A Computational Tool for Defining Conservation of Molecular Initiating Events Across Species

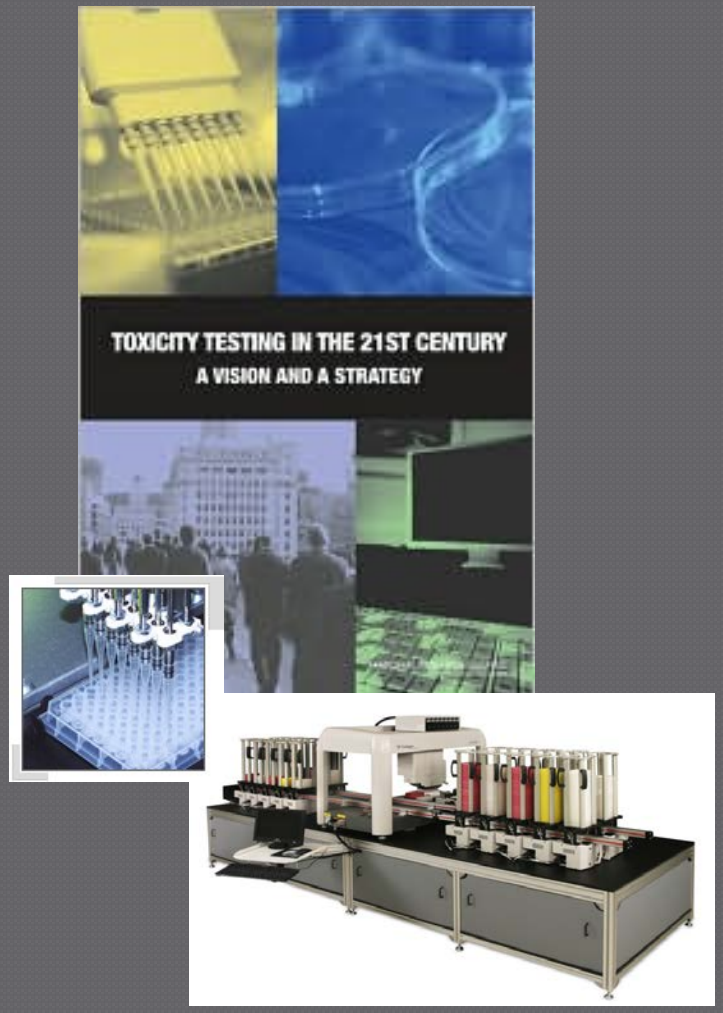
U.S. EPA
Carlie LaLone
8/22/13

Chemical Safety for Sustainability

Problem Statement

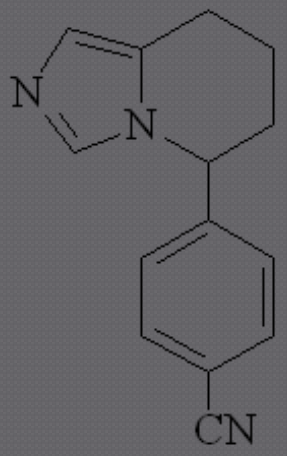
Although chemicals are essential to modern life, **we lack innovative, systematic, effective, and efficient approaches and tools to inform decisions that reduce negative environmental and societal impacts of chemicals** while increasing economic value.

Toxicity Testing in the 21st Century



- "Transform toxicity testing **from a system based on whole-animal testing to one founded primarily on in vitro [or in silico] methods** that evaluate changes in biologic processes using cells, cell lines, or cellular components, preferably of human origin"
- "The vision emphasizes the development of **suites of predictive, high-throughput assays** ….." (p. 7)
- "The mix of tests in the vision include tests that **assess critical mechanistic endpoints involved in the induction of overt toxic effects rather than the effects themselves**." (p. 121)

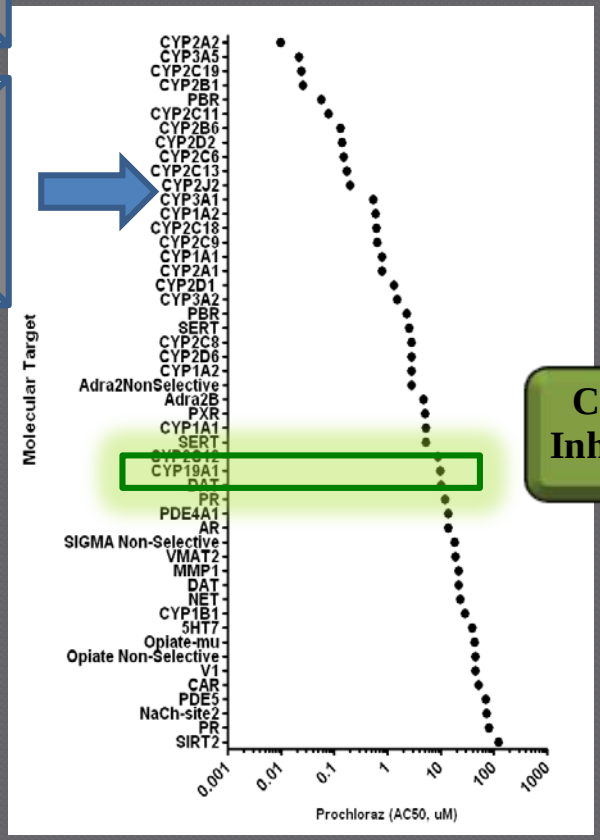
Increase the efficiency and speed of chemical evaluations



Inherency Tools

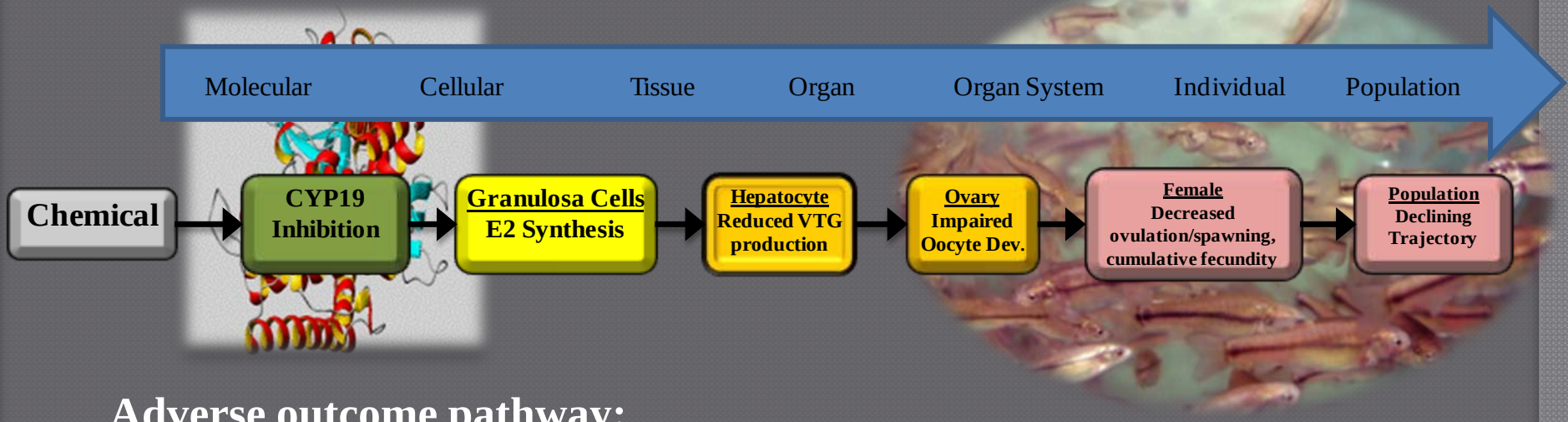
High Throughput Screening

CYP19 Inhibition



CYP19 Inhibition

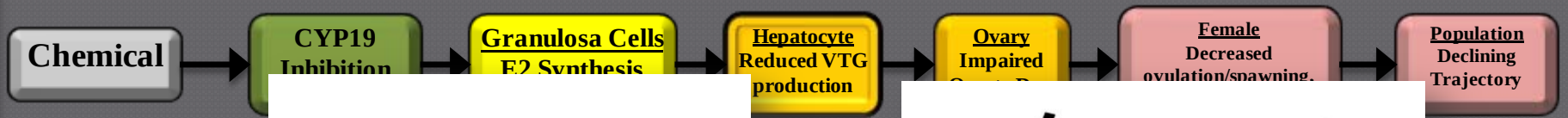
Identify putative molecular initiating events: chemical - biomolecule

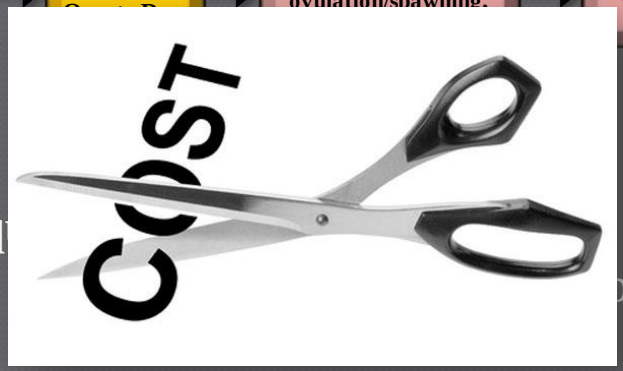


Adverse outcome pathway:

- Links molecular initiating event to adverse outcome relevant to risk assessment
- Hazard prediction from inherency or toxicity-pathway data

Application of 21st C Paradigm to Ecotoxicology – challenge of species extrapolation





rget Seq

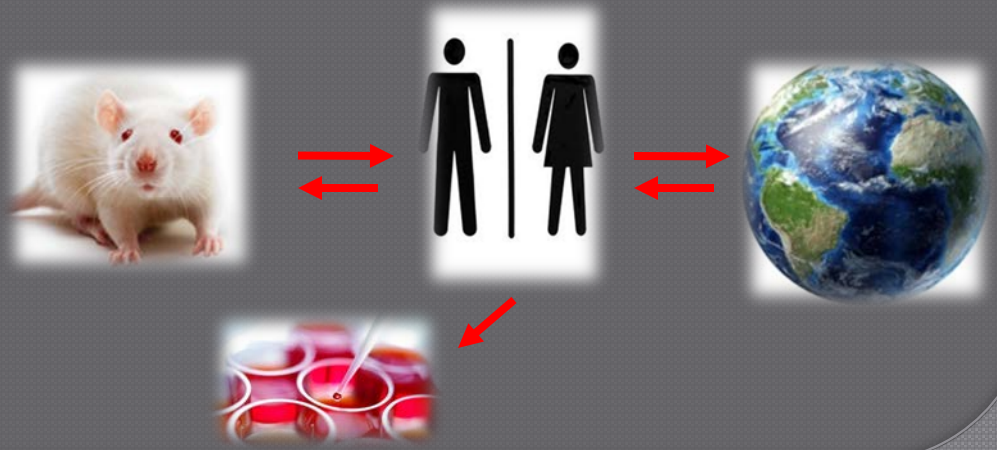
tes rapid

rotein

tion for predicting species applicability

Taxonomic applicability domain

- **Inherency tools**
- **Molecular screening data**
- **AOPs (to some extent)**



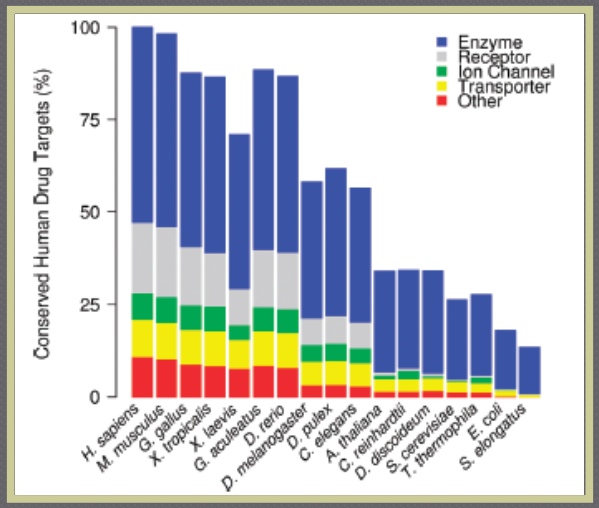
Chemicals with well defined MIEs

- Pharmaceuticals and Pesticide
  - Designed to act on specific molecular targets to provide therapeutic benefits or exterminate pests
    - Molecular targets
      - Genomic information and translated protein sequence information
- How do we translate this knowledge to potential effects on ecological species?
- Molecular Target Similarity– Species Extrapolation

Molecular Target Similarity: Species Extrapolation

- Ability to identify organism classes with differing sensitivity based on molecular target homology (Kostich and Lazorchak, Science of the Total Environment. 2008, 389, 329-339)
- Distribution of orthologs

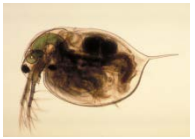
(Gunnarsson et al. Environ Sci Technol. 2008; 42(15):5807-13)



- Which proteins are conserved and to what degree?

Example Supporting Sequence-based Intrinsic Susceptibility Predictions

- Estrogen Receptor (OW/ORD Emerging Contaminants Workgroup, 2008)

Animal Kingdom	Genus	Common name		Chronic value (ng/L)
Vertebrates, aquatic	*Danio*	Zebrafish		<1.1
	Pimephales	Fathead minnow		1.5
	Oryzias	Medaka		3.2
	Oncorhynchus	Rainbow trout		<16
Invertebrates, aquatic	*Potamopyrgus*	Snail		50
	Gammarus	Freshwater shrimp		>7,600
	Daphnia	Water flea		45,000
	Tisbe	Copepod		>100,000
	Chironomus	Midge		320,000
	Brachionus	Rotifer		800,000

- Moving from qualitative understanding of molecular target conservation to quantitative measures
  - New tools and technologies have emerged

Developing an Automated Computational Methodology

Protein Sequence Similarity Tool

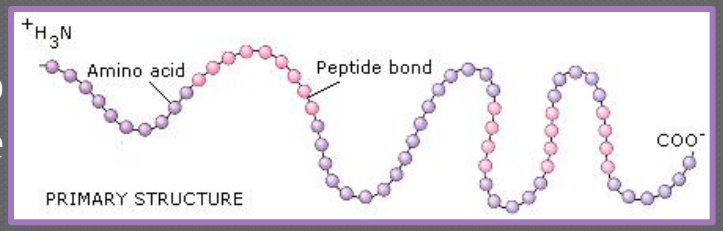
Predict relative intrinsic susceptibility

- Relative in that it is dependent on which species is selected as the target (query species)
- Intrinsic susceptibility can be defined as the vulnerability (or lack thereof) of an organism to chemical insult due to its inherent biological composition
  - Receptor/enzyme (protein) available for the chemical to act upon

Strategic Approach for Assessing Protein Similarity

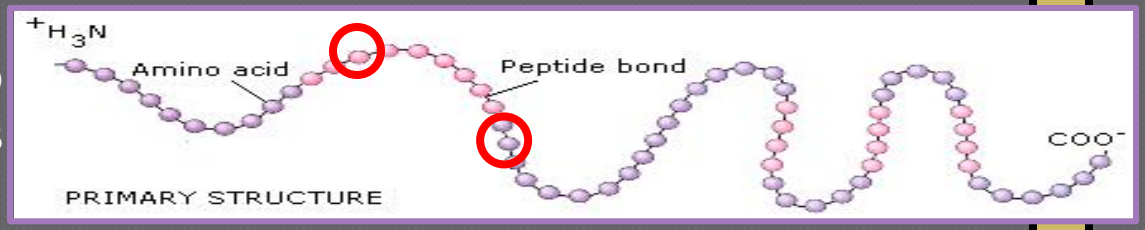
Low Level of Complexity

Primary Amino Acid Sequence

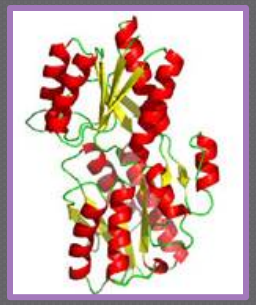


Conserved Functional Domains

Individual Amino Acid Residue Queries



Tertiary Protein Structure

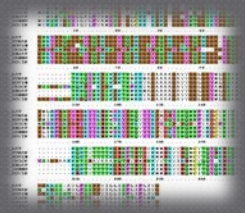


High level of Complexity

Computational Assessment of Protein Similarity: Quantitative Metrics

- Assume that presence of molecular target in non-target species is one critical route via which a chemical could cause adverse effects
- Target species vs. Non-target species (NCBI)
  - Align amino acid sequences and conserved domains
  - Assume greater similarity = greater likelihood
    - interact with molecular target in non-target species

vs.



All Species in NCBI Protein Database

Linking Data Sources

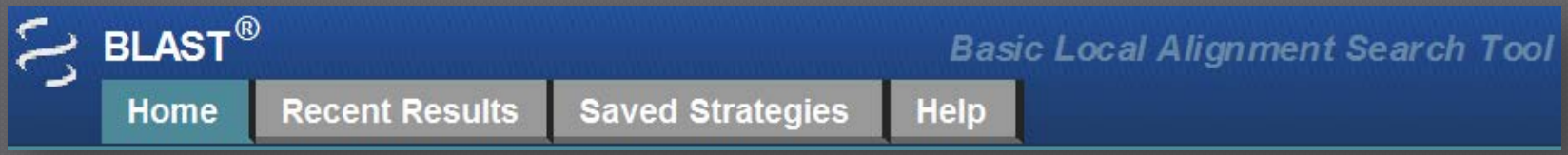
PPDB

DRUGBANK
Open Data Drug & Drug Target Database

VSDB

t3db
Toxin, Toxin-Target Database

- Pesticide Properties DataBase, T3DB, Veterinary Substances DataBase, & DrugBank : identifies molecular targets for pesticide/pharmaceutical
- Link to NCBI GenBank: Protein accession



- Molecular Target Similarity Tool
  - Automated BLASTp and Conserved domain database
  - Query target species protein accession against all organism classes for vertebrates, invertebrates, and plants

Output from Sequence Analysis

Number of Proteins in NCBI Database	Organism Class	Species	NCBI Protein Accession	Protein Name			% Similarity	Number of Conserved Domains
35,859	Insecta	*Aedes aegypti*	ACB37024.1	voltage-gated para-like sodium channel			100.0	4
115	Insecta	*Culex pipiens pallens*	BAI77918.1	para-sodium channel	3964	0	88.8	4
38,393	Insecta	*Culex quinquefasciatus*	BAI77917.1	para-sodium channel			88.8	4
30,028	Insecta	*Drosophila virilis*	EDW65213.1	paralytic	3410	0	76.4	3
324	Insecta	*Periplaneta americana*	ACX44801.1	voltage-gated sodium channel Nav1 alpha subunit	3218	0	72.1	4
16,324	Insecta	*Danaus plexippus*	EHJ74501.1	voltage-gated sodium channel alpha subunit	3212	0	72.0	4
15,129	Insecta	*Harpegnathos saltator*	EFN86793.1	Sodium channel protein para	3212	0	72.0	4
31,591	Branchiopoda	*Daphnia pulex*	EFX81393.1	hypothetical protein DAPPUDRAFT_50150	2652	0	59.4	3
20	Malacostraca	*Cancer borealis*	ABL10360.2	voltage-gated sodium channel	2523	0	56.5	3
41,714	Arachnida	*Ixodes scapularis*	EEC03677.1	skeletal muscle sodium channel alpha subunit, putative	2416	0	54.1	3
212	Arachnida	*Haplopelma schmidti*	ABH12275.1	putative voltage-gated sodium channel	2273	0	50.9	2
25,335	Mammalia	*Callithrix jacchus*	XP_002749429.1	PREDICTED: sodium channel protein type 2 subunit alpha isoform 2	1640	0	36.7	2
25,661	Mammalia	*Macaca mulatta*	XP_001100011.1	PREDICTED: sodium channel protein type 2 subunit alpha-like isoform 6	1639	0	36.7	2
596,962	Mammalia	*Homo sapiens*	NP_066287.2	sodium channel protein type 2 subunit alpha isoform 1	1639	0	36.7	2
40,939	Mammalia	*Sus scrofa*	XP_003133492.1	PREDICTED: 1			5.7	2
39,590	Mammalia	*Pan troglodytes*	XP_001153476.1	PREDICTED:			5.7	2

Most similar to target species

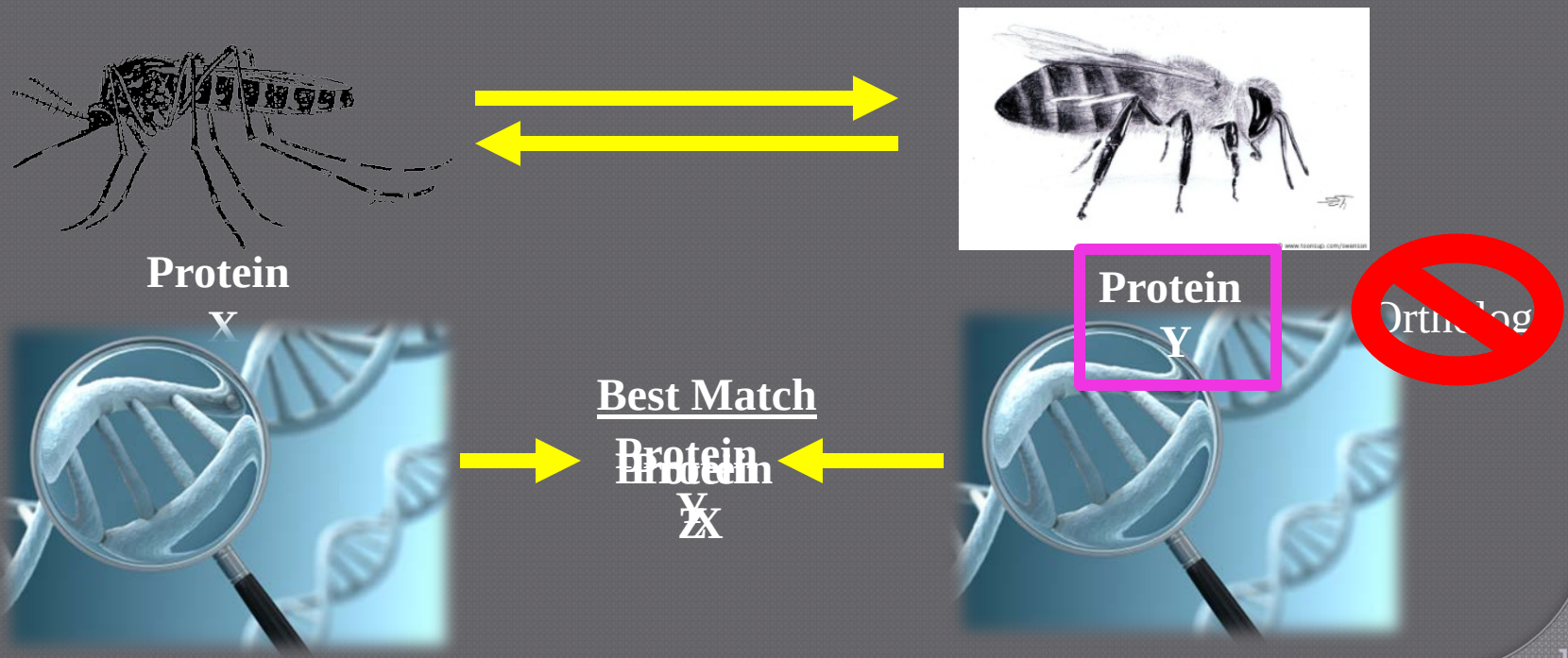
Species	NCBI Protein Accession	Protein Name	Bit score	E-value	% Similarity
Aedes aegypti	ACB37024.1	voltage-gated para-like sodium channel	4464	0	100.0
Homo sapiens	NP_066287.2	sodium channel protein type 2 subunit alpha isoform 1	1639	0	36.7

Species	NCBI Protein Accession	Protein Name	Bit score	E-value	% Similarity	Number of Conserved Domains
Aedes aegypti	ACB37024.1	voltage-gated para-like sodium channel	4464	0	100.0	4
Drosophila erecta	EDV46862.1	GG19319	3432	0	76.9	0

Least similar to target species

Setting the Susceptibility Cut-off

- Ortholog identification
  - Ortholog = A sequence diverged after a speciation event
- Reciprocal best hit BLAST

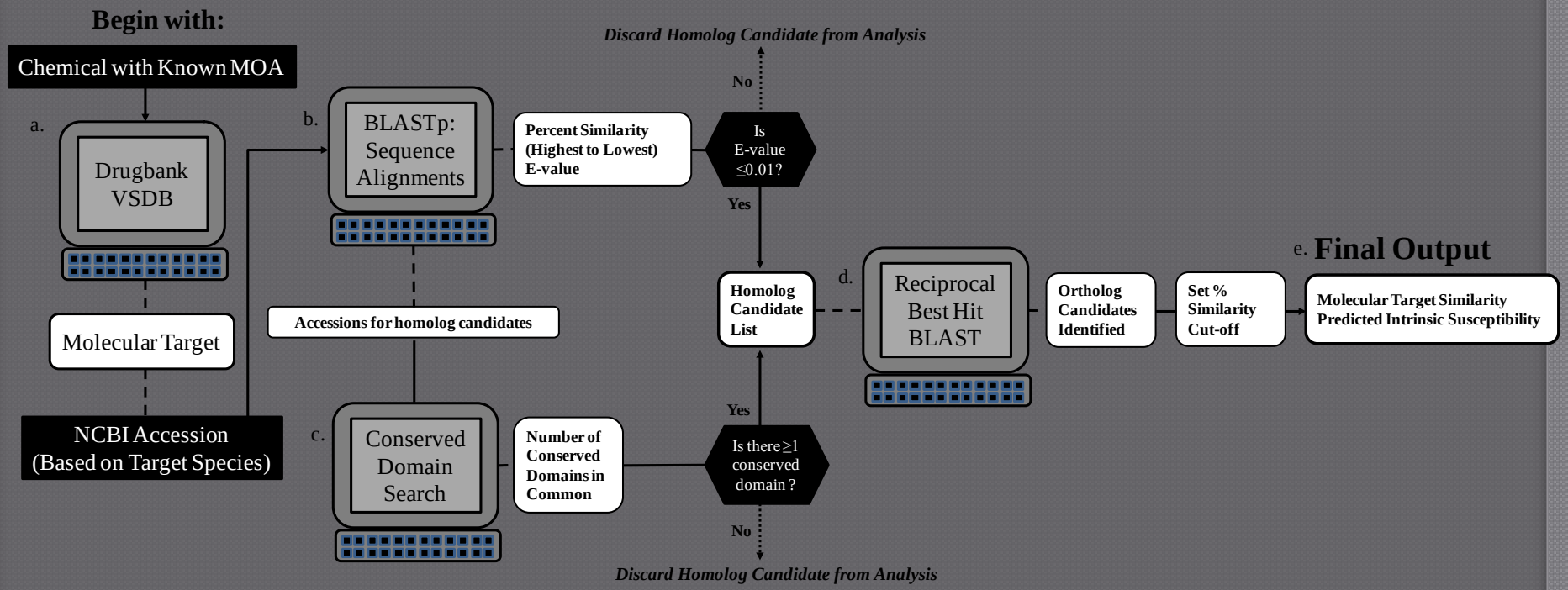


Orthologs Identified

- Set susceptibility cut-off
  - Protein with lowest percent similarity that was identified as an ortholog

Taxonomic Lineage	Species	Reciprocal Best Hit	NCBI Protein Accession	Protein Name	Bit score	E-value	% similarity
Insecta	*Aedes aegypti*	Target Species	ACB37022	voltage-gated para-like sodium channel	4449	0	100
Insecta	*Harpegnathos saltator*	Ortholog	EFN86793	Sodium channel protein para	3223	0	72.4
Arachnida	*Ixodes scapularis*	-	XP_002407119	skeletal muscle sodium channel alpha subunit, putative	2417	0	54.3
Insecta	*Solenopsis invicta*	Ortholog	EFZ17857	hypothetical protein SINV_07049	1551	0	34.9
Ascidiacea	*Ciona intestinalis*	-	XP_002123673	similar to sodium channel, voltage-gated, type III, alpha	1403	0	31.5
Gastropoda	*Aplysia californica*	-	NP_001191637	sodium channel alpha-subunit SCAP1	890	0	20.0
Chondrichthyes	*Heterodontus francisci*	-	ADV73289	voltage-dependent sodium channel 2	843	0	18.9

Analysis Overview

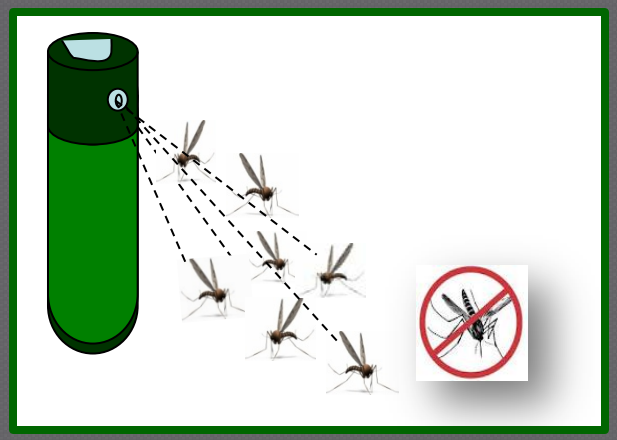


Case Examples: Illustration of Concept

Ethinylestradiol

17β-trenbolone

Permethrin

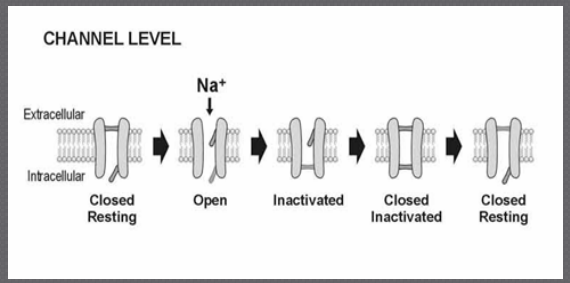


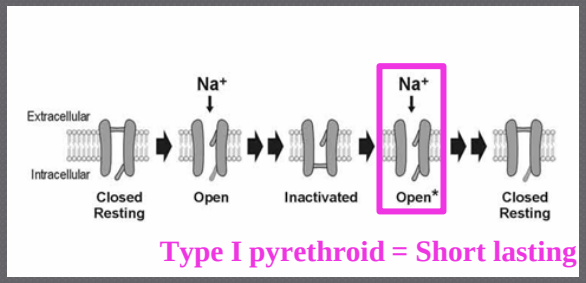
LaLone et al. Molecular Target Sequence Similarity as a Basis for Species Extrapolation to Assess the Ecological Risk of Chemicals with Known Modes of Action. 2013. Submitted.

Permethrin-Insecticide

- Molecular Target: Voltage-gated sodium channel

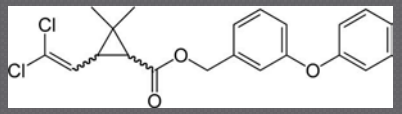
Exposure/ Target Tissue Dose	Molecular Initiating Event	Cellular Response	Organ Response	Organism Response	Population Response
→	VGSC Alterations →	Altered Neuronal Firing Rates →	Disrupted Neuronal Networks and Pathways →	Clinical Signs →	Dead Insects/ Altered Non-target Insect Populations





VGSC Slide material: Ke Dong, Michigan State University

- Acute neurotoxicity: Muscle spasm, paralysis, death
  - Target Species: Lice, ticks, fleas, mites, scabies (www.drugbank.ca), mosquitoes (U.S. EPA, Registration Eligibility Decision, RED)



Tool: Example Query

Protein Target 1: Voltage-gated para-like sodium channel

Multiple Protein Targets — Click to Add more protein targets

Protein Target 2:

Target Species : Aedes aegypti

Multiple Target Species — Click to Add more target species

Target Species:

Choose Analysis:

☑ **Full Search** ☐ **My Species** Apis mellifera ☐ **Program Office Species of Interest**

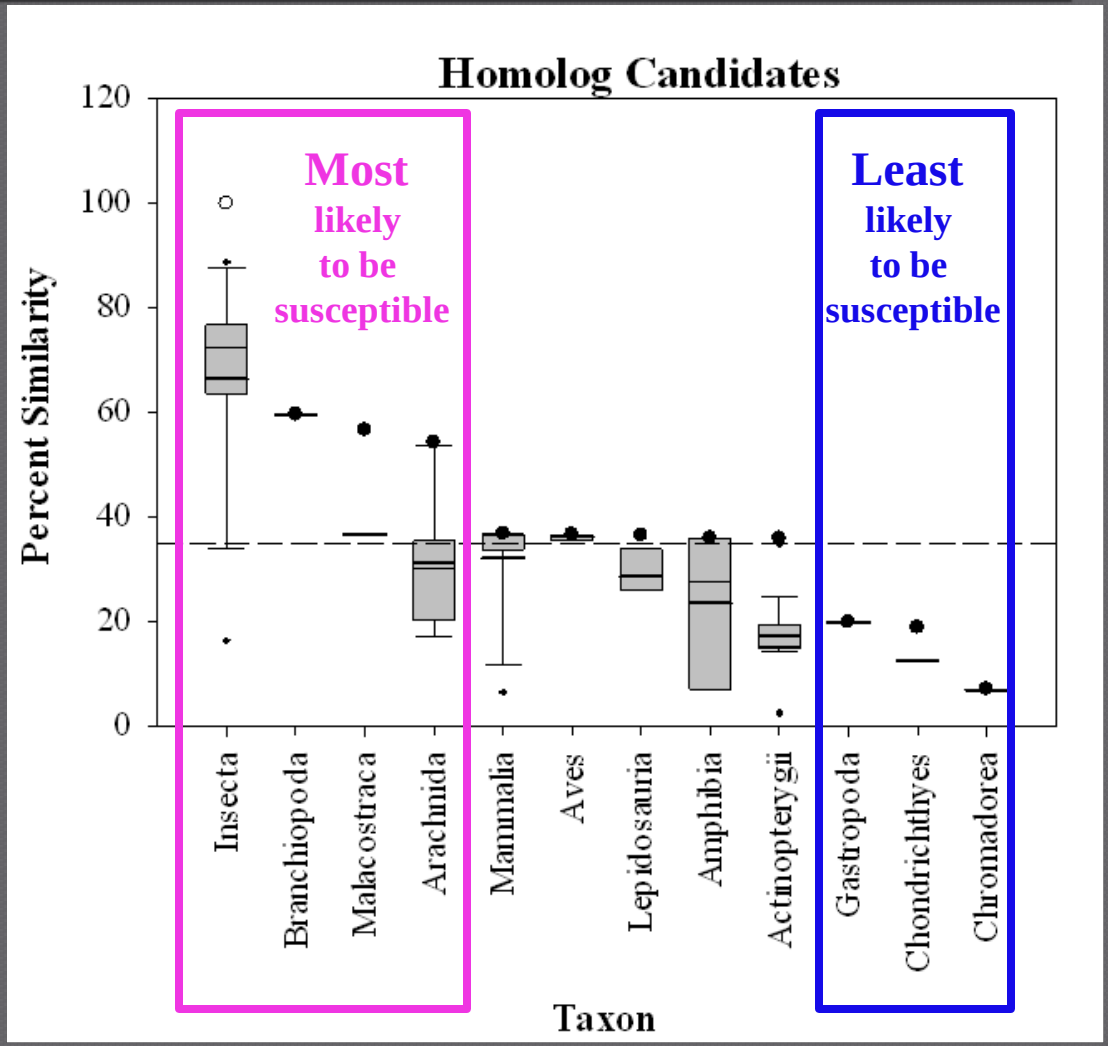
Click to Add more non- target species

Align Sequences

Conservation Across Species

Target Species: *Aedes aegypti*
Molecular Target: Voltage-gated para-like
sodium channel

Define taxonomic applicability domain



Comparing Relative Intrinsic Species Susceptibility Predictions to Empirical Toxicity Data

Mining the ECOTOX database

Summarizes Toxicity Data for Individual Chemicals

Chemical (Analytical measurements, Chemical purity)

Chemical	CAS
Permethrin	52645-53-1
(+-)-cis-Permethrin	61949-76-6
(+-)-trans-Permethrin	61949-77-7
(+)-cis-Permethrin	54774-45-7
(-)-cis Permethrin	54774-46-8
(+)-trans Permethrin	51877-74-8
(-)-trans Permethrin	54774-47-9

Species (Exclude Resistant/Susceptible Strains, Pesticide exposed)

- All Aquatic Taxa (Identify Life Stage)

Endpoint (Desired Acute Response in Target Species = Death)

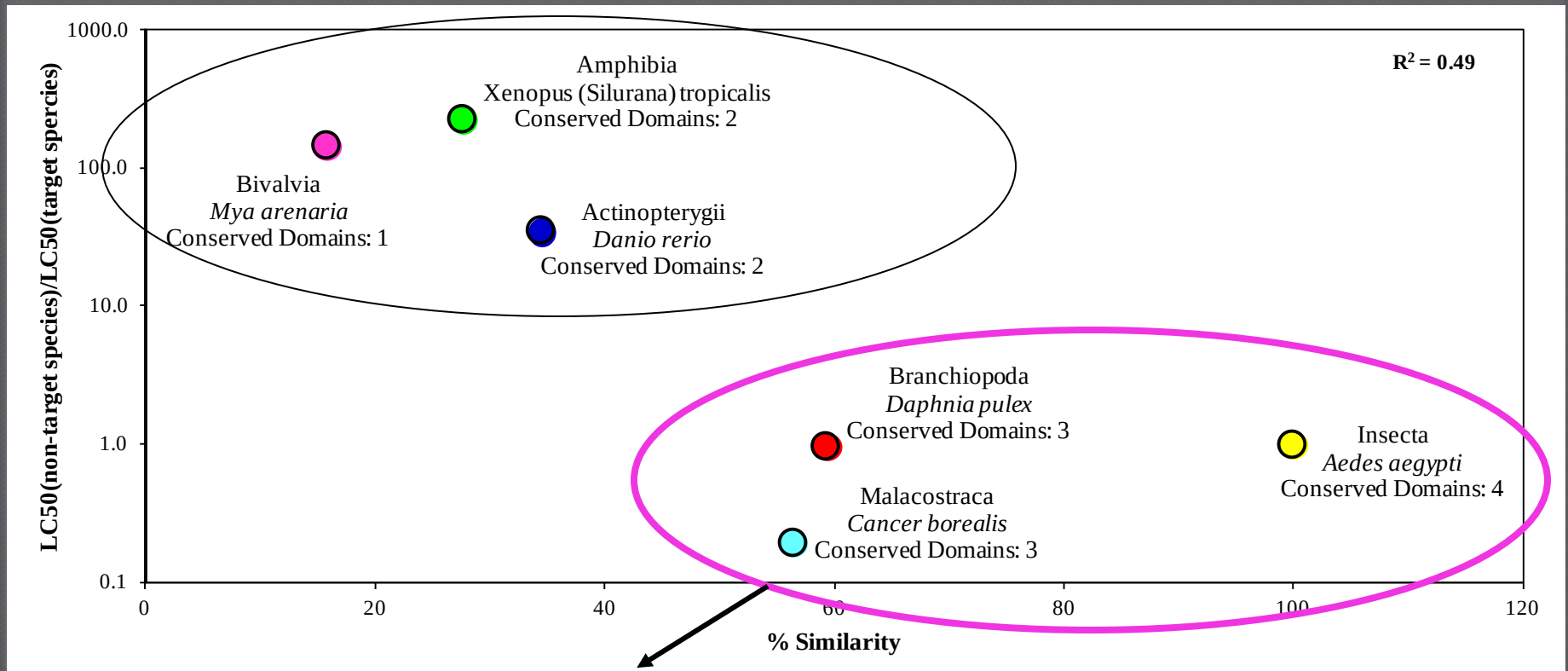
- EC50/LC50
- Mortality / Survivability
- Acute durations (48 hr invertebrate/96 hr vertebrate)
- Water Exposure (No feeding or diet)

Quick Database Query

Advanced Database Query

Compare to Homology Data

Test of concept



MIE likely to be relevant

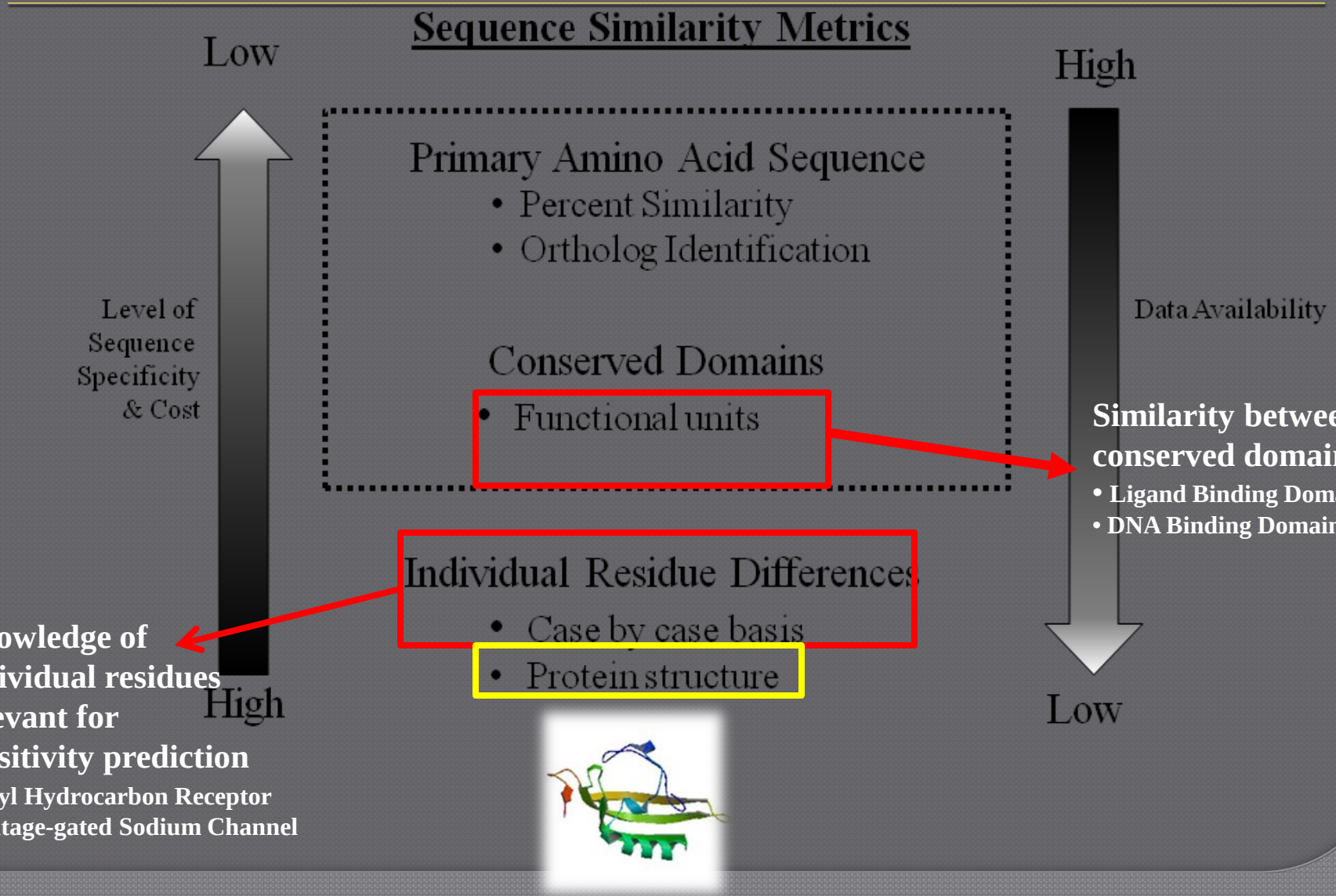
Exposure/ Target Tissue Dose → VGSC Alterations → Altered Neuronal Firing Rates → Disrupted Neuronal Networks and Pathways → Clinical Signs → Dead Insects/ Altered Non-target Insect Populations

Molecular Initiating Event

Taking a Closer Look

Delving into sequence comparisons

Logic behind development

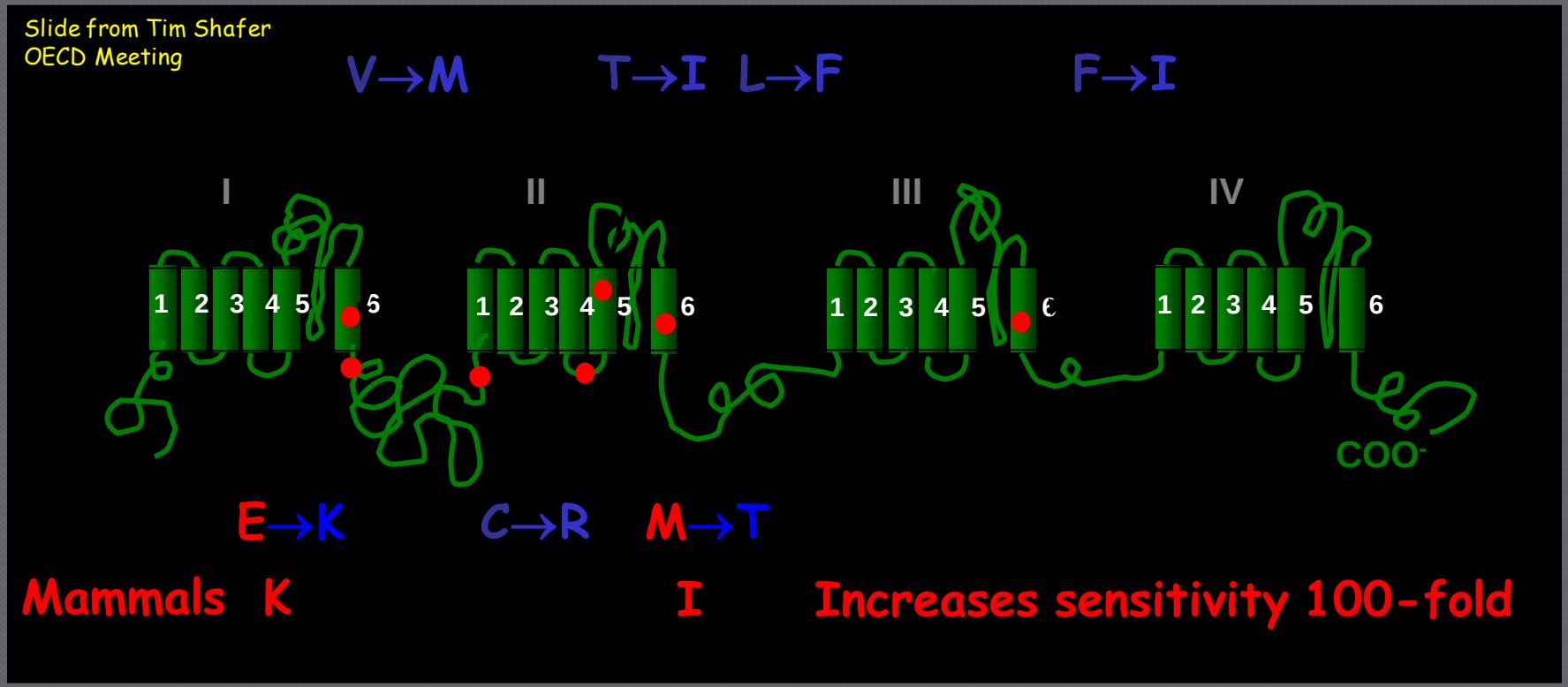


Conserved Domains: The Next Level

Original Query
Target Species: *Aedes aegypti*
Molecular Target: Voltage-gated para-like sodium channel
NCBI Accession: ACB37024.1
Bit Score: 4464
Total Conserved Domains: 4
Organism Class: Insecta

Organism Class	Non-Target Species	Protein Name	Calculated % similarity	Common Domains	Conserved domain IDs	**Domain type**	Sequence coverage of conserved domains	Conserved Domain % Similarity
Insecta	*Apis mellifera*	paralytic	73	3	pfam11933 pfam08016 pfam00520	Unknown function Polycystin cation channel Ion transport protein	1-222, same 217-423, same 1-194, same	61 92 93

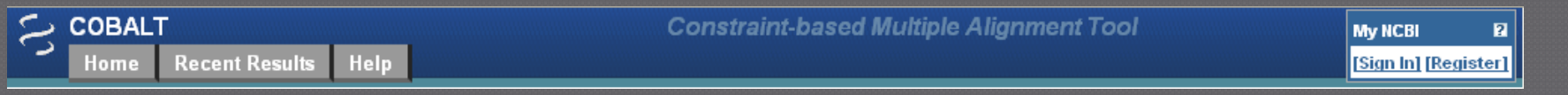
Knowledge of individual residues



Resistance and sensitivity: Single amino acid residue substitutions

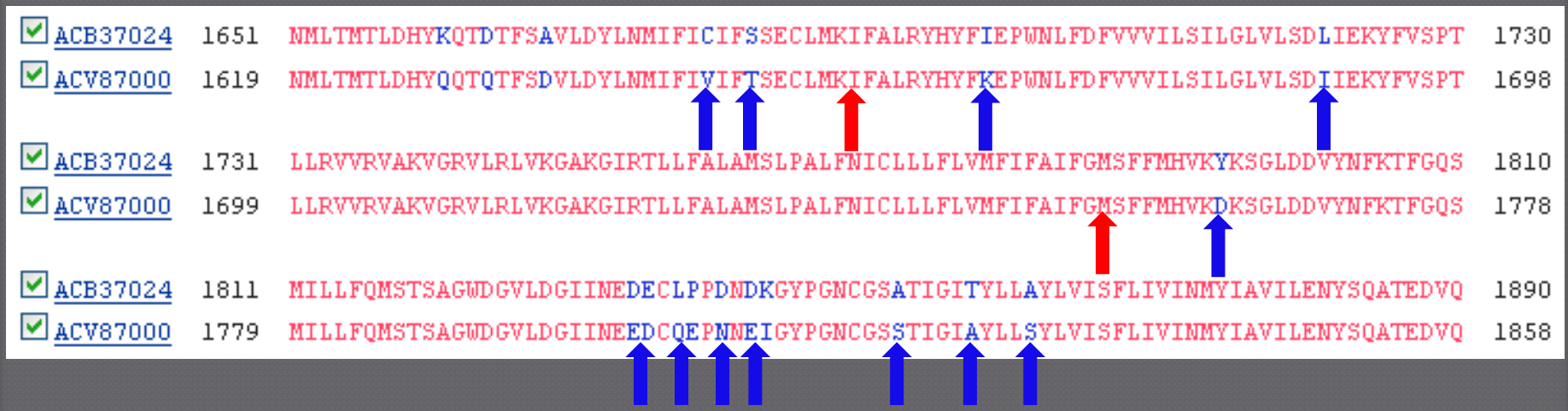
- Targeted mutagenesis
- Voltage clamp techniques

Targeted sequence comparisons



vs.

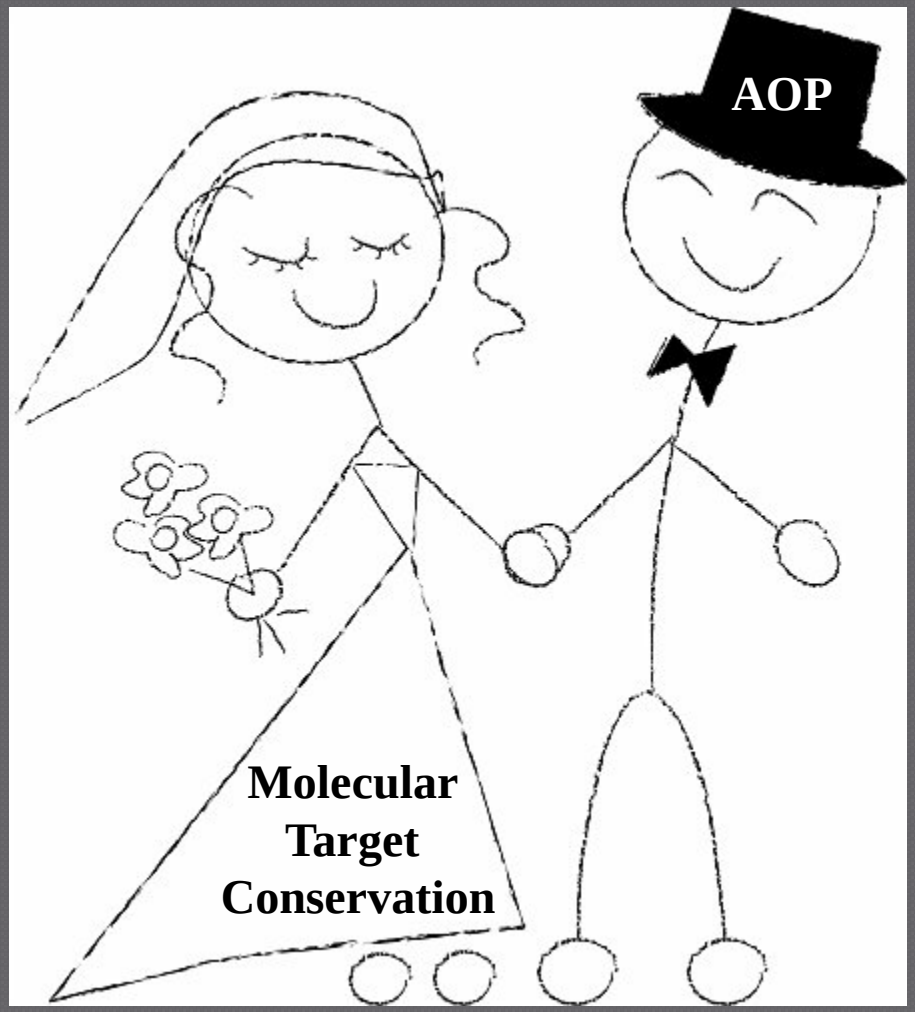
vs. Species A, B, C…..



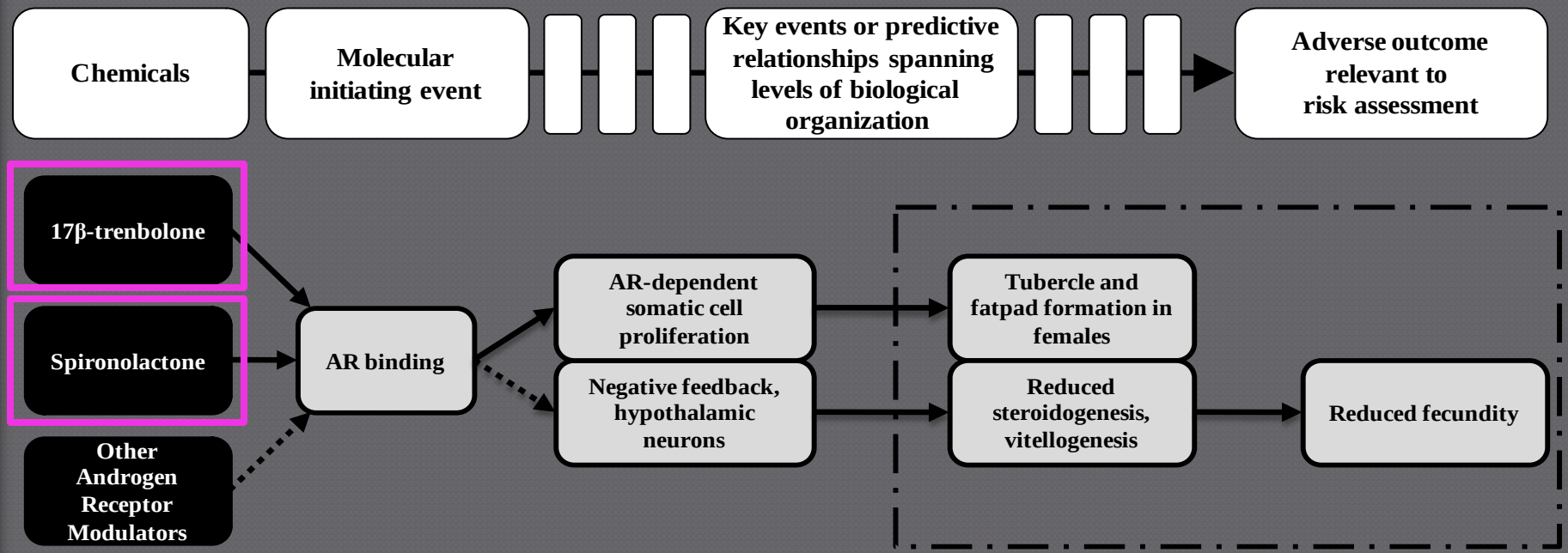
Automate the process of querying a/multiple specific residue position/s

- Use knowledge for species susceptibility predictions
- Hypotheses generation

Knowledge of molecular target conservation complements the AOP construct



Established AOP



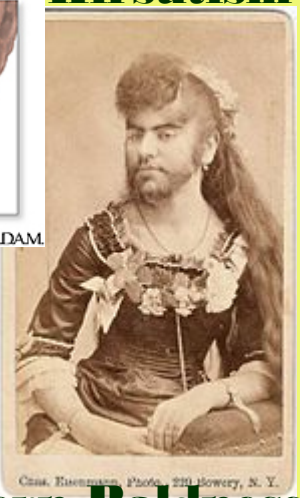
Androgen Receptor

Conserved Molecular Target

Humans

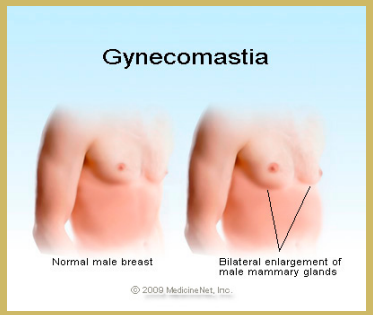
Acne

Hirsutism



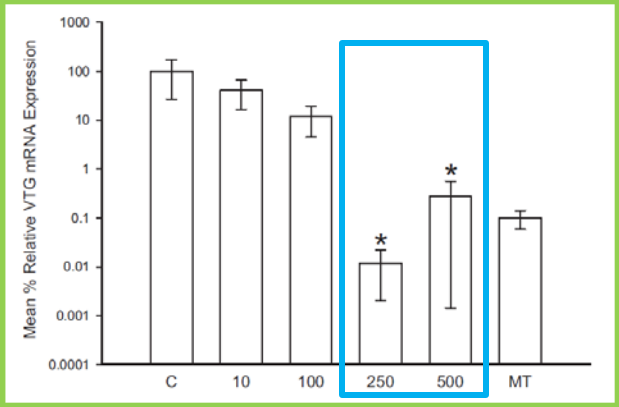
Female Pattern Baldness

Side Effect

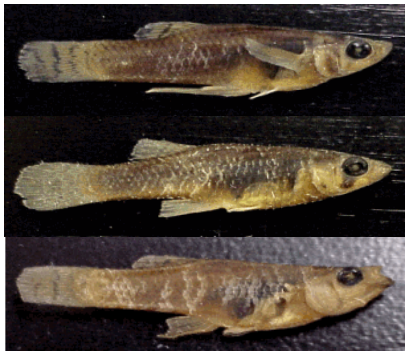


Western Mosquitofish

Decrease Vitellogenin mRNA



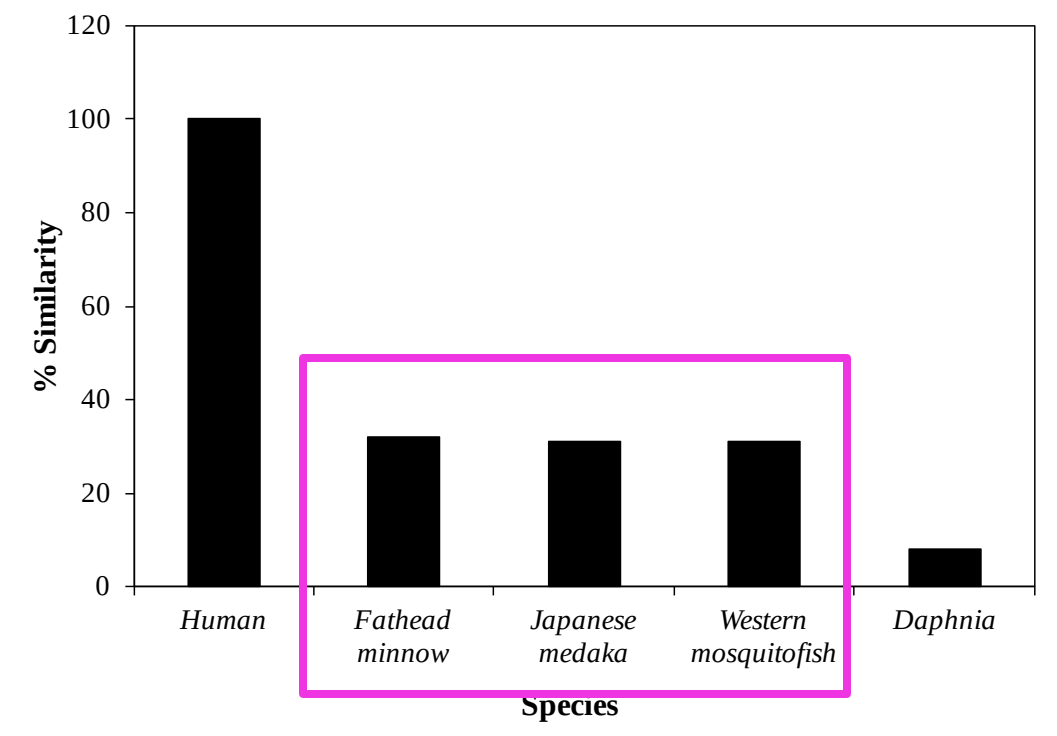
Masculinization



Environmental Toxicology and Chemistry, Volume 30, Issue 6, pages 1376-1382, 2011

Observed Adverse Effect in Aquatic Species

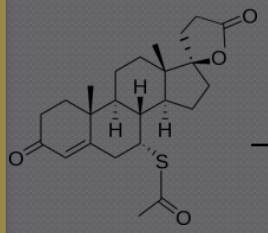
Molecular Target Conservation: Species Susceptibility Prediction



- Hypothesis:
  - Based on homology to the human androgen receptor
    - Small fish likely to be susceptible
    - Invertebrates unlikely to be susceptible

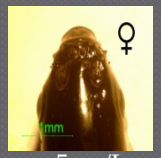
AR Activation in Small Fish

Spironolactone



Androgen Receptor Activation

Tubercle Formation

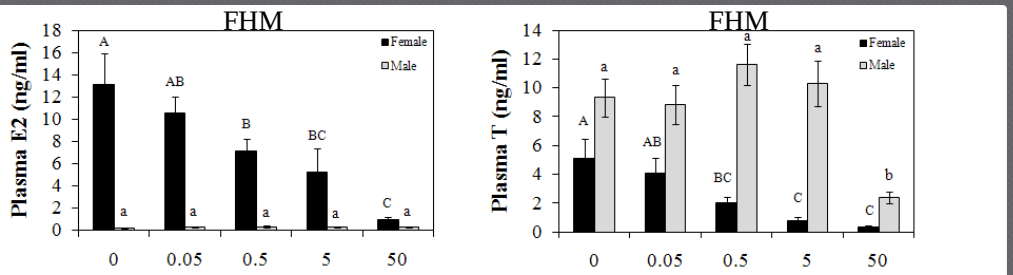


5 µg/L

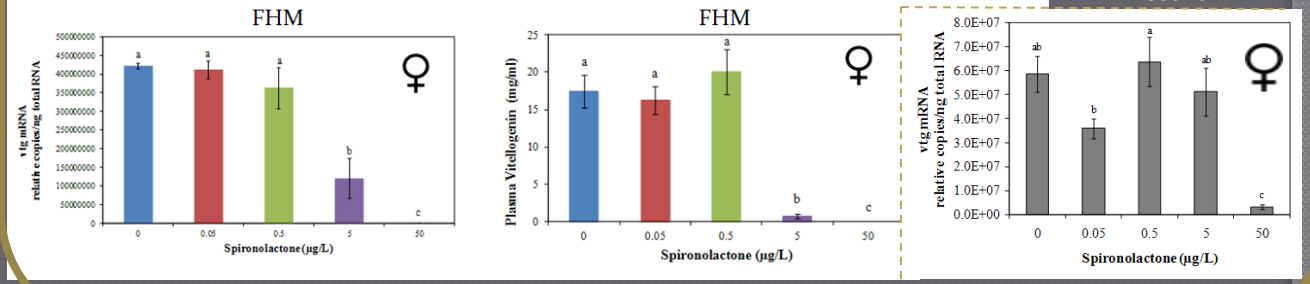
Papillary Process Formation

5 µg/L

Reduced Steroidogenesis



Reduced vtg expression/synthesis

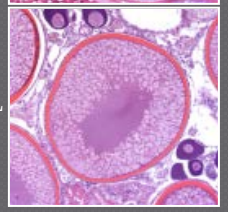


FHM

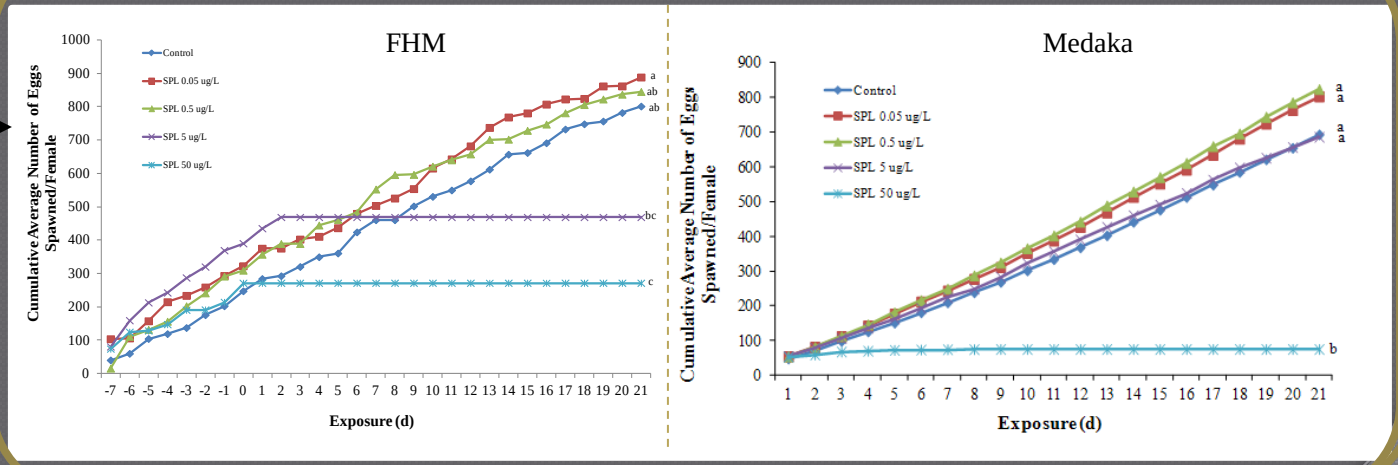
Reduced deposition of vtg in oocyte

Control

50 µg/L



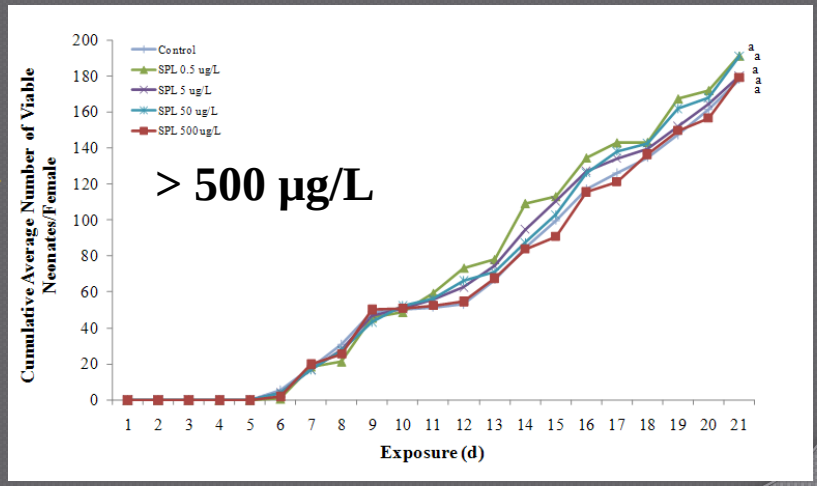
Reduced Fecundity



Small fish species sensitive

Secondary Sex **0.5 µg/L** → **5 µg/L**

Reproduction **5 µg/L** → **50 µg/L**



Key Findings

- Evidence that MIE conservation and knowledge of sequence similarity across species can be useful for defining taxonomic domain of applicability for AOP
- Consistency of AOP across small fish species
  - Predictable adverse effects
- Further examination of functional domains, individual residue queries, and protein structure between species may enhance predictive utility

Progress

- Manuscripts:
  - Molecular Target Sequence Similarity as a Basis for Species Extrapolation to Assess the Ecological Risk of Chemicals with Known Modes of Action. Submitted
  - Cross species sensitivity to a novel androgen agonist of environmental concern, spironolactone. ET&C 2013. Published online
- Sequence Similarity Tool:
  - Transferrable tool
    - Automated with capabilities described throughout talk
      - Primary protein sequence, conserved domains, and individual residue query capabilities

Current Analyses: Sequence Similarity Tool

- Honey bee sensitivity- focus on pesticide MIE nAChR
- Acetylcholine esterase
- Avian AOP for Conazoles – CYP51
- 27 Pharmaceutical case study
- Estrogen receptor-sequence similarity vs. binding of in vitro recombinant ER across species

Future Directions: Sequence Similarity Tool

- ***Test*** *the predictive utility:*
  - Establishing quantitative relationships between target similarity and initiation of responses using comparative in vitro systems
  - Confirmation that in silico predictions correspond with in vivo responses
  - Use tool to provide cross-species insights as to ADME to support PBPK modeling
- *Improve the tool:* Develop automated computational methods for assessing tertiary structure across species

Acknowledgments

- USEPA (NHEERL)– Duluth, MN
  - C. LaLone, G. Ankley, D. Villeneuve, C. Russom, J. Berninger, J. Tietge, J. Cavallin, M. Severson
- USEPA (NCEA) – RTP, NC
  - Lyle Burgoon
- Computer Science Corporation
  - H. Helgen, D. Lane, S. Watala
- USEPA-RTP, NC
  - V. Wilson, E. Gray, P. Hartig
- USEPA (NERL)– Athens, GA
  - T. Collette, D. Ekman
- USEPA-Chicago, IL (GLNPO)
  - T. Smith