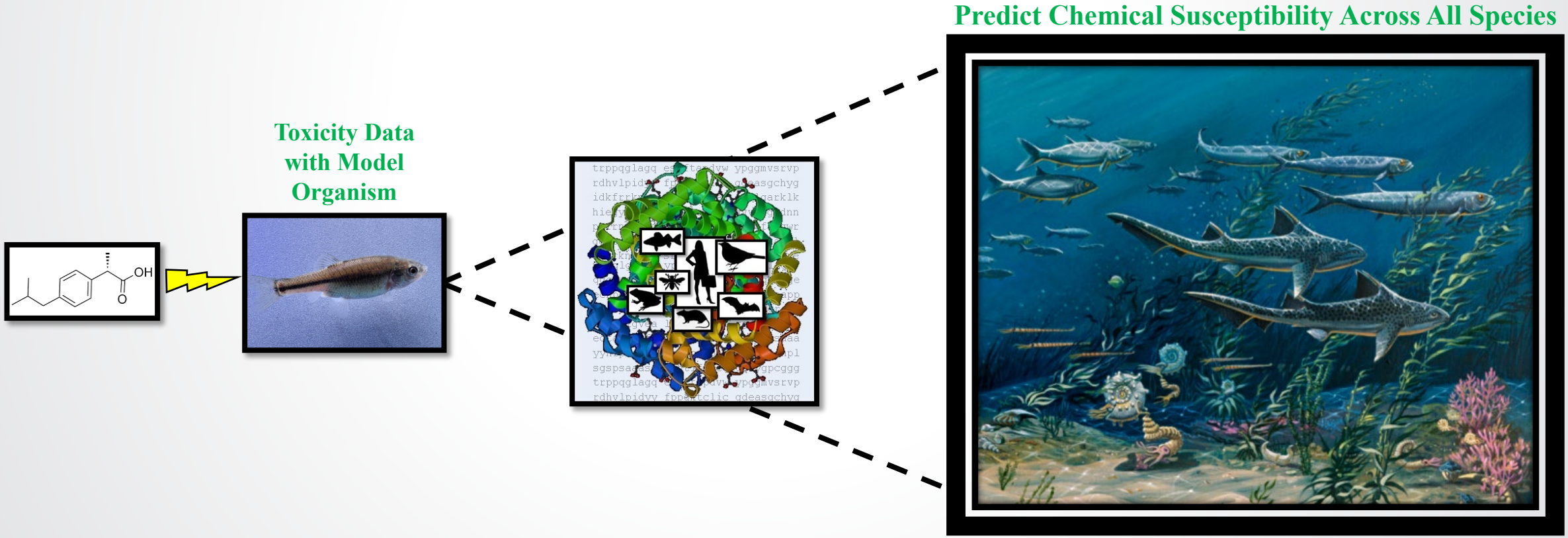
Species Sensitivities: Evaluating Conservation Using the SeqAPASS Tool

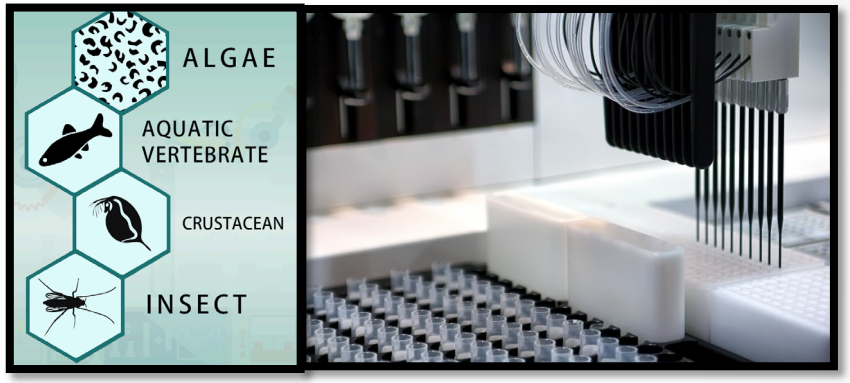
Carlie A. LaLone, USEPA, Great Lakes Toxicology and Ecology Division, Duluth, MN

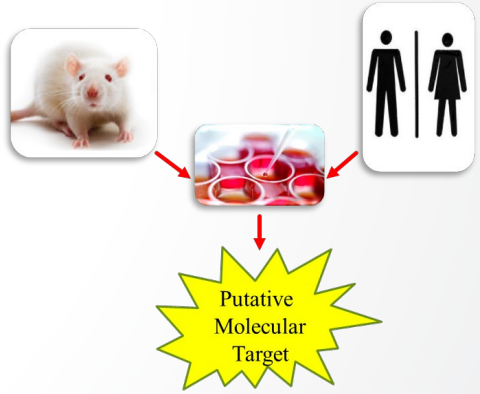


*Content does not necessarily reflect US Environmental Protection Agency (USEPA) position or policy.

New Approach Methods for Regulatory Decision-Making

- An umbrella term
  - Systems biology
  - In silico and bioinformatics
  - Omics
  - High throughput screening
  - In chemico
  - In vitro

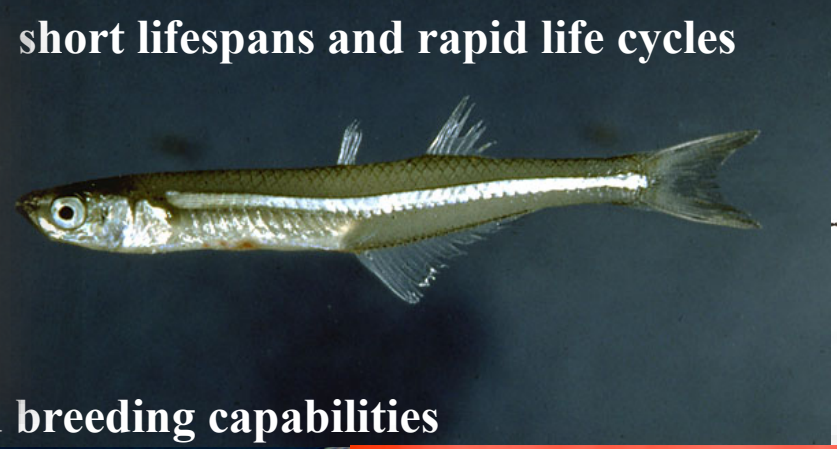




Molecular Initiating Event → Cellular Responses → Organ Responses → Organism Responses → Population Responses

cheap and readily available

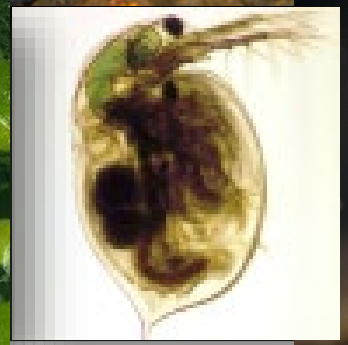
short lifespans and rapid life cycles



easy maintenance and good breeding capabilities

ability to control diet and surroundings

requires least space and time-consuming care



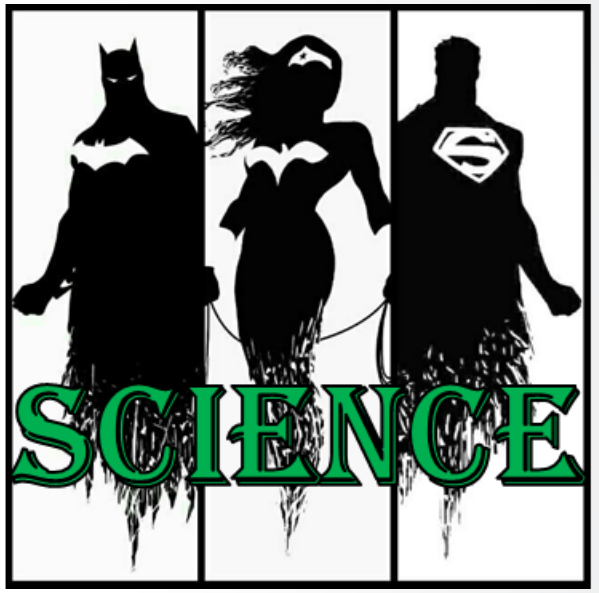
Species Extrapolation

What is it?

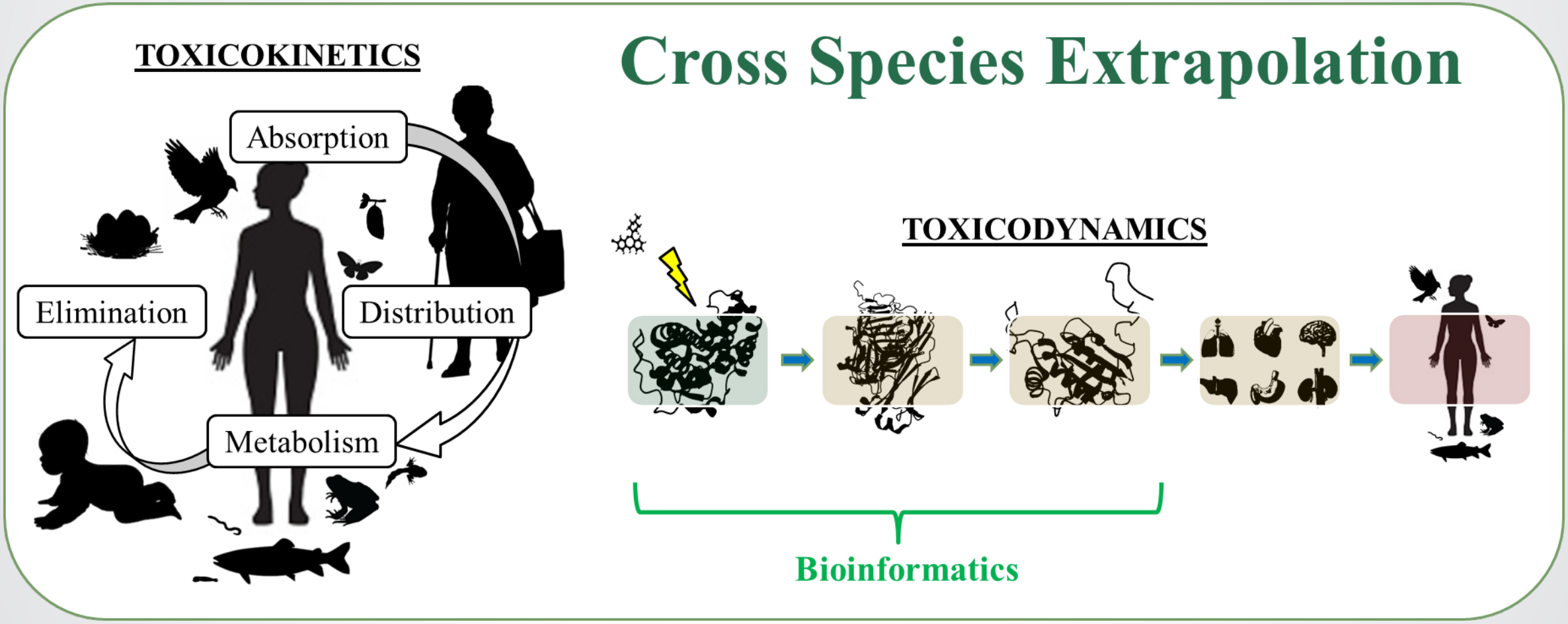
- Using existing knowledge about one species to estimate, predict, project, or infer the effect, impact, or trajectory of another species
  - For chemical safety typically dealing with toxicity

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



Simplify Complexity

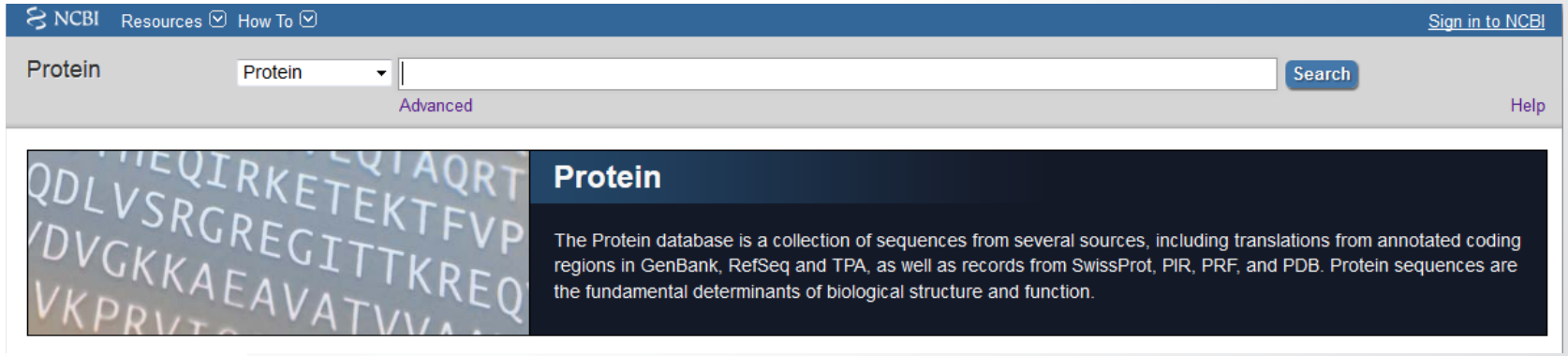
Where could we begin in understanding species similarities and differences?

Look for existing, expanding data that does not require the destruction of live organisms

Sequence and structural data: New tools and technologies have emerged

- Improved sequencing technologies
- Large databases of sequence data

NCBI: 210,703,648 Proteins representing 113,002 Organisms



Bioinformatics

- Combines mathematics, information science, and biology to answer biological questions

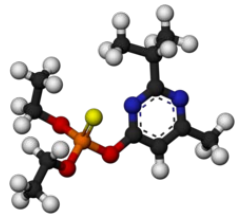
- Developing methodology and analysis tools to explore large volumes of biological data
  - Query, extract, store, organize, systematize, annotate, visualize, mine, and interpret complex data
    - Usually pertains to DNA and amino acid sequences

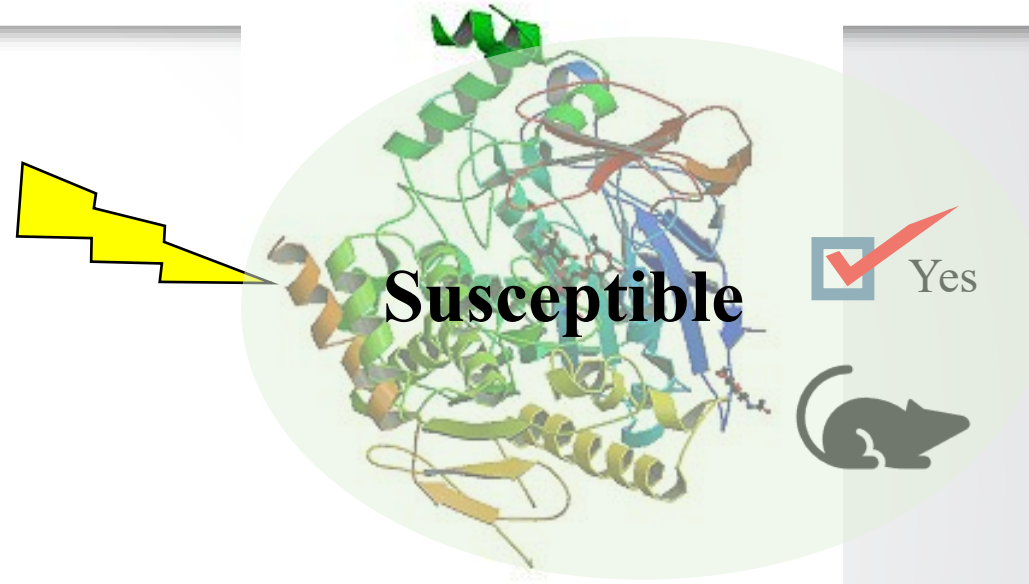
Let the computers do the work

Considering chemical sensitivity?

Factors that make a species sensitive

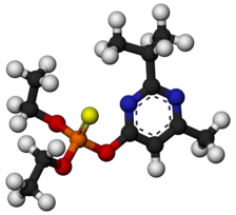
- Exposure
- Dose
- ADME
- Target receptor availability
- Life stage
- Life history
- etc.
- etc.





Susceptible

Yes



Not Susceptible

No

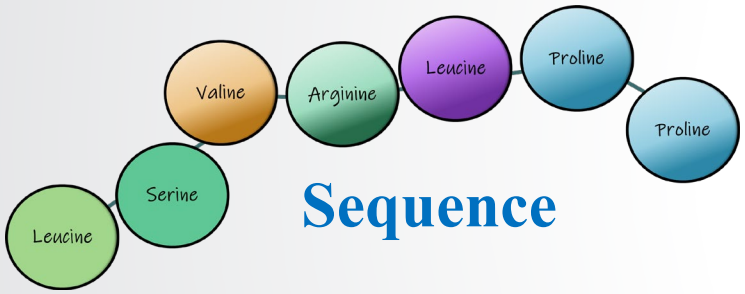
Simple question to address:

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)

Begin Simple and Advance as the Science Advances

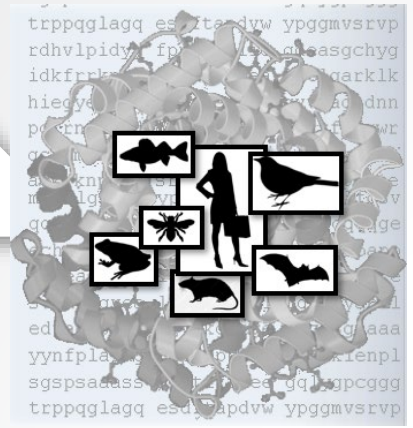


Sequence

Structure

Function

Consider sequence and structural attributes to understand protein conservation across species



TOXICOLOGICAL SCIENCES, 153(2), 2016, 228–245

OXFORD SOT Society of Toxicology www.toxsci.oxfordjournals.org

doi: 10.1093/toxsci/kfw119
Advance Access Publication Date: June 30, 2016
Research article

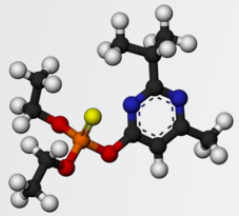
Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): A Web-Based Tool for Addressing the Challenges of Cross-Species Extrapolation of Chemical Toxicity

Carlie A. LaLone,*,[1] Daniel L. Villeneuve,* David Lyons,† Henry W. Helgen,‡ Serina L. Robinson,§,[2] Joseph A. Swintek,¶ Travis W. Saari,* and Gerald T. Ankley*

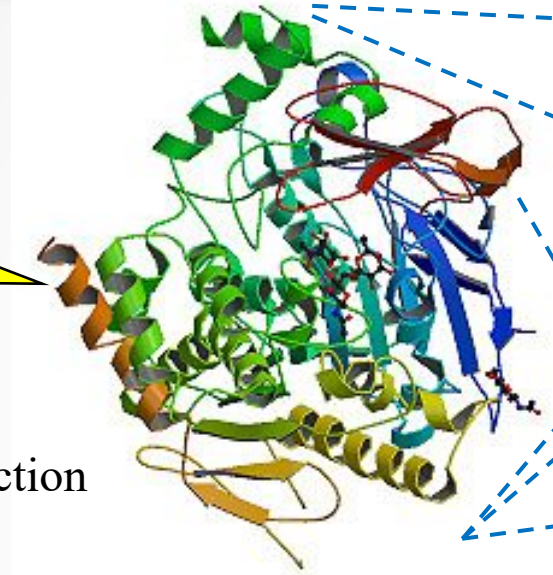
https://seqapass.epa.gov/seqapass/

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)

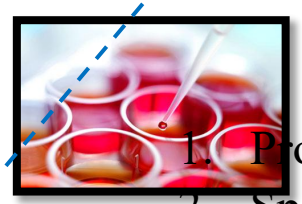
What information is required for a SeqAPASS query?



Chemical-Protein Interaction



Knowledge of a sensitive or targeted species

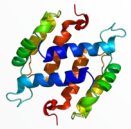


Knowledge of the model organism used in an *in vitro* assay

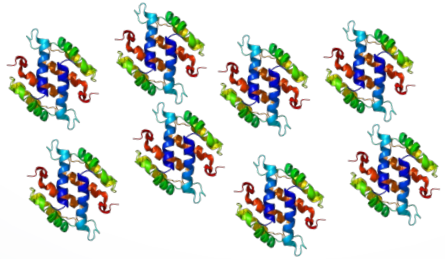
1. Protein
2. Species

Knowledge of the species for which the Key Event was developed

Chemical Molecular Target in Target Species



≅



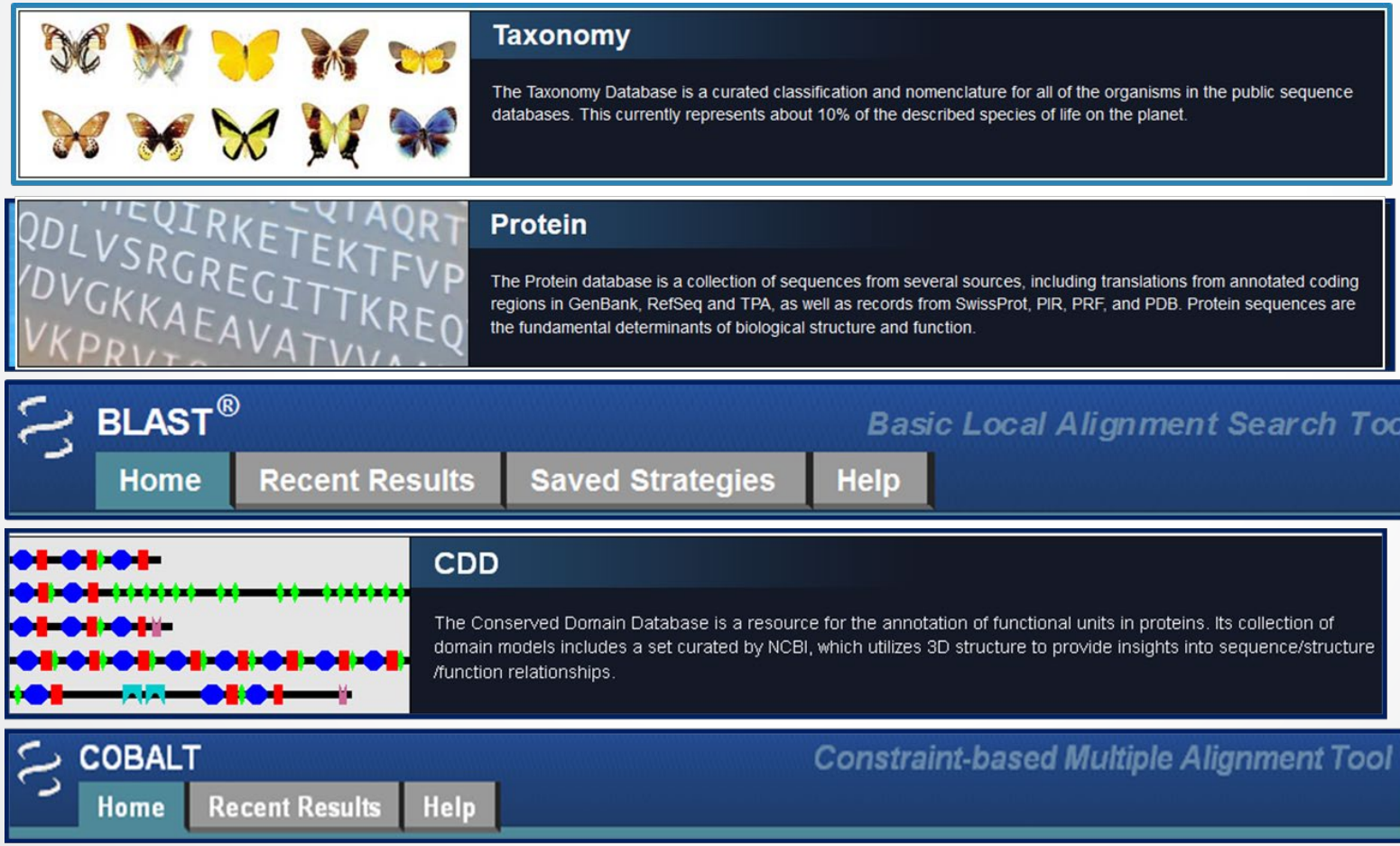
Compare to Millions of Proteins From Thousands of Species

Greater similarity = Greater likelihood that chemical can act on the protein

Line of Evidence: Predict Potential Chemical Susceptibility Across Species

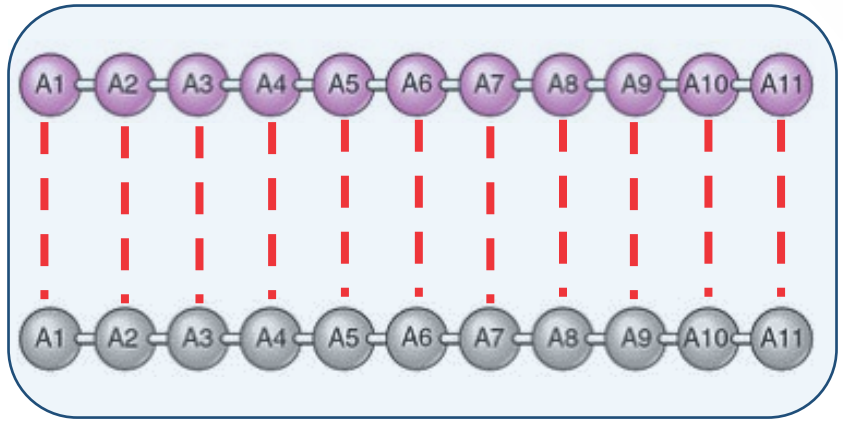
Available Databases and Tools

National Center for Biotechnology Information
Established in 1988: a division of National Library of Medicine at NIH



SeqAPASS: Level 1

Primary Amino Acid Sequence Alignment

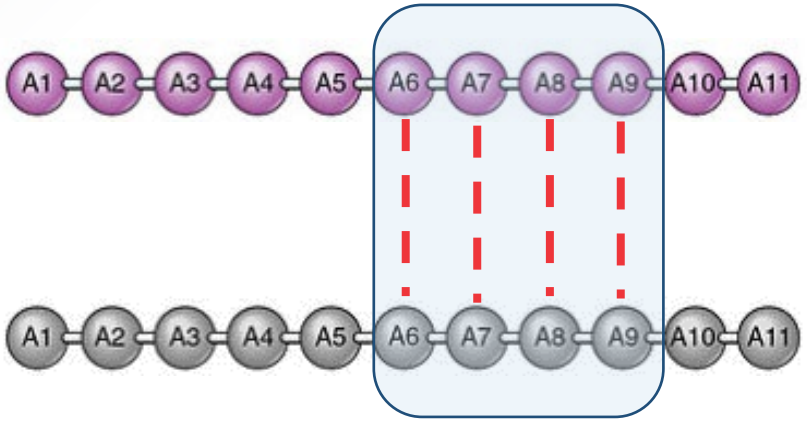


	Bit Score	Percent Similarity
Query	1241.9	100
Hit	1229.5	99.0
Hit	1223.0	98.5
Hit	1111.3	89.5
Hit	862.4	69.4

$$\text{Percent Similarity} = \frac{\text{Hit Bit Score}}{\text{Query Bit Score}} \times 100$$

SeqAPASS: Level 2

Functional Domain Sequence Alignment

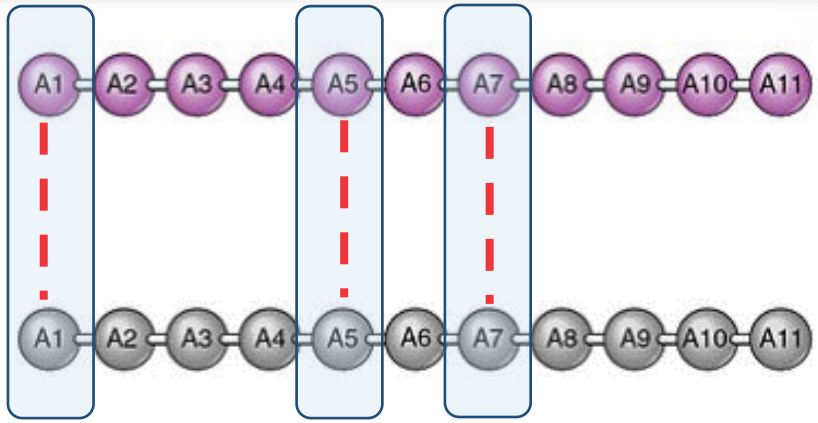


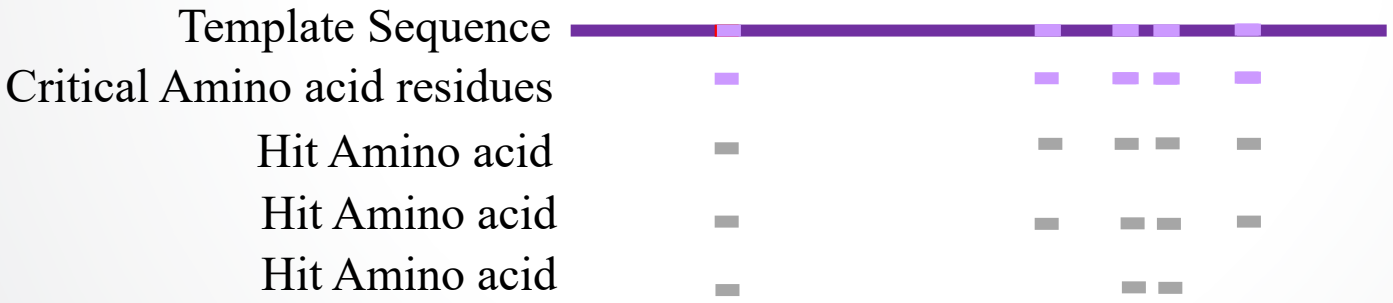
Query Sequence	Primary Amino Acid Sequence	Bit Score	Percent Similarity
	Query Sequence domain	482.6	100
	Hit domain	471.9	97.8
	Hit domain	303.5	62.9
	Hit domain	100.1	20.7

$$\text{Percent Similarity} = \frac{\text{Hit Bit Score}}{\text{Query Bit Score}} \times 100$$

SeqAPASS: Level 3

Critical Amino Acid Sequence Alignment



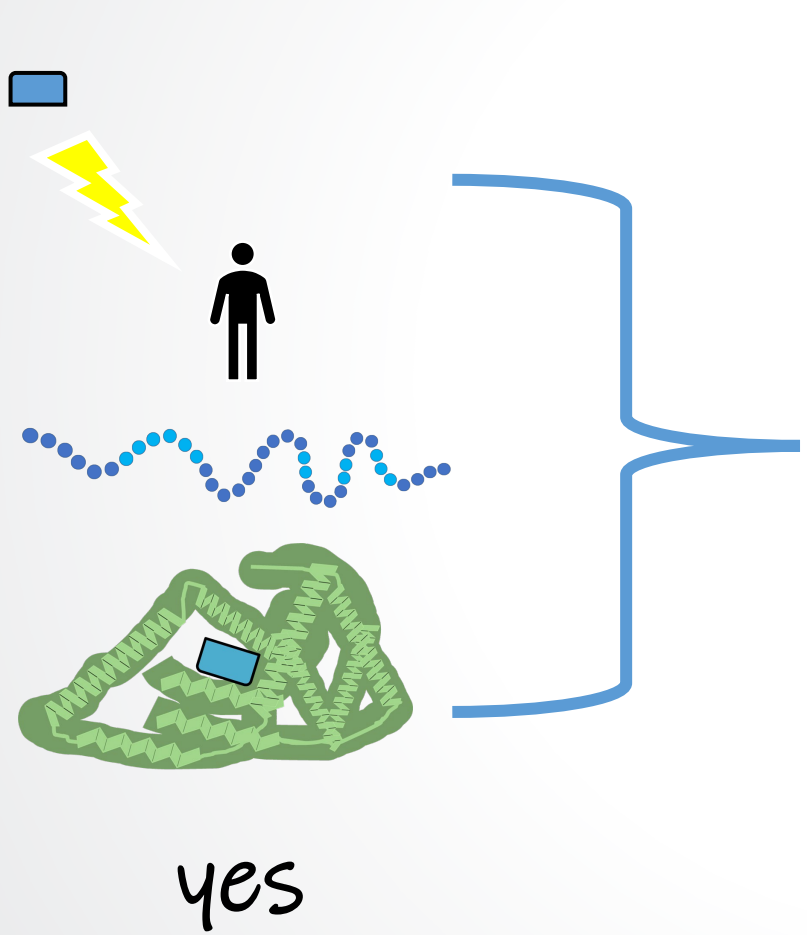


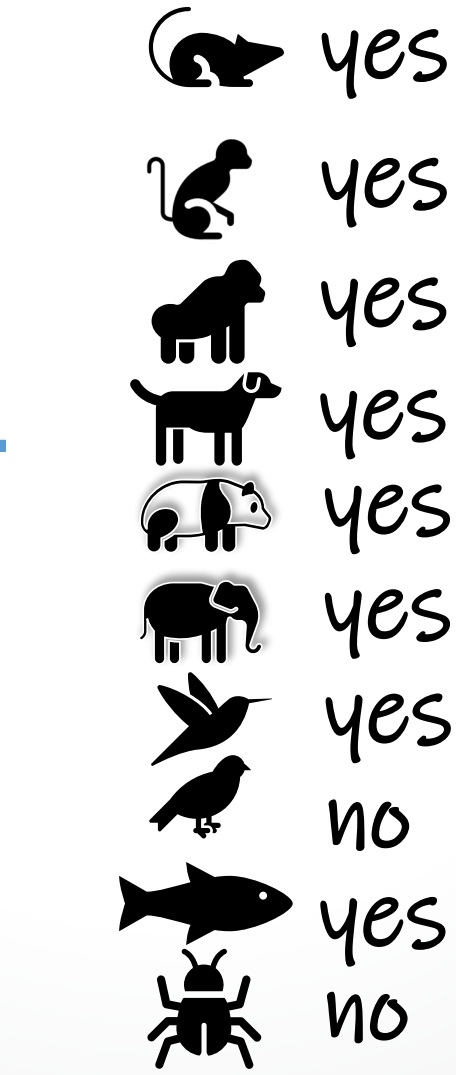
Rules to automate susceptibility prediction:

- Same side-chain classification as template
- MW as measure of size 30g/mol or less from template

Exact Match
Partial Match
Not a Match

SeqAPASS Predicts Likelihood of Similar Susceptibility based on Sequence Conservation:



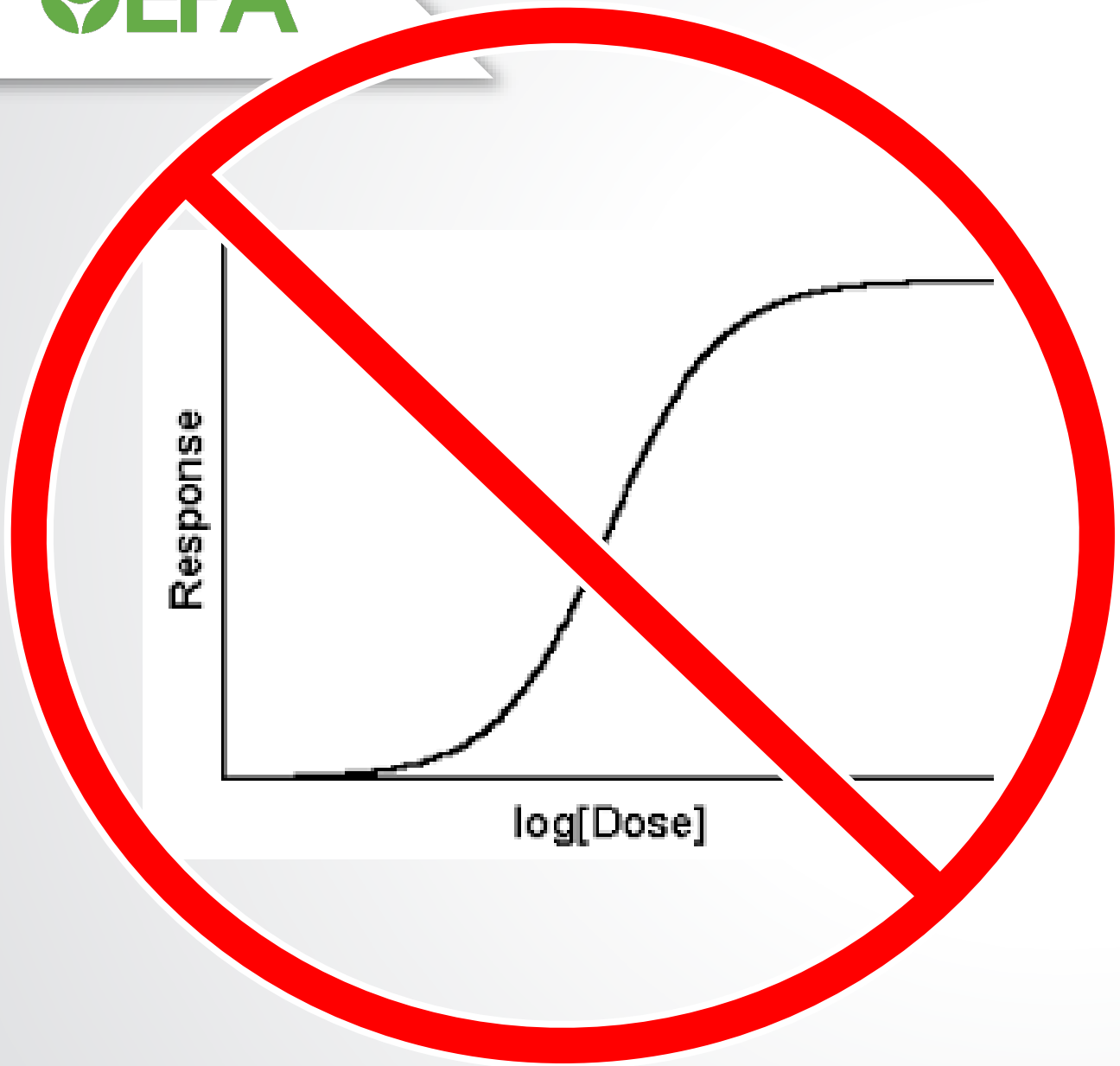


Line(s) of evidence indicate

- The protein is conserved
- The protein is NOT conserved

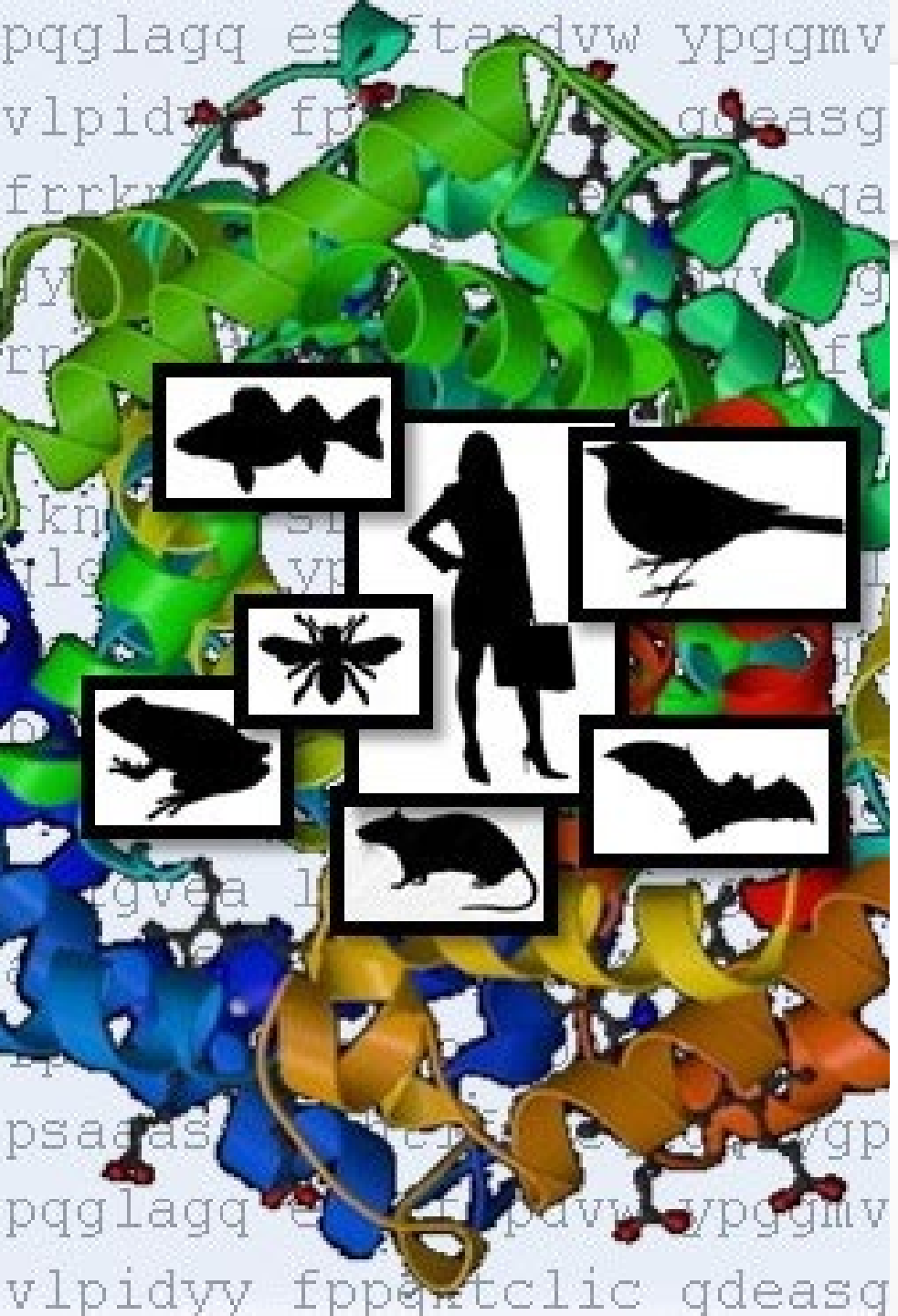
Predictions for 100s-1000s of species rapidly

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.

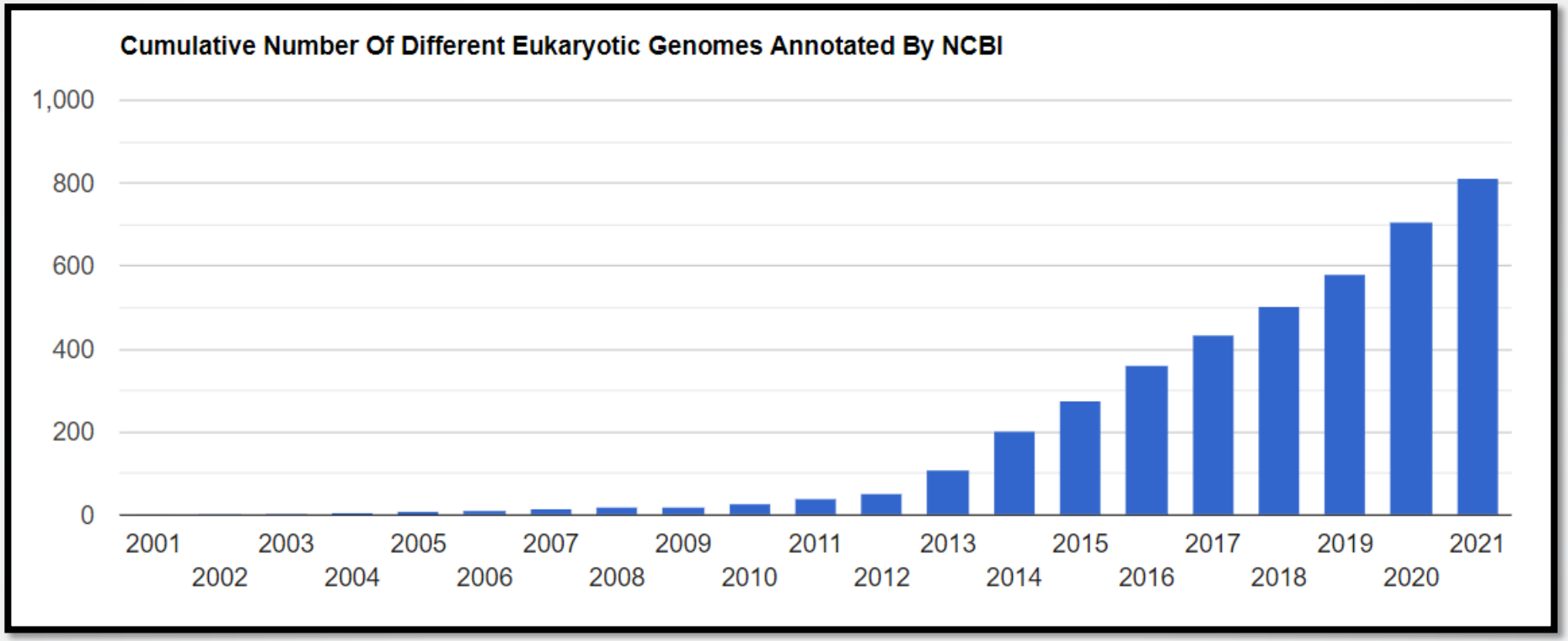


Strengths of SeqAPASS

- **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


Application of SeqAPASS
MIE
KE
KE
KE
AO

Challenges for Marine Species



2.2 million eukaryotic marine species
Sequencing and quality annotation needs

Acknowledgements

U.S. EPA, ORD

Marissa Jensen (University of Minnesota Duluth)
Sally Mayasich (University of Wisconsin)
Donovan Blatz (ORISE)
Monique Hazemi (ORISE)
Sara Vliet (US EPA)
Jon Doering (U of Lethbridge)
Colin Finnegan (Iowa State University)

GDIT

Thomas Transue
Cody Simmons
Audrey Wilkinson
Wilson Menendez

SeqAPASS v6.0 (Released Sept. 2021)

LaLone.Carlie@epa.gov

https://seqapass.epa.gov/seqapass/