The Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) Tool: Catalyzing a Change in Species Extrapolation

Carlie A. LaLone, Ph.D.
Research Bioinformaticist
Chemicals make up the world around us – necessary for our modern society
Toxicity Testing to Understand Chemical Safety

• Regulatory decision-making

US EPA Examples:
- Clean Air Act
- Clean Water Act
- Resource Recovery Act
- Endangered Species Act
- Food Quality Protection Act
- Endocrine Disruptor Screening Program
- Federal Insecticide, Fungicide, and Rodenticide Act
- Frank R. Launtenberg Chemical Safety for the 21st Century Act
- Comprehensive Environmental Response, Compensation, and Liability Act
- Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses
easy maintenance and good breeding capabilities

short lifespans and rapid life cycles

ability to control diet and surroundings

requires least space and time-consuming care

cheap and readily available
Species Extrapolation

What is it?
- Using existing knowledge about one species to estimate, predict, project, or infer the effect, impact, or trajectory of another species
  - For chemical safety typically dealing with toxicity

Why is it important:
- Limited or no toxicological data for the animal or plant species of interest – reliance on surrogate (model organisms)
  - Impractical to generate new data for all species
- Testing resources are limited
  - International interest to reduce animal use
  - Ever-increasing demand to evaluate more chemicals in a timely and sometimes expedited manner
- Sensitivity of species must be estimated based on scientifically-sound methods of cross-species extrapolation
  - Immense diversity of species in the wild
  - Important challenge for species listed under the Endangered Species Act
Sensitivity to Chemical Perturbation

**TOXICOKINETICS**
- Absorption
- Distribution
- Metabolism
- Elimination

**TOXICODYNAMICS**

- Cross Species Extrapolation
- Bioinformatics

Simplify Complexity
Considering chemical sensitivity?

Factors that make a species sensitive
- Exposure
- Dose
- ADME
  - Target receptor availability
- Life stage
- Life history
- etc.
- etc.

Simple question to address:
Is the known chemical target available in a species for a chemical to act upon?
Yes or No
Likely susceptible or Not likely susceptible (at least through the known mechanism)
Predictive Approaches

Start simple

**Sequence**

```
MTMTLHTKASSMALLHDIQ6NELEPLVRPLKLIEPLPLGE
VYLDSSKPVVRPEAARFNEAAABAOAQQYGGOTLYG
PGSEAAAGSNGLGGFFPLNLVSPSLMLHLHPPQIPSFLLG
PHGQDYPVYLYNESPSTYRTVEAFPGPAFPRPSIDRRRQGR
ERLASTDNKGMAMESEKETYCVCNDYASAYYAYGYV2SCEGC
GAFRRGSGHNYDMPATNGCTIKNKIRKSCAQCIKRL
KYEVMGMMGKGNVRGDRGRLKHLHRQDQDGGRGGEVG
SAGDMRAANLWPSPLMKRSKNSLALSLTDQMVSALLAPAEPILYSEYQTFRFPEASMGLTTNADRELWLMINWAKV
PGFVLTLHCQ2HLLECAWLEILMLGMLWWR5MENHPGLLFLP
PAMELDRNNGCGVCGAMYEFDMLATSSRFMVMYNLGEEF
VCLSKILLNSGYTYFLSSLTSKSEKOHIIHRVLDKIDTLHLM
```

**Structure**

**Function**

Bioinformatics
Chemical-Protein Interaction:

Natural Ligands/Chemicals

Critical amino acids

Functional domain (e.g. Ligand binding domain)

Similarity across species at the molecular level
Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)

https://seqapass.epa.gov/seqapass/
SeqAPASS Predicts Likelihood of Similar Susceptibility based on Sequence Conservation:

- The protein is conserved
- The protein is NOT conserved

Line(s) of evidence indicate:
- The protein is conserved
- The protein is NOT conserved
SeqAPASS DOES NOT predict the degree of sensitivity/susceptibility:

Factors that make a species sensitive
• Exposure
• Dose
• ADME
• Target receptor availability
• Life stage
• Life history
• etc.
• etc.
Strengths of SeqAPASS

New tools and technologies
• Improved sequencing technologies
• Large databases of sequence data

• Publicly available to all
• Lines of evidence for conservation for 100s-1000s of species rapidly
• Takes advantage of well-established tools and databases
• Streamlined, consistent, transparent, and published methods
  • Case examples to demonstrate applications
• Guides users to appropriate input
• Evolves as bioinformatics approaches become more user friendly
  • Smart automation or semi-automation

NCBI: 197,232,209 Proteins representing 108,257 Organisms
Applications of Bioinformatics: Case Studies

• Extrapolate adverse outcome pathway knowledge across species
  • Define the taxonomic relevance: Apis vs Non-Apis bees

• Extrapolate high throughput screening data
  • Chemicals that target human estrogen receptor alpha, androgen receptor, steroidogenic enzymes, thyroid axis proteins
  • All ToxCast Assay targets

• Predict relative intrinsic susceptibility
  • Pesticides
  • Endangered Species Act
  • Derivation of Aquatic Life Criteria

• Predict chemical bioaccumulation across species
  • Chemicals of concern: PFAS

• Generate research hypotheses Strobilurin fungicides

• Prioritization strategies Pharmaceuticals
Where do we go from here?

• Connect bioinformatics with general informatics including systematic methods
  • ECOTOX Knowledgebase
  • Systematic literature review for WOE
• Advance structural evaluations (computing power and storage)
  • Move from docking to virtual screening
• Specific laboratory studies to support bioinformatics
  • Site-directed mutagenesis
International Consortium to Advance Cross Species Extrapolation in Regulation ICACSER

1. Define the taxonomic domain of applicability
2. Define the global regulatory landscape/need
3. Develop a bioinformatics toolbox
4. Communicate a shared scientific vision

Interested in Learning more or Joining: Contact LaLone.Carlie@epa.gov or Geoff.Hodges@unilever.com

Steering Committee:
- Carlie LaLone (US EPA)
- Geoff Hodges (Unilever)
- Nil Basu (McGill U)
- Steve Edwards (RTI)
- Fiona Sewell (NC3Rs)
- Michelle Embry (HESI)
- Patience Browne (OECD)
Other ORD Research: High throughput transcriptomics for Ecotoxicology
Problem Statement:
Tens of thousands of chemicals are currently in use and hundreds more are introduced to the market every year. Only a small fraction has been thoroughly evaluated for potential risks to human health and the environment.

- Need for high-throughput assays to evaluate chemical safety.
- Provide adequate coverage of relevant pathways through which chemicals can cause toxicity.
High-Throughput Transcriptomics Platform for Screening Environmental Chemicals


Whole human transcriptome

- Number of mammalian studies have shown short-term transcriptomics-based PODs are predictive of apical potency.
- Generally, within ½ log.
- Health protective points of departure.

Toxicological Sciences, Volume 181, Issue 1, May 2021, Pages 68–89, https://doi.org/10.1093/toxsci/kfab009
Ecotoxicology Perspective

- Humans are just a tiny fraction of the biological diversity we are charged to protect.
- Many genes/pathways are conserved
- Unique physiology in other kingdoms, phyla, classes…
- How do we assure those pathways are covered?
High throughput assays for three major trophic levels of aquatic ecosystems

- Primary producers (e.g., algae)
- Primary consumers (e.g., zooplankton, aquatic inverts)
- Secondary consumers (e.g., fish)

Commonly used for GHS classification and labeling of chemicals for environmental hazard

Aquatic organisms highly vulnerable to exposure
Incorporating transcriptomics as assessment endpoint

Load organisms → 24 h exposure → Homogenate/extract → RNA → Whole transcriptome

Phenotypic anchoring → Curve fitting → Distribution of BMDs → BMD Express

Point of departure → [tPOD] Rough estimate per chemical ~$5K for one species

10th centile BMD

Rough estimate per chemical ~$5K for one species
Evaluating the approach

[tPOD] ≤ [Most sensitive chronic endpoint]

[tPOD] > [Most sensitive chronic endpoint]

[tPOD] <<<< [Most sensitive chronic endpoint]

Approach is being explored with cell-lines for eco species as well.
1. Generate transcriptomic PODs for ≈ 20 chemicals
   • Initial focus on fathead minnow
   • Parallel assays with additional taxa – for future analyses

2. Compare tPODs with available acute and chronic toxic toxicity data

3. Compare tPODs with in vitro-derived PODs
Acknowledgements

**U.S. EPA, ORD**
- Dan Villeneuve
- Kevin Flynn
- Adam Biales
- Donovan Blatz (ORISE)
- Sara Vliet (ORISE)
- Sally Mayasich (ORISE)
- Marissa Jensen (Univ. Minnesota Duluth)

**GDIT**
- Thomas Transue
- Cody Simmons
- Audrey Wilkinson

**University of Pittsburgh**
- Carla Ng
- Weixiao Cheng

Newly Released: SeqAPASS v6.0

LaLone.Carlie@epa.gov
https://seqapass.epa.gov/seqapass/