR	androgen receptor	PUBMED
			unspecified	CXCR4	chemokine (C-X-C motif) receptor 4	PUBMED
			inhibition	ERBB2	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)	PUBMED
			activation	IGF1R	insulin-like growth factor 1 receptor	PUBMED
			inhibition	IGFBP3	insulin-like growth factor binding protein 3	PUBMED
			inhibition	NOTCH3	Notch homolog 3 (Drosophila)	PUBMED
			inhibition	PLAU	plasminogen activator, urokinase	PUBMED
			activation	SLC2A1	solute carrier family 2 (facilitated glucose transporter), member 1	PUBMED
			activation	SOD2	superoxide dismutase 2, mitochondrial	PUBMED
			activation	BCAS3	breast carcinoma amplified sequence 3	PUBMED
			activation	TFF1	trefoil factor 1	PUBMED
			inhibition	ABCG2	ATP-binding cassette, sub-family G (WHITE), member 2	PUBMED
			activation	MUC1	mucin 1, cell surface associated	PUBMED
7	Estriol	1	activation	ERBB2	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)	PUBMED
8	Estrone	2	activation	ERBB2	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)	PUBMED
			inhibition	ABCG2	ATP-binding cassette, sub-family G (WHITE), member 2	PUBMED
9	Ethynodial diacetate	1	transported by	ABCB1	ATP-binding cassette, sub-family B (MDR/TAP), member 1	QSAR

Single Chemical vs. “Products” or mixtures

- Phytochemicals
  - The WHO estimates that ~80% of global population relies on traditional herbal medicines as part of standard healthcare
  - In USA, 1 in 5 individuals regularly consume dietary supplements
    - Out-of-pocket spending on non-vitamin, non-mineral natural products, 1/3 of pharmaceuticals
- Product information and ratings: GoodGuide
- Industry specific product / chemical usage and ratings: SciVera maps auto industry

Figure 2. Potential drug-herb interactions between *KangAi Pian* and chemotherapeutic agents.



From: Chan, et al. (2010) Curr Opin Drug Disc Devel 13(1) 50-65

Conclusion

- Current developing practices for filling data gaps has long history with validation
- Physico-chemical properties accurately predicted
- Databases linking chem-biol interactions are limited in scope and size
  - Genotoxicity linked to carcinogenicity generally accurate predictions
  - Non-gentoxic carcinogenicity more complex
- QSARs always limited by applicability domain (chemical space) of the training sets
- Interactions with key biological pathways important to identify targets and mechanisms
- Additional screening is mandatory