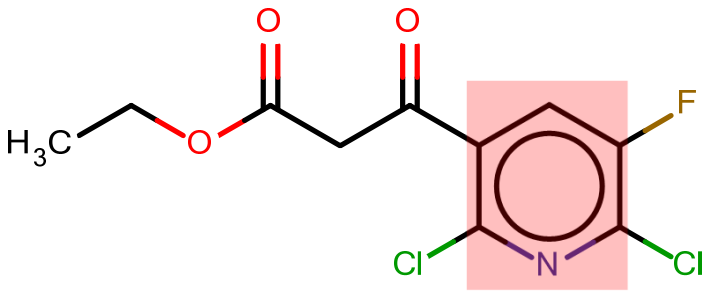
Alarms about Structural Alerts*



Alexander Tropsha

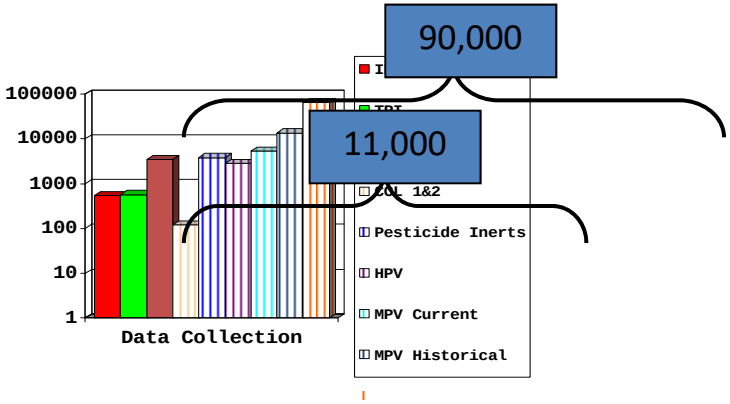
Laboratory for Molecular Modeling

UNC Eshelman School of Pharmacy

*Alves et al, Alarms about structural alerts. Green Chem, 2016, DOI: 10.1039/C6GC01492E, accepted paper

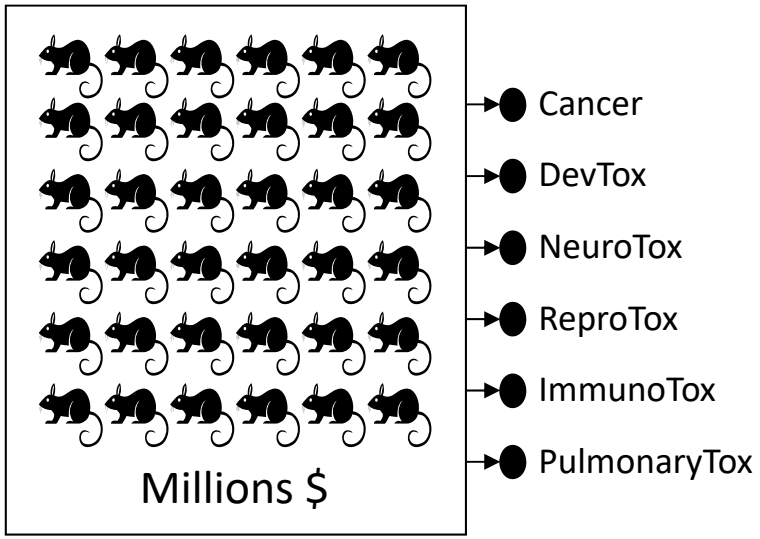
Challenges of current toxicity testing

Too Many Chemicals



...and not enough data.

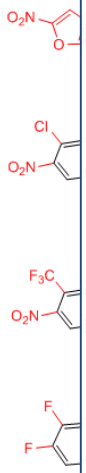
Too High a Cost



Too many endpoints
Too many mechanisms

Slide courtesy of Dr. Ann Richard, EPA

Future of Chemical Toxicity Testing



THE NATIONAL

July 2007

REPORT IN BRIEF

Toxicity Testing in the 21st Century: A Vision and a Strategy

Advances in molecular biology, biotechnology, and other fields are paving the way for major improvements in how scientists evaluate the health risks posed by potentially toxic chemicals found at low levels in the environment. These advances would make toxicity testing quicker, less expensive, and more directly relevant to human exposures. They could also reduce the need for animal testing by substituting more laboratory tests based on human cells. This National Research Council report creates a far-reaching vision for the future of toxicity testing.

Toxicity tests on laboratory animals are conducted to evaluate chemicals—including medicines, food additives, and industrial, consumer, and agricultural chemicals—for their potential to cause cancer, birth defects, and other adverse health effects. Information from toxicity testing serves as an important part of the basis for public health and regulatory decisions concerning toxic chemicals. Current test methods were developed incrementally over the past 50 to 60 years and are conducted using laboratory animals, such as rats and mice. Using the results of animal tests to predict human health effects involves a number of assumptions and extrapolations that remain controversial. Test animals are often exposed to higher doses than would be expected for typical human exposures, requiring assumptions about effects at lower doses or exposures. Test animals are typically observed for overt signs of adverse health effects, which provide little information about biological changes leading to such health effects. Often controversial uncertainty factors must be applied to account for differences between test animals and humans. Finally, use of animals in testing is expensive and time consuming, and it sometimes raises ethical issues.

Today, toxicological evaluation of chemicals is poised to take advantage of the on-going revolution in biology and biotechnology. This revolution is making it increasingly possible to study the effects of chemicals using cells, cellular components, and tissues—preferably of human origin—rather than whole animals. These powerful new approaches should help to address a number of challenges facing the

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

National Academy of Sciences • National Academy of Engineering • Institute of Medicine • National Research Council

POLICYFORUM

TOXICOLOGY

Transforming Environmental Health Protection

We propose a shift from primarily in vivo animal studies to in vitro assays, in vivo assays with lower organisms, and computational modeling for toxicity assessments.

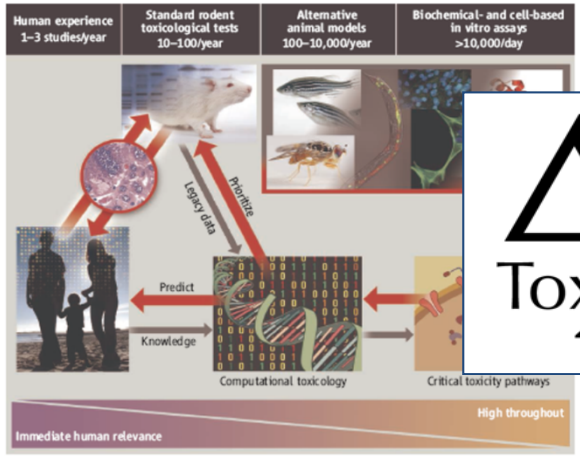
Francis S. Collins,[1*†] George M. Gray,[2*] John R. Bucher[3*]

In 2005, the U.S. Environmental Protection Agency (EPA), with support from the U.S. National Toxicology Program (NTP), funded a project at the National Research Council (NRC) to develop a long-range vision for toxicity testing and a strategic plan for implementing that vision. Both agencies wanted future toxicity testing and assessment paradigms to meet evolving regulatory needs. Challenges include the large numbers of substances that need to be tested and how to incorporate recent advances in molecular toxicology, computational sciences, and information technology; to rely increasingly on human as opposed to animal data; and to offer increased efficiency in design and costs (*1*–*5*). In response, the NRC Committee on Toxicity Testing and Assessment of Environmental Agents produced two reports that reviewed current toxicity testing, identified key issues, and developed a vision and implementation strategy to create a major shift in the assessment of chemical hazard and risk (*6*, *7*). Although the NRC reports have laid out a solid theoretical rationale, comprehensive and rigorously gathered data (and comparisons with historical animal data) will determine whether the hypothesized improvements will be realized in practice. For this purpose, NTP, EPA, and the National Institutes of Health Chemical Genomics Center (NCGC) (organizations with expertise in experimental toxicology, computational toxicology, and high-throughput technologies, respectively) have established a collaborative research program.

EPA, NCGC, and NTP Joint Activities

In 2004, the NTP released its vision and roadmap for the 21st century (*1*), which established initiatives to integrate high-throughput screening (HTS) and other automated screening assays into its testing program. In 2005, the EPA established the National Center for Computational Toxicology (NCCT). Through these initiatives, NTP and EPA, with the NCGC, are promoting the evolution of toxicology from a predominantly observational science at the level of disease-specific models in vivo to a predominantly predictive science focused on broad inclusion of target-specific, mechanism-based, biological observations in vitro (*1*, *4*) (see figure, below).

Toxicity pathways. In vitro and in vivo tools are being used to identify cellular responses after chemical exposure expected to result in adverse health effects (*7*). HTS methods are a primary means of discovery for drug development, and screening of >100,000 compounds per day is routine (*8*). However, drug-discovery HTS methods traditionally test compounds at one concentration, usually between 2 and 10 μM, and tolerate high false-negative rates. In contrast, in the EPA, NCGC, and NTP combined effort, all compounds are tested at as many as 15 concentrations, generally ranging from ~5 nM to ~100 μM, to generate a concentration-response curve (*9*). This approach is highly reproducible, produces significantly lower false-positive and false-negative rates than the traditional HTS methods (*9*), and facilitates multiassay comparisons. Finally, an informatics platform has been built to compare results among HTS screens; this is being expanded to allow comparisons with historical toxicologic NTP and EPA data (http://ncgc.nih.gov/pub/openhts). HTS data collected by EPA and NTP, as well as by the NCGC and other Molecular Libraries Initiative centers (http://mli.nih.gov/), are being made publicly available through Web-based databases [e.g., PubChem (http://pubchem.ncbi.nlm.nih.gov)]. In addition,



Transforming toxicology. The studies we propose will test whether high-throughput and computational toxicology approaches can yield data predictive of results from animal toxicity studies, will allow prioritization of chemicals for further testing, and can assist in prediction of risk to humans.

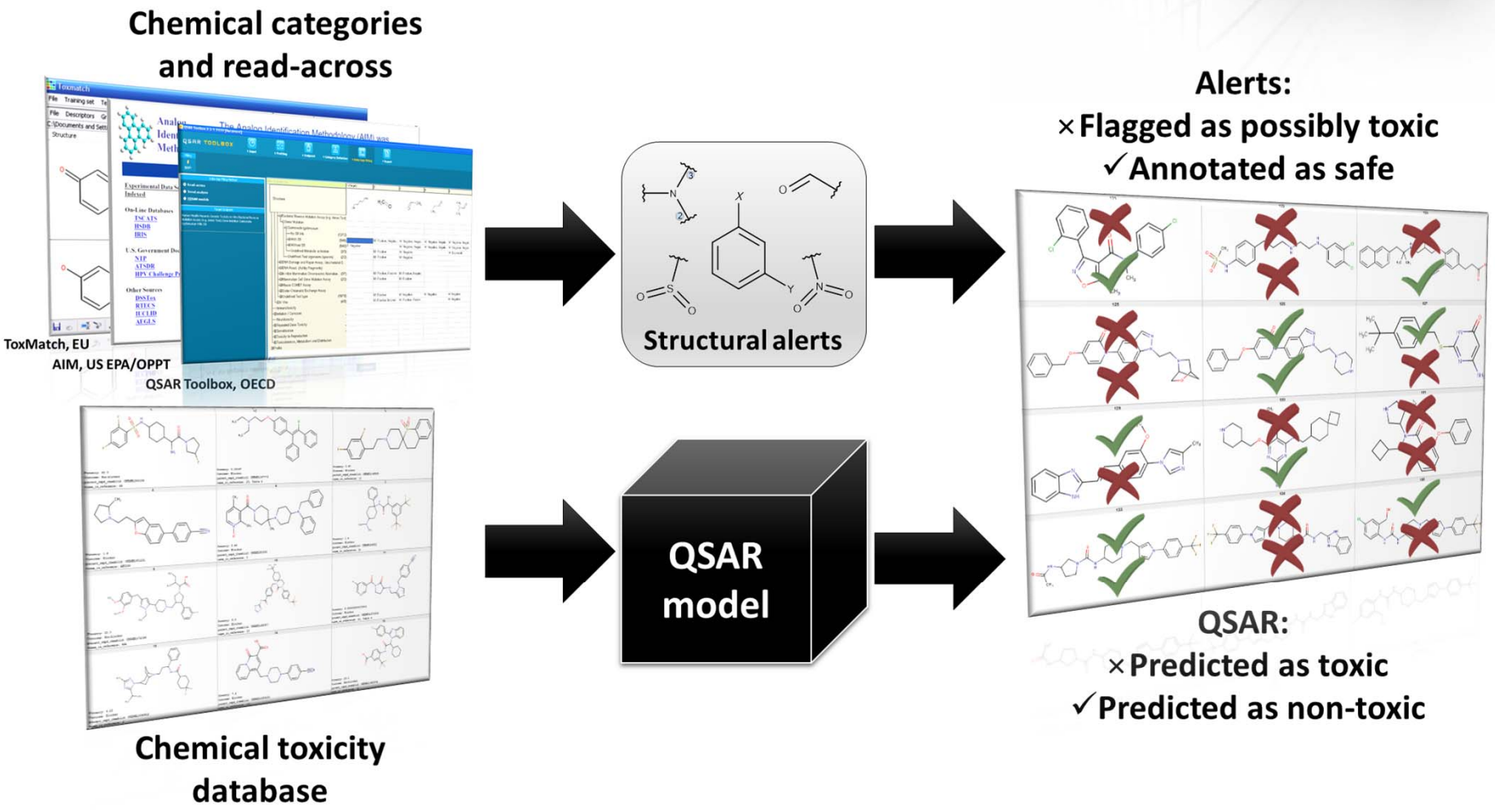
[1]Director, National Human Genome Research Institute (NHGRI), National Institutes of Health, Bethesda, MD 20892; [2]Assistant Administrator for the Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC 20460; [3]Associate Director, U.S. National Toxicology Program, National Institute of Environmental Health Sciences (NIEHS), Research Triangle Park, NC 27709, USA.

*The views expressed here are those of the individual authors and do not necessarily reflect the views and policies of their respective agencies.

†Author for correspondence. E-mail: francisc@mail.nih.gov

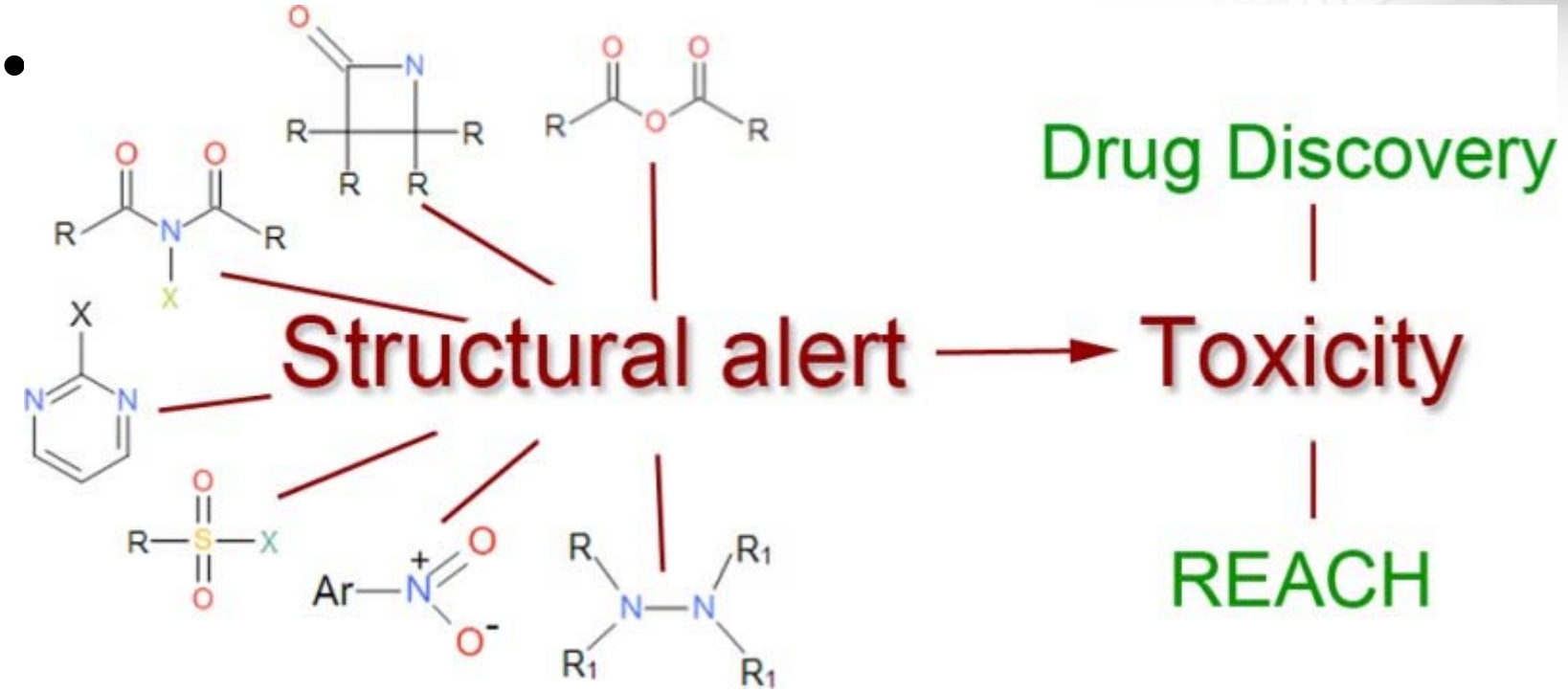
Tox21

Structural alerts and QSAR-based predictions in chemical safety assessment.



*Alves et al, Alarms about structural alerts. Green Chem, 2016, DOI: 10.1039/C6GC01492E

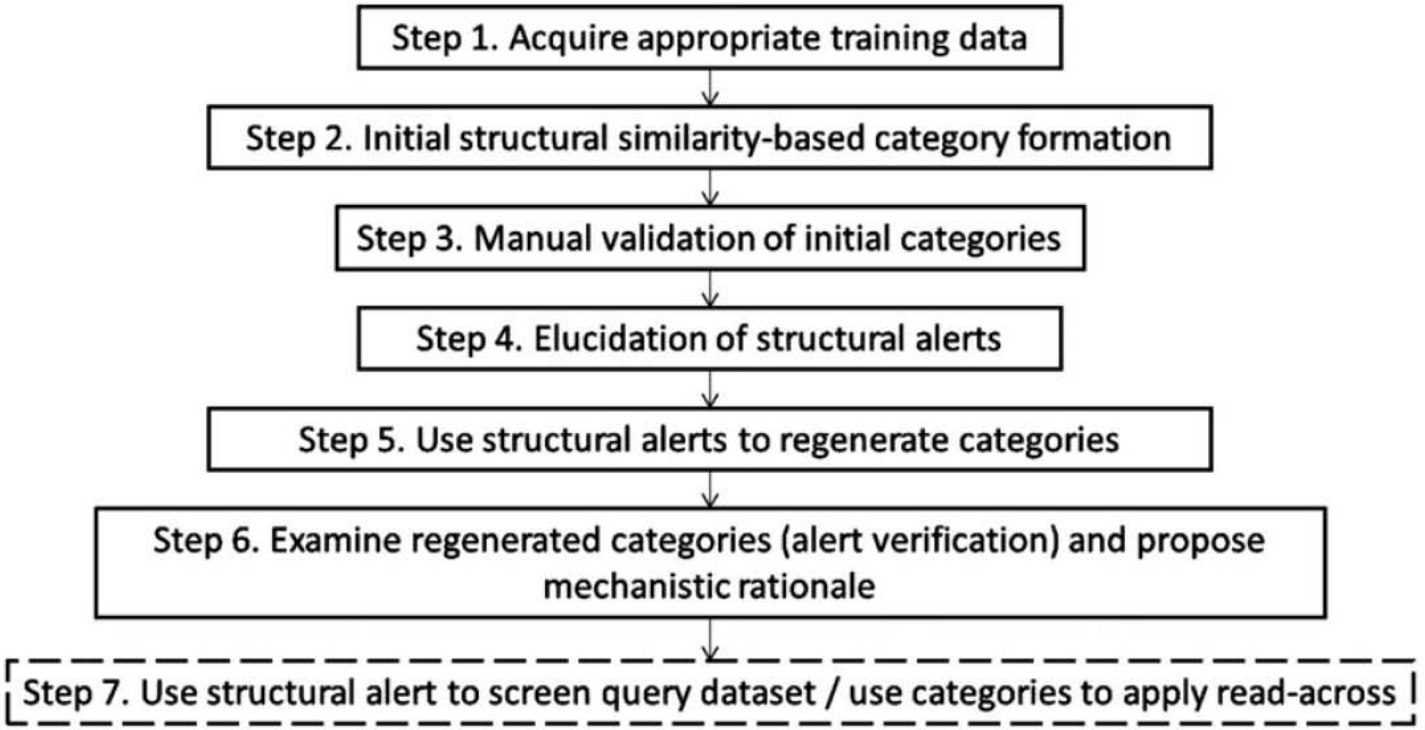
Structural Alerts: A Popular Concept in Chemical Toxicology



*Structural alerts are "molecular patterns that are associated with particular types of toxicity or ADRs either directly or after undergoing of a metabolic activation in vivo"**

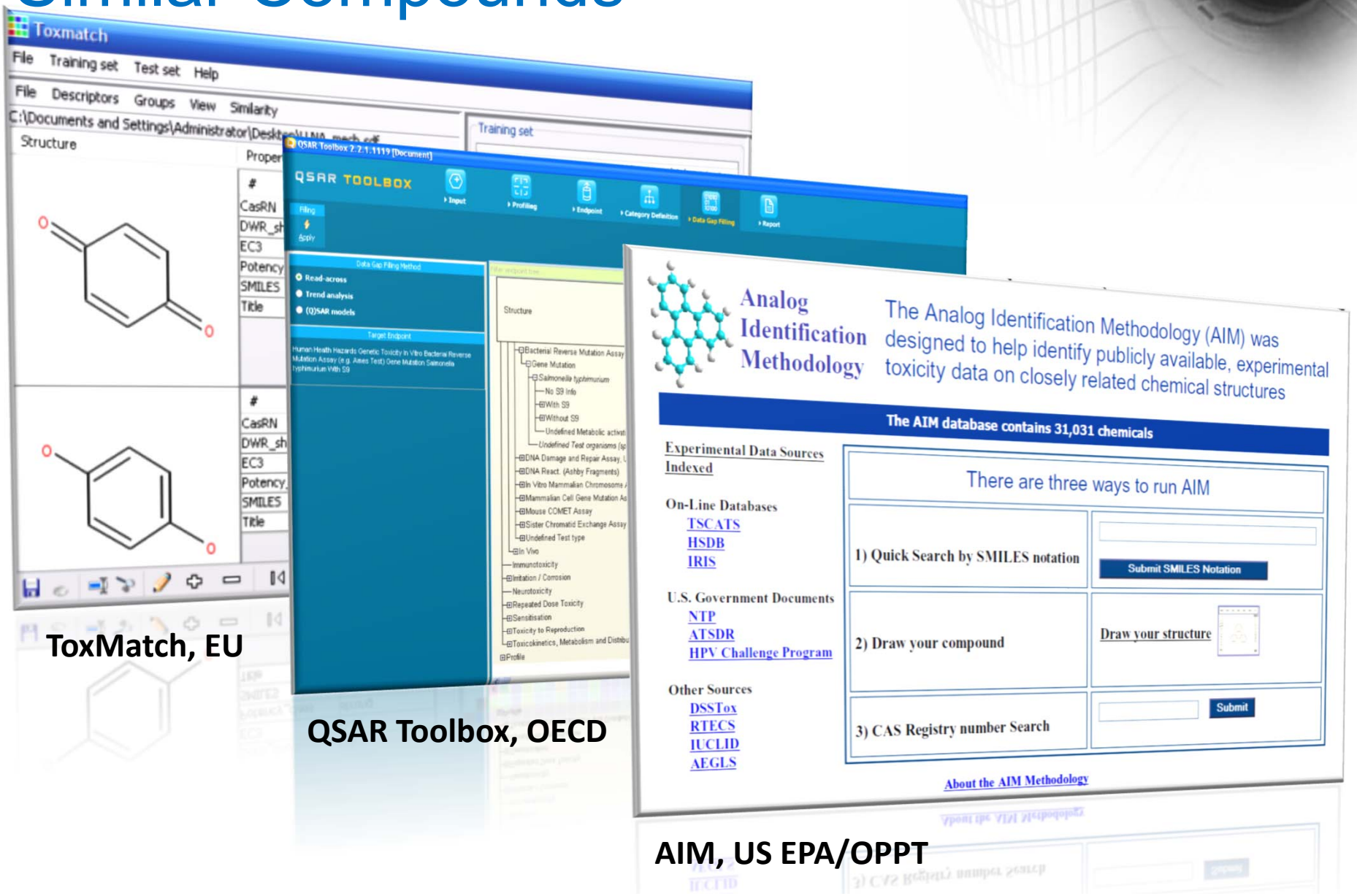
*Image and definition from Sushko et al, J Chem Inf Model. 2012 Aug 27; 52(8): 2310–2316.

Common strategy for developing alerts



Hewitt et al, Hepatotoxicity: A scheme for generating chemical categories for read-across, structural alerts and insights into mechanism(s) of action. *Crit Rev Toxicol*, **2013**; 43(7): 537–558

Chemical Read-Across: Learning from Similar Compounds



ToxMatch, EU

QSAR Toolbox, OECD

AIM, US EPA/OPPT

OECD QSAR Toolbox

OECD
BETTER POLICIES FOR BETTER LIVES

Français

Follow us
E-mail Alerts Blogs

OECD Home | About | Countries | Topics | Statistics | Newsroom

OECD Home › Chemical safety and biosafety › Assessment of chemicals › Grouping of Chemicals: Chemical Categories and Read-Across

Assessment of chemicals

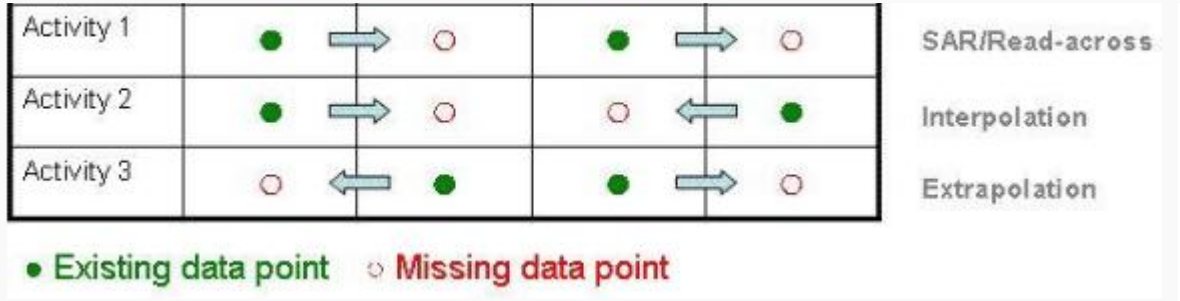
In the **read-across approach**, endpoint information for one chemical (the source chemical) is used to predict the same endpoint for another chemical (the target chemical), which is considered to be "similar" in some way (usually on the basis of structural

Quantitative read-across involves:

- the identification of a chemical substructure or mode or mechanism of action that is common to two substances (which are considered to be analogues); and
- the assumption that the known value of a property for one substance can be used to estimate the unknown value of the same property for another substance.

In both cases, expert judgement is needed and some justification should be provided.

across.htm



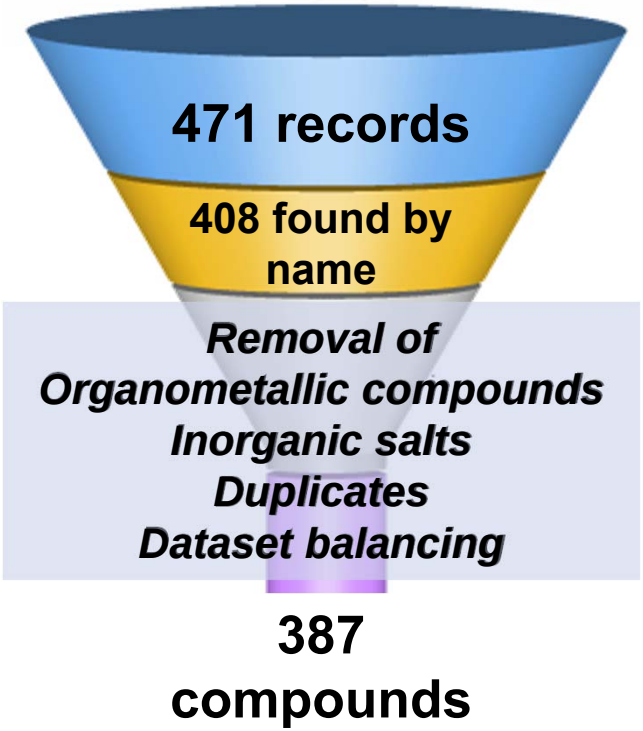
Skin Sensitizers: commonly identified using toxicity alerts (OECD QSAR Toolbox)

Source

ICCVAM (Interagency Coordinating Committee on the Validation of Alternative Methods) report 2009

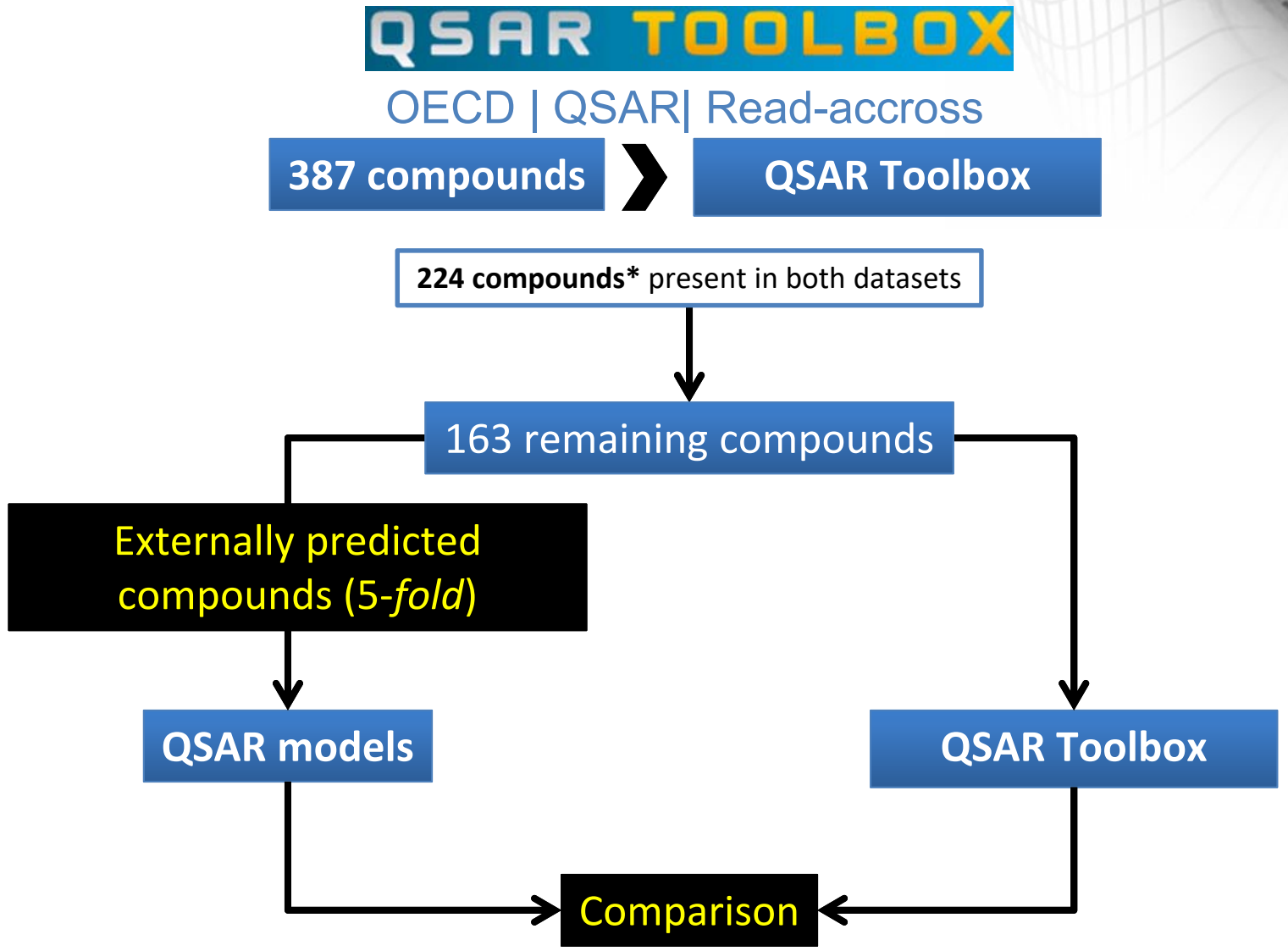
Vehicle type	Non-sensitizer	Sensitizer	Total
ACE	14	31	45
AOO	51	178	229
dH_2O	2	2	4
DMF	40	27	67
DMSO	16	15	31
PG	6	8	14
Pluronic L92 (1%)	2	5	7
Others	4	7	11
Total	**135**	**273**	**408**

Abbreviations: AOO, acetone&olive oil (4:1 by volume); ACE, acetone; DMF, dimethyl formamide; DMSO, dimethyl sulfoxide; PG, propylene glycol.



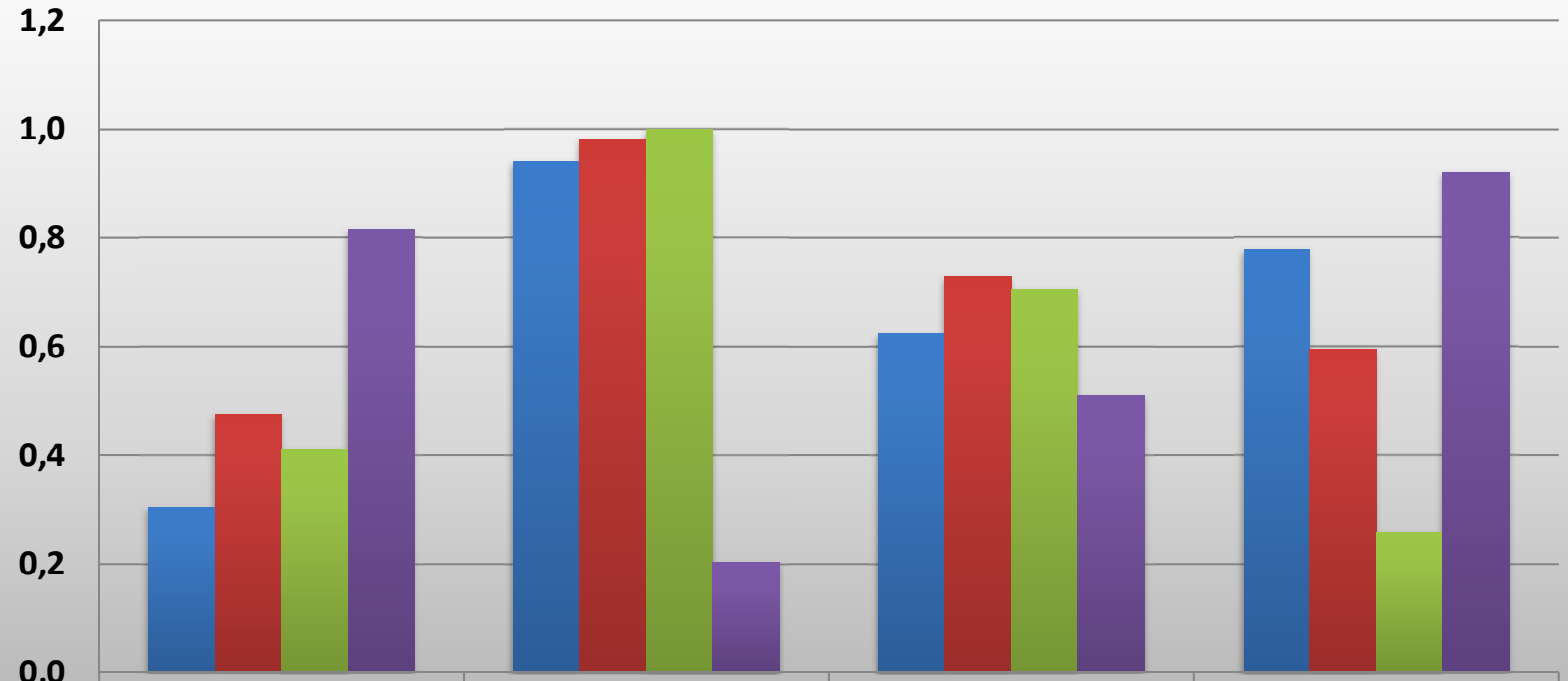
254 compounds were retained for QSAR modeling:
127 sensitizers and 127 non-sensitizers
133 remaining sensitizers were used as external validation set

Workflow for comparing QSAR versus OECD QSAR Toolbox



*These compounds had 94% concordance with ICCVAM report

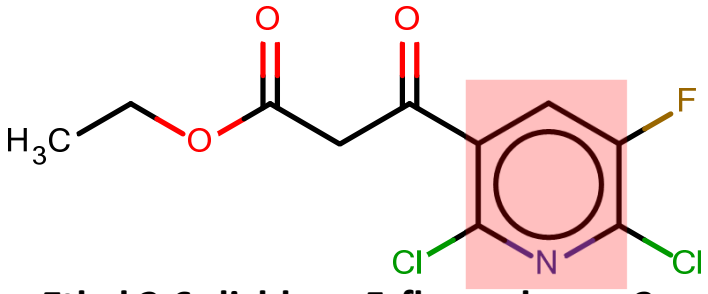
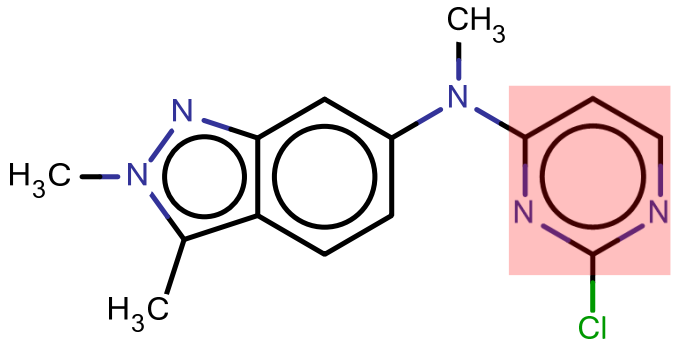
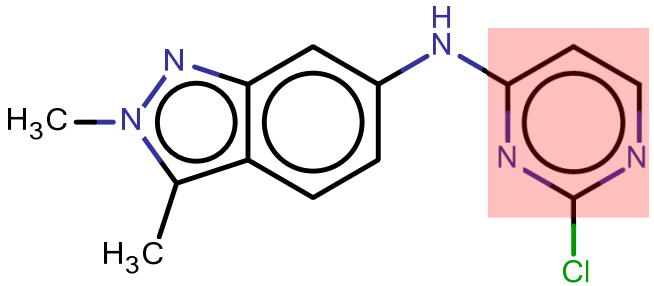
Comparison between QSAR Models and the Toolbox



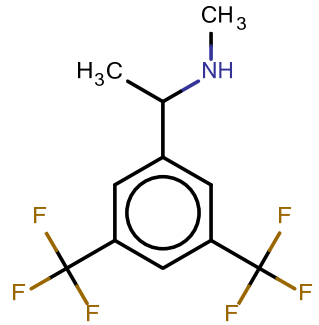
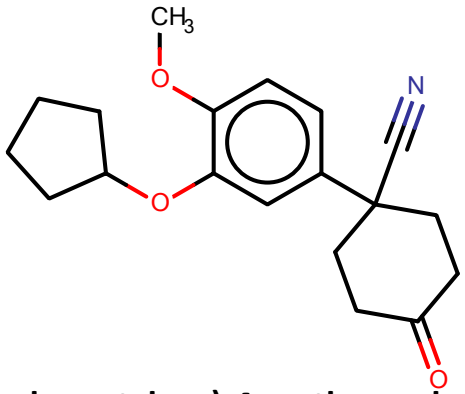
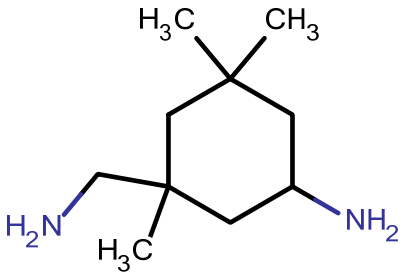
	Sensitivity	Specificity	CCR	Coverage
Consensus no AD	0,31	0,94	0,62	0,78
Consensus	0,48	0,98	0,73	0,60
Consensus Rigor	0,41	1,00	0,71	0,26
QSAR Toolbox	0,82	0,20	0,51	0,92

Models were built using **Random Forest** approach – 5-fold External CV results
* Applicability Domain wasn't considered in this model

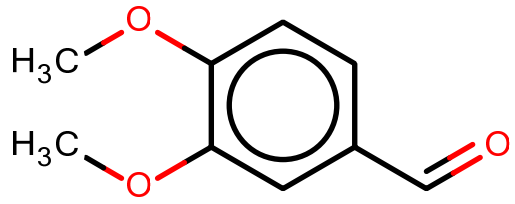
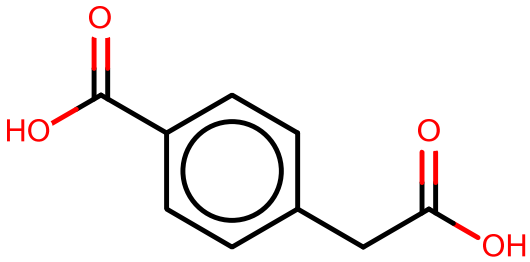
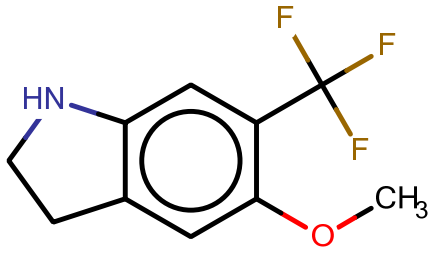
ALERTS vs. QSAR: ACTIVATED PYRIDINE/PYRIMIDINE

	QSAR Toolbox	QSAR	Experiment
 Ethyl 2,6-dichloro-5-fluoro-b-oxo-3-pyridinepropanoate	✓ *Contains Activated Pyridine* → Sensitizer	Non Sensitizer	Non Sensitizer
 N-(2-Chloro-4-pyrimidinyl)-N,2,3-trimethyl-2H-indazol-6-amine	✓ *Contains Activated Pyridine* → Sensitizer	Non Sensitizer	Non Sensitizer
 N-(2-Chloro-4-pyrimidinyl)-2,3-dimethyl-2H-indazol-6-amine	✓ *Contains Activated Pyridine* → Sensitizer	Non Sensitizer	Non Sensitizer

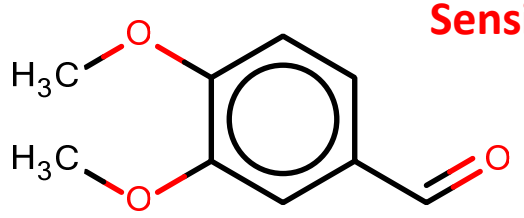
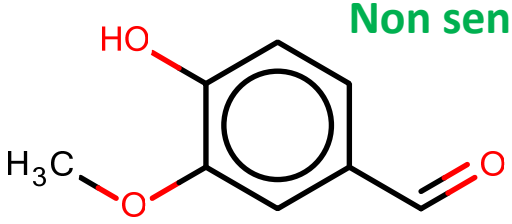
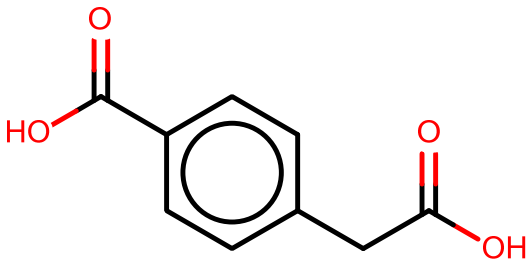
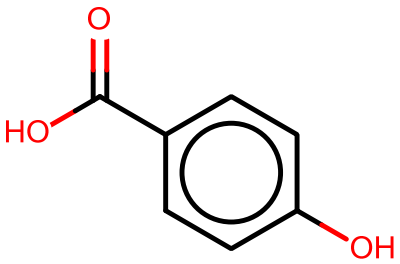
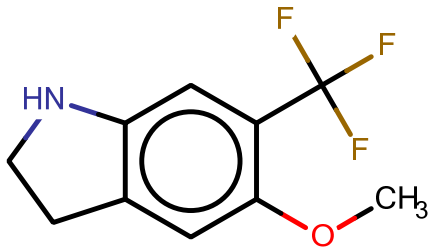
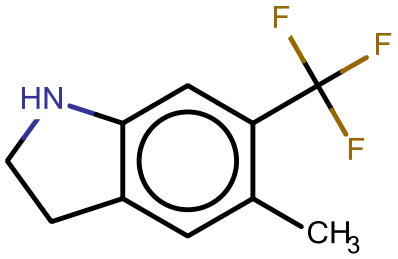
ALERTS vs. QSAR: NO PROTEIN BINDING ALERTS

	QSAR Toolbox	QSAR	Experiment
 1-[3,5-Bis(trifluoromethyl)phenyl]-N-methylethanamine	*No alert* → Sensitizer	Non Sensitizer	Non Sensitizer
 1-[3-(Cyclopentyloxy)-4-methoxy-phenyl]-4-oxocyclohexane carbonitrile	*No alert* → Sensitizer	Non Sensitizer	Non Sensitizer
 3-Aminomethyl- 3,5,5-trimethylcyclohexyl amine	*No alert* → Non sensitizer	Sensitizer	Sensitizer

MISPREDICTED COMPOUNDS

	QSAR Toolbox	QSAR	Experiment
 Veratraldehyde	*No alert* → Non sensitizer	Non Sensitizer	Sensitizer
 4-Carboxyphenylacetate	*No alert* → Non sensitizer	Non Sensitizer	Sensitizer
 5-Methoxy-6-trifluoromethyl-2,3-dihydro-1H-indole	*No alert* → Non sensitizer	Non Sensitizer	Sensitizer

MISPREDICTED COMPOUNDS

	FIRST NEIGHBOR	Tanimoto Score
Sensitizer  Veratraldehyde	Non sensitizer  Vanillin	0.92
Sensitizer  4-Carboxyphenylacetate	Non sensitizer  4-Hydroxybenzoic acid	0.70
Sensitizer  5-Methoxy-6-trifluoromethyl-2,3-dihydro-1H-indole	Non sensitizer  5-Methyl-6-(trifluoromethyl)indoline	0.81

OECD QSAR Toolbox (categories, read across): predict or alert?

From reviewer's critique of our manuscript: "Novel computational tools to predict chemically-induced skin reactions. Part I: QSAR Models of Skin Sensitization and their application to identify potentially hazardous compounds" (TAAP, 2015)

… I don't think the authors have properly understood the function of the read-across facilities implemented in OECD Toolbox…

… providing tools for implementation of read across in the Toolbox does not guarantee adequate predictions.

… the Toolbox is not a model that can be compared with other models, but should rather be considered as an instrument for generation of models …

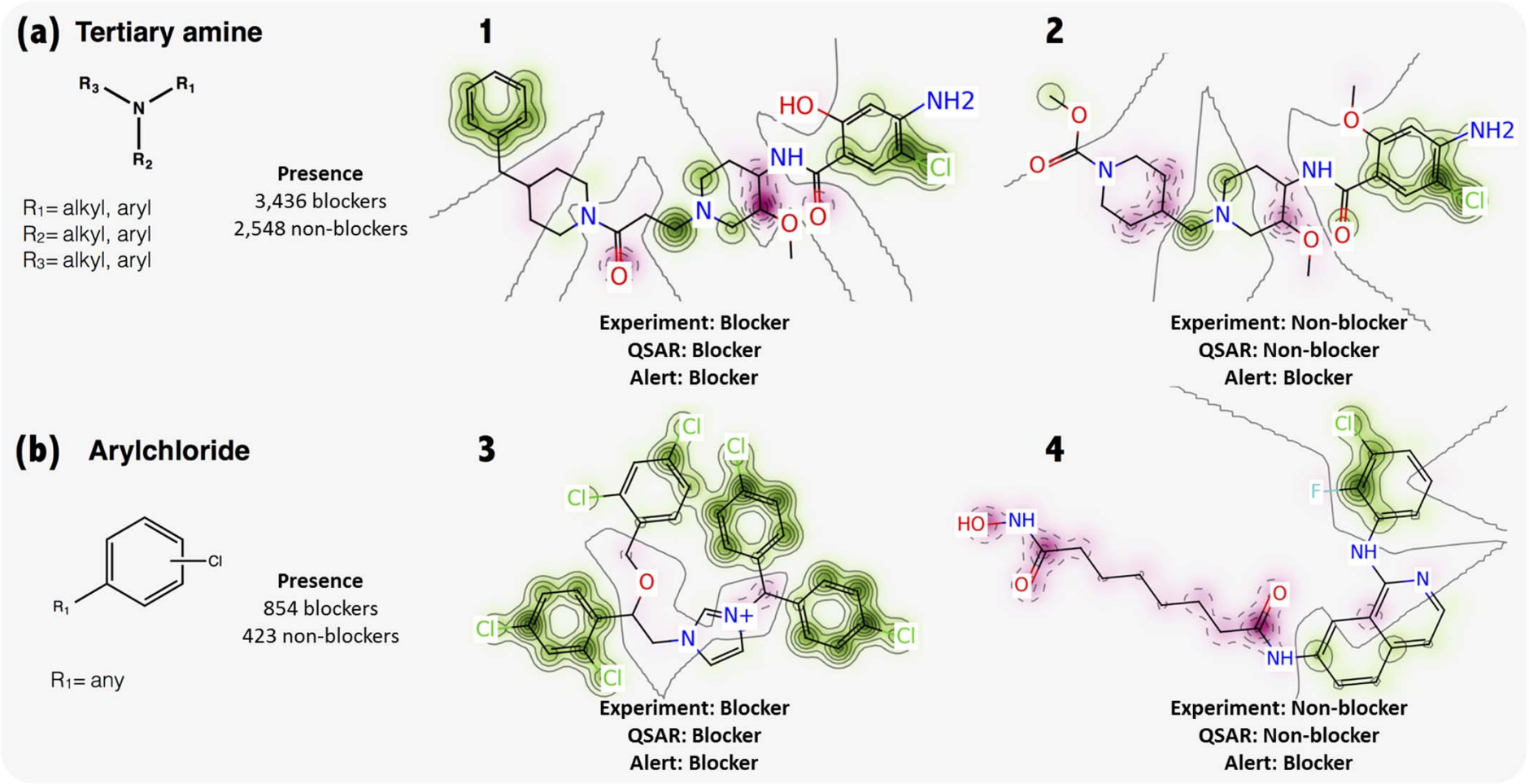
…. I strongly recommended that the comparison with the Toolbox (and analysis of the Toolbox system) should be removed.

From Hewitt et al, Hepatotoxicity: a scheme for generating chemical categories for read-across, structural alerts and insights into mechanism(s) of action. *Crit Rev Toxicol. 2013 Aug;43(7):537-58*

It must be stressed that we are not aiming to develop a model for predicting hepatotoxicity; rather we are detailing a scheme capable of generating mechanistically supported structural alerts suitable for identifying chemicals with hepatotoxic potential

Chemical Alerts of Toxicity: what are they for, really?

ALERTS vs. QSAR: TERTIARY AMINE / ARYLCHORIDE



Braga, R. C.; *Curr. Top. Med. Chem.* **2014**, *14* (11), 1399–1415.

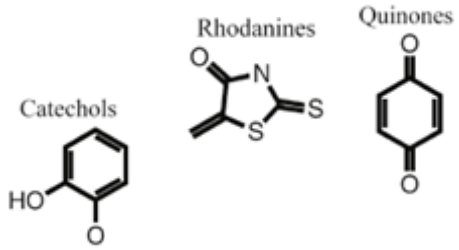
Alerts based toxicity estimate for withdrawn and marketed drugs

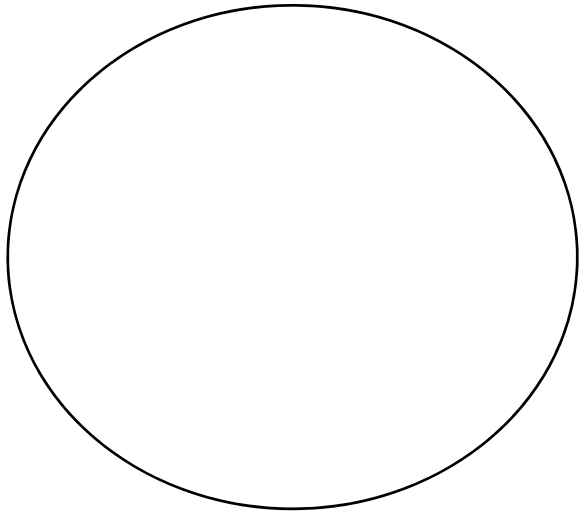
Drug Name	State	QSAR prediction[2]*	Toxic hazard classification by Cramer (extension)	Toxic hazard classification by Cramer (original)	Carcinogenicity (genotox and nongenotox) alerts by ISS	DNA alerts for AMES, MN and CA by OASIS v.1.3	In vitro mutagenicity (Ames test) alerts by ISS	In vivo mutagenicity (Micronucleus) alerts by ISS
Amineptine	withdrawn	unsafe	**High (Class III)**	**High (Class III)**	No alert found	No alert found	No alert	**Alerts**
Duract	withdrawn	unsafe	**High (Class III)**	**High (Class III)**	No alert found	No alert found	No alert	**Alerts**
Vioxx	withdrawn	unsafe	**High (Class III)**	**High (Class III)**	No alert found	No alert found	No alert	**Alerts**
Astemizole	withdrawn	unsafe	**High (Class III)**	**High (Class III)**	**Alerts**	No alert found	No alert	**Alerts**
Cerivastatin	withdrawn	unsafe	**High (Class III)**	**High (Class III)**	No alert found	No alert found	No alert	**Alerts**
Chlormezanone	withdrawn	unsafe	**High (Class III)**	**High (Class III)**	**Alerts**	No alert found	No alert	**Alerts**
Fenfluramine	withdrawn	unsafe	**High (Class III)**	**High (Class III)**	No alert found	No alert found	No alert	No alert
Flosequinan	withdrawn	unsafe	**High (Class III)**	**High (Class III)**	**Alerts**	No alert found	**Alerts**	**Alerts**
Glafenine	withdrawn	unsafe	**High (Class III)**	**High (Class III)**	**Alerts**	No alert found	No alert	**Alerts**
Grepafloxacin	withdrawn	unsafe	**High (Class III)**	**High (Class III)**	**Alerts**	No alert found	No alert	**Alerts**
Mibefradil	withdrawn	unsafe	**High (Class III)**	**High (Class III)**	**Alerts**	No alert found	No alert	**Alerts**
Troglitazone	withdrawn	unsafe	**High (Class III)**	**High (Class III)**	No alert found	No alert found	No alert	**Alerts**
Ximelagatran	withdrawn	unsafe	**High (Class III)**	**High (Class III)**	No alert found	**Alerts**	No alert	**Alerts**
Aspirin	marketed	safe	Low (Class I)	Low (Class I)	No alert found	No alert found	No alert	**Alerts**
Ibuprofen	marketed	safe	Low (Class I)	Low (Class I)	No alert found	No alert found	No alert	**Alerts**
Valtrex	marketed	safe	**High (Class III)**	**High (Class III)**	No alert found	No alert found	No alert	**Alerts**
Microzide	marketed	safe	**High (Class III)**	**High (Class III)**	**Alerts**	No alert found	No alert	**Alerts**
Neurontin	marketed	safe	**High (Class III)**	**High (Class III)**	No alert found	No alert found	No alert	**Alerts**
Enoxaparin	marketed	safe	**High (Class III)**	**High (Class III)**	No alert found	No alert found	No alert	**Alerts**
Lyrica	marketed	safe	Low (Class I)	Low (Class I)	No alert found	No alert found	No alert	**Alerts**

*Zakharov, Lagunin, Poroikov. Chem. Res. Toxicol., 2012, 25, 2378–2385.

Pan-Assay Interference Compounds

- **Assay interference** is a source of error in drug screening.
- A true screening hit exhibits its effect (inhibition or activation) through direct binding with a protein.
  - false positives are often interspersed among these true hits.
- The measured effect of false positives does not depend specific interactions with a protein.
  - Interference mechanisms, such as auto-fluorescence, hydrogen peroxide production, metal chelation, and chemical aggregation





Frequent Hitters

Thorne, N., Auld, D. S. & Inglese, J. *Curr Opin Chem Biol* **14,** 315–324 (2010).

Journal of Medicinal Chemistry Article

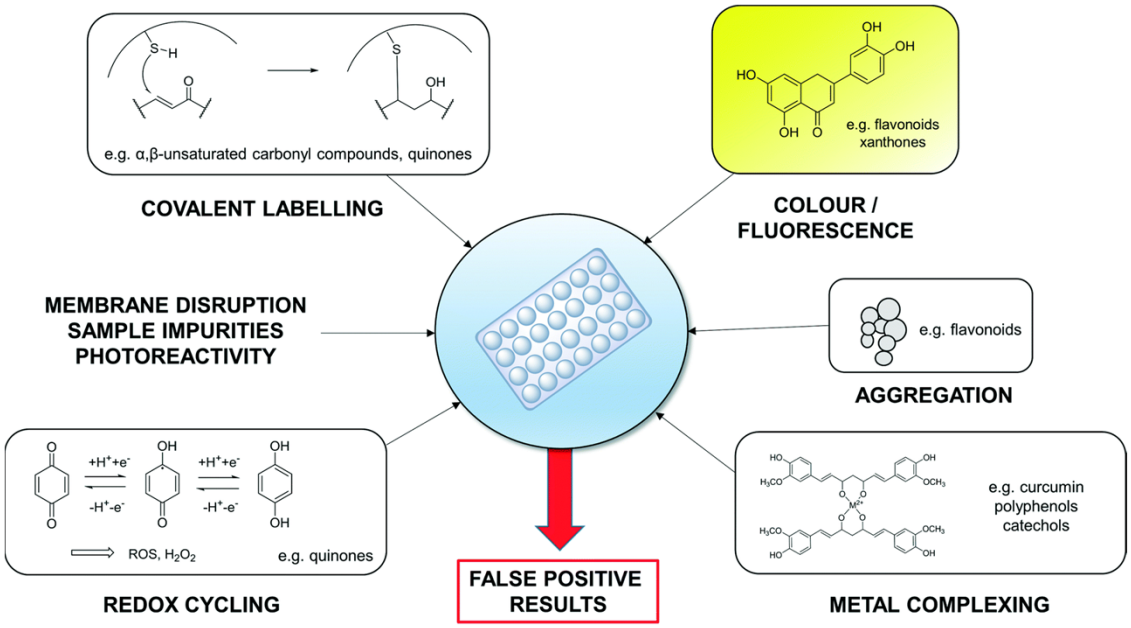
J. Med. Chem. **2010,** *53,* 2719–2740 **2719**
DOI: 10.1021/jm901137j

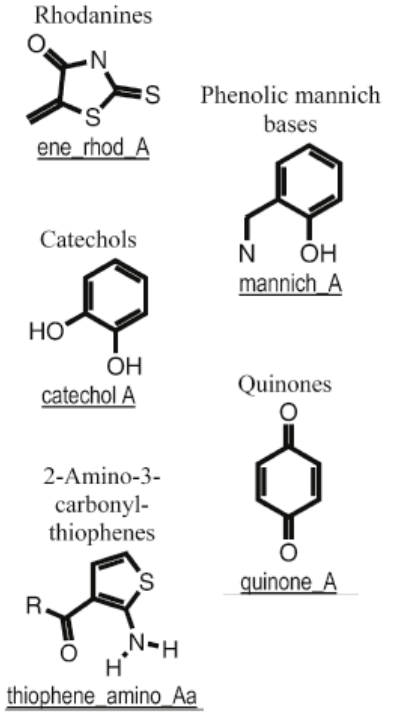
New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays

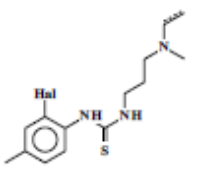
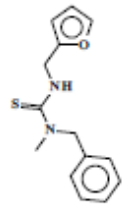
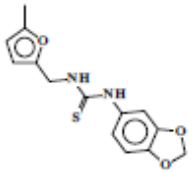
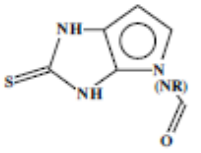
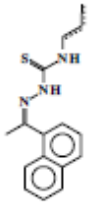
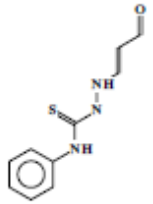
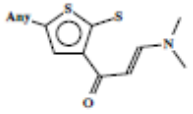
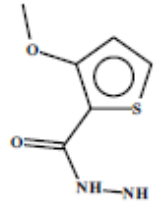
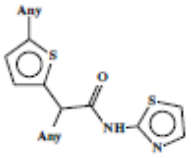
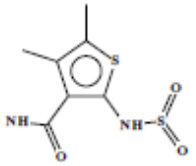
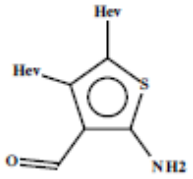
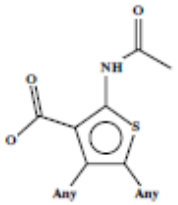
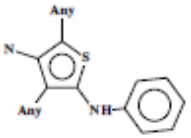
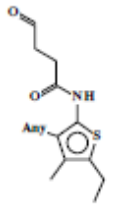
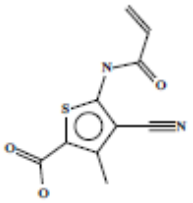
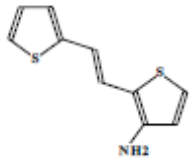
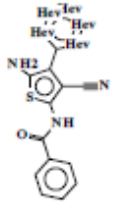
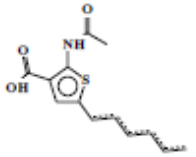
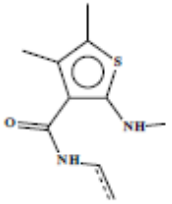
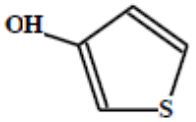
Jonathan B. Baell*,†,‡ and Georgina A. Holloway†,‡

PAINS = Pan Assay Interference Compounds

- Compounds with certain substructures are likely to be false positives
- 480 substructural PAINS alerts





 461:thio_urea_M(1)	 462:thio_urea_N(1)	 463:thio_urea_O(1)	 464:thio_urea_P(1)
 465:thio_urea_Q(1)	 466:thio_urea_R(1)	 467:thiophene_C(3)	 468:thiophene_D(2)
 469:thiophene_E(2)	 470:thiophene_F(1)	 471:thiophene_amino_Aa(4-	 472:thiophene_amino_Ab(4-
 473:thiophene_amino_B(12	 474:thiophene_amino_C(7)	 475:thiophene_amino_D(3)	 476:thiophene_amino_E(2)
 477:thiophene_amino_F(2)	 478:thiophene_amino_G(2)	 479:thiophene_amino_H(2)	 480:thiophene_hydroxy(28)

Learning from PAINful lessons

Daniel A. Erlanson*

Feeling Nature's PAINS: Natural Products, Natural Product Drugs, and Pan Assay Interference Compounds (PAINS)

Jonathan B. Baell*

PAINS in the Assay: Chemical Mechanisms of Assay Interference and Promiscuous Enzymatic Inhibition Observed during a Sulfhydryl-Scavenging HTS

Jayme L. Dahlin,[†,‡] J. Willem M. Nissink,[§] Jessica M. Strasser,[∥] Subhashree Francis,[∥] LeeAnn Higgins,[⊥] Hui Zhou,[#] Zhiguo Zhang,[#] and Michael A. Walters*[,∥]

Pan Assay Interference Compounds (PAINS) and Other Promiscuous Compounds in Antifungal Research

Miniperspective

Martin Pouliot and Stephane Jeanmart*

Assay Interference by Chemical Reactivity

Jayme L. Dahlin, MD, PhD
Jonathan Baell, PhD
Michael A. Walters, PhD

Screening Technologies and Their Application to Drug Discovery

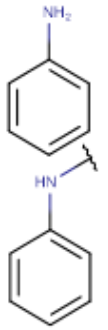
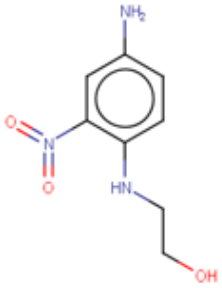
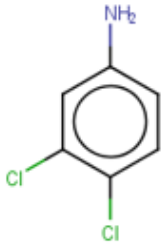
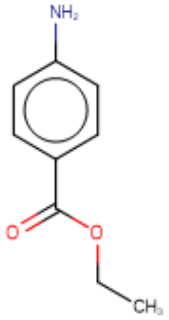
Jonathan B. Baell

Activity artifacts in drug discovery and different facets of compound promiscuity

Jürgen Bajorath

Alerts are … well … just Alerts

- Structural Alerts are used in toxicity to identify potentially toxic compounds

Alerts	Compounds		
			
Aromatic amines	**2-(4-Amino-2nitro-phenylamino)-ethanol** Sensitizer	**3,4-dichloroaniline hydrochloride** Sensitizer	**Benzocaine** Non-sensitizer

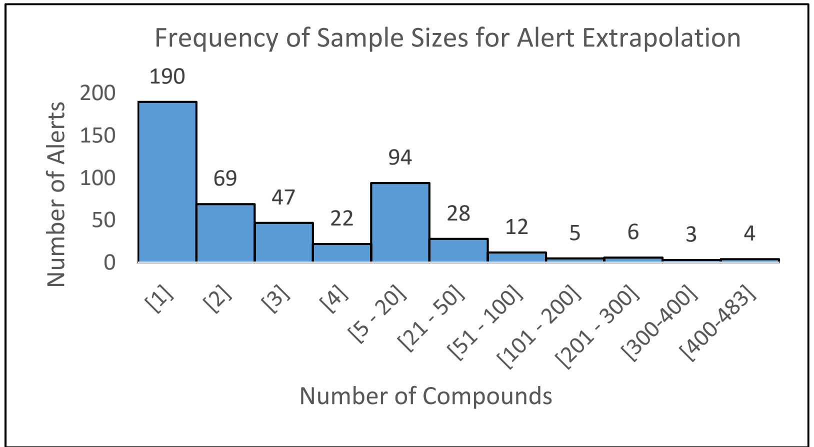
- Structural Alerts are generally overly sensitive (false positives)
- Our group has shown that QSAR models has better accuracy at predicting toxicity than alerts alone.

Alves, V. M. *et al*. *Toxicol. Appl. Pharmacol.* **284,** 273–80 (2015).

The Problem with PAINS

- What does this all mean?
  - Data analysis of PAINS substructures and overextrapolation
  - Limited applicability of alerts for assay interference
  - Just how "pain"-ful are PAINS?
  - What evidence is sufficient to declare the results of a study to be invalid simply on the basis of a substructure?

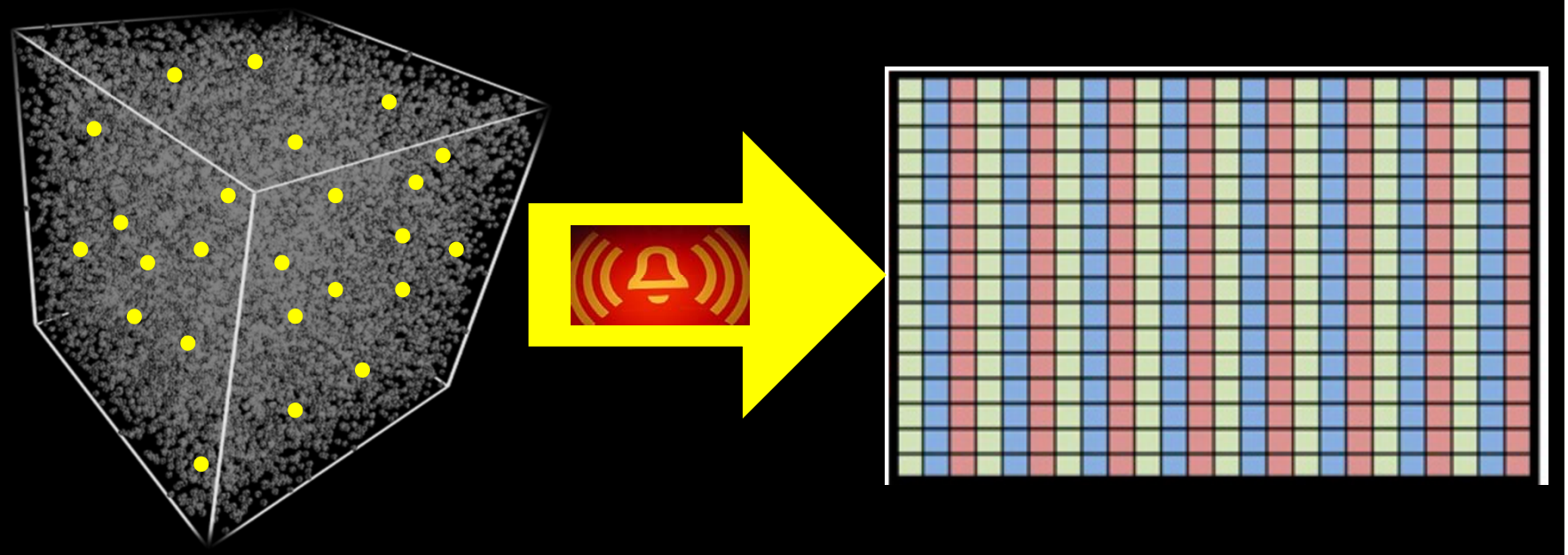
- How were original PAINS substructures derived?
  - Generalization from only six HTS campaigns
  - No structures of false hits
  - All six campaigns were directed at protein-protein interactions
  - 3 of the 6 targets were kept confidential
  - Only AlphaScreen technology was used
  - Limited sample sizes



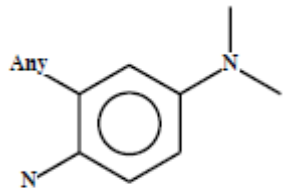
The Extrapolative Power of 1. A histogram showing the frequency of the 480 PAINS alerts and the number of compounds used to derive them. 190 PAINS alerts were extrapolated from only one representative compound. Only 18 PAINS alerts were extrapolated from sample sizes for greater than 100 compounds per alert.

Random PAINS

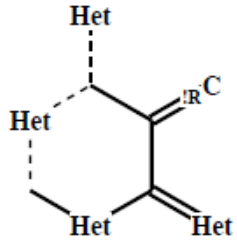
- What is the “pan-assay” activity of PAINS compared to non-PAINS?
  - PubChem Promiscuity
    - All assays
    - All beta-lactamase, luciferase, and fluorescence-based assays



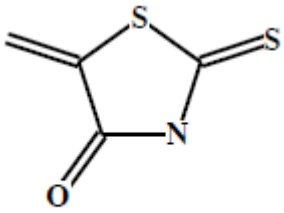
- Calculate the frequency of activity across all assays



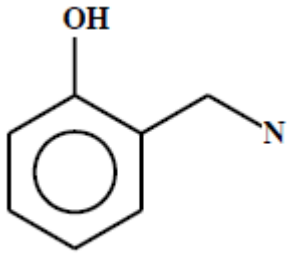
anil_di_alk_A(478)



ene_six_het_A(483)



ene_rhod_A(235)



mannich_A(296)

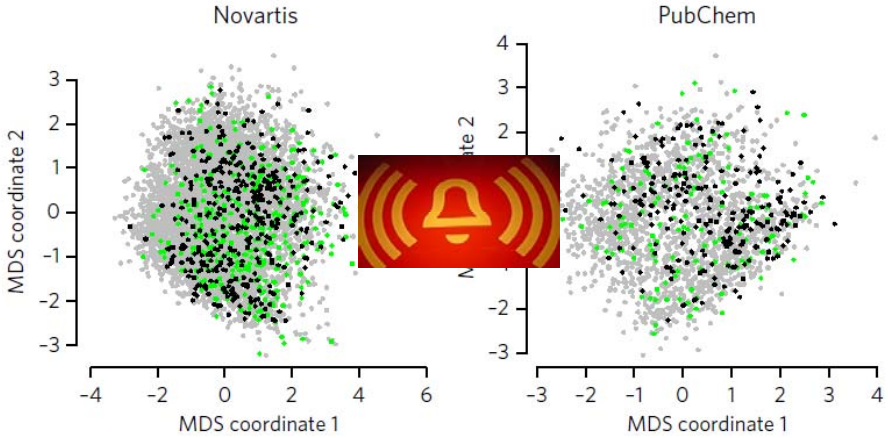
		Percent Active (Average)			
	# of cpds	All Assays	Luciferase	Beta-lactamase	Fluorescence
Random-PAINS	14,611	3% (562)	3% (95)	1% (12)	2% (329)
Random-NoPAINS	58,722	1% (550)	2% (93)	0.6% (13)	0.8% (321)
Anil_di_alk_A(478)	2,598	1% (552)	2% (93)	1% (11)	1% (323)
Ene_six_het_A(483)	1,315	2% (603)	1% (100)	1% (12)	2% (357)
Ene_rhod_A(235)	1,109	3% (544)	3% (92)	1% (9)	3% (320)
Mannich_A(296)	927	3% (580)	4% (98)	1% (12)	2% (339)

Table 1. Pan-assay activity of compounds in PubChem. The average assay activity for PAINS and non-PAINS across all assays in PubChem, highlighting luciferase-, beta-lactamase-, and fluorescence-based assays. The average number of assays in which the compounds were tested are provided in the parenthesis.

PAINS in Dark Chemical Matter

Dark Chemical Matter

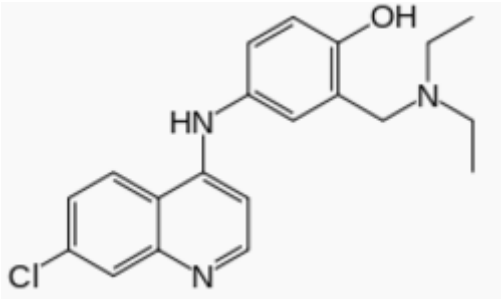
- Small molecules that have never shown biological activity despite having been exhaustively tested in HTS assays
- DCM is a potential starting point for the optimization of selective compounds
- ~ 140,000 DCM from a Novartis and PubChem collection tested in at least 100 assays



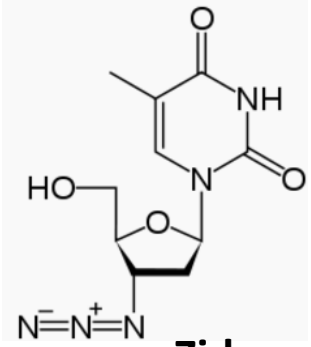
- PAINS substructures can be found in DCM.
  - ~4,500 compounds
- PAINS can be biologically inert!

PAINS in Approved Drugs

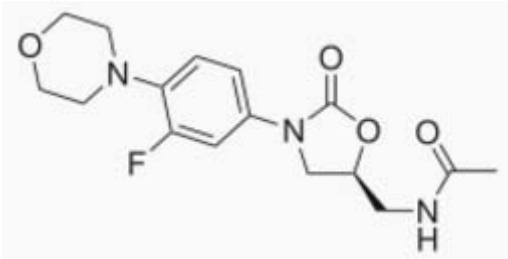
- **76 drugs possess PAINS substructures**
  - 21 individual PAINS alerts types
- 19 are part of the WHO's List of Essential Medicine



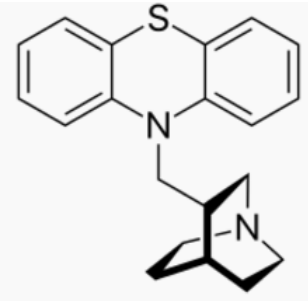
Amodiaquine
Anti-malarial
mannich_A(296)



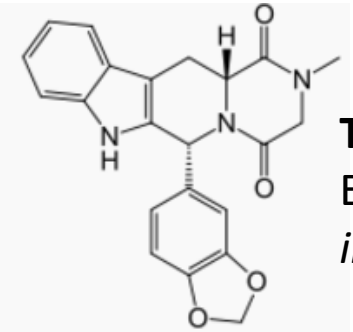
Zidovudine
Anti-retroviral
azo_A(324)



Linezolid
Antibiotic
anil_di_alk_A(478)

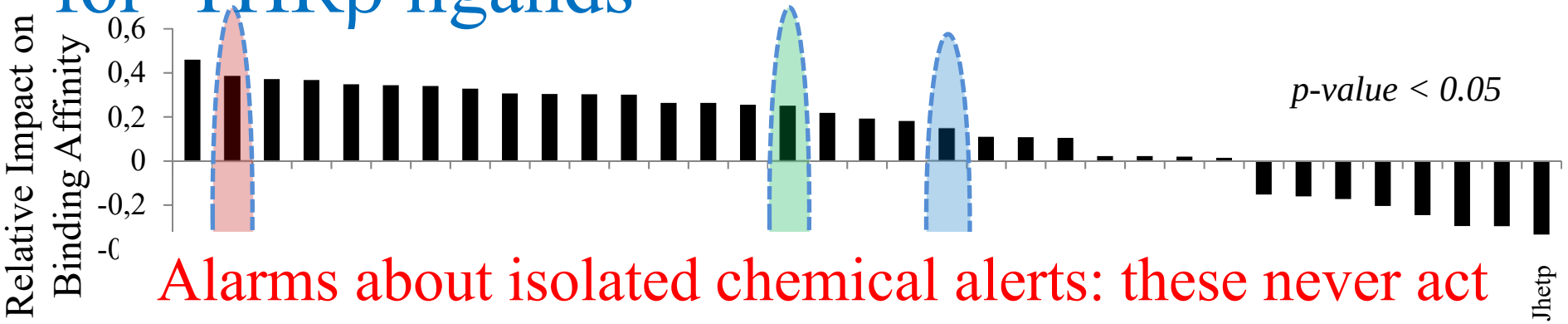


Mequitazine
Antihistamine
het_thio_666_A(13)

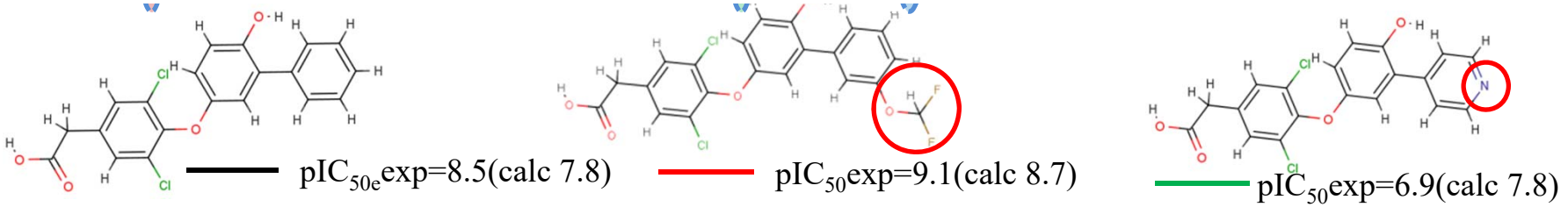


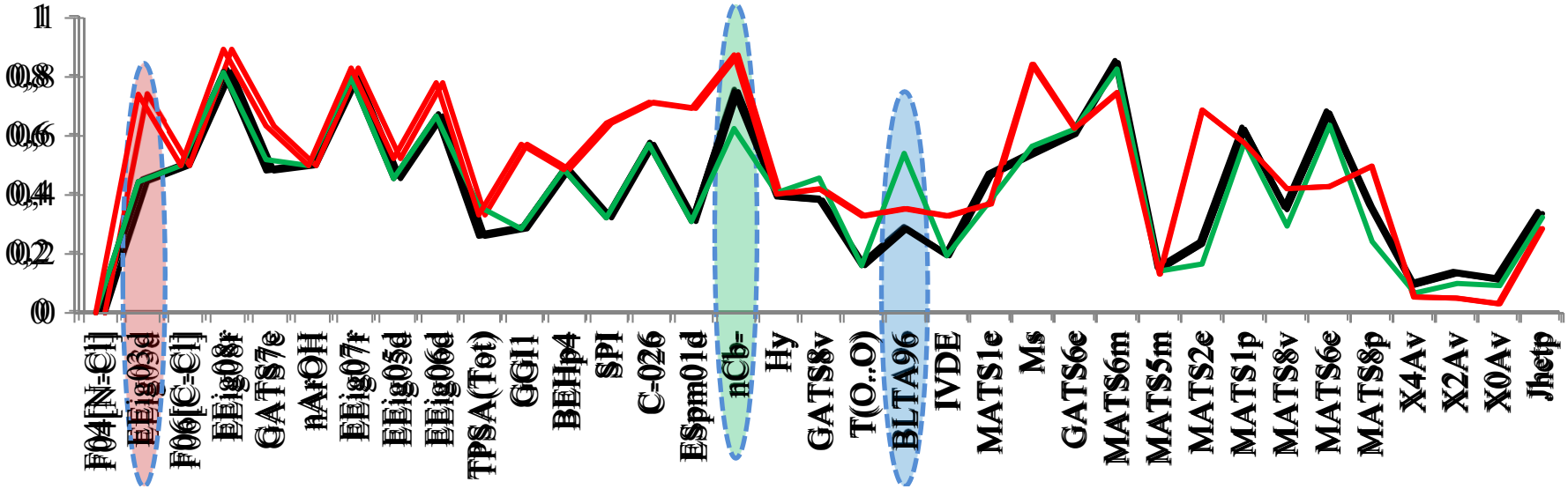
Tadalafil
Erectile disfunction
indol_3yl_alk(461)

Relative descriptor influence: QSAR Models for THRβ ligands

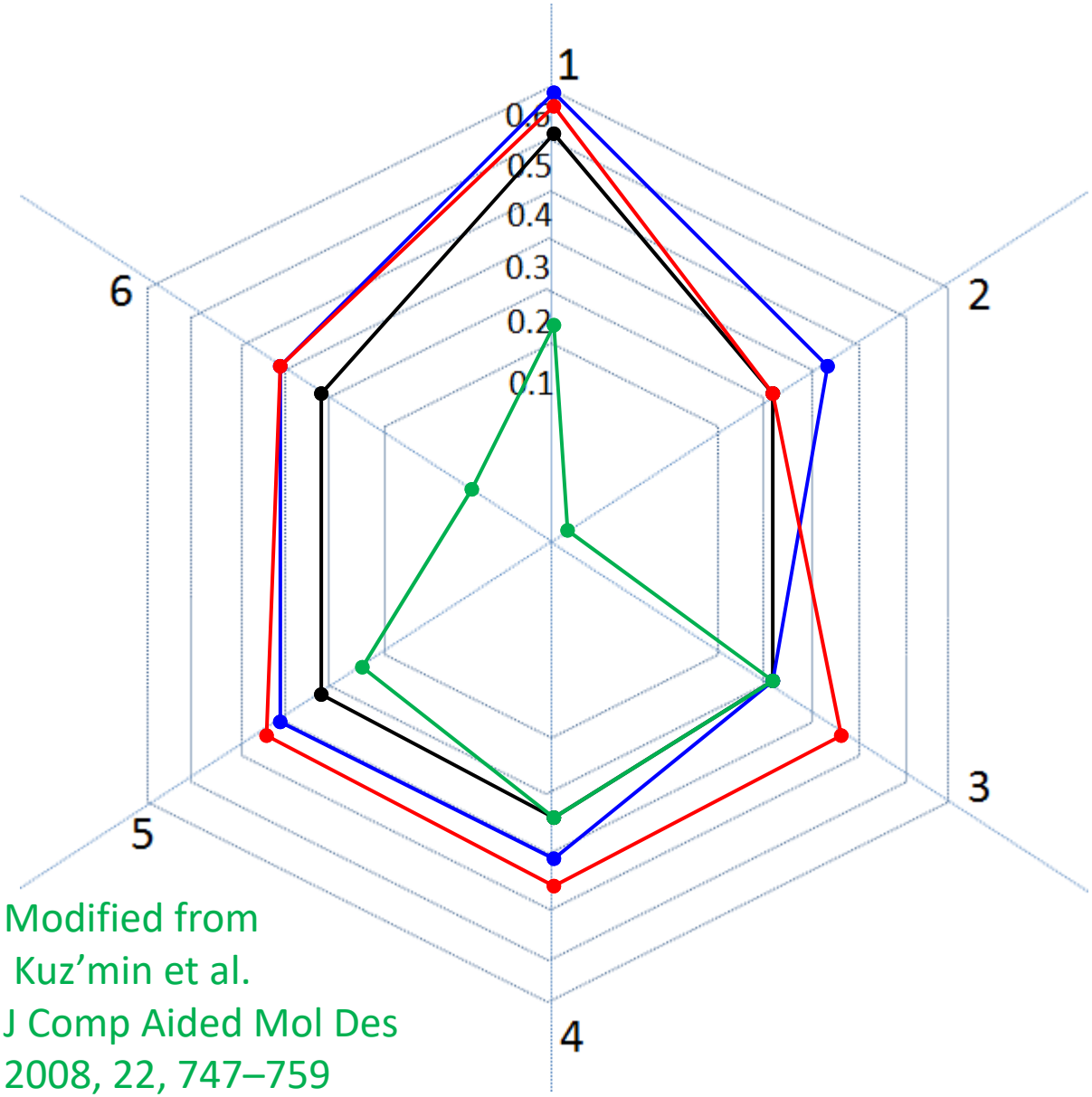


Alarms about isolated chemical alerts: these never act independently from the rest of the structure!

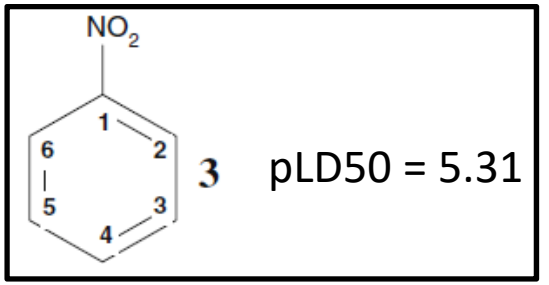




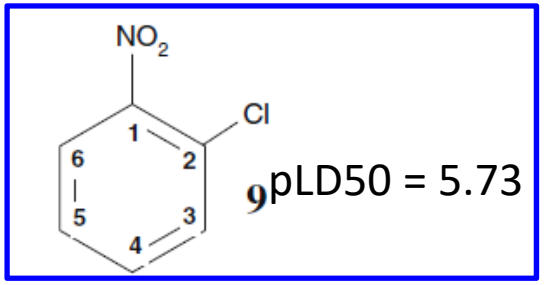
Relative influence of structural fragments on toxicity of chlorosubstituted nitrobenzenes



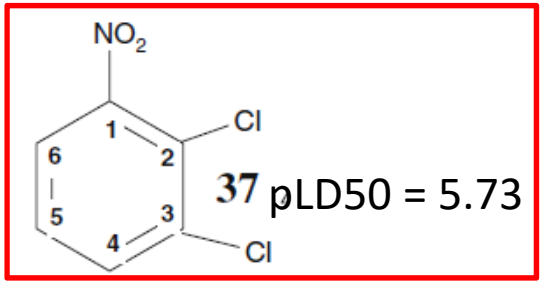
Modified from
Kuz'min et al.
J Comp Aided Mol Des
2008, 22, 747–759



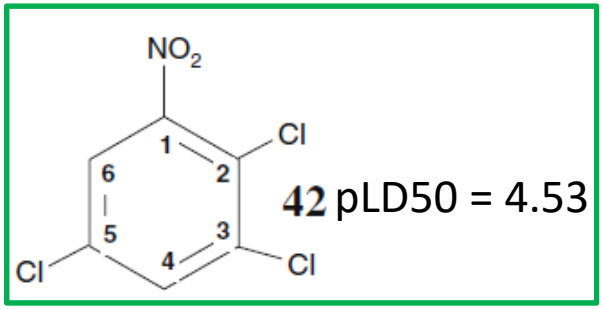
3 pLD50 = 5.31



9 pLD50 = 5.73

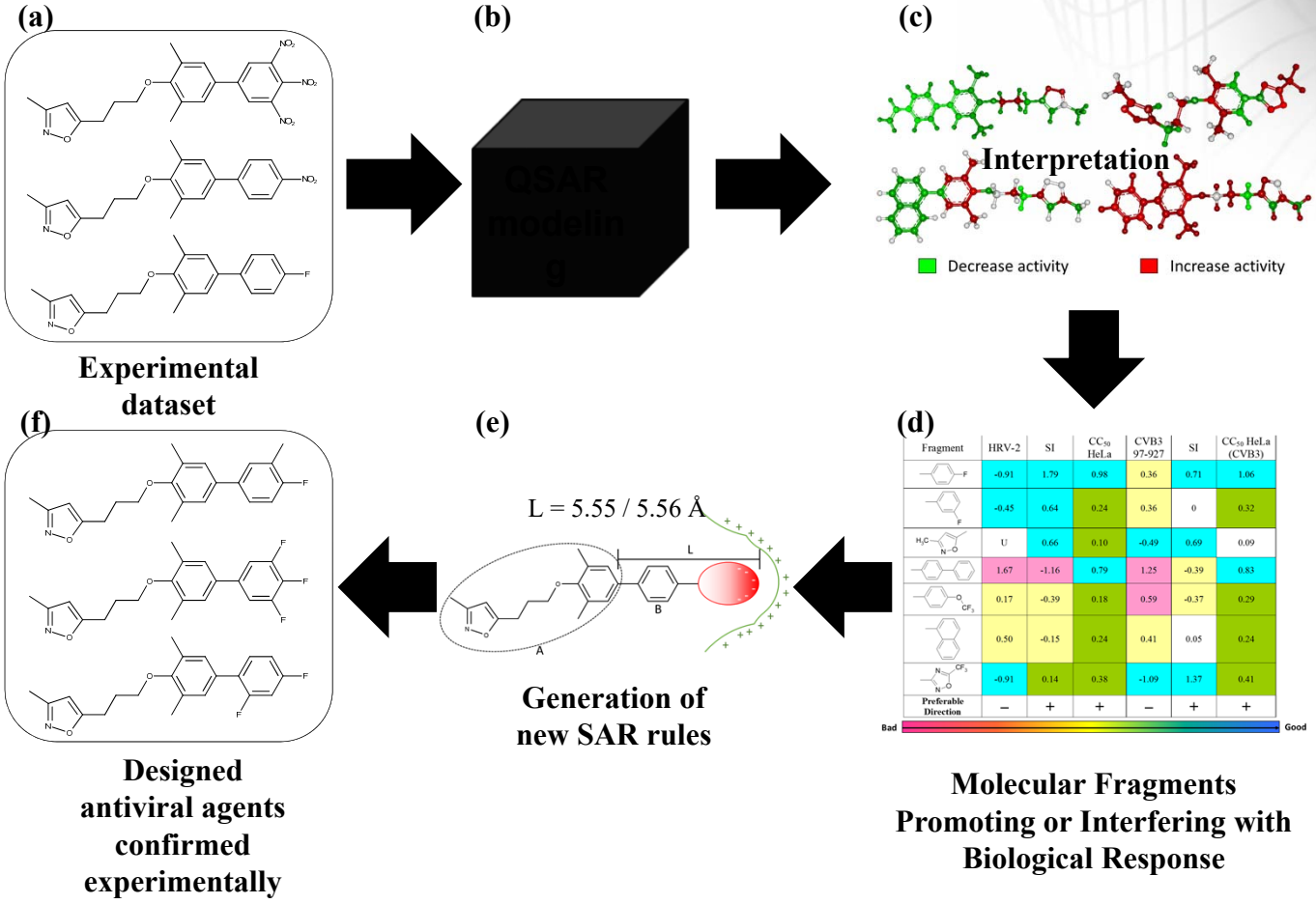


37 pLD50 = 5.73



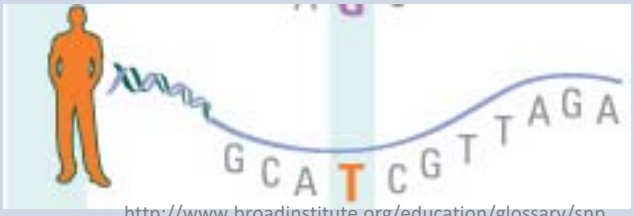
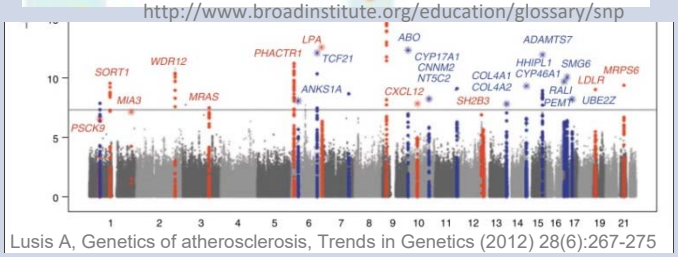
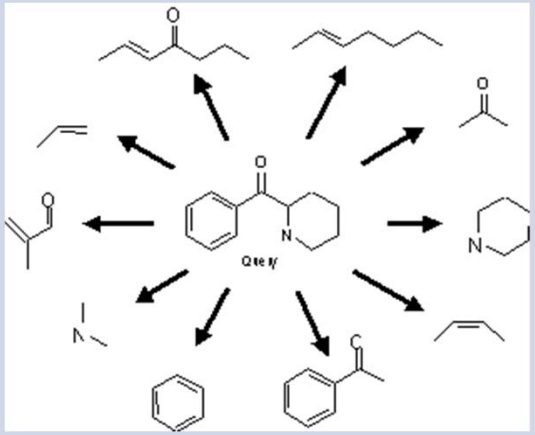
42 pLD50 = 4.53

An example of drug design based on descriptor interpretation



*Alves et al, Alarms about structural alerts. Green Chem, 2016, DOI: 10.1039/C6GC01492E

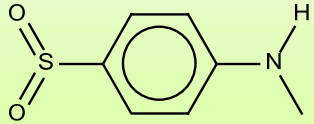
QSAR model interpretation based on Chemistry-Wide Association Studies (CWAS)

	GWAS	(Q)SAR
Samples	Patients	Compounds
Response	Phenotype (disease/no disease)	Activity (active/inactive)
Features	Single Nucleotide Polymorphisms (SNPs)  http://www.broadinstitute.org/education/glossary/snp  Lusis A, Genetics of atherosclerosis, Trends in Genetics (2012) 28(6):267-275	Chemical descriptors (e.g. fragments)  http://www.aldrichmarketselect.com/support/similarityOverview.asp
Objectives	Identify SNPs/loci associated with phenotype Predict phenotype from SNPs	Identify substructure associated with activity Predict activity from structure

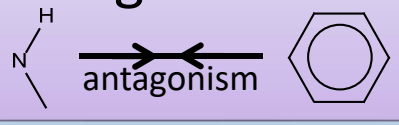
CWAS: develop and employ QSAR models using GWAS framework

Significant fragments

Structural alerts (combined fragments)

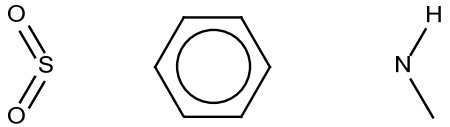


Mutually influencing fragments

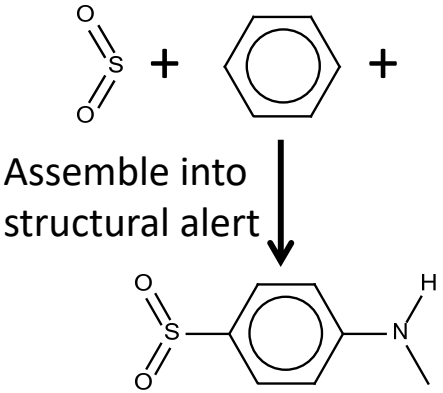


CWAS: study how chemical structures are associated with activity

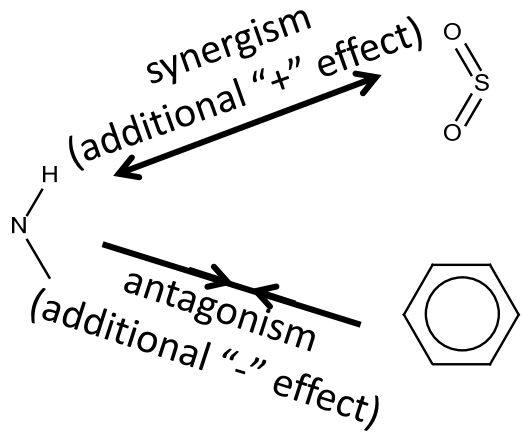
Significant fragments



Co-occurring fragments



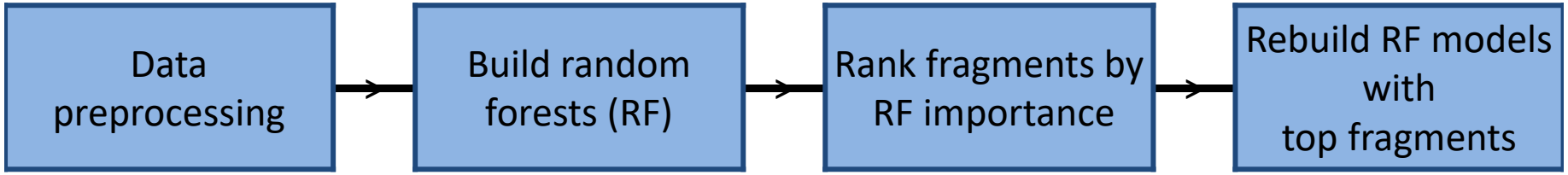
Fragment-fragment interactions associated with activity



Modeling and identifying important fragments

Ames data set
5,439 compounds
2,121 mutagenic
3,318 non-mutagenic

967 fragments → 76 fragments

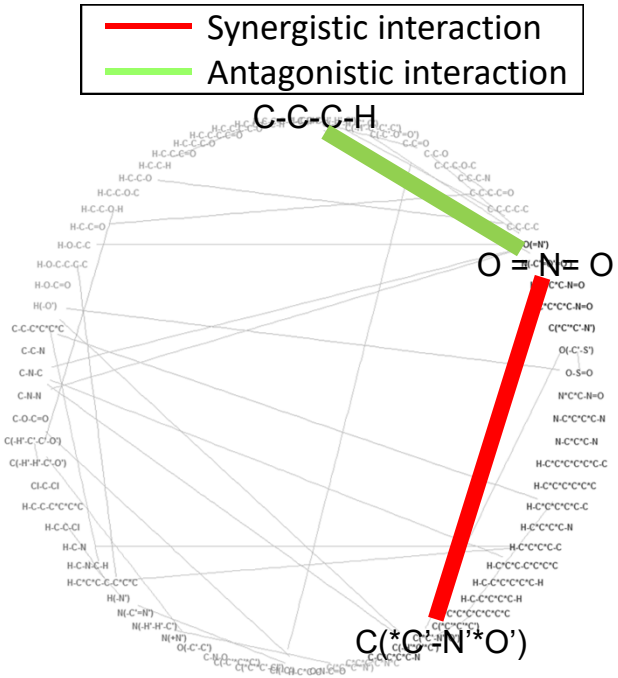


Chemical curation
Remove invariant, highly correlated fragments

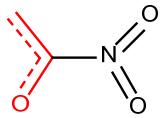
	Full model (967 fragments)		Reduced model (76 fragments)
Specificity	0.92 ±0.009		0.92 ±0.009
Sensitivity	0.78 ±0.005	Slightly improved	0.81 ±0.005
Balanced Accuracy	0.85 ±0.005		0.87 ±0.005
AUC	0.91 ±0.004		0.94 ±0.003

Results from 5-fold external cross validation

Nitro's mutagenic effect is:
increased by furan (synergism)
decreased by primary alkanes(antagonism)

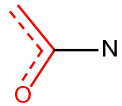


Synergistic influence

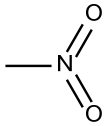


100% mutagenic
79:0

+



94% mutagenic
79:5



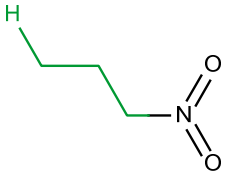
84% mutagenic ("penetrance")
620:118

+

C – C – C – H

29% mutagenic
785:1884

Antagonistic influence

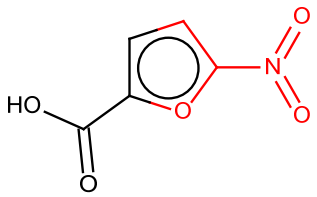


69% mutagenic
100:46

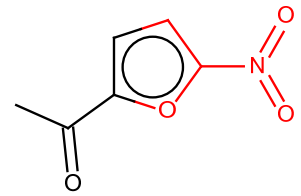
Number of mutagenic compounds : Number of non-mutagenic compounds

Nitro compounds are active when paired with aromatic rings
inactive when paired with primary alkanes

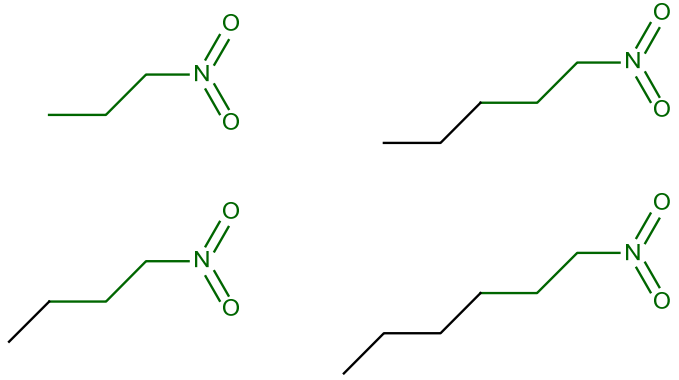
Examples



645-12-5
5-nitro-2-furanoate
Mutagenic

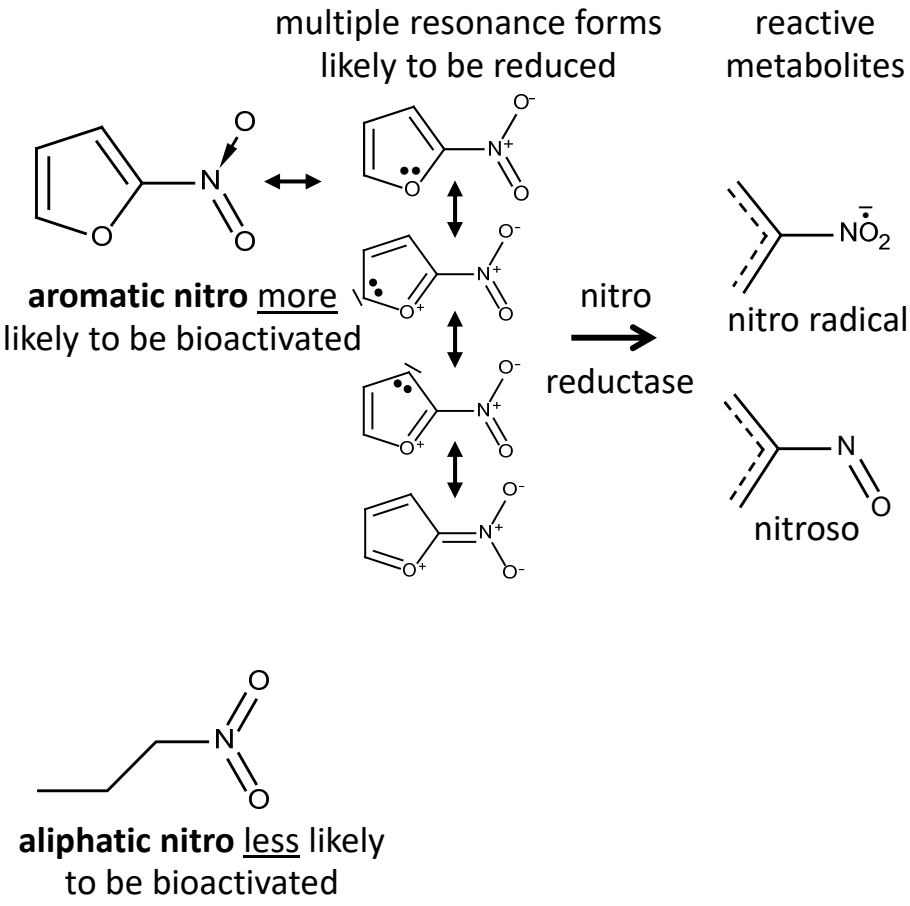


5275-69-4
2-acetyl-5-nitrofuran
Mutagenic



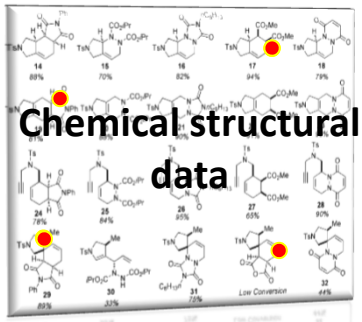
nitroalkanes (primary)
Nitro(prop – hex)ane
Non-mutagenic

Mechanism

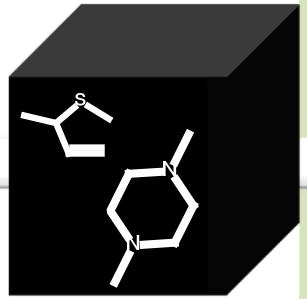


Benigni 2011 *Chem Rev*
Helguera 2006 *Toxicol*
McCalla 1983 *Env Mutagen*

Marrying SAR and QSAR in CWAS: Deriving alerts from validated QSAR models



Chemical structural data



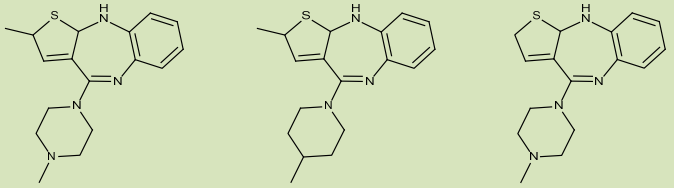
QSAR model

PREDICT

Virtual screening

Mutagenic

94% AUC



Non-mutagenic

76 significant fragments

Structural alerts

Mutually influencing fragments

INTERPRET

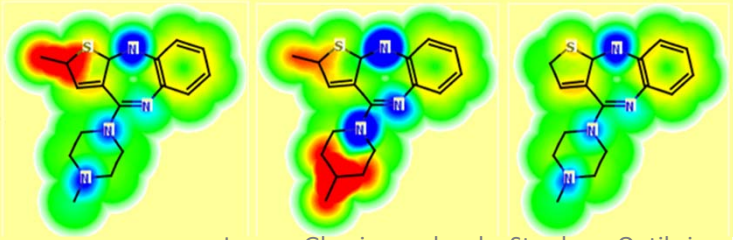
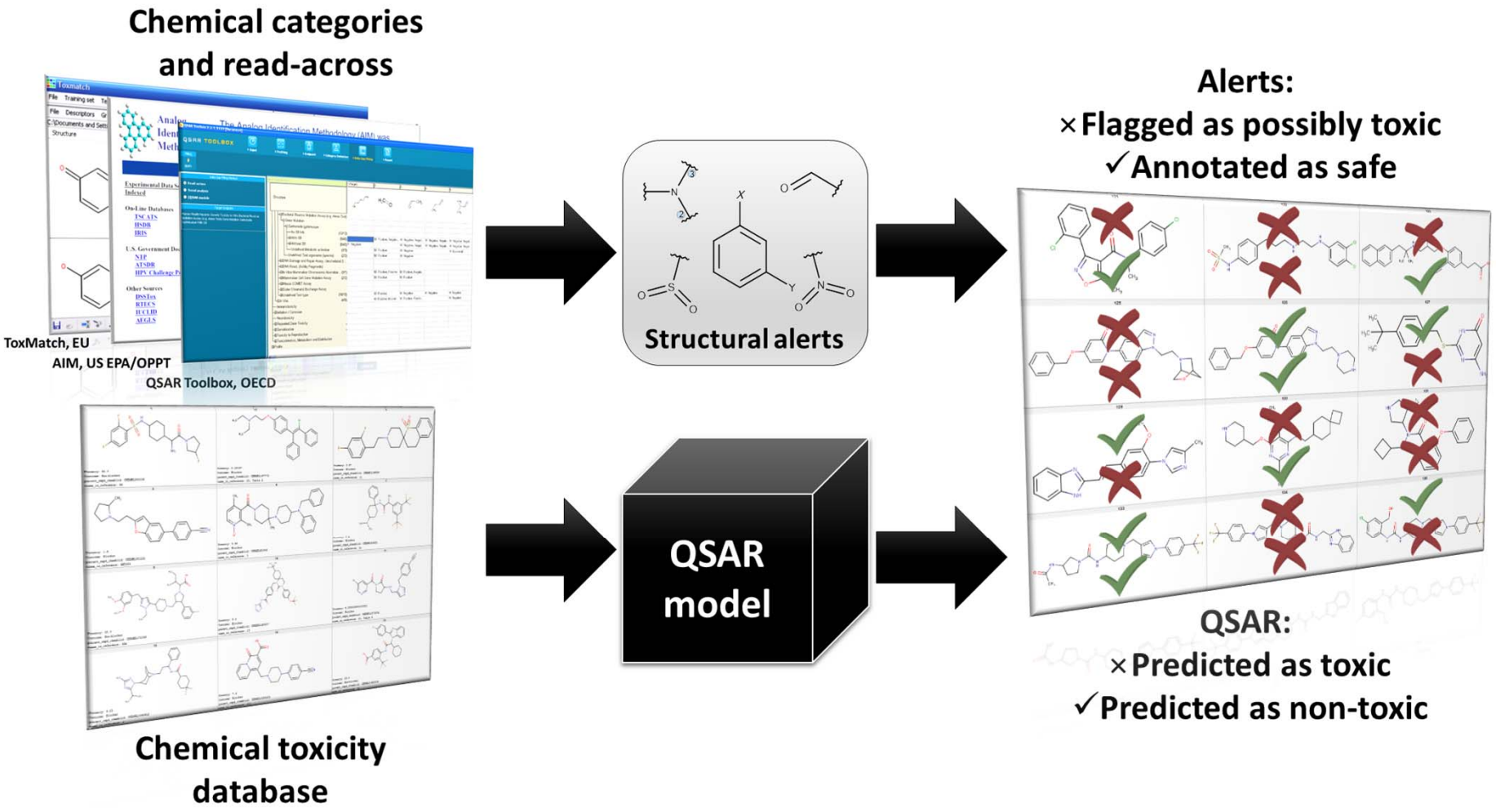


Image: Glowing molecule, Stardrop, Optibrium

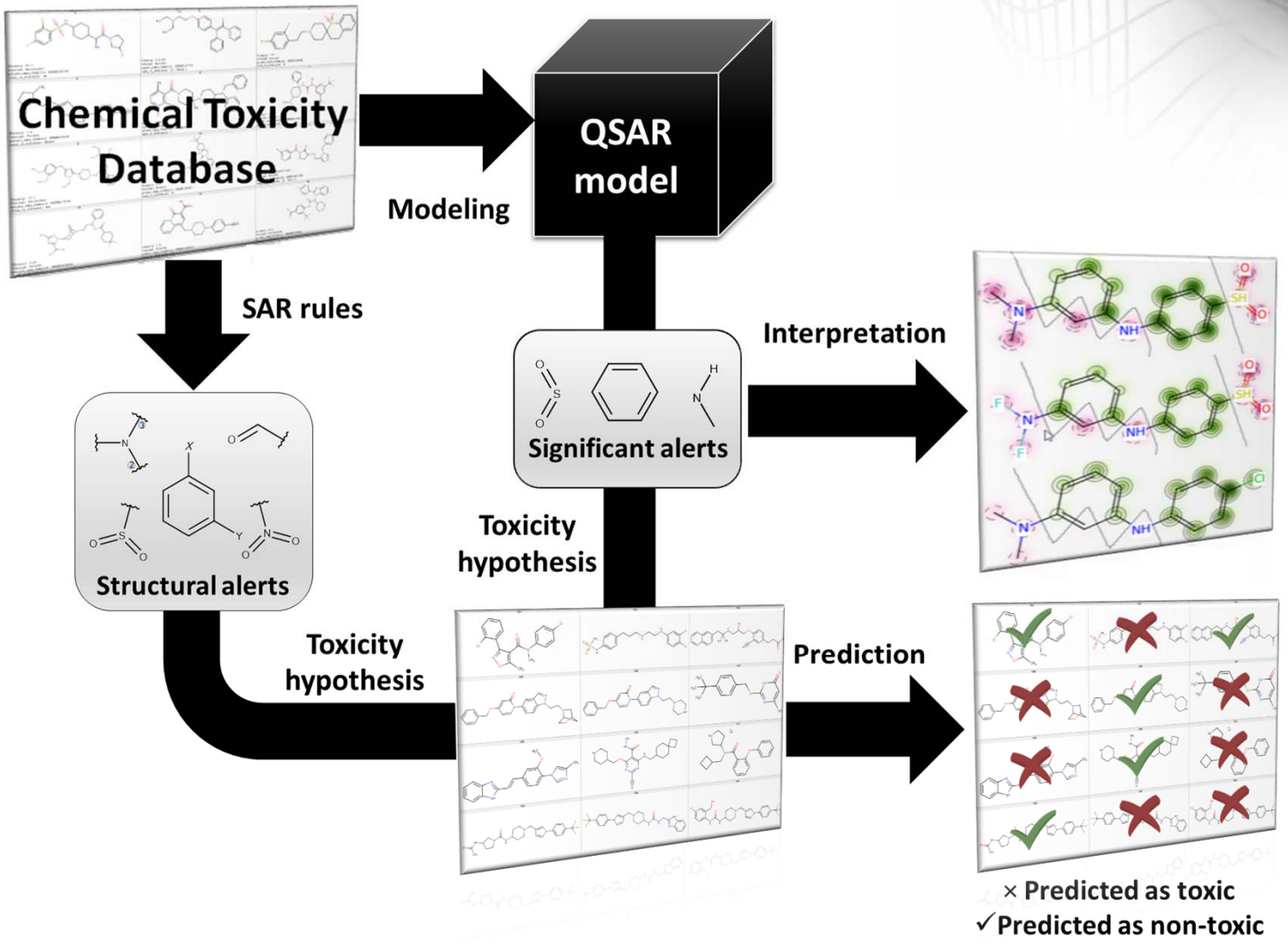
Data-driven drug design

Contrasting alerts- and QSAR-based predictions in chemical safety assessment.



*Alves et al, Alarms about structural alerts. Green Chem, 2016, DOI: 10.1039/C6GC01492E

Concept and strategy consolidation: Combining structural alerts and QSAR models.



*Alves et al, Alarms about structural alerts. Green Chem, 2016, DOI: 10.1039/C6GC01492E

Conclusions

- Although transparent and mechanistically interpretable, “isolated” alerts derived from training sets have limited predictive power;
- The influence of “alerts” on the compound depends on their structural environment;
- Mutually dependent alerts derived from externally validated QSAR models (cf. CWAS) afford higher predictivity;
- Any alert should be viewed as a structural hypothesis of chemical action; its predictive power should be confirmed by QSAR predictions and, if possible, by experimental validation

The Laboratory for Molecular Modeling

Principal Investigator
Alexander Tropsha

Research Professors
Alexander Golbraikh
Denis Fourches (now NCSU)
Eugene Muratov

Graduate students
Stephen Capuzzi
Sherif Faraq
Vinicius Alves
Yen Low (now Netflix)

Postdoctoral Fellows
Olexander Isayev
Regina Politi

Adjunct Members
Weifan Zheng, Shubin Liu

Collaborators
Ivan Rusyn
Diane Pozefsky

MAJOR FUNDING

NIH
- R01-GM66940
- R01-GM068665

NSF
- ABI 9179-1165

EPA (STAR awards)
- RD832720
- RD833825
- RD834999

ONR