

# Green Chemistry in Higher Education

October 26, 2010

Berkeley, CA



## Computational Toxicology Coursework at UC Berkeley

Dale Johnson, PharmD, PhD  
Dept of Nutritional Science and Toxicology  
College of Natural Resources  
UC Berkeley



# Computational Toxicology at UC Berkeley

- Focus on chemicals in the environment, consumer products, and chemicals as therapeutics
- Undergraduate Molecular Toxicology major
  - 2006 -2010
    - Required 4 credit course – individual/group projects, independent study and honors research
    - ~10% of students received internships at FDA
    - 9 students with publications, 10 students with publications pending, 6 posters presented by students at national meetings, 1 student co-authored book chapter
  - 2011 →
    - Splitting educational concept into 2 courses
    - NST 121 Computational Toxicology (3 credits)
      - Toolbox creation, environmental and disease related issues
    - NST 115 Principles of Drug Action (2 credits)
      - Therapeutics and new data sources

# 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 (biomarkers)
  - 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

# Course Structure and Resources

- **UC Berkeley computer laboratory**
  - College of Natural Resources; Geospatial Innovation Facility (GIF); Mulford
  - Full administrative and technical support
- **Tool Box Creation**
  - A combination of free on-line resources and commercial software
  - Major software contribution from **Genego, Inc.**
  - Converting chemical and biologic data into usable formats
  - Student proficiency exercises
- **Tutorials and resources**
  - Toxicology tutorials for non-toxicology majors
  - Software tutorials and user manuals
  - Extensive links to software, datasets, environmental and chemical information
- **Environmental or therapeutic challenge**
  - **2010 student example**
  - Create *in silico* methods to identify and prioritize chemicals of concern that may increase the risk of human breast cancer
  - Innovate to solve chemical-related disease issues in new ways

## Challenges: 2006 → 2011

- Toxicity data exists in different databases, different formats, and not always compatible for *in silico* modeling
- Difficult to select and combine chemical and toxicity data from multiple sources
- Difficult to integrate public and “in-house” data, and to incorporate predictions from various applications
- Development, validation, application, and interpretation of QSAR models difficult for most toxicologists

**However.....**

# The Good News

- Continued increased availability of larger and better curated public databases
- Increase in open-source predictive tools
- Now very close to providing:
  - Flexible framework that integrates existing data sources, predictive solutions, and emerging developments
  - Integration of chemical-biological data acquisition, filtering, and processing
    - Interactions between proteins, genes, networks, and chemicals
    - Possible metabolic transformations of chemical
    - Mutants and variants of proteins that define population or individual susceptibility

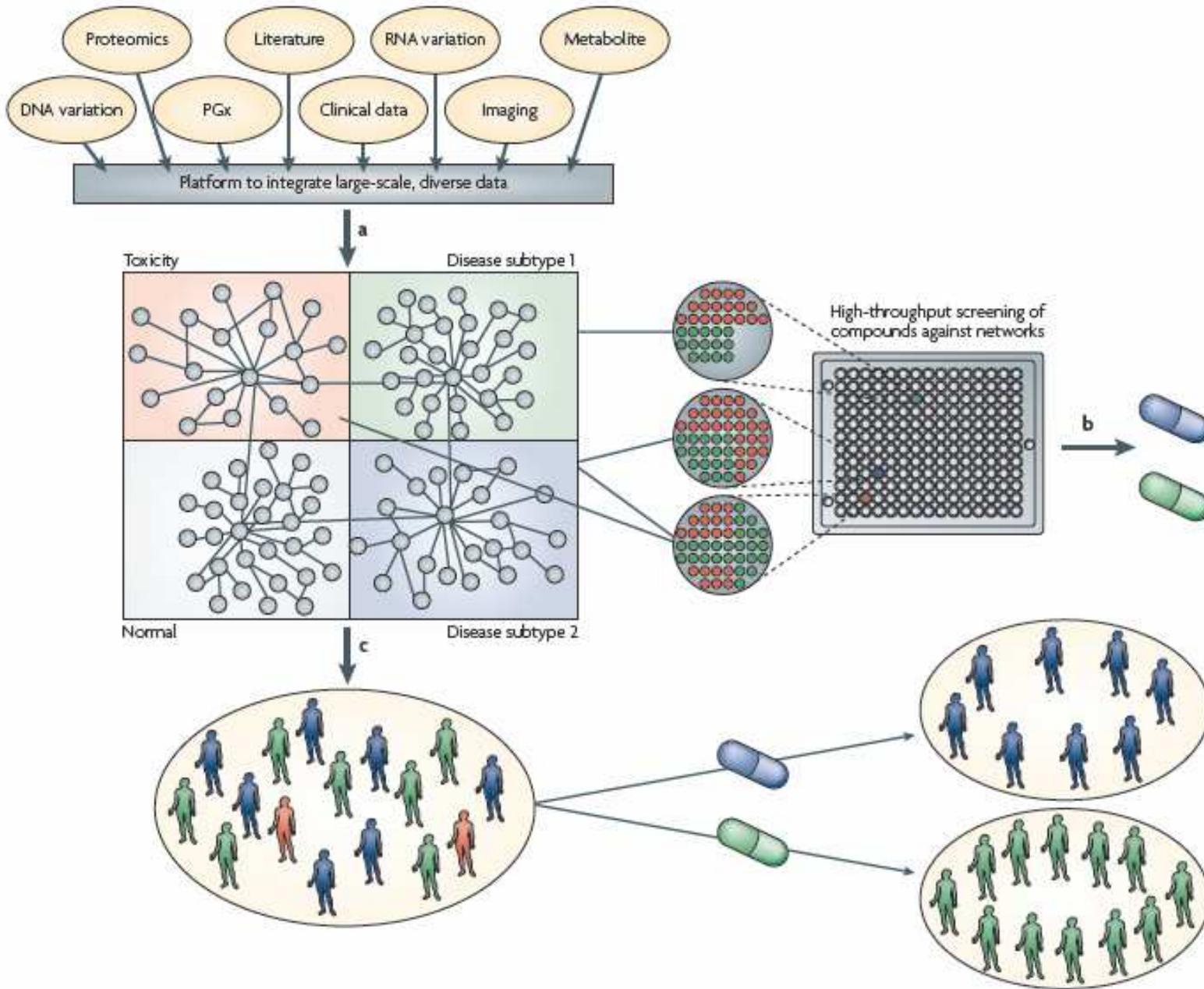
**Why it is important to view computational toxicology from both the therapeutic and environmental sides**

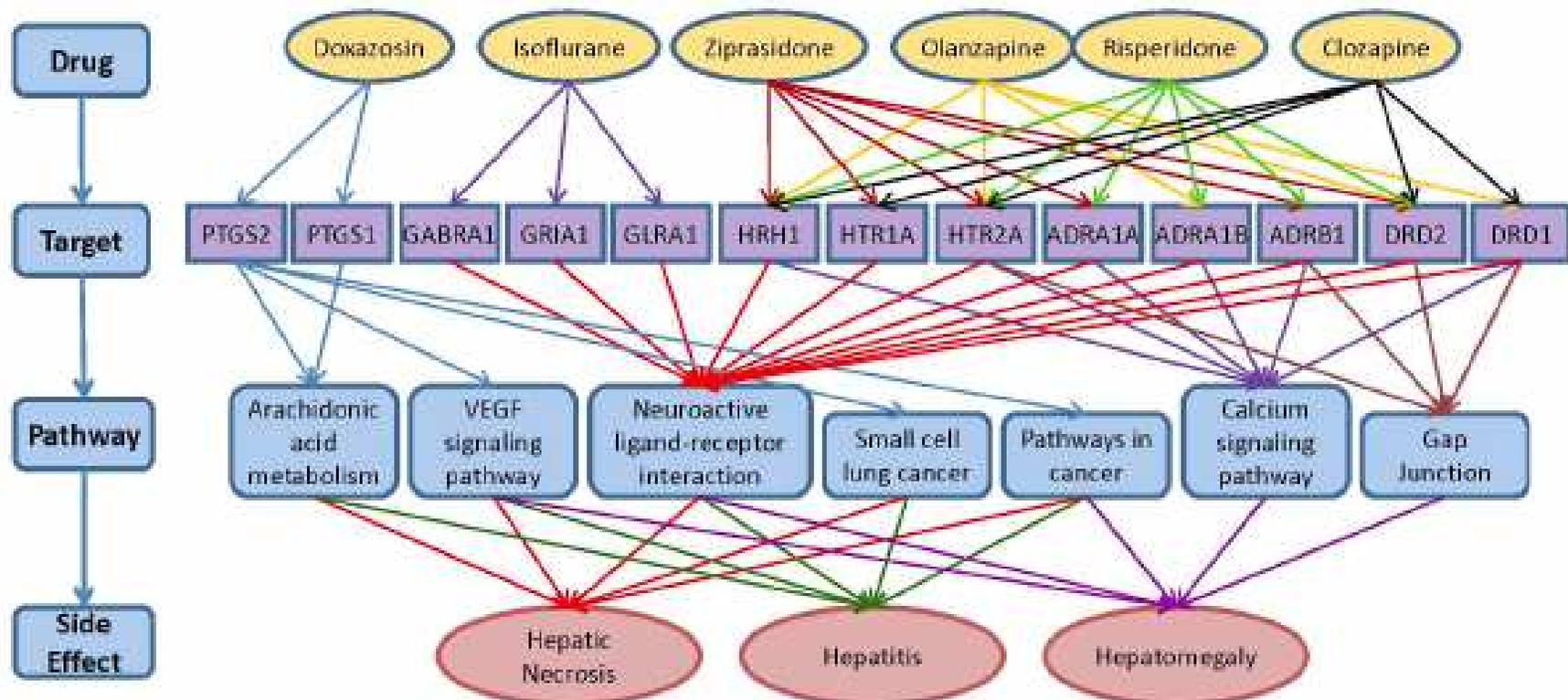
# Computational Toxicology had early roots in combinatorial chemistry

- 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, solid state, *in silico*)
- 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

# 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

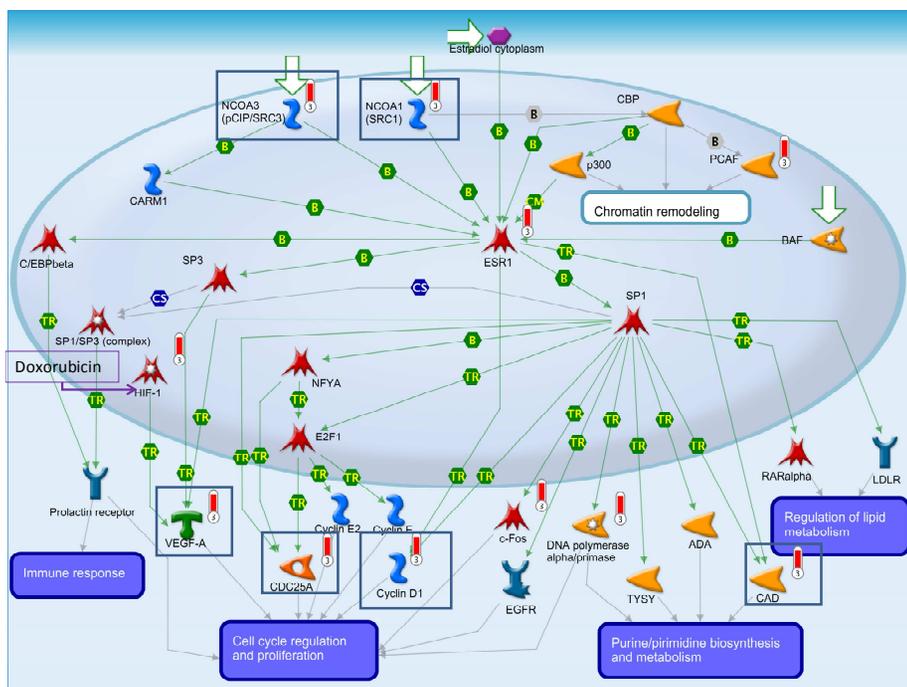




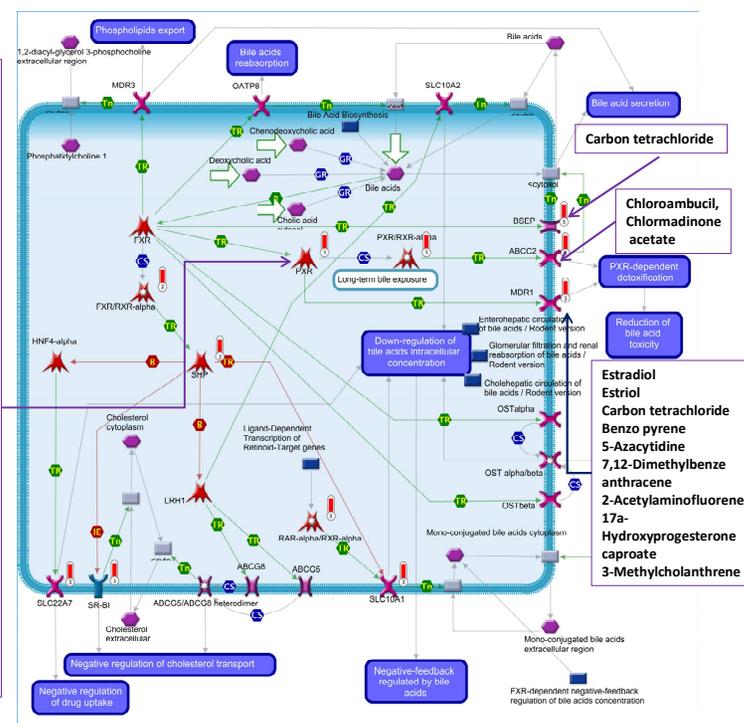
**Figure 8** Associating pathways with hepatotoxic effects. The drugs that are associated with hepatotoxicity-related side effects are associated with their targets using DrugBank. The targets are associated with pathways using KEGG to establish association chains between pathways and side-effects.



# 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  
 Ethynodiol diacetate  
 Griseofulvin  
 Lynestrol  
 Medroxyprogesterone acetate  
 Megestrol acetate  
 Mestranol  
 Methyleugenol  
 Norethisterone  
 Norethynodrel  
 Norlestrin  
 Ochrotoxin 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)

# Methods to fill “data gaps”

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

# Application to Green Chemistry Curriculum

- Toxicology tutorials for non-toxicology majors
  - <http://sis.nlm.nih.gov/enviro/toxtutor.html>
- User guides and tutorials for:
  - Structural alerts, analog identification, categorization, data sources, QSARs
- Flexible integration framework
  - OpenTox [www.opentox.com](http://www.opentox.com)
  - Hardy B, et al J. Cheminformatics (2010) 2:7
- Integration of chemical-biological and systems bio information
  - Genego, Inc. [www.genego.com](http://www.genego.com)