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

U.S. EPA
Carlie LaLone
8/22/13
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.
“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

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

Taxonomic applicability domain
- Inherency tools
- Molecular screening data
- AOPs (to some extent)
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
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


Which proteins are conserved and to what degree?
Example Supporting Sequence-based Intrinsic Susceptibility Predictions

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

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

- 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
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

All Species in NCBI Protein 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

<table>
<thead>
<tr>
<th>Species</th>
<th>NCBI Protein Accession</th>
<th>Protein Name</th>
<th>Bit score</th>
<th>E-value</th>
<th>% Similarity</th>
<th>Number of Conserved Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Aedes aegypti</em></td>
<td>ACB370241</td>
<td>voltage-gated para-like sodium channel</td>
<td>4464</td>
<td>0</td>
<td>100.0</td>
<td>4</td>
</tr>
<tr>
<td><em>Homo sapiens</em></td>
<td>NP_066287.2</td>
<td>sodium channel protein type 2 subunit alpha isoform 1</td>
<td>1639</td>
<td>0</td>
<td>36.7</td>
<td>3</td>
</tr>
<tr>
<td><em>Drosophila erecta</em></td>
<td>EDV46862.1</td>
<td>voltage-gated para-like sodium channel</td>
<td>3431</td>
<td>0</td>
<td>76.9</td>
<td>3</td>
</tr>
</tbody>
</table>

**Most similar to target species**

**Least similar to target species**
Ortholog identification
- Ortholog = A sequence diverged after a speciation event

Reciprocal best hit BLAST
Set susceptibility cut-off
  • Protein with lowest percent similarity that was identified as an ortholog

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

Begin with:
- **Chemical with Known MOA**
  - **Drugbank VSDB**
  - **Molecular Target**
  - **NCBI Accession (Based on Target Species)**

**a.**
- **BLASTp: Sequence Alignments**
  - Percent Similarity (Highest to Lowest)
  - E-value

**b.**
- **Accessions for homolog candidates**
- **Conserved Domain Search**
  - Number of Conserved Domains in Common

**c.**
- **Homolog Candidate List**
  - **Reciprocal Best Hit BLAST**
  - **Ortholog Candidates Identified**
  - **Set % Similarity Cut-off**

**d.**
- **Is there ≥1 conserved domain?**
  - **Yes**
  - **Molecular Target Similarity Predicted Intrinsic Susceptibility**
  - **No**
  - **Discard Homolog Candidate from Analysis**

**e.**
- **Final Output**
  - **Is E-value ≤0.01?**
    - **Yes**
    - **Final Output**
    - **No**
    - **Discard Homolog Candidate from Analysis**
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.
Molecular Target: Voltage-gated sodium channel

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

Species (Exclude Resistant/Susceptible Strains, Pesticide exposed)

Endpoint (Desired Acute Response in Target Species = Death)

**Chemical (Analytical measurements, Chemical purity)**

<table>
<thead>
<tr>
<th>Chemical</th>
<th>CAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Permethrin</td>
<td>52645-53-1</td>
</tr>
<tr>
<td>(+)-cis-Permethrin</td>
<td>61949-76-6</td>
</tr>
<tr>
<td>(+)-trans-Permethrin</td>
<td>61949-77-7</td>
</tr>
<tr>
<td>(+)-cis-Permethrin</td>
<td>54774-45-7</td>
</tr>
<tr>
<td>(-)-cis Permethrin</td>
<td>54774-46-8</td>
</tr>
<tr>
<td>(+)-trans Permethrin</td>
<td>51877-74-8</td>
</tr>
<tr>
<td>(-)-trans Permethrin</td>
<td>54774-47-9</td>
</tr>
</tbody>
</table>

**Species (Exclude Resistant/Susceptible Strains, Pesticide exposed)**

- All Aquatic Taxa (Identify Life Stage)
- EC50/LC50
- Mortality / Survivability
  - Acute durations (48 hr invertebrate/96 hr vertebrate)
  - Water Exposure (No feeding or diet)

**Endpoint (Desired Acute Response in Target Species = Death)**

Compare to Homology Data
Evidence that quantitative measures of sequence similarity can be used to predict susceptibility to insecticides. Aqueous solutions of insecticides are likely to be relevant molecular initiating events (MIEs) for aquatic organisms. VGSC alterations lead to disrupted neuronal networks and pathways, altered neuronal firing rates, and clinical signs, including dead insects and altered non-target insect populations.
Taking a Closer Look

Delving into sequence comparisons
Logic behind development

Sequence Similarity Metrics

Primary Amino Acid Sequence
- Percent Similarity
- Ortholog Identification

Conserved Domains
- Functional units

Individual Residue Differences
- Case by case basis
- Protein structure

Low

Knowledge of individual residues relevant for sensitivity prediction
- Aryl Hydrocarbon Receptor
- Voltage-gated Sodium Channel

High

Level of Sequence Specificity & Cost

Data Availability

Similarity between conserved domains
- Ligand Binding Domain
- DNA Binding Domain

Low
**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

<table>
<thead>
<tr>
<th>Organism Class</th>
<th>Non-Target Species</th>
<th>Protein Name</th>
<th>Calculated % similarity</th>
<th>Common Domains</th>
<th>Conserved domain IDs</th>
<th>Domain type</th>
<th>Sequence coverage of conserved domains</th>
<th>Conserved Domain % Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insecta</td>
<td><em>Apis mellifera</em></td>
<td>paralytic</td>
<td>73</td>
<td>3</td>
<td>pfam11933 pfam08016 pfam00520</td>
<td>Unknown function Polycystin cation channel Ion transport protein</td>
<td>1-222, same 217-423, same 1-194, same</td>
<td>61 92 93</td>
</tr>
</tbody>
</table>
Knowledge of individual residues

Resistance and sensitivity: Single amino acid residue substitutions
- Targeted mutagenesis
- Voltage clamp techniques

Mammals

K  I

Increases sensitivity 100-fold
Targeted sequence comparisons

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

Chemicals

Molecular initiating event

Key events or predictive relationships spanning levels of biological organization

Adverse outcome relevant to risk assessment

Established mechanistic linkage with quantitative or semi-quantitative data

Plausible linkage with limited data

Predicted apical outcomes following AR activation in small fish

- 17β-trenbolone
- Spironolactone
- Other Androgen Receptor Modulators

AR binding

AR-dependent somatic cell proliferation

Negative feedback, hypothalamic neurons

Tubercle and fatpad formation in females

Reduced steroidogenesis, vitellogenesis

Reduced fecundity

- Molecular event or predictive relationship spanning levels of biological organization relevant to risk assessment

- Chemicals

- Adverse outcome relevant to risk assessment

- Predicted apical outcomes following AR activation in small fish
Androgen Receptor
Conserved Molecular Target

Female Pattern Baldness

Western Mosquitofish

Decrease Vitellogenin mRNA

Masculinization

Observed Adverse Effect in Aquatic Species
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

Plasma T2 (ng/ml)
- Control
- SPL 0.05 µg/L
- SPL 0.5 µg/L
- SPL 5 µg/L
- SPL 50 µg/L

Reduced Steroidogenesis

Plasma T (ng/ml)
- Control
- SPL 0.05 µg/L
- SPL 0.5 µg/L
- SPL 5 µg/L
- SPL 50 µg/L

Reduced vtg expression/synthesis

FHM Medaka

Reduced Fecundity

FHM
- 50 µg/L
- Control

Reduced deposition of vtg in oocyte

FHM Medaka

5 µg/L

Medaka

50 µg/L
Small fish species sensitive

Secondary Sex
Reproduction

0.5 µg/L  5 µg/L  50 µg/L

> 500 µg/L
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
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
- 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