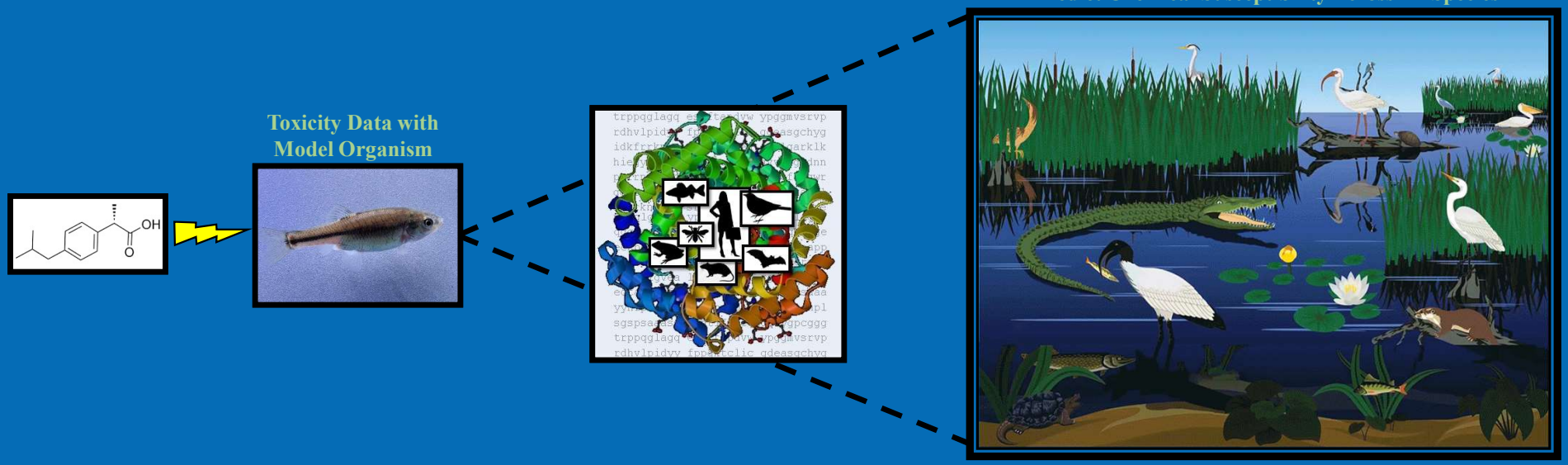
EPA
United States
Environmental Protection
Agency

A vision for the role of bioinformatics in species extrapolation for chemical safety

Carlie A. LaLone, Ph.D.
Research Bioinformaticist



Office of Research and Development
Center for Computational Toxicology and Exposure, Great Lakes Toxicology and Ecology Division

The views expressed in this presentation are those of the authors and do not necessarily reflect the views or policies of the US EPA

April 30th, 2021

Vision Overview Questions

- What species do we rely on for toxicity testing and why?
- Why consider predictive and computational approaches?
- How can bioinformatics help for chemical safety evaluations?
- What tools are available now and moving forward?
- How do we incorporate bioinformatics in decision making?

EPA
United States
Environmental Protection
Agency

EPA United States Environmental Protection Agency

Toxicity Testing to Understand Chemical Safety

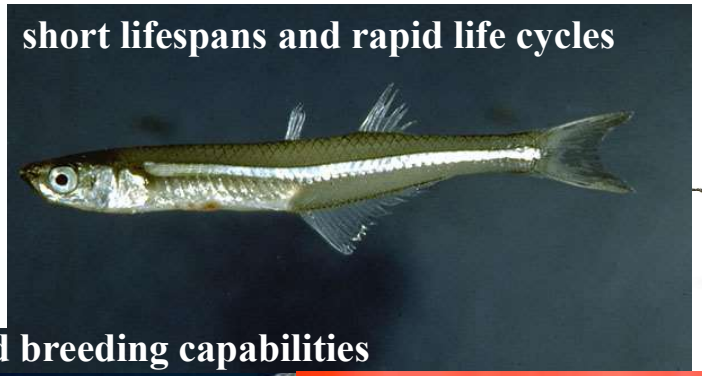
- Regulatory decision-making

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. Lautenberg 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

cheap and readily available

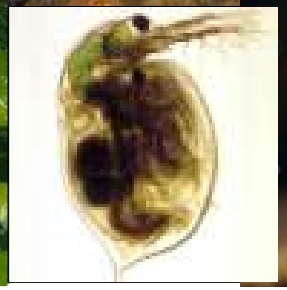
short lifespans and rapid life cycles



easy maintenance and good breeding capabilities

ability to control diet and surroundings

requires least space and time-consuming care



EPA United States Environmental Protection Agency

Species Extrapolation

What is it?

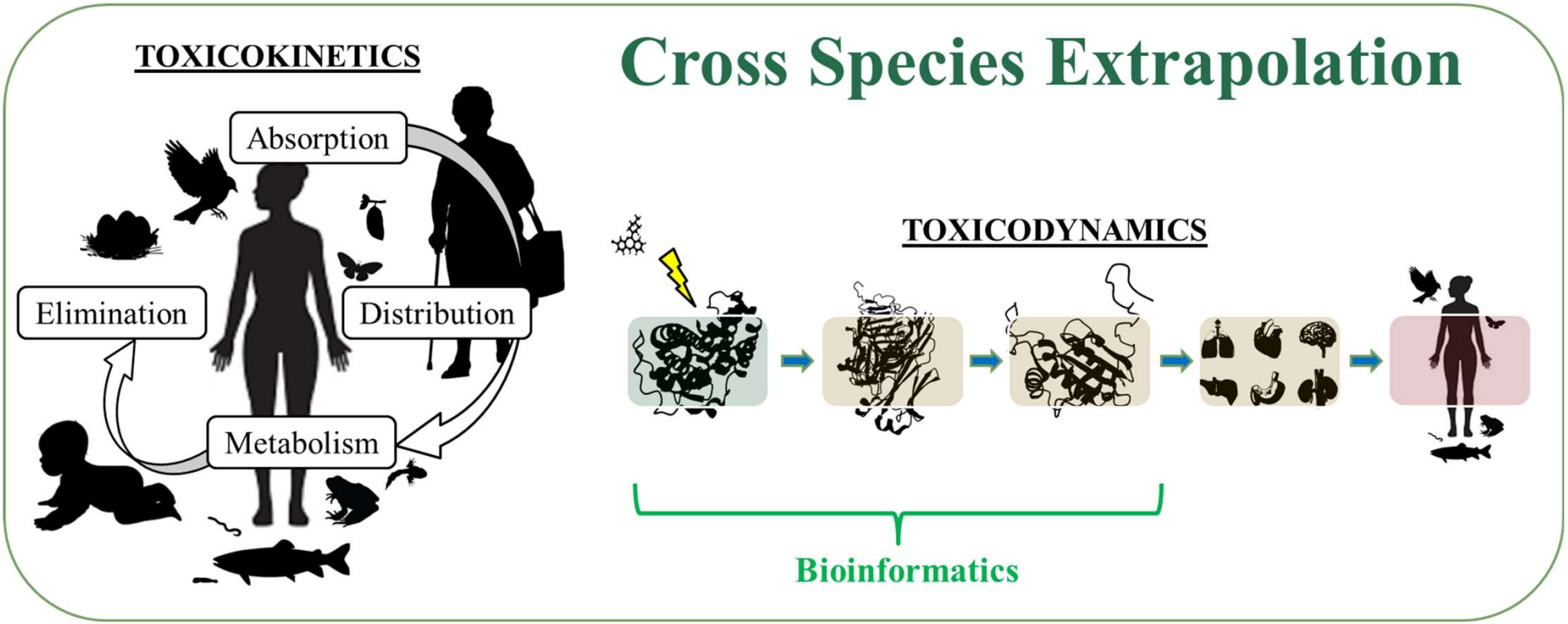
- Using existing knowledge about one species to estimate, predict, project, or infer the effect, impact, or trajectory of another species
  - For chemical safety typically dealing with toxicity

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

EPA United States Environmental Protection Agency

Sensitivity to Chemical Perturbation



Simplify Complexity

Bioinformatics

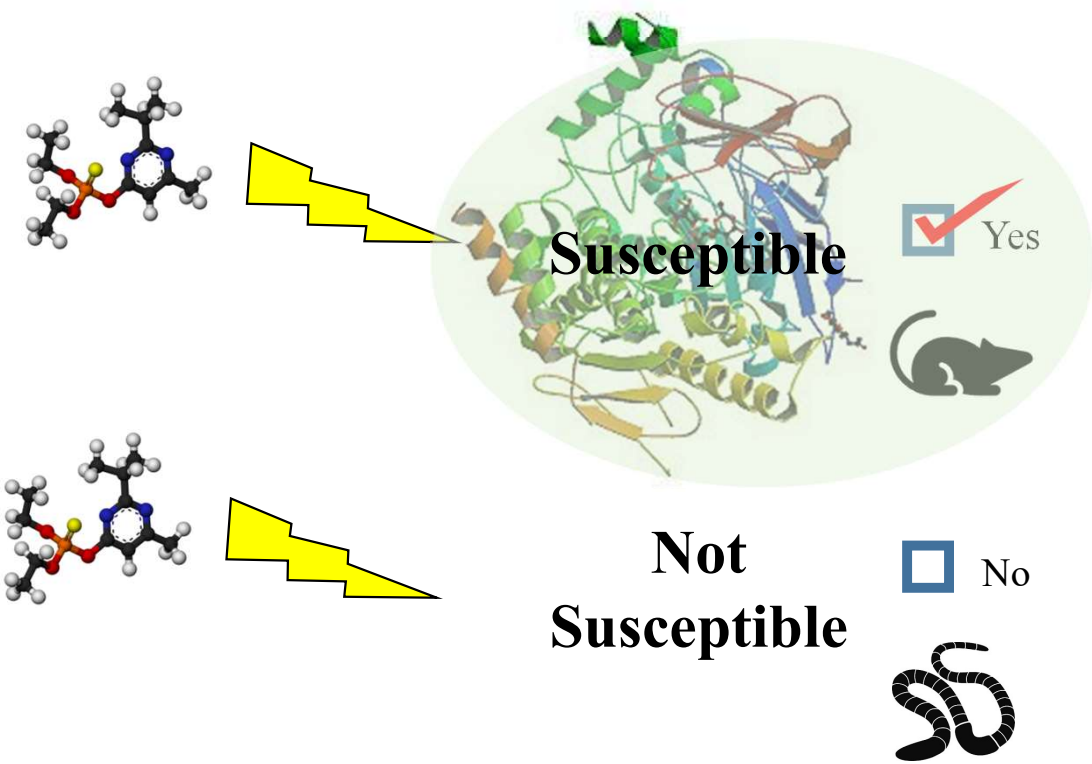
- Combines mathematics, information science, and biology to answer biological questions

- 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

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?

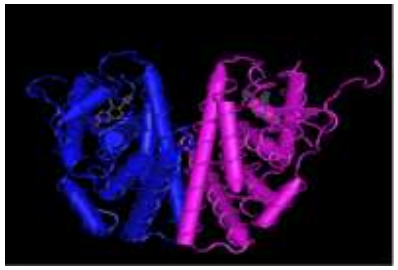
Yes or No

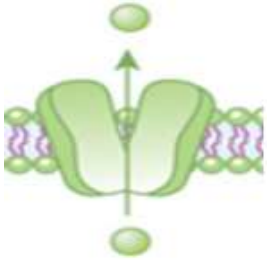
Likely susceptible or Not likely susceptible (at least through the known mechanism)

EPA United States Environmental Protection Agency

Sequence → Structure → Function

MTMTLHTKASGMALLHQIQGNELEPLNRPQLKIPLERPLGE
VYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYGQTGLPYG
PGSEAAAFGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQ
PHGQQVPYYLENEPSGYTVREAGPPAFYRPNSDNRRQGGR
ERLASTNDKGSMAMESAKETRYCAVCNDYASGYHYGVWSC
EGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLR
KCYEVGMMKGGIRKDRRGGRMLKHKRQRDDGEGRGEVG
SAGDMRAANLWPSPLMIKRSKKNSLALSLTADQMVSALLA
EPPILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKV
PGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFA
PNLLLDRNQGKCVEGMVEIFDMLLATSSRFRMMNLQGEEF
VCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITDTLIHLM

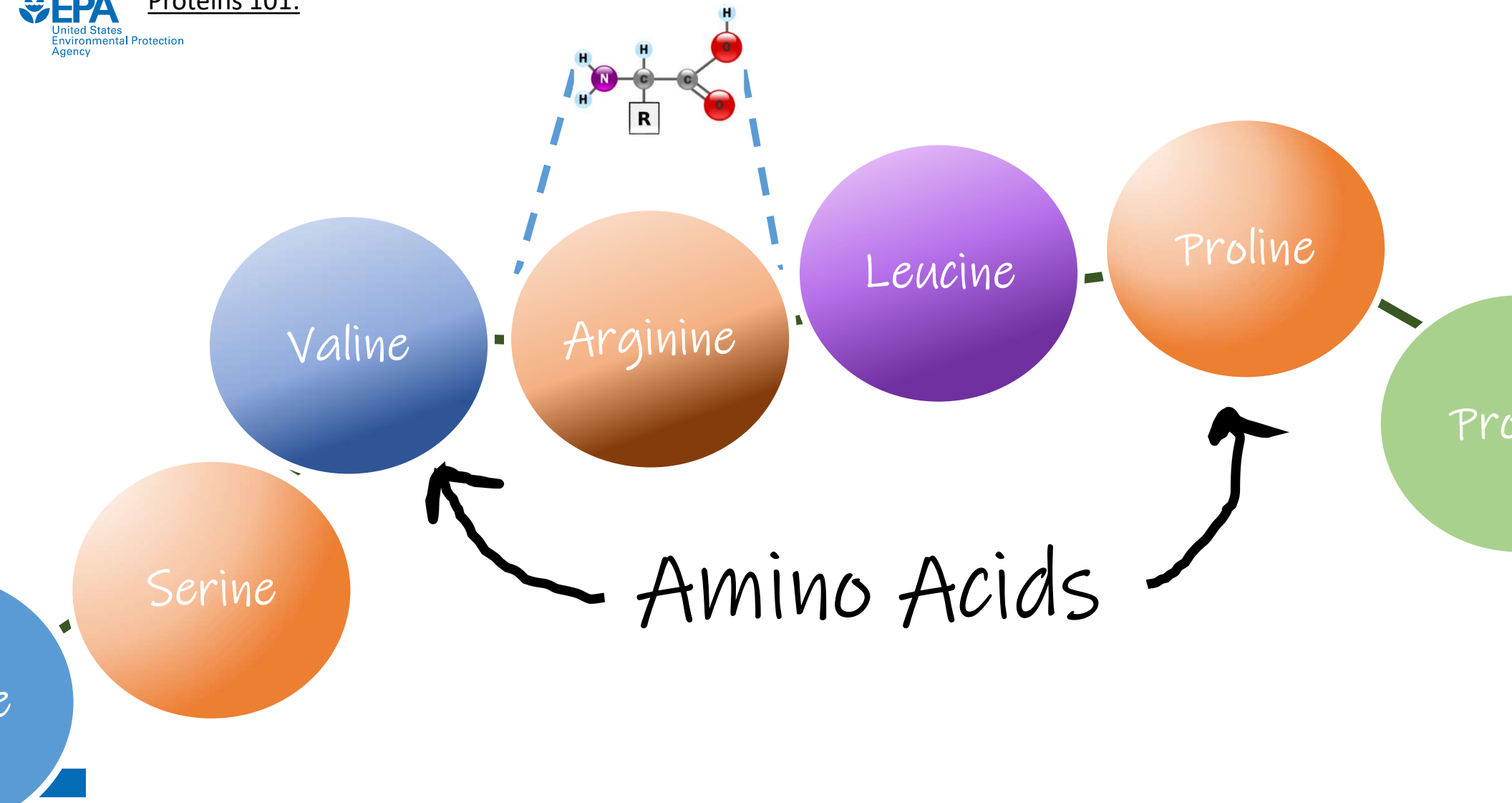




Bioinformatics

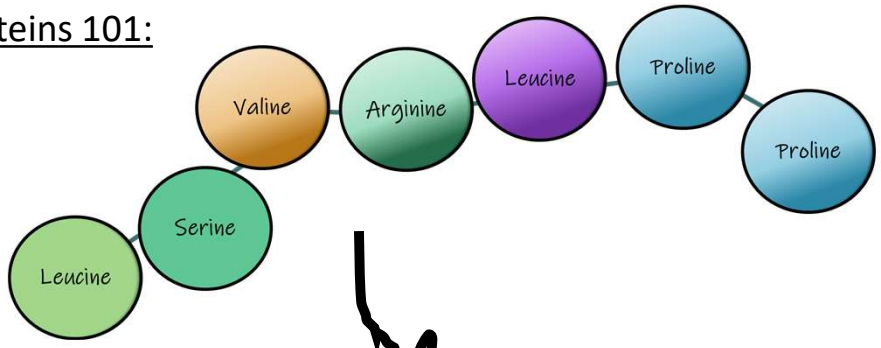
EPA United States Environmental Protection Agency

Proteins 101:



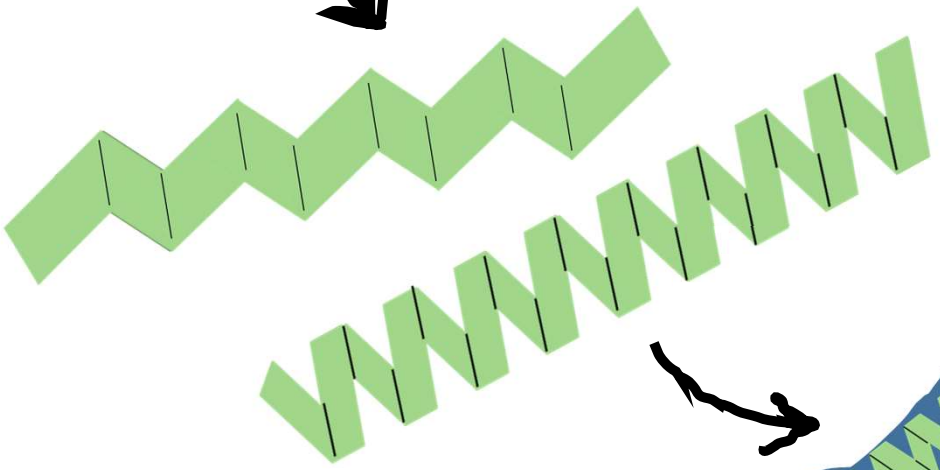
United States Environmental Protection Agency

Proteins 101:



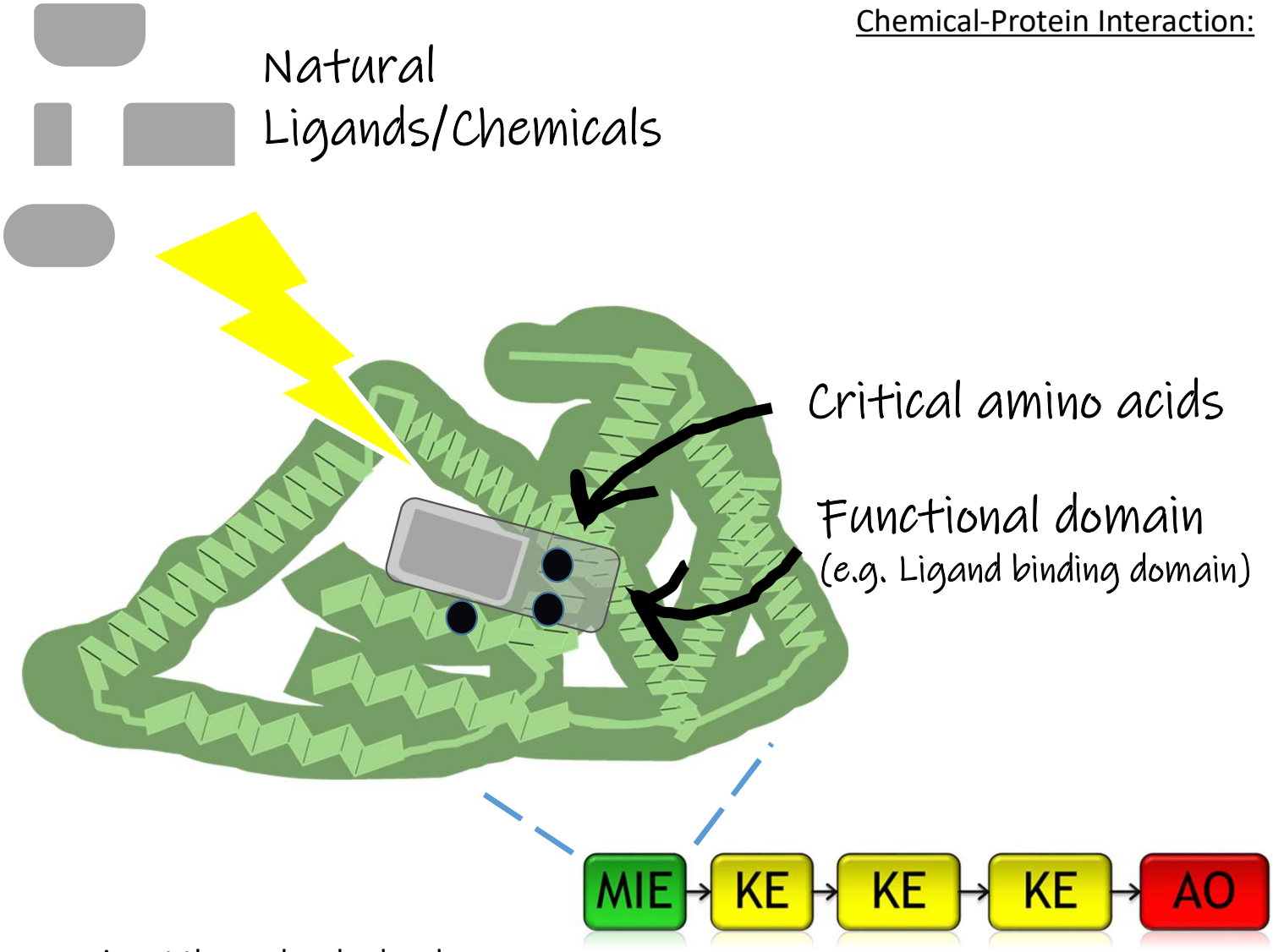
Primary amino acid sequence

Secondary Structure



Tertiary Structure

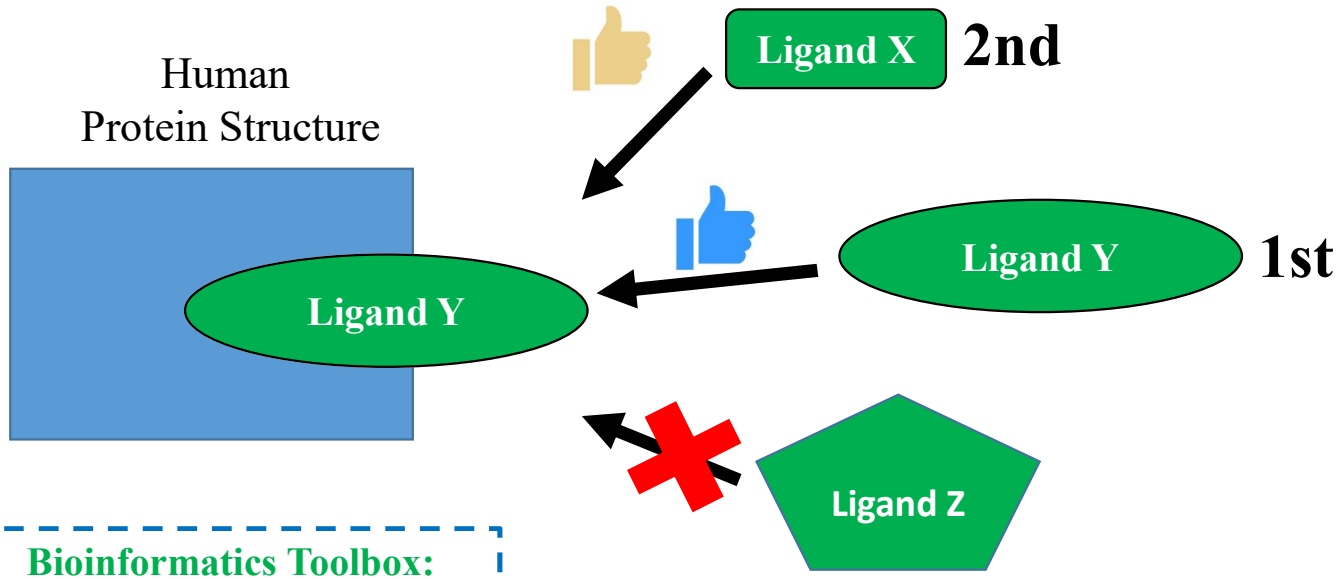
Chemical-Protein Interaction:



Similarity across species at the molecular level

Advances in Drug Discovery/Development

Structure derived from X-ray crystallography



Bioinformatics Toolbox:

Molecular modeling
Molecular docking
Virtual screening
Molecular dynamic simulations
Functional prediction

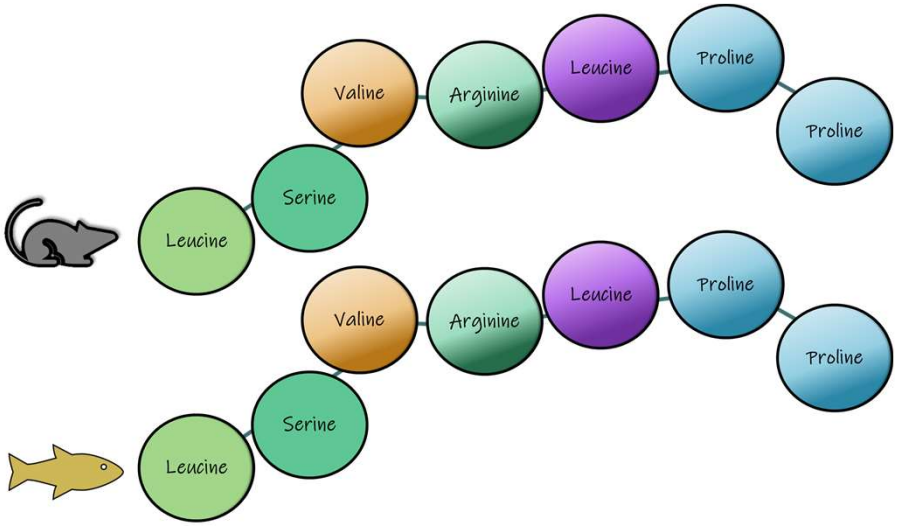
How to begin using bioinformatics for species extrapolation?

- Start simple, include target audiences and always look to future state of the science during development (prepare for what's to come)
- Stay motivated because working to change minds is rewarding
  - "Well, no one will question the author's bravery" – respected colleague and internal reviewer

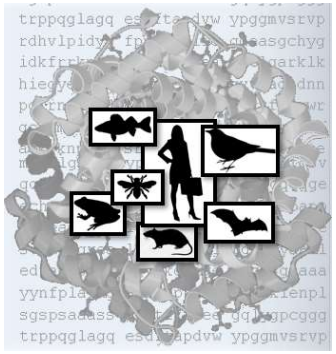
Environ. Sci. Technol. **2008,** *42,* 5807–5813

Evolutionary Conservation of Human Drug Targets in Organisms used for Environmental Risk Assessments

LINA GUNNARSSON,[†]
ALEXANDRA JAUHIAINEN,[†,⊥]
ERIK KRISTIANSSON,[†,‡,§]
OLLE NERMAN,[†,⊥] AND
D. G. JOAKIM LARSSON*[,†]

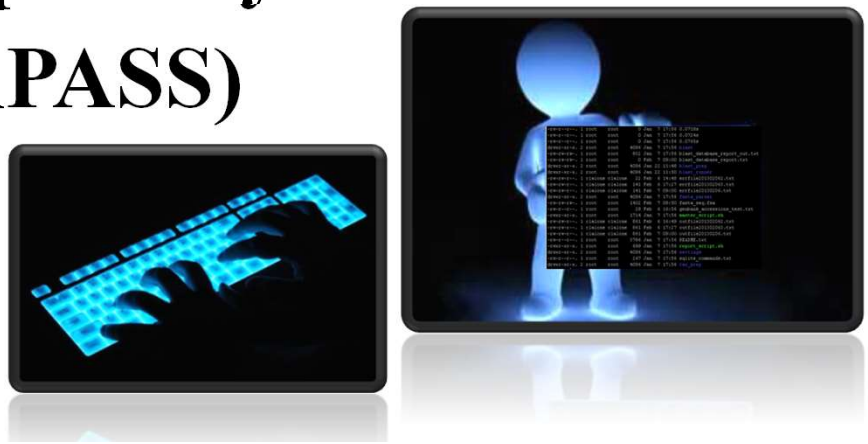


EPA United States Environmental Protection Agency



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

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)



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

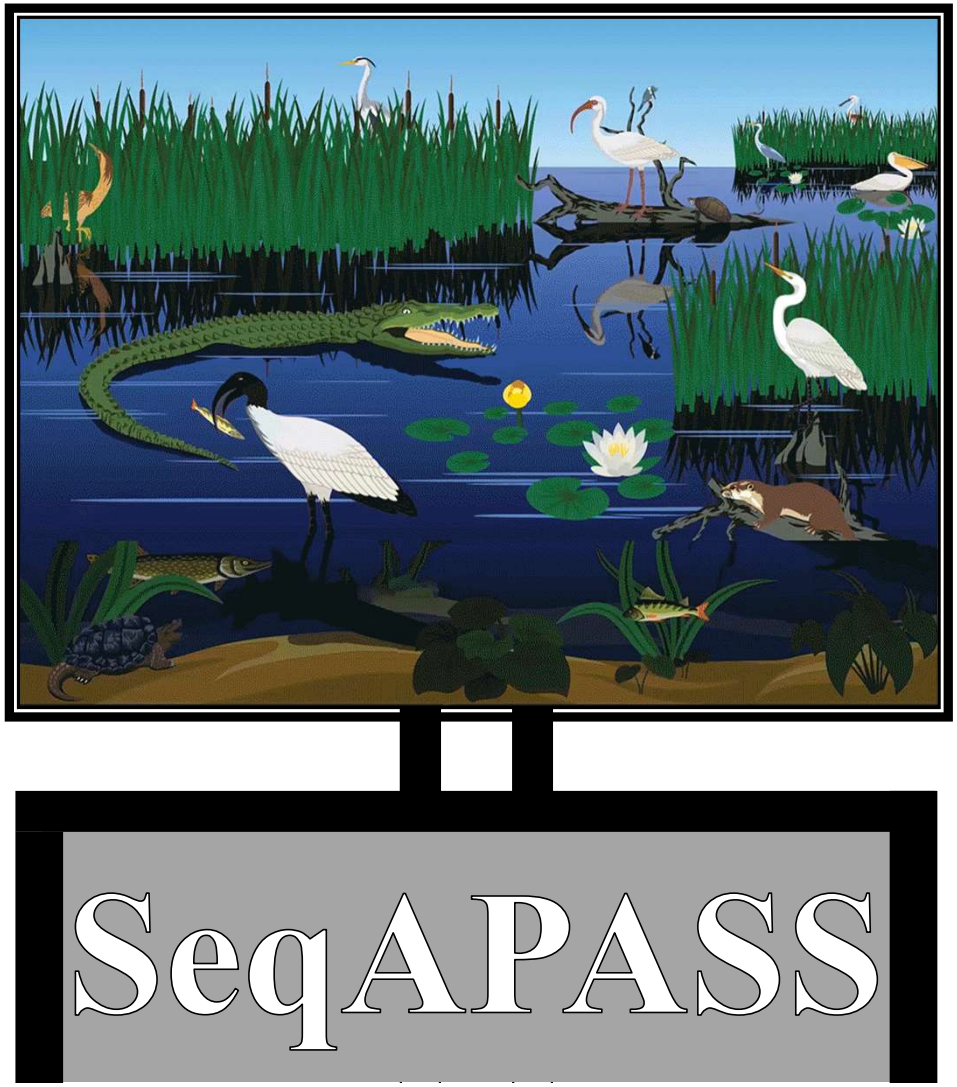
TOXICOLOGICAL SCIENCES, 153(2), 2016, 228–245

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

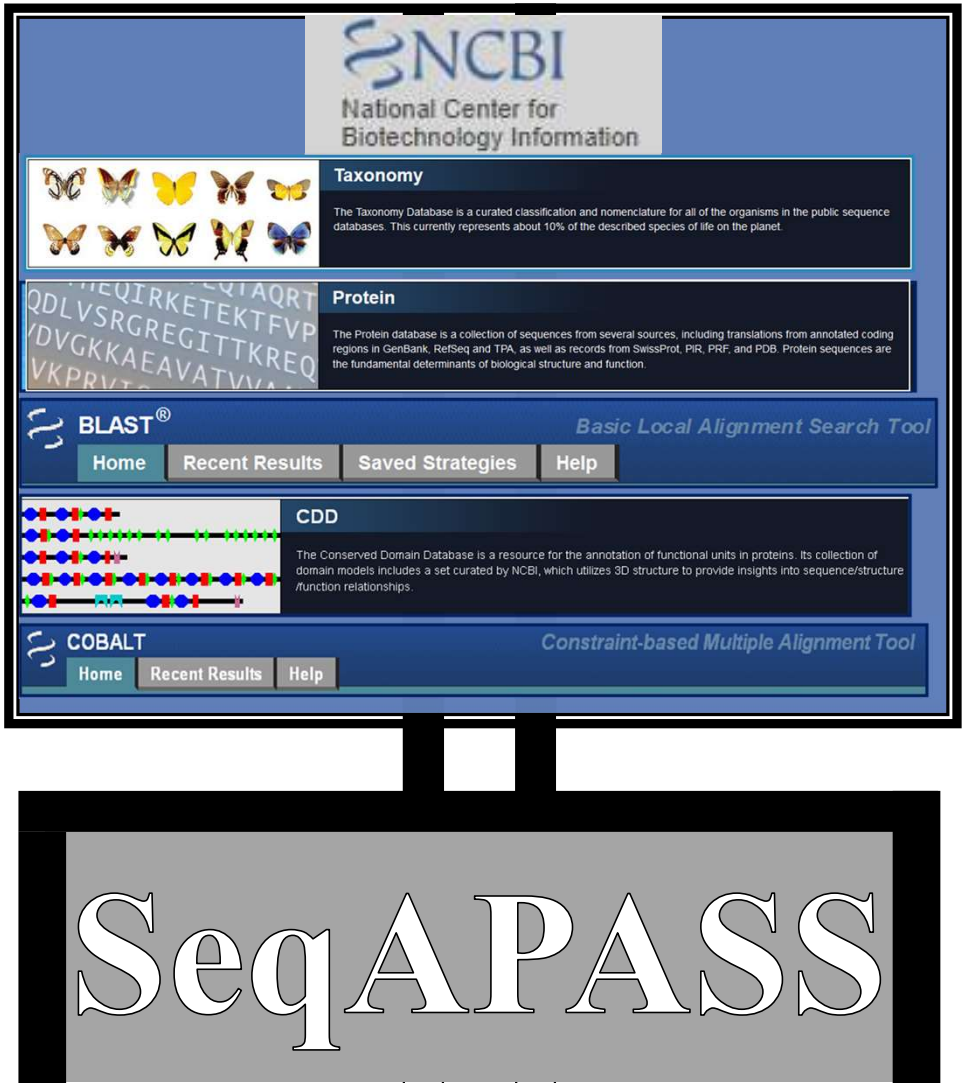
Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): A Web-Based Tool for Addressing the Challenges of Cross-Species Extrapolation of Chemical Toxicity

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*

Animation by: Miguel Moravec (EPA CSS) & Andrew Patterson



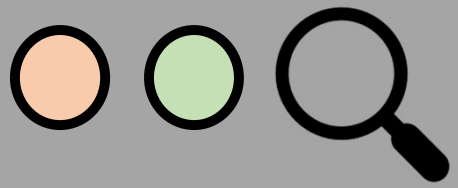
EPA United States Environmental Protection Agency

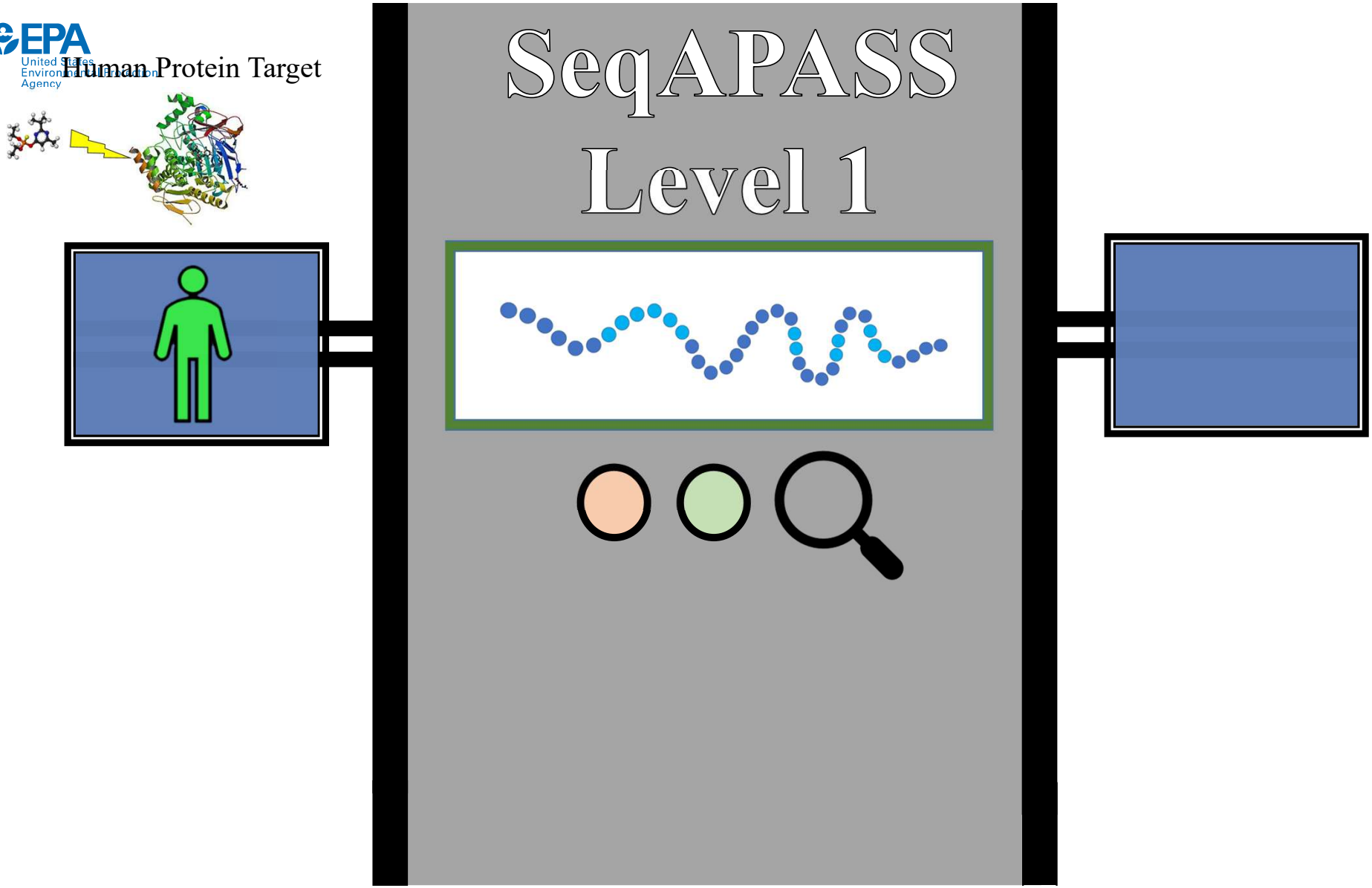


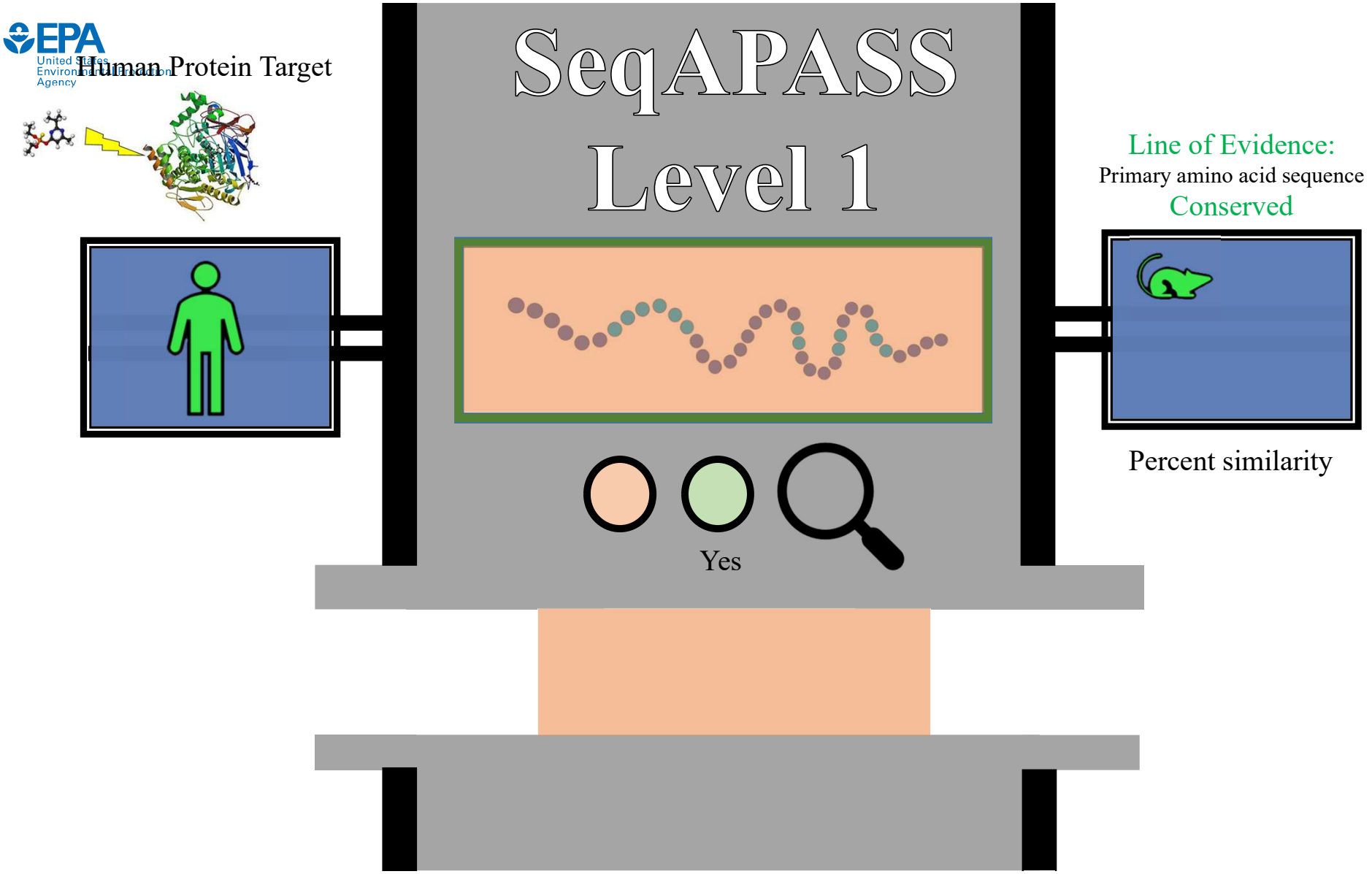
EPA United States Environmental Protection Agency

SeqAPASS

Level 1

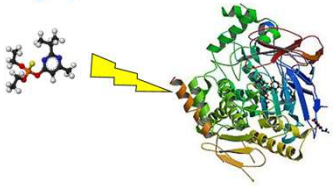



Human Protein Target
SeqAPASS
Level 1


Human Protein Target
SeqAPASS
Level 1
Line of Evidence:
Primary amino acid sequence
Conserved
Yes
Percent similarity

SeqAPASS Level 1

Human Protein Target

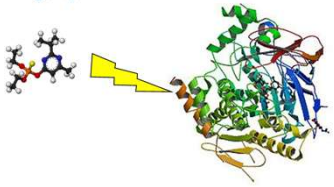


Line of Evidence:
Primary amino acid sequence
Conserved

Yes

SeqAPASS Level 1

Human Protein Target

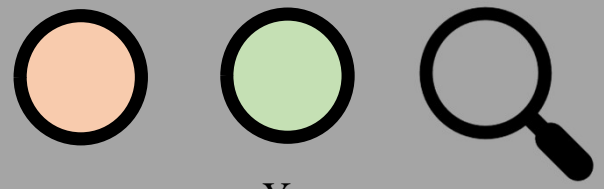


Line of Evidence:
Primary amino acid sequence
Conserved

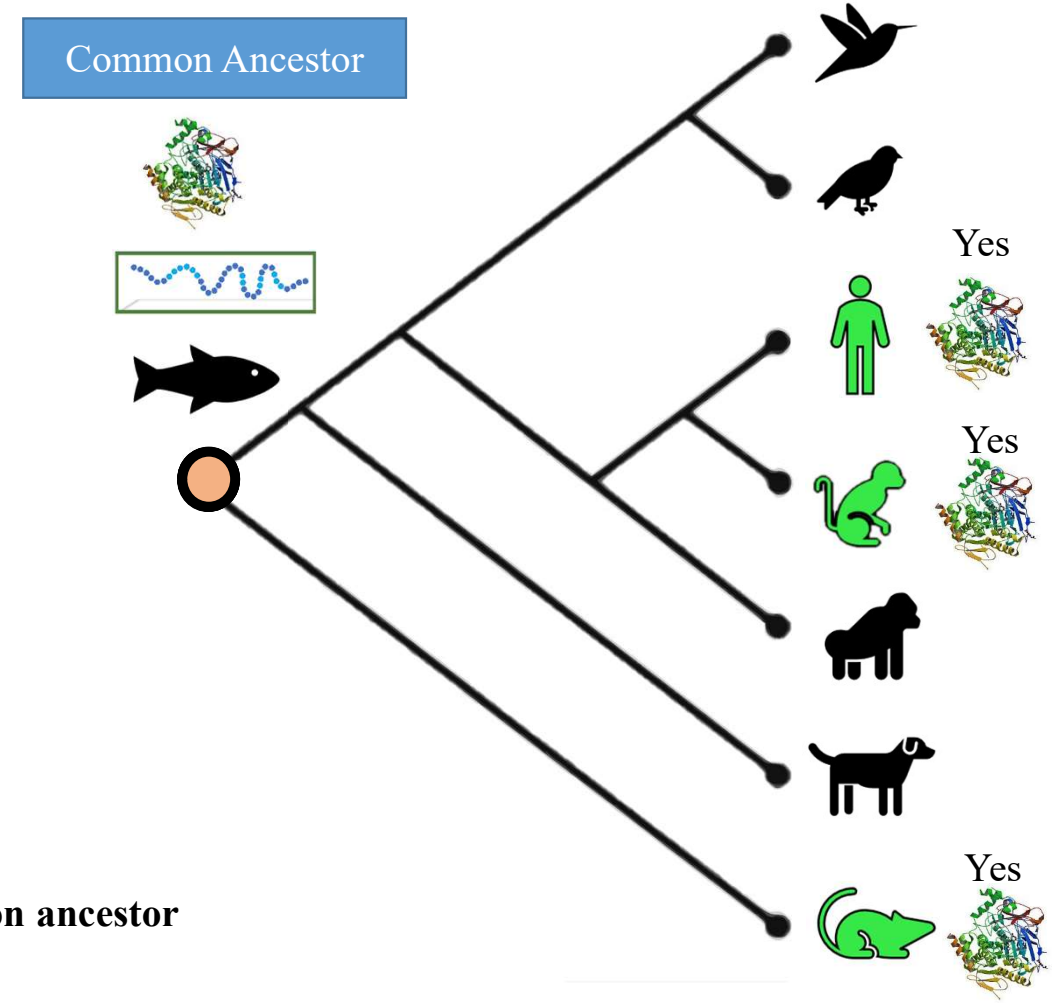
No

SeqAPASS Level 1

Ortholog Candidate Identification



Yes



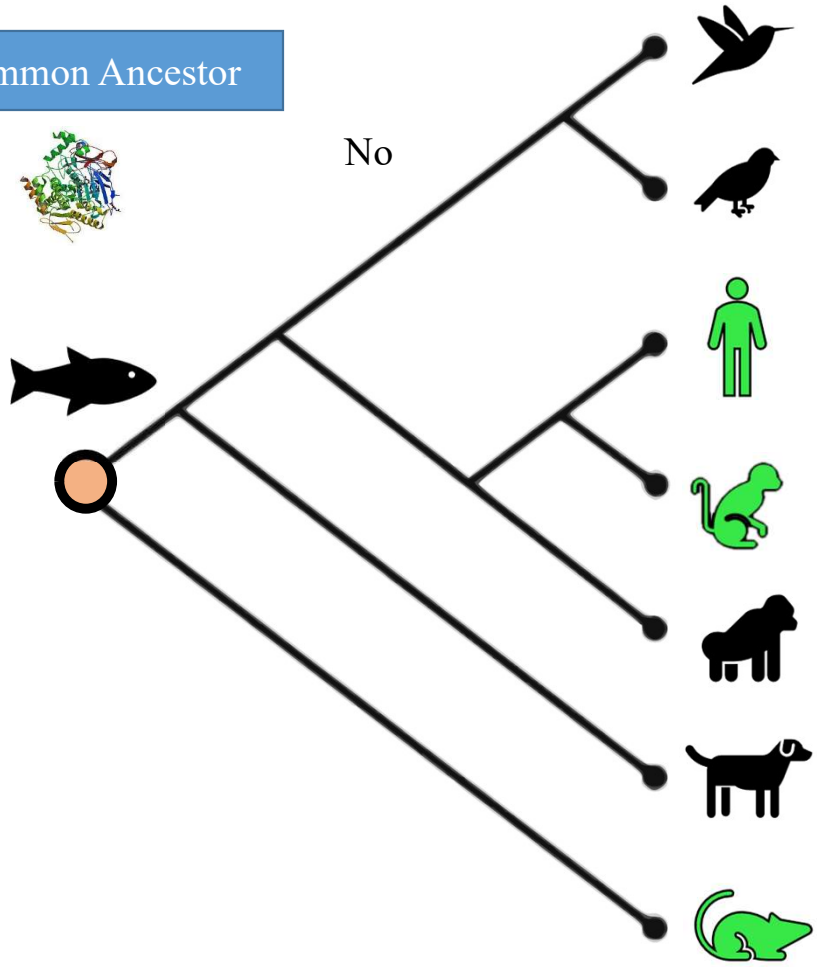
Proteins in different species that evolved from a common ancestor

Typically maintain similar function

SeqAPASS Level 1

Ortholog Candidate Identification

No



SeqAPASS Level 1

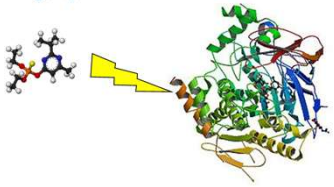
Common Name	Ortholog Candidate	Cut-off	Percent Similarity
Human	Y	33.15	100
Florida manatee	Y	33.15	98.8
Mallard	Y	33.15	82.29
Rock pigeon	Y	33.15	80.93
Green anole	Y	33.15	80.65
Pacific transparent sea squirt	Y	33.15	33.15
Yesso scallop	N	33.15	32.87
Purple sea urchin	N	33.15	26.05
Human whipworm	N	33.15	23.53
Bed bug	N	33.15	21.62

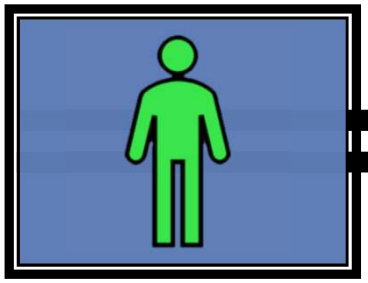
Lowest % Similarity that is still an ortholog

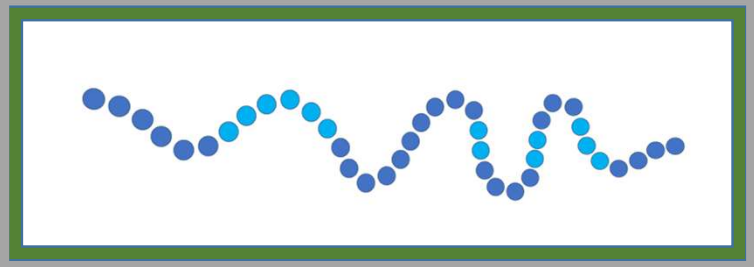
Example:
Susceptibility Cut-off: Set at 33.15
Above cut-off: More likely to be susceptible base on similar **FUNCTION**

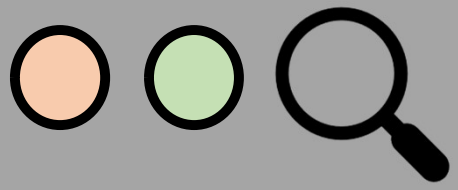
SeqAPASS Level 1

Human Protein Target

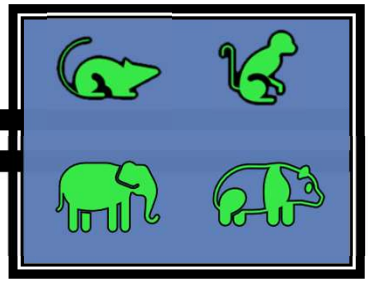




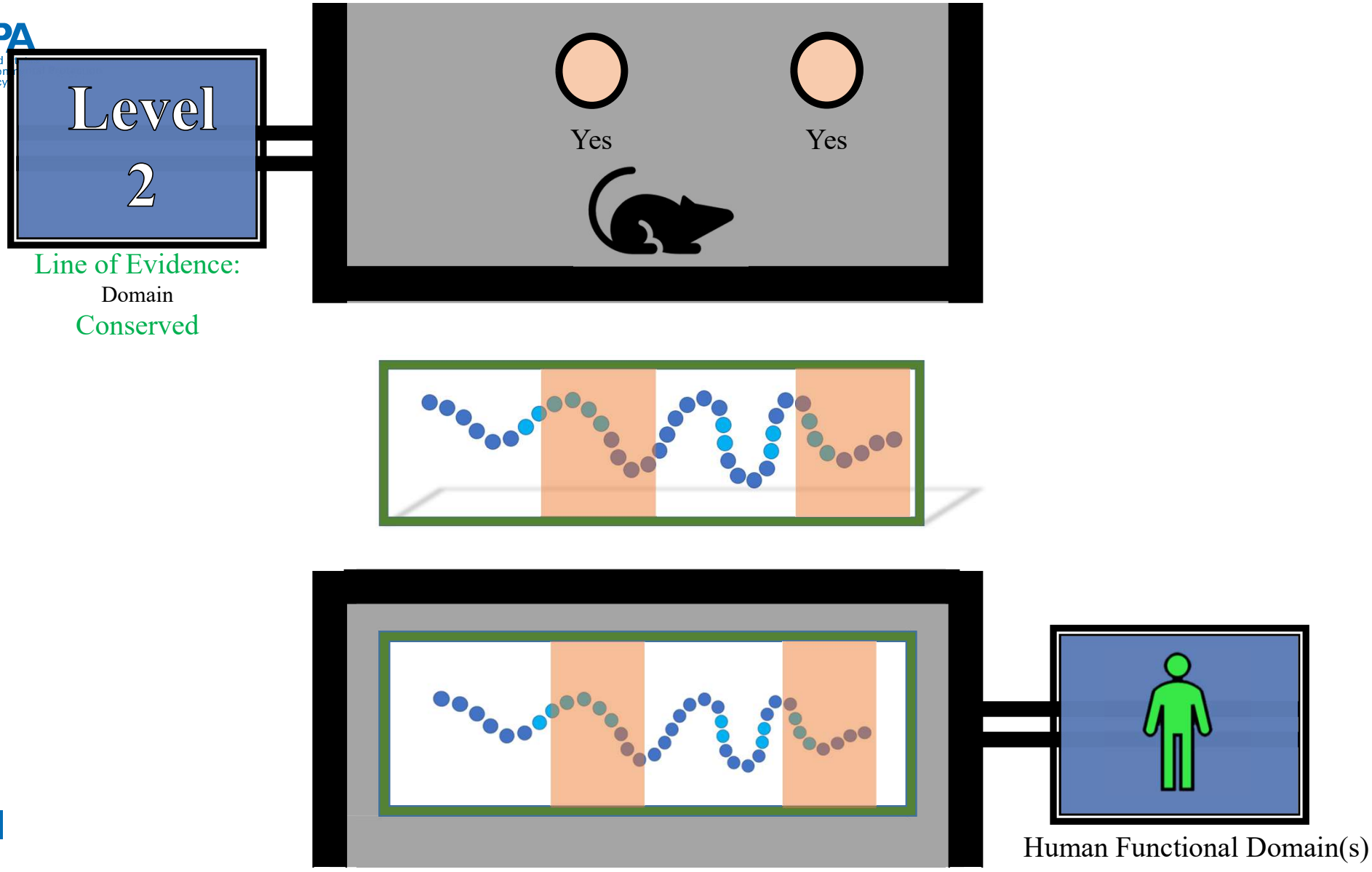


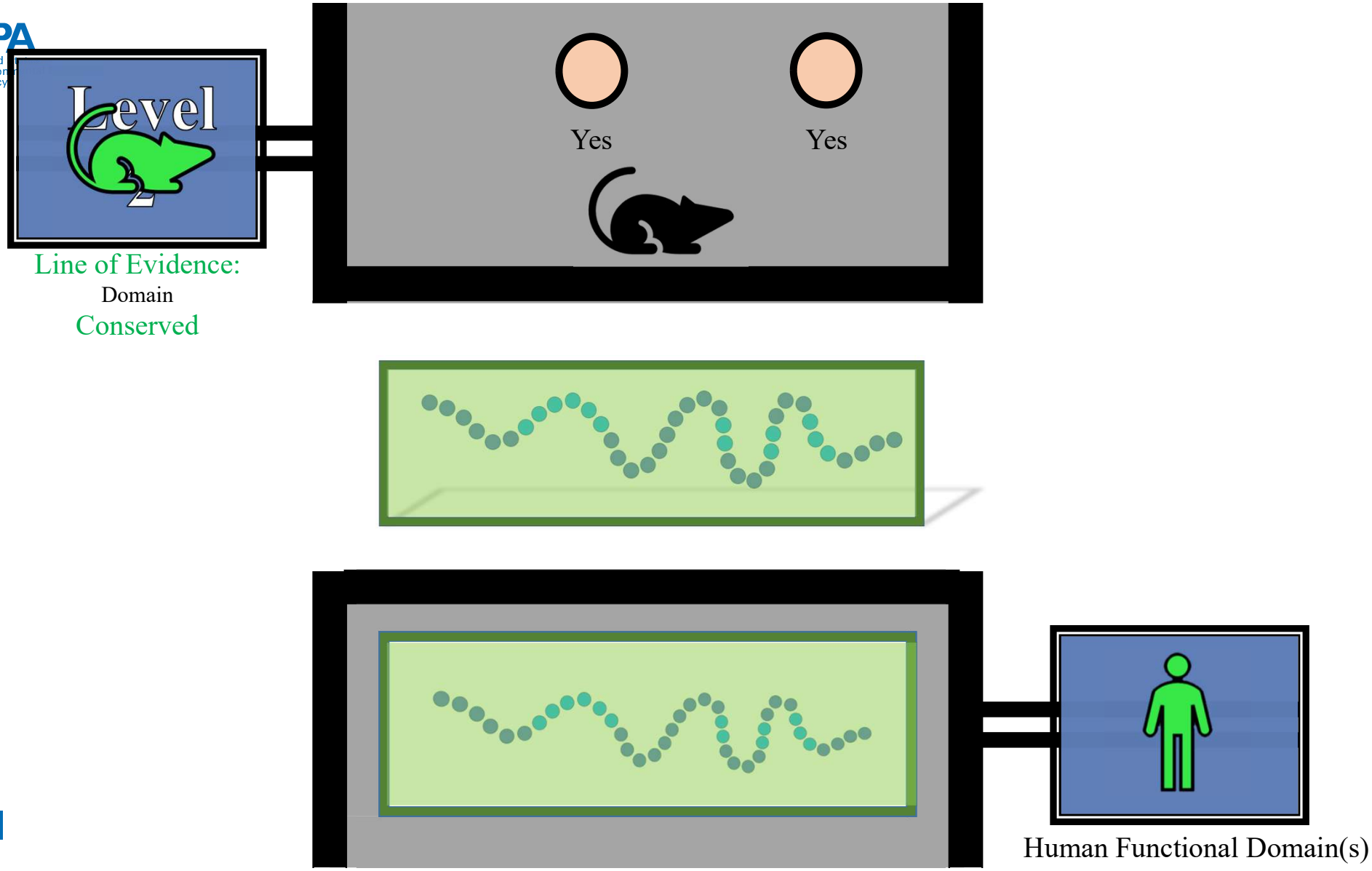


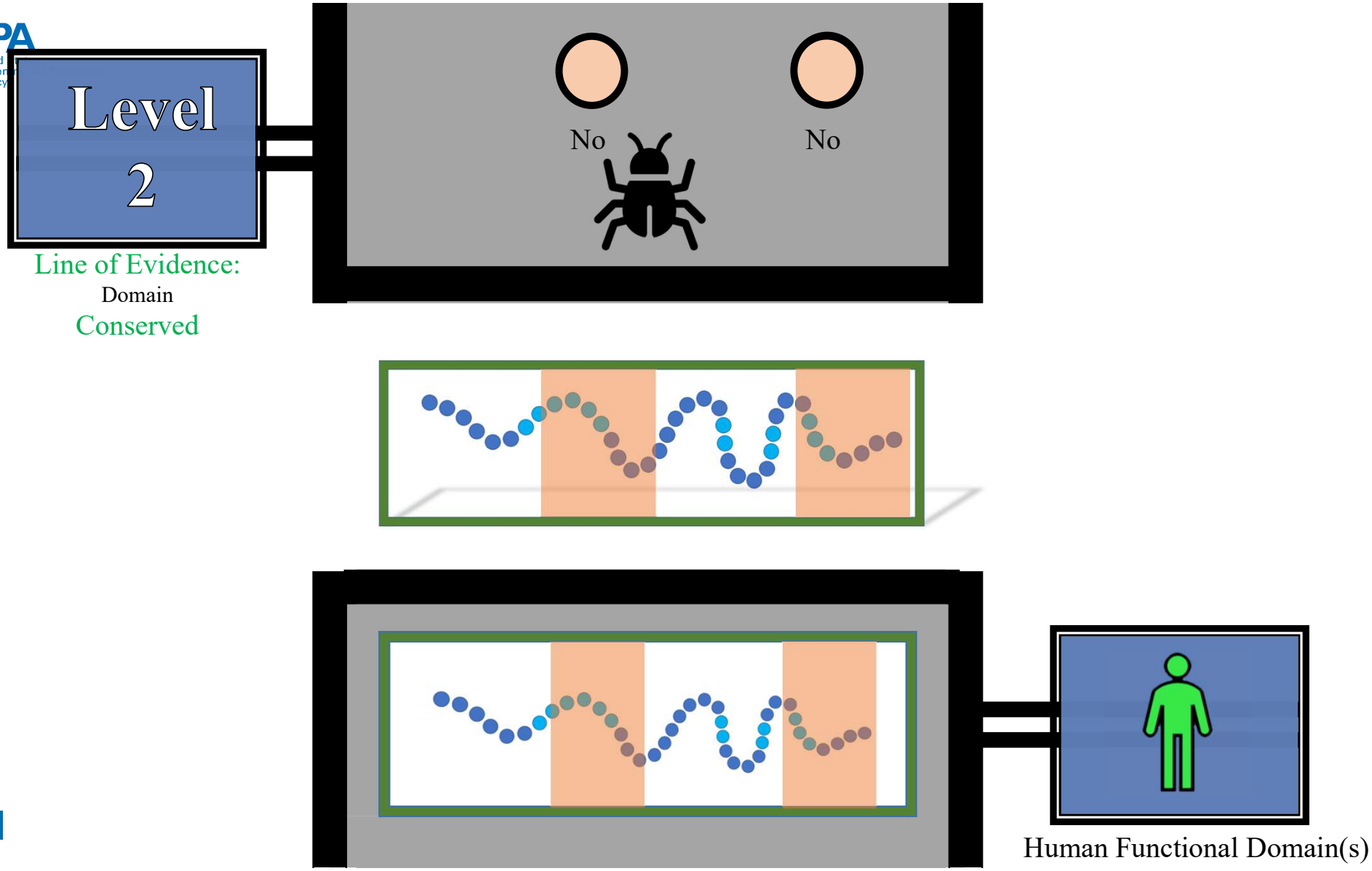
Line of Evidence:
Primary amino acid sequence
Conserved

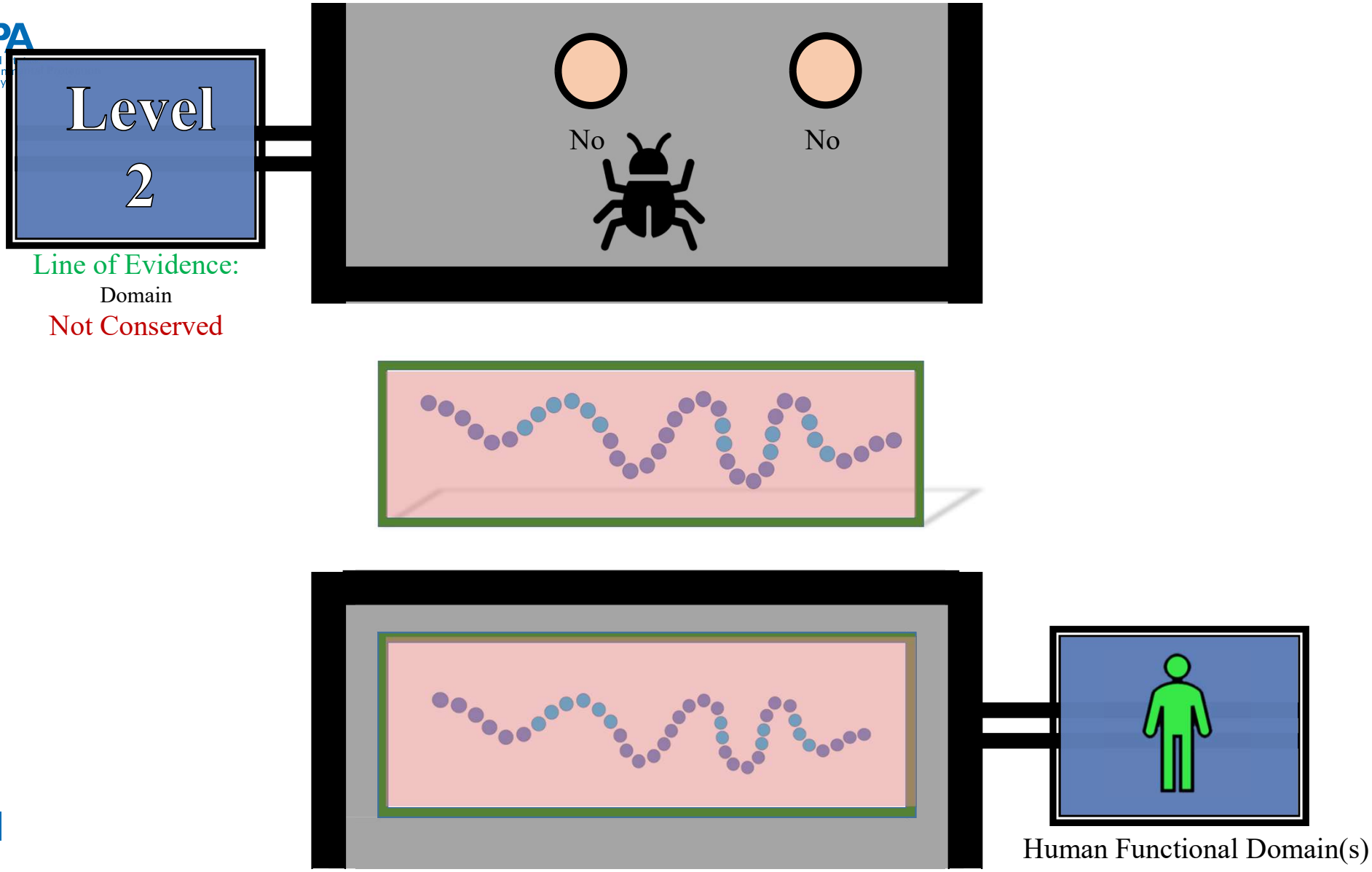


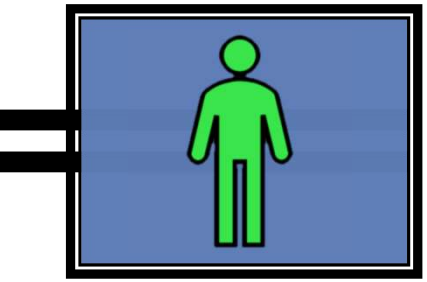
Hundreds to Thousands of Species


Level 2
Yes
Yes
Line of Evidence:
Domain
Conserved
Human Functional Domain(s)

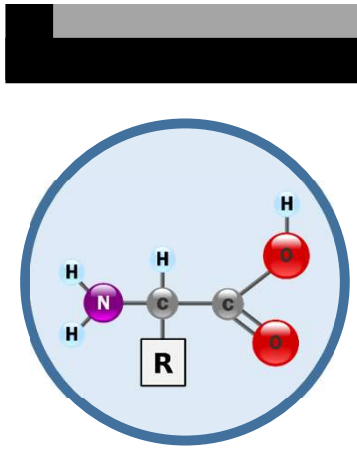

Level 2
Yes
Yes
Line of Evidence:
Domain
Conserved
Human Functional Domain(s)

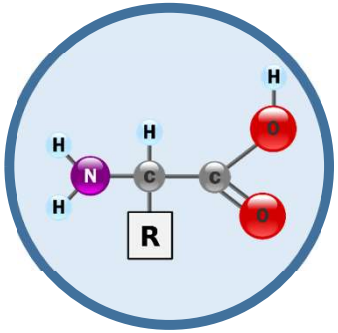

Level 2
No
No
Line of Evidence:
Domain
Conserved
Human Functional Domain(s)

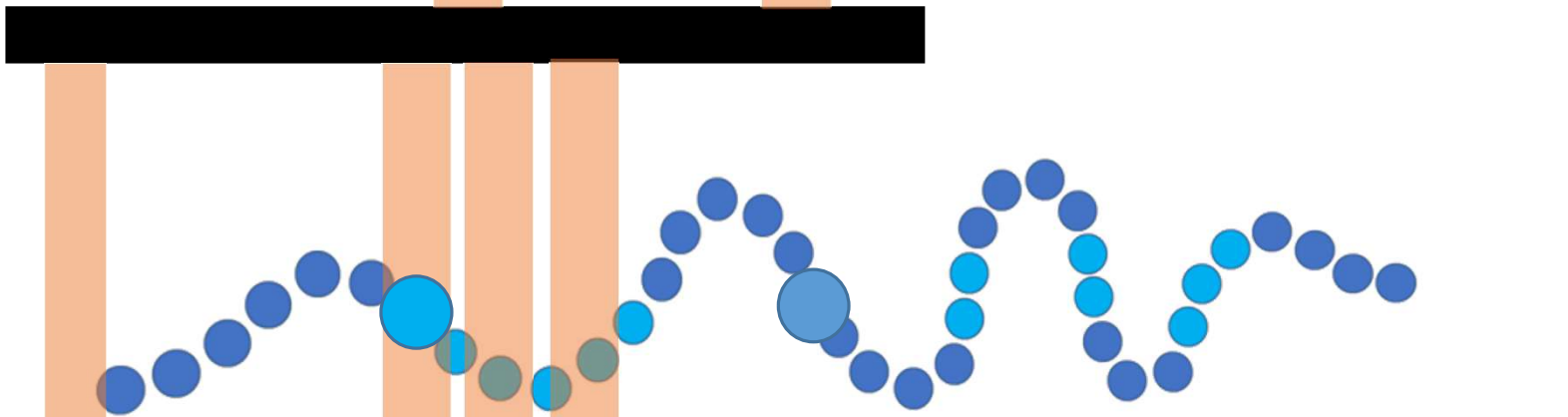

Level 2
No
No
Line of Evidence:
Domain
Not Conserved
Human Functional Domain(s)

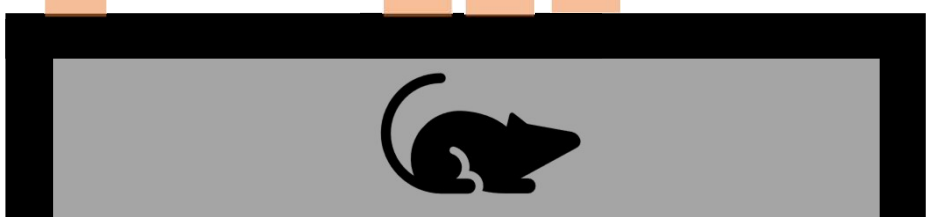


Human Critical Amino Acids

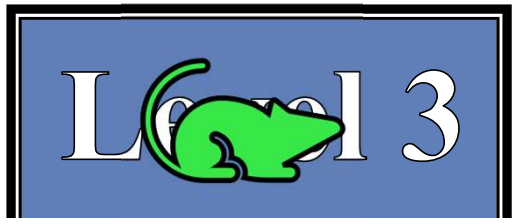


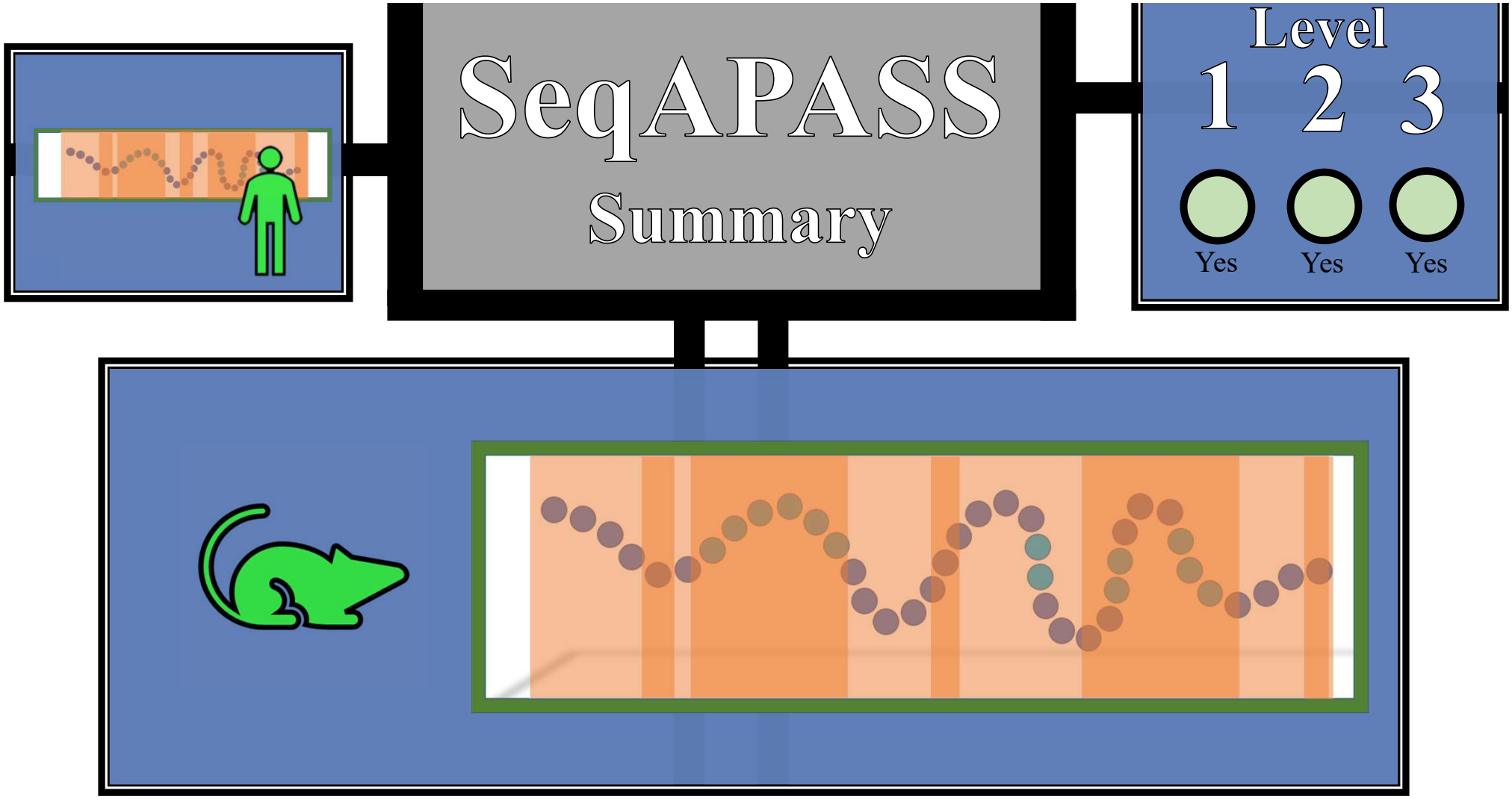






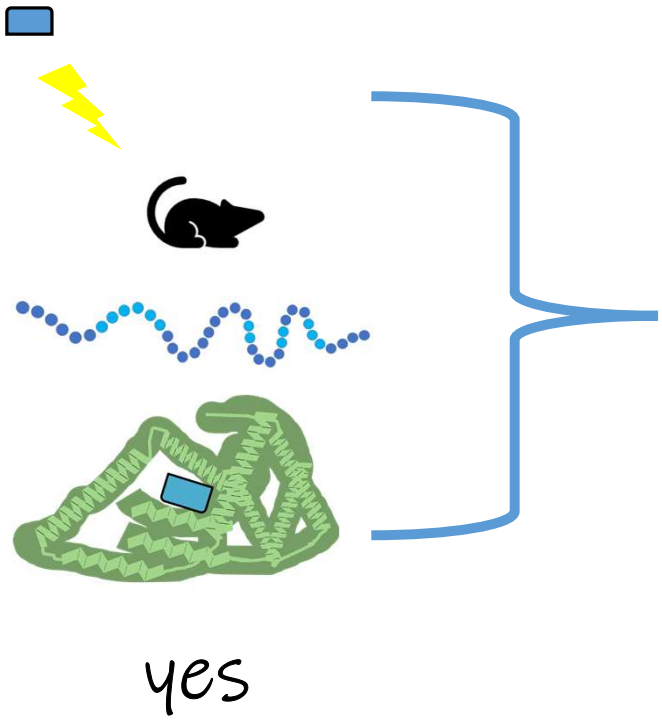
Line of Evidence: Conserved

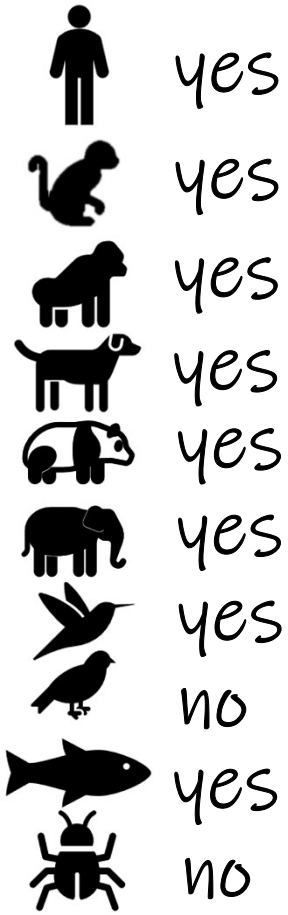




Gather Lines of Evidence for Conservation of Protein Target:
Susceptibility Prediction: Yes or No

SeqAPASS Predicts Likelihood of Similar Susceptibility based on Sequence Conservation:

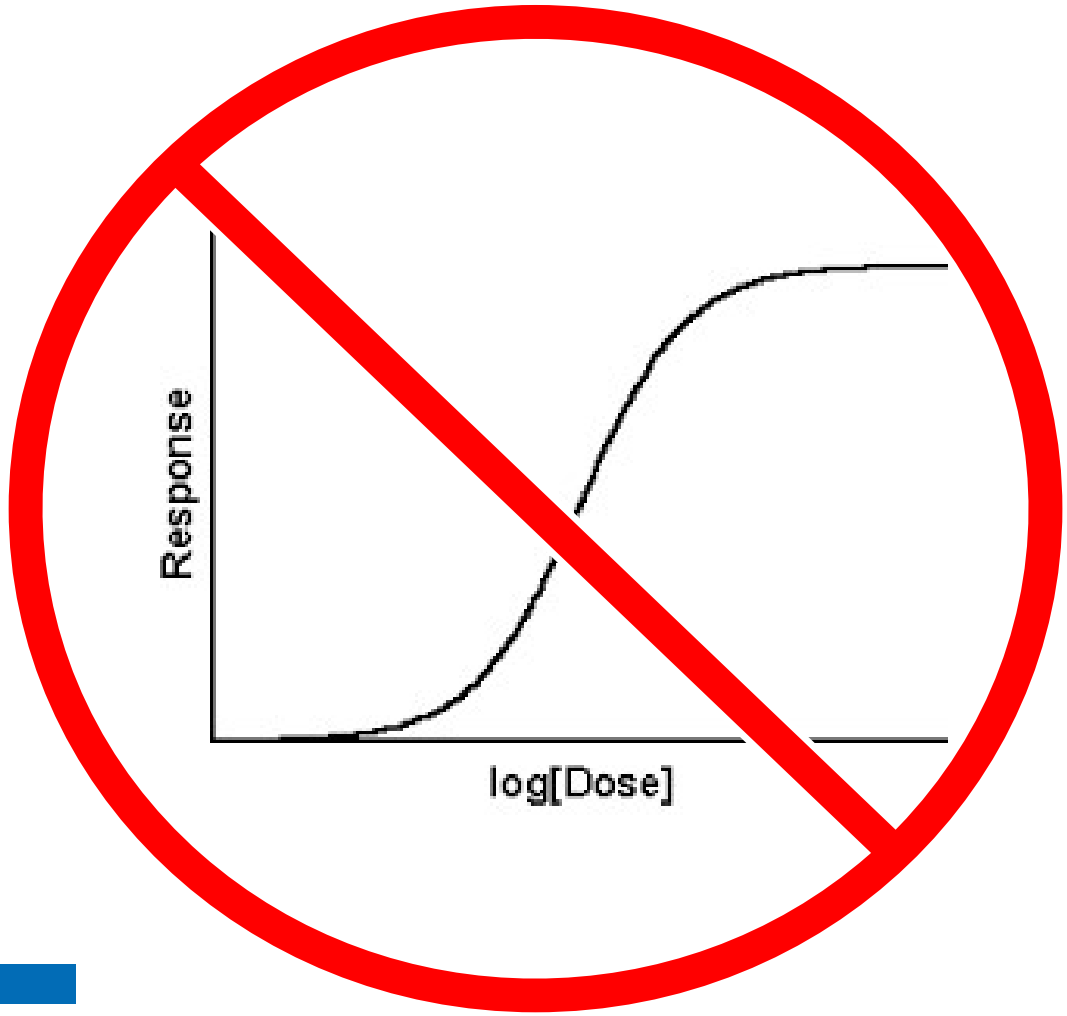




Line(s) of evidence indicate

- The protein is conserved
- The protein is NOT conserved

SeqAPASS DOES NOT Predicts the degree of sensitivity/susceptibility:



Factors that make a species sensitive

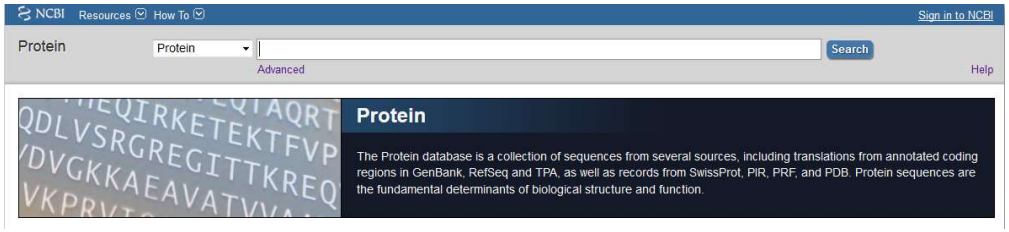
- Exposure
- Dose
- ADME
- Target receptor availability
- Life stage
- Life history
- etc.
- etc.

EPA United States Environmental Protection Agency

Strengths of SeqAPASS

New tools and technologies have emerged

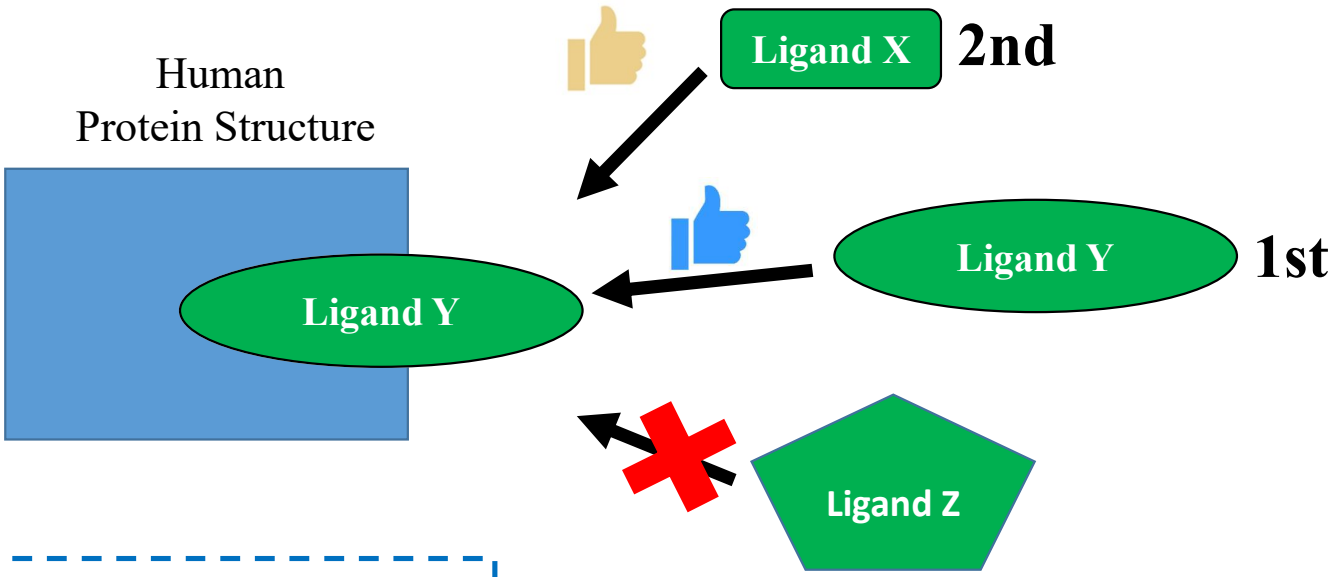
- Improved sequencing technologies
- Large databases of sequence data



- 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

Advances in Drug Discovery/Development

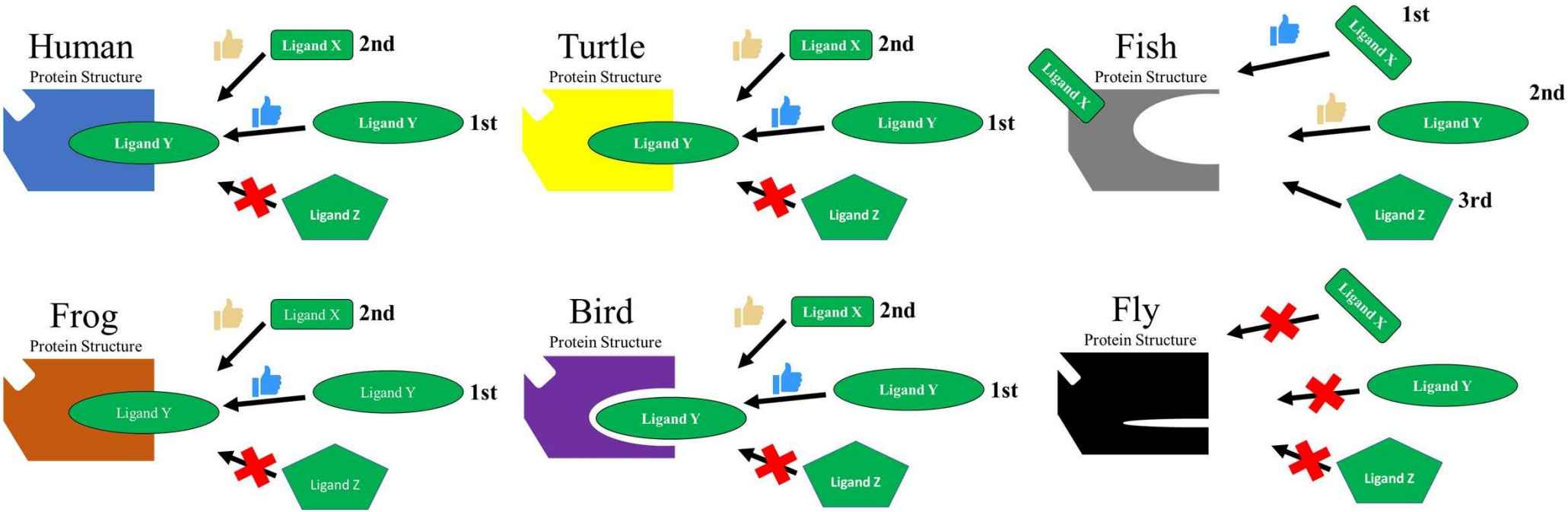
Structure derived from X-ray crystallography



Bioinformatics Toolbox:

Molecular modeling
Molecular docking
Virtual screening
Molecular dynamic simulations

Application to Species Extrapolation



Bioinformatics Toolbox:

Molecular modeling

Molecular docking

Virtual screening

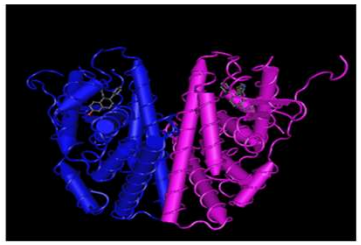
Molecular dynamic simulations

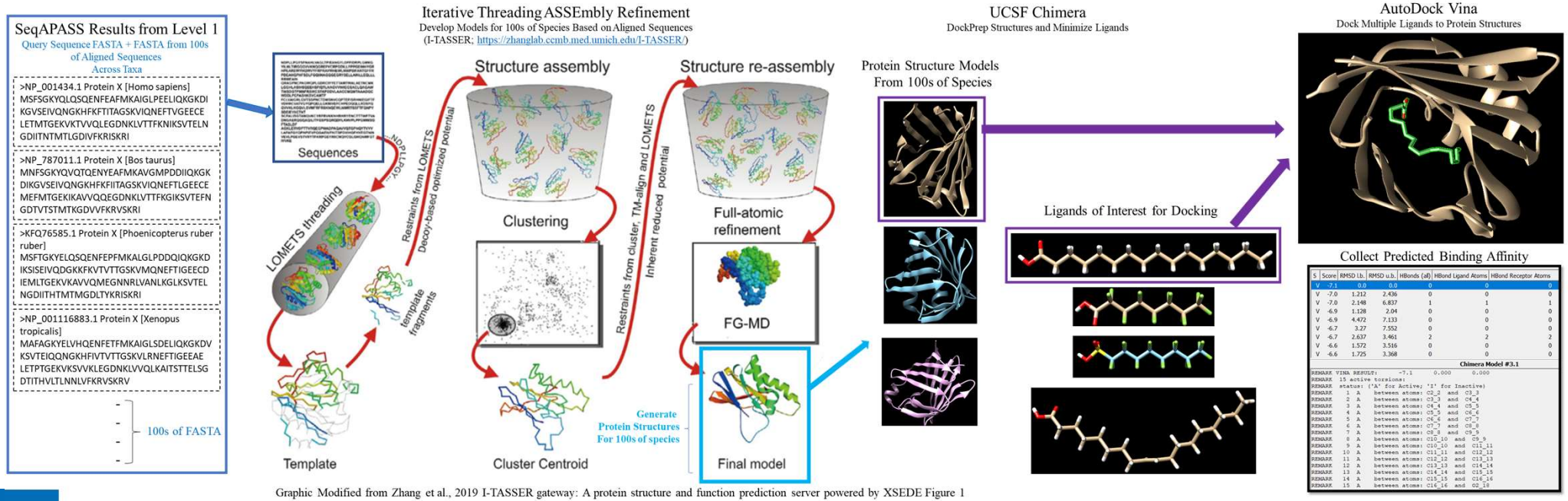
EPA United States Environmental Protection Agency

Sequence

MTMTLHTKASGMALLHQIQGNELEPLNRPQLKIPLERPLGE
VYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYGQTGLPYG
PGSEAAAFGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQ
PHGQQVPYYLENEPSGYTVREAGPPAFYRPNSDNRRQGGR
ERLASTNDKGSMAMESAKETRYCAVCNDYASGYHYGVWSC
EGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLR
KCYEVGMMKGGIRKDRRGGRMLKHKRQRDDGEGRGEVG
SAGDMRAANLWPSPLMIKRSKKNSLALSLTADQMVSALLA
EPPILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKV
PGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFA
PNLLLDRNQGKCVEGMVEIFDMLLATSSRFRMMNLQGEEF
VCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITDTLIHLM

Structure



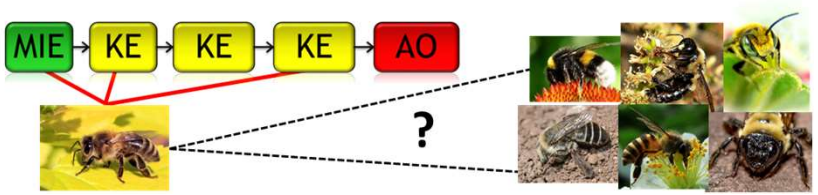


Graphic Modified from Zhang et al., 2019 I-TASSER gateway: A protein structure and function prediction server powered by XSEDE Figure 1

Predicting Binding Affinity

Applications of Bioinformatics: Case Studies

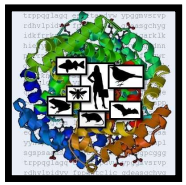
- **Extrapolate adverse outcome pathway knowledge across species**
  - Define the taxonomic relevance: Apis vs Non-Apis bees



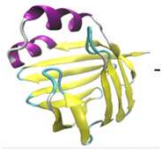
- **Extrapolate high throughput screening data**
  - Chemicals that target human estrogen receptor alpha, androgen receptor, steroidogenic enzymes, thyroid axis proteins
  - All ToxCast Assay targets
- **Predict relative intrinsic susceptibility**
  - Pesticides
  - Endangered Species Act
  - Derivation of Aquatic Life Criteria
- **Predict chemical bioaccumulation across species**
  - Chemicals of concern: PFAS
- **Generate research hypotheses** Strobilurin fungicides
- **Prioritization strategies** Pharmaceuticals

EPA United States Environmental Protection Agency

Predict chemical bioaccumulation across species

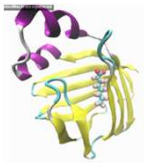


Sequence comparisons SeqAPASS



Homology modeling (Phyre2) or PDB

Molecular docking (Autodock vina)



Molecular dynamics (AMBER 14)

LFABP structures across 7 different species

- Liver-type fatty acid binding protein (LFABP) is used as protein proxy for bioaccumulation assessment
- human and rat LFABP structures are available in Protein Data Bank (PDB)
- chicken, zebrafish, rainbow trout, Japanese medaka, and fathead minnow structures were generated using Phyre2

9 PFAS structures

- 6 PFCAs: PFBA(C4), PFPA(C5), PFHxA(C6), PFHpA(C7), PFOA(C8), PFNA(C9)
- 3 PFSAs: PFBS(C4), PFHxS(C6), PFOS(C8)

By integrating SeqAPASS and the molecular dynamics workflow, our approach:

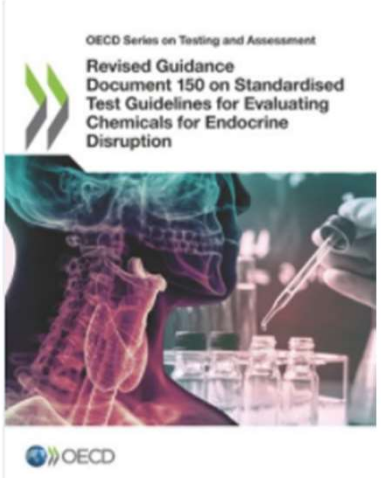
- Provides insights into the bioaccumulation potential across different species from the evaluation of both the structure and function of the critical protein LFABP
- Suggests that rat, chicken, zebrafish and rainbow trout are better representative species than Japanese medaka and fathead minnow for predicting bioaccumulation and toxicity in humans

Weixiao Cheng, Jon A Doering, Carlie LaLone, Carla Ng, Integrative Computational Approaches to Inform Relative Bioaccumulation Potential of Per- and Polyfluoroalkyl Substances Across Species, Toxicological Sciences, Volume 180, Issue 2, April 2021, Pages 212–223

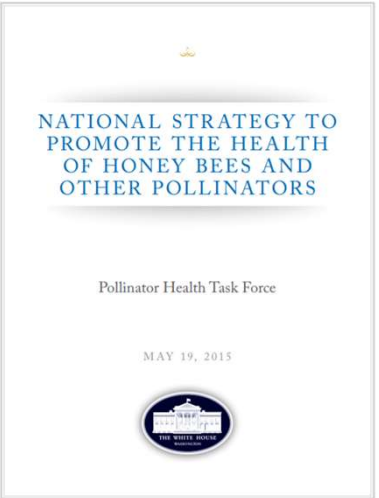
United States Environmental Protection Agency

Bioinformatics in Species Extrapolation for Chemical Safety Evaluation

- The Organisation for Economic Co-operation and Development (OECD) published the Revised Guidance Document 150 on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption



- Appendices of the White House National Strategy to Promote the Health of Honey Bees and Other Pollinators



Pillars for Taking on a New Approach

TIME/CHAMPIONS

CONFIDENCE

TRAINING

SUPPORT

Acknowledgements

U.S. EPA, ORD

Donovan Blatz (ORISE)
Sara Vliet (ORISE)
Sally Mayasich (ORISE)
Marissa Jensen (Univ. Minnesota Duluth)

GDIT

Thomas Transue
Cody Simmons
Audrey Wilkinson

SeqAPASS v5.0

LaLone.Carlie@epa.gov

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