Addressing Challenges in Cross-Species Extrapolation of Chemical Toxicity Information: Application of the SeqAPASS Tool

Lead Instructor:

Carlie A. LaLone, Ph.D.

Great Lakes Toxicology and Ecology Division

Office of Research and Development

U.S Environmental Protection Agency

Instructors:

Donovan Blatz

Sara Vliet, PhD

Oak Ridge Institute for Science and Education

SeqAPASS Training

Training format:

- Overview presentation (LaLone) — Video 1
- How to request a user account to SeqAPASS (Blatz) — Video 2
- Demonstration of SeqAPASS functionality (LaLone & Blatz) — Video 3
- Case Examples (Vliet) — Video 4

Goals for training:

- Introductory training for new/beginner SeqAPASS user
- Understand why SeqAPASS was developed
- Gain a basic understanding of how SeqAPASS evaluations work
- Understand the general workflow of SeqAPASS
- Independently submit a SeqAPASS query
- Independently collect output from SeqAPASS
- Develop box-plot visualizations

United States Environmental Protection Agency

Available Knowledge

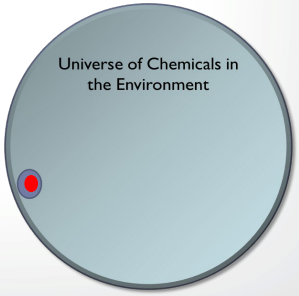
Science-based, Streamlined, Transparent, Publicly Accessible PIPELINE

Predict Susceptibility

Chemical Safety Evaluation

- Protect human health and the environment
  - Ensure that chemicals in the marketplace are reviewed for safety

- Challenging mission:
  - Tens of thousand of chemicals are currently in use and hundreds are introduced annually
  - Many have not been thoroughly evaluated for potential risk to human health and the environment
    - *Chemicals tested across species: Even more sparse*



Reduce Animal Testing at the US EPA

- EPA Administrator Andrew Wheeler signed directive (Sept. 10th 2019) to reduce animal testing
  - Calls for the Agency to:
    - Reduce its request for, and funding of, mammal studies by 30% by 2025
      - That is ~5 years from today!
    - Eliminate all mammal study requests and funding by 2035
      - That is ~15 years from today

How do we get there?

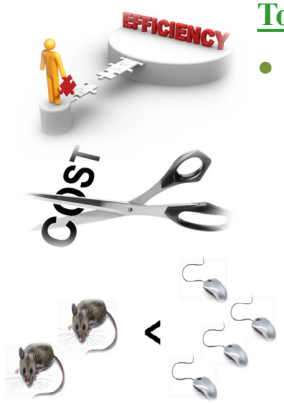
NAMs

Transformation of Toxicity Testing

Historically:

Whole animal test

- **Observe Toxic Outcome**
  - **Examples**
    - **tumor development**
    - **mortality**
  - **Resource intensive**



Toxicity Testing in the 21st Century:

- ***In vitro* and *in silico* methods**
  - **Pathway-based approaches**
  - **Focus on disturbance of the biological pathway**
    - **Predictive of the observable toxic effects**

New Approach Methods (NAMs)

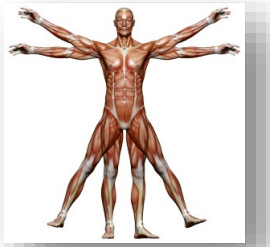
- Informatics
- High throughput
- Systems biology
- OMICs

Enabled by evolution of the science and technology

Model Organisms for Toxicity Testing

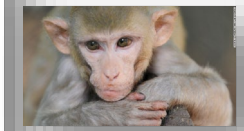
- **Assumed** that sensitivity of species to a chemical is a function of their relatedness

- **Human Health Risk Assessment**



Cannot Test

≈

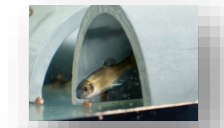
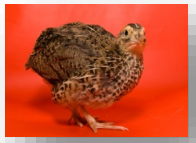


Use of Surrogates

- **Ecological Risk Assessment**

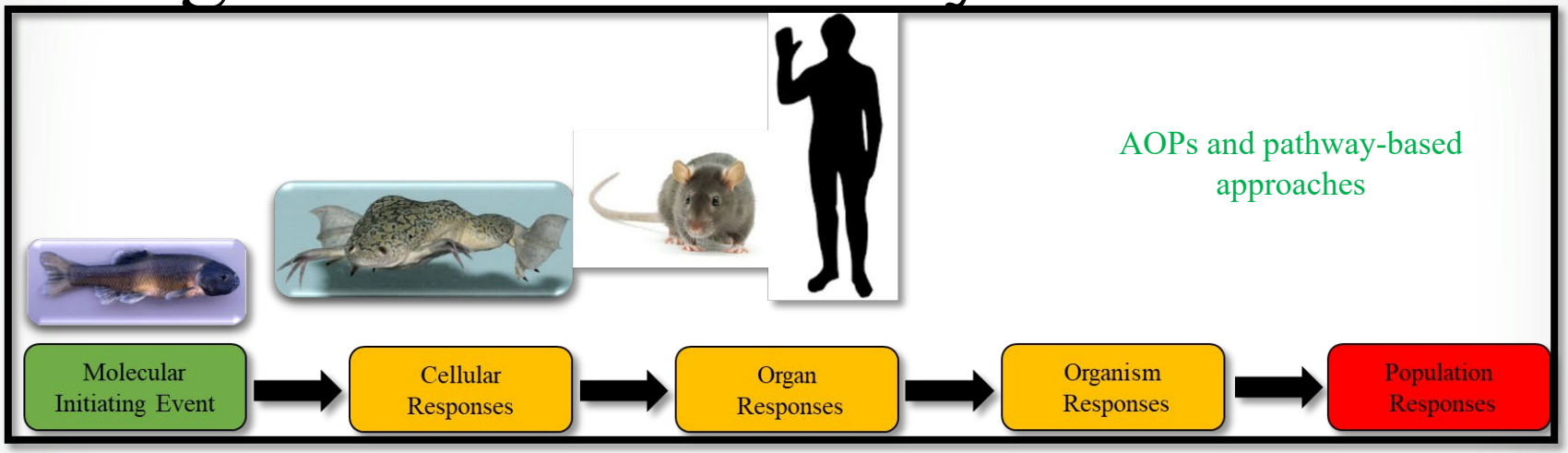
Cannot Test

≈

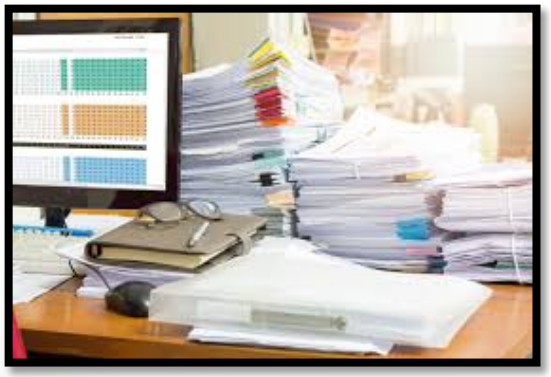
Representative species across a diversity of organism classes

Surrogates in 21st Century Chemical Safety



AOPs and pathway-based approaches

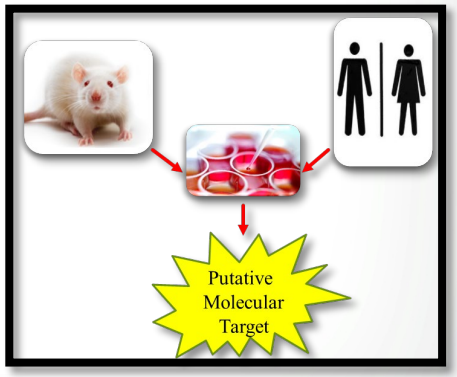
AI and Text Mining of Available Toxicity Literature



Is the model organism
A good surrogate for the
Species we are trying to protect?
Maybe

Could we gather evidence rapidly
to help us understand this question?
Absolutely!!!

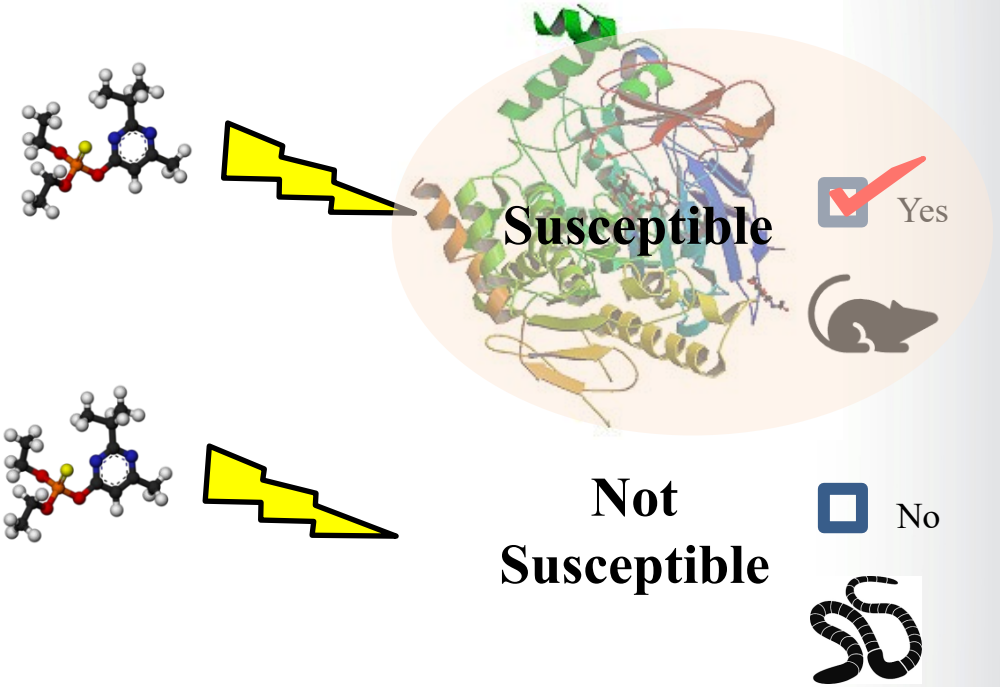
HTS



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

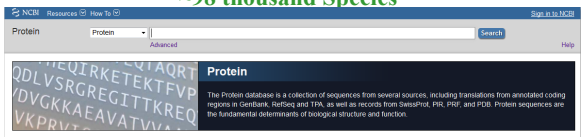
Likely susceptible or Not likely susceptible (at least through the known mechanism)

New Approach Methods: Species Extrapolation

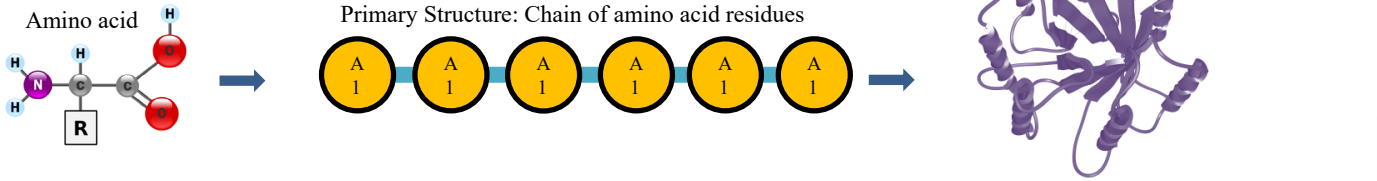
New tools and technologies have emerged

- Improved sequencing technologies
- Large databases of sequence data

As of this week
~161 million Proteins
~98 thousand Species



- **Focus on the molecular machine: The Protein**
  - **Large biomolecule assembled from amino acids encoded in genes**



  - **Many functions (e.g., catalyze reactions, structural/mechanical functions, cell signaling, immune response, etc.)**

- **Evaluate protein similarity between species**
  - **Moving away from empirical testing and qualitative understanding of molecular target (protein) conservation to quantitative measures**

EPA
United States
Environmental Protection
Agency

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

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

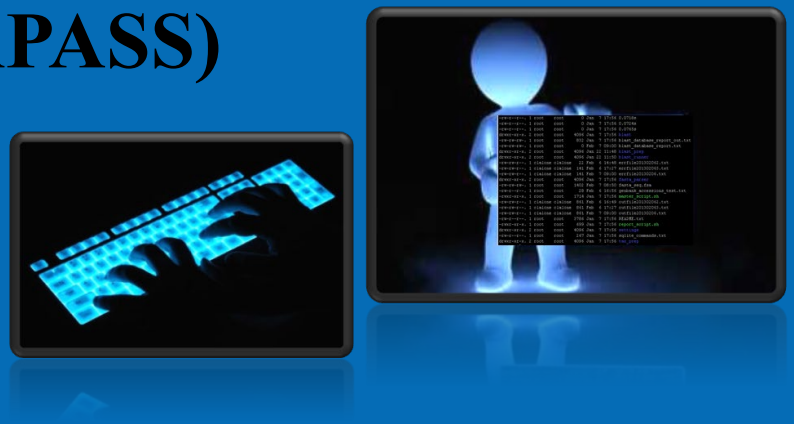
doi: 10.1093/toxsci/kfw119
Advance Access Publication Date: June 30, 2016
Research article

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

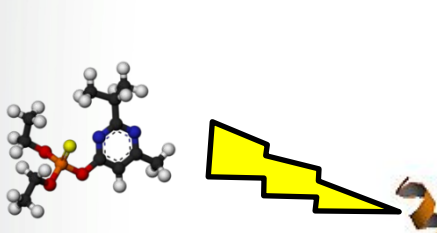
Carlie A. LaLone,*,[1] Daniel L. Villeneuve,* David Lyons,† Henry W. Helgen,‡ Serina L. Robinson,§,[2] Joseph A. Swintek,¶ Travis W. Saari,* and Gerald T. Ankley*

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

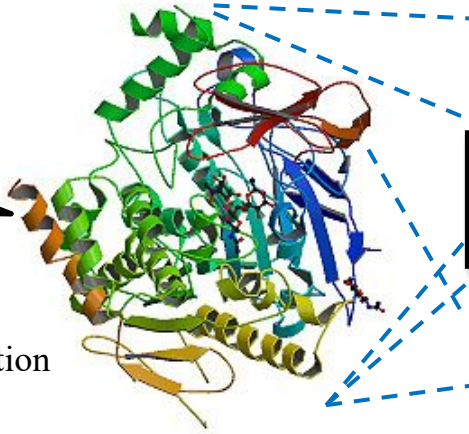
Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)



What information is required for a SeqAPASS query?



Chemical-Protein Interaction



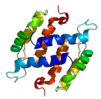
Knowledge of a sensitive or targeted species

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

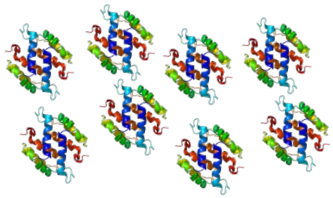
1. Protein
2. Species

Knowledge of the species for which the Key Event was developed

Chemical Molecular Target in Target Species



≅

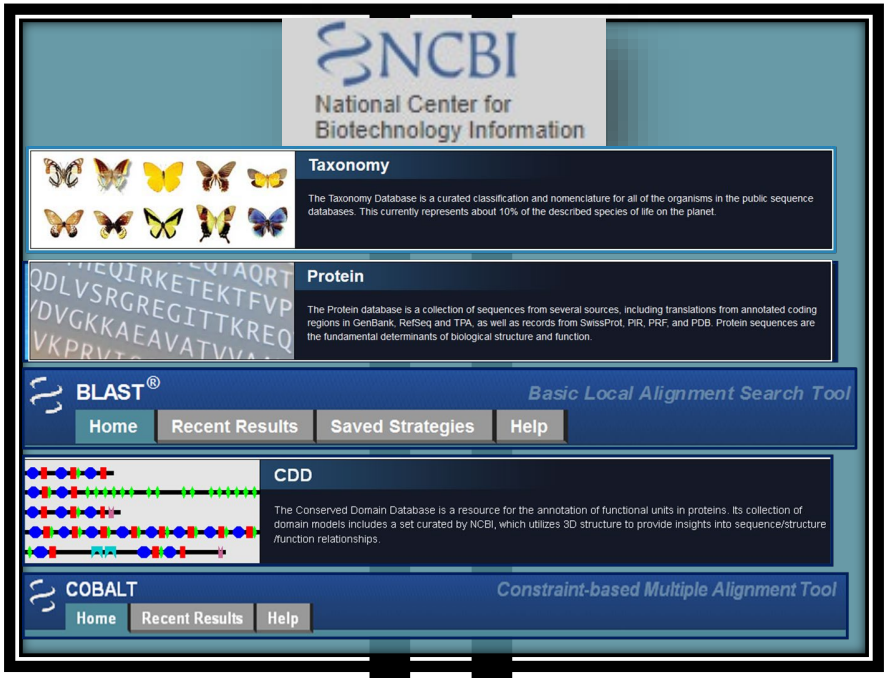


Compare to Millions of Proteins From Thousands of Species

Greater similarity = Greater likelihood that chemical can act on the protein
Line of Evidence: Predict Potential Chemical Susceptibility Across Species

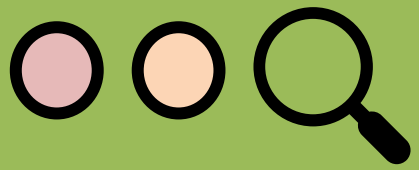
Animation by: Miguel Moravec (EPA CSS) & Andrew Patterson

SeqAPASS


NCBI
National Center for
Biotechnology Information
Taxonomy
The Taxonomy Database is a curated classification and nomenclature for all of the organisms in the public sequence databases. This currently represents about 10% of the described species of life on the planet.
Protein
The Protein database is a collection of sequences from several sources, including translations from annotated coding regions in GenBank, RefSeq and TPA, as well as records from SwissProt, PIR, PRF, and PDB. Protein sequences are the fundamental determinants of biological structure and function.
BLAST®
Basic Local Alignment Search Tool
Home
Recent Results
Saved Strategies
Help
CDD
The Conserved Domain Database is a resource for the annotation of functional units in proteins. Its collection of domain models includes a set curated by NCBI, which utilizes 3D structure to provide insights into sequence/structure /function relationships.
COBALT
Constraint-based Multiple Alignment Tool
Home
Recent Results
Help

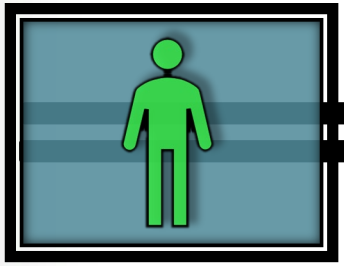
SeqAPASS

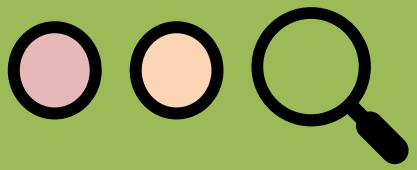
SeqAPASS Level 1

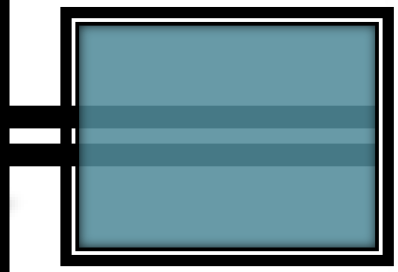


Human Protein Target

SeqAPASS Level 1







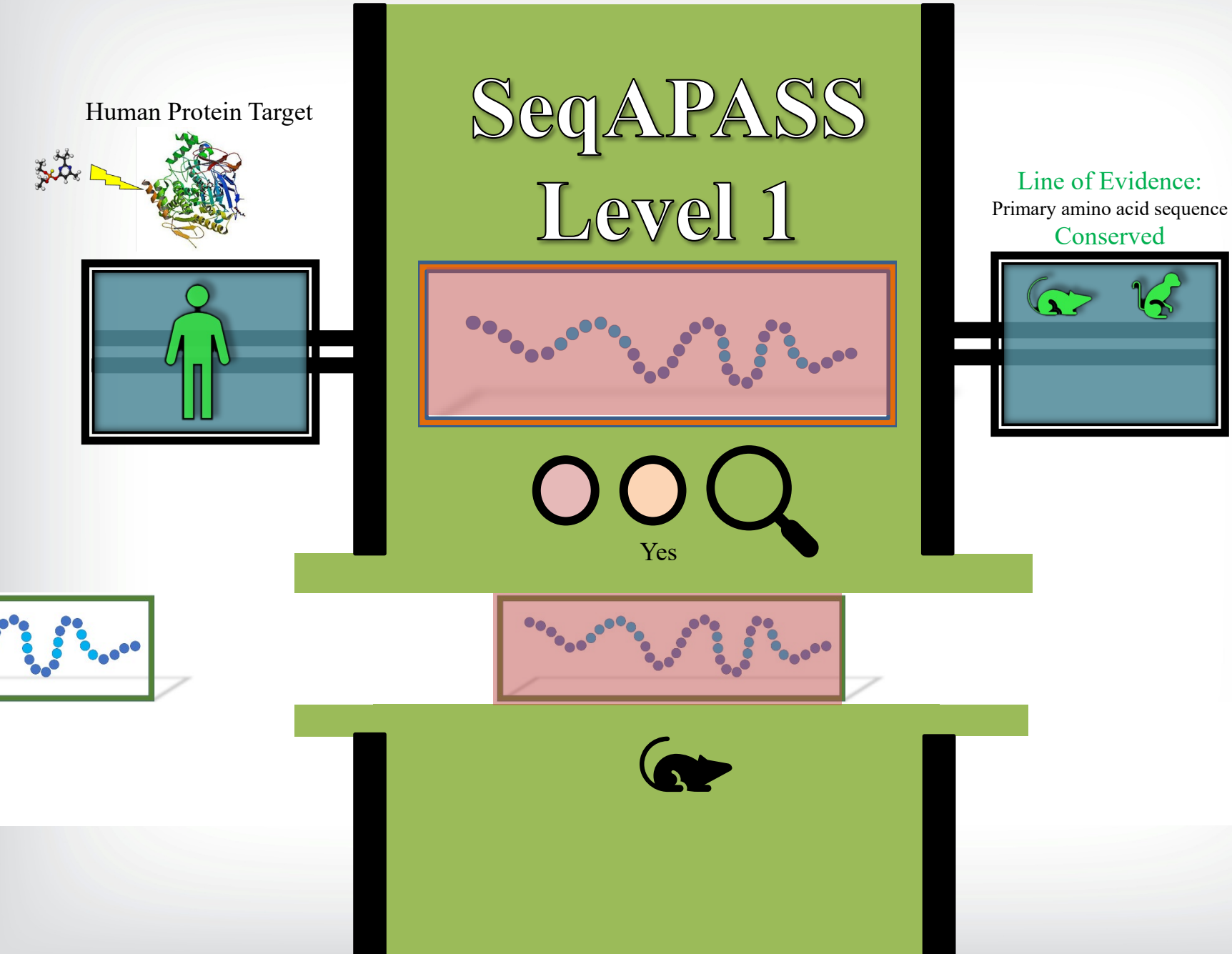
SeqAPASS Level 1

Human Protein Target

Line of Evidence:
Primary amino acid sequence
Conserved

Yes

Percent similarity


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

SeqAPASS Level 1

Human Protein Target

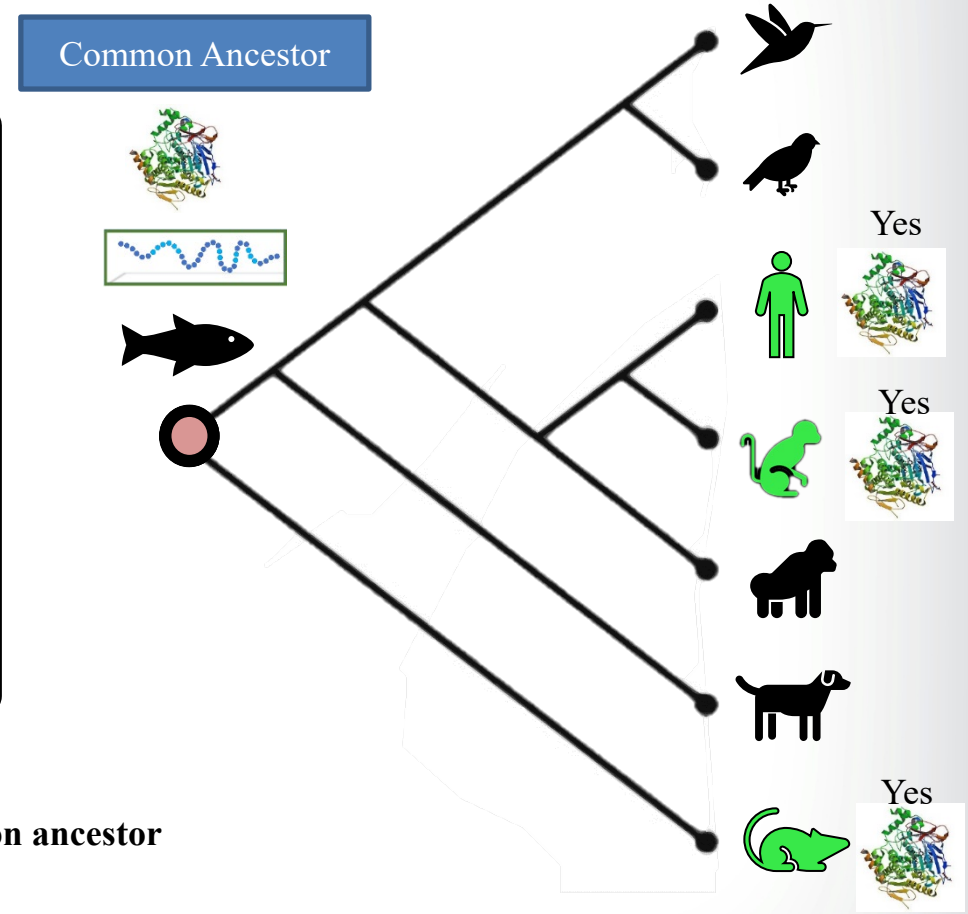
Line of Evidence:
Primary amino acid sequence
Conserved

No

SeqAPASS Level 1

Ortholog Candidate Identification

Yes



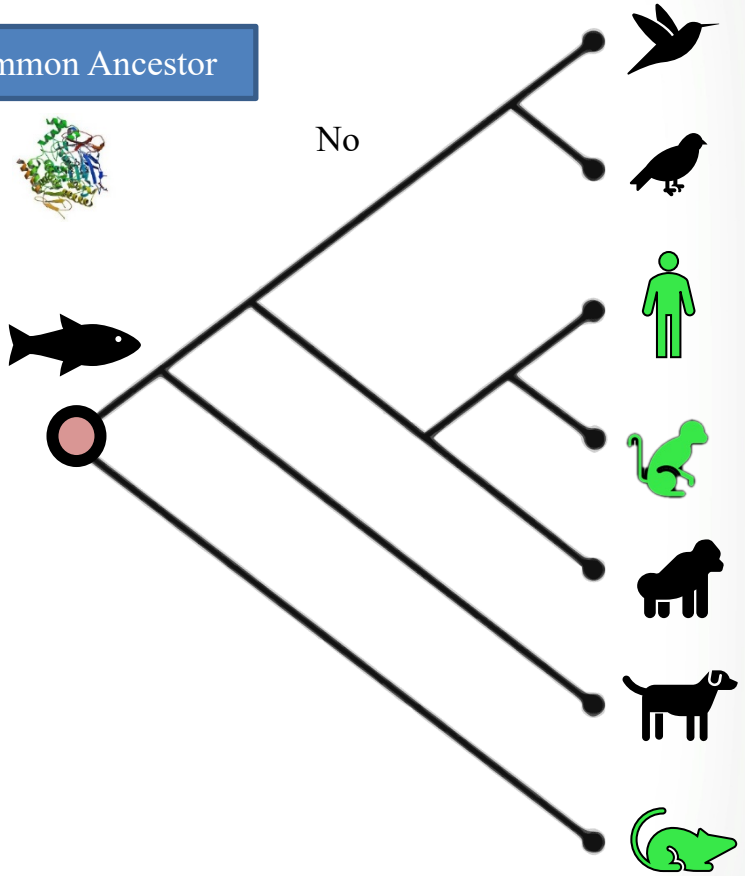
Proteins in different species that evolved from a common ancestor

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

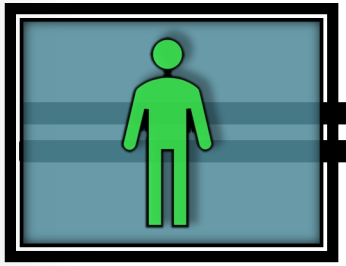
Example:

Susceptibility Cut-off: Set at 33.15

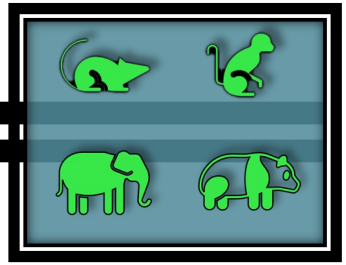
Above cut-off: More likely to be susceptible base on similar **FUNCTION**

SeqAPASS Level 1

Human Protein Target



Line of Evidence:
Primary amino acid sequence
Conserved



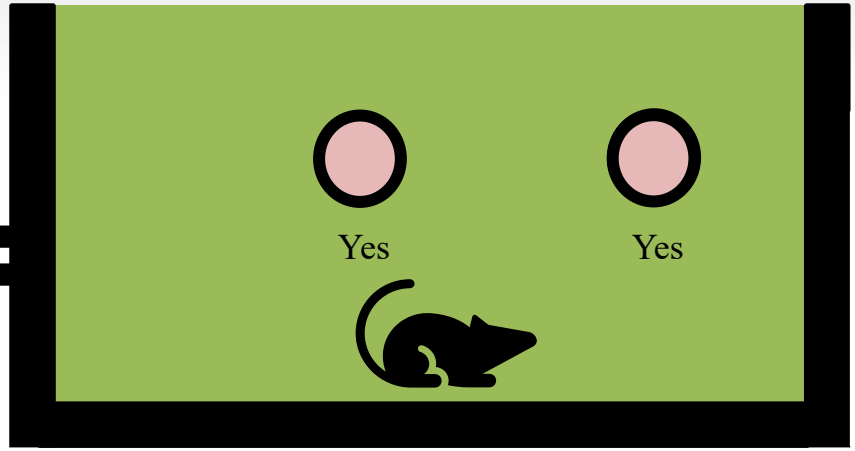
Hundreds to Thousands of Species

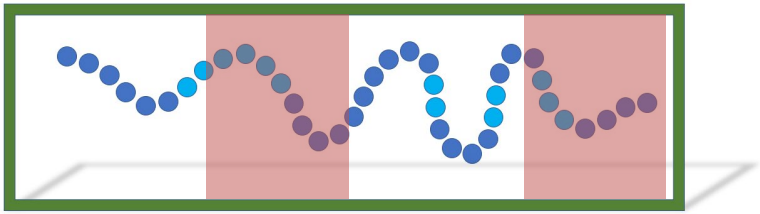
Level 2

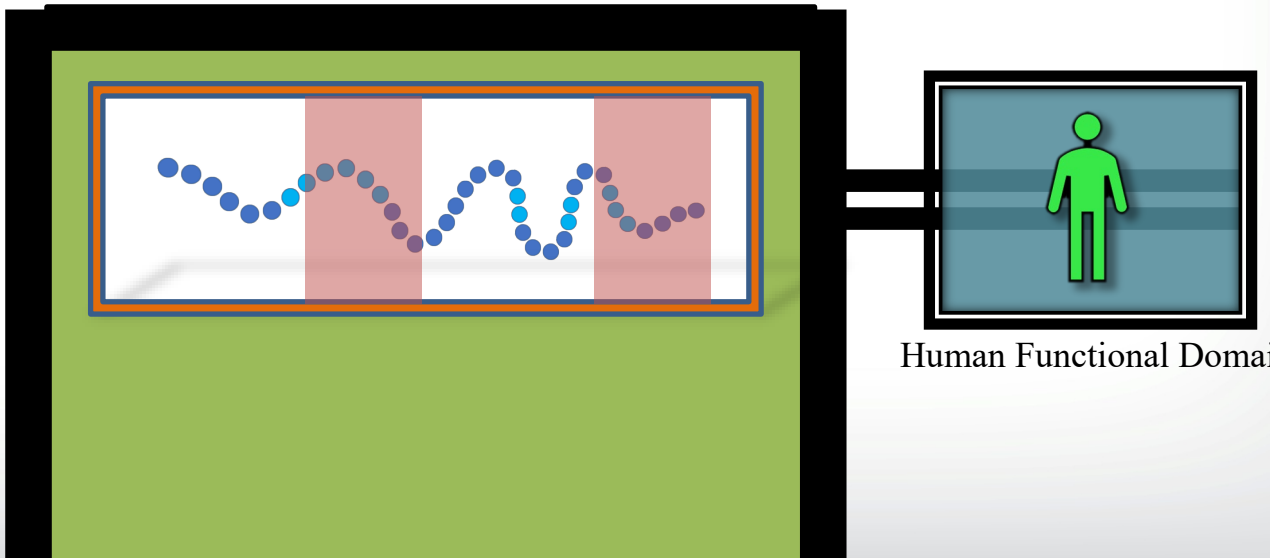
Line of Evidence:

Domain

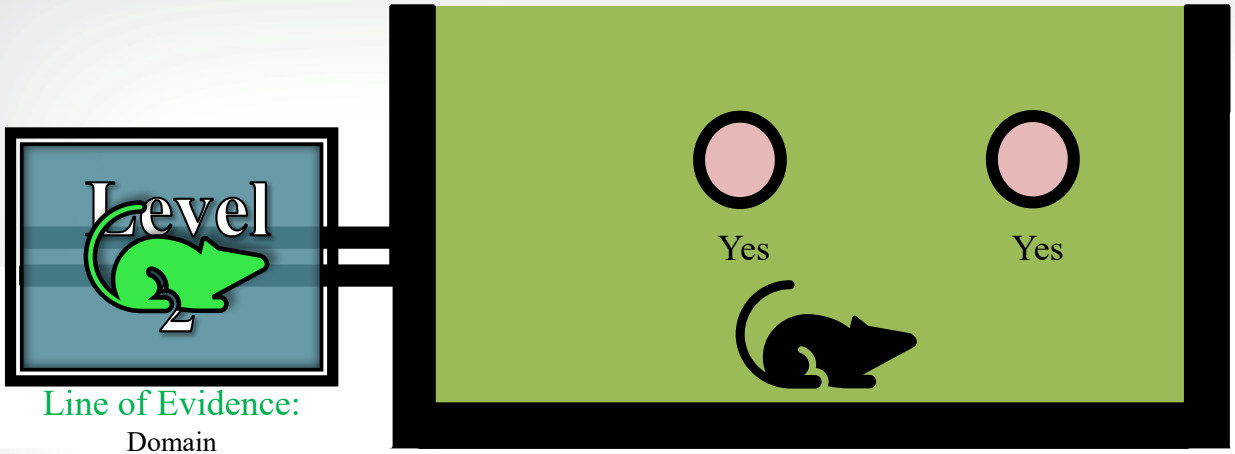
Conserved







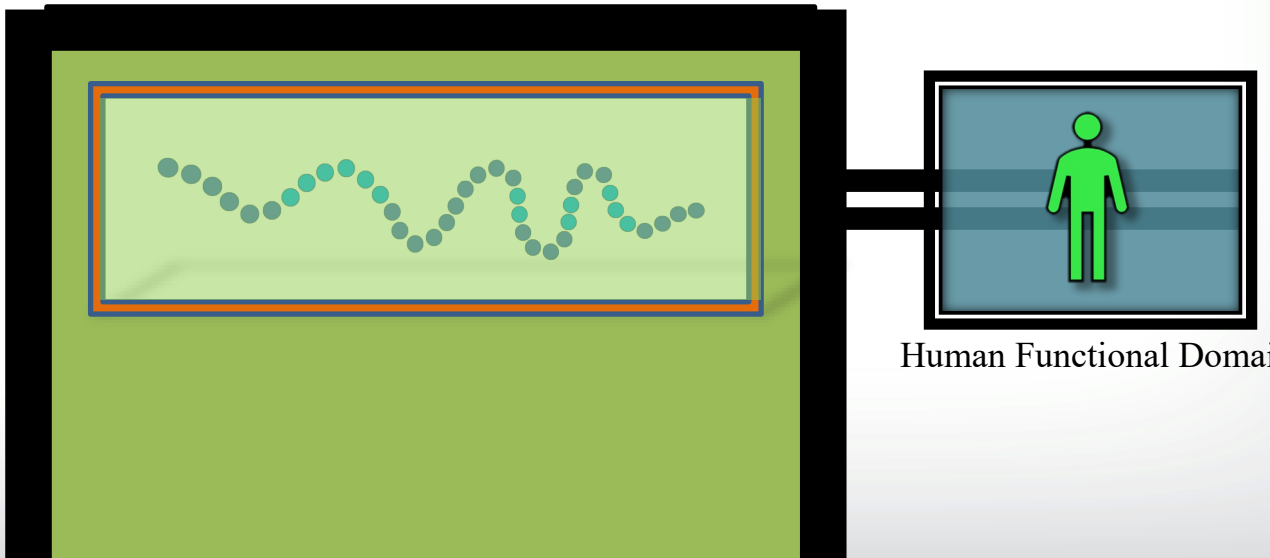
Human Functional Domain(s)



Line of Evidence:

Domain

Conserved



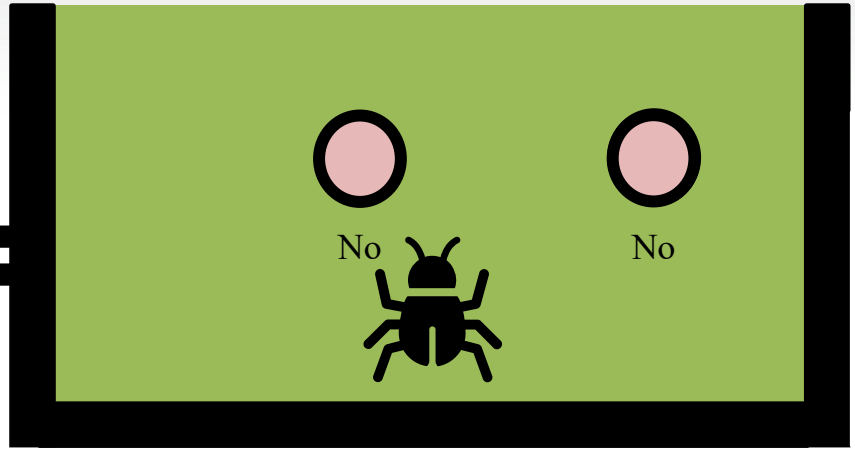
Human Functional Domain(s)

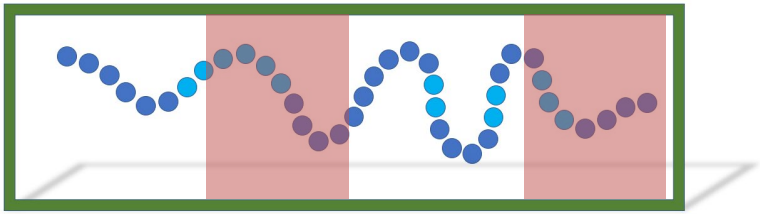
Level 2

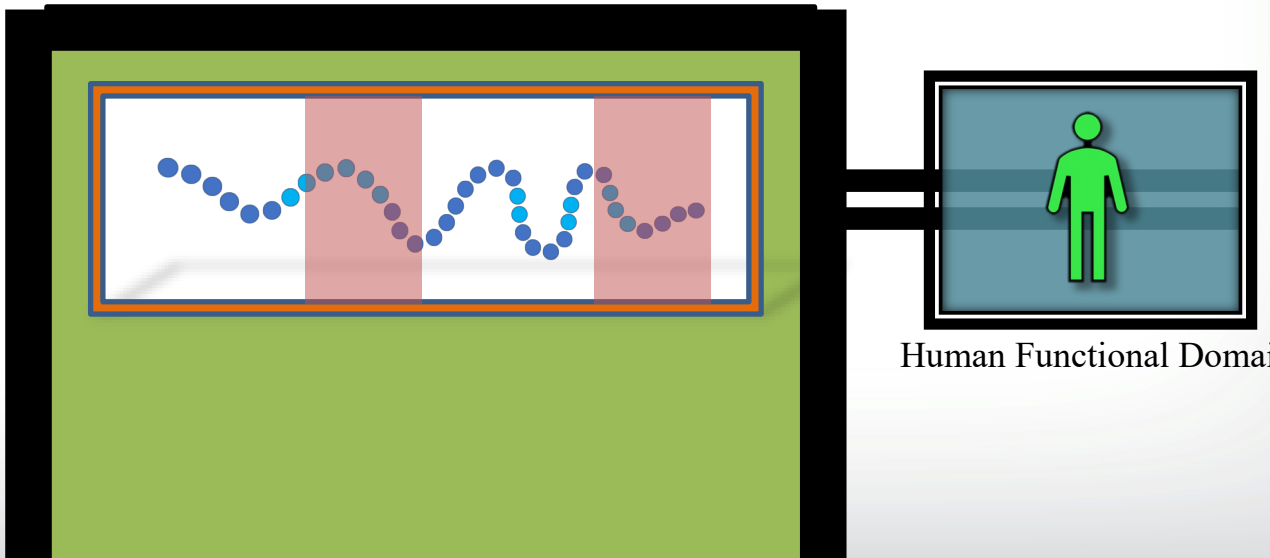
Line of Evidence:

Domain

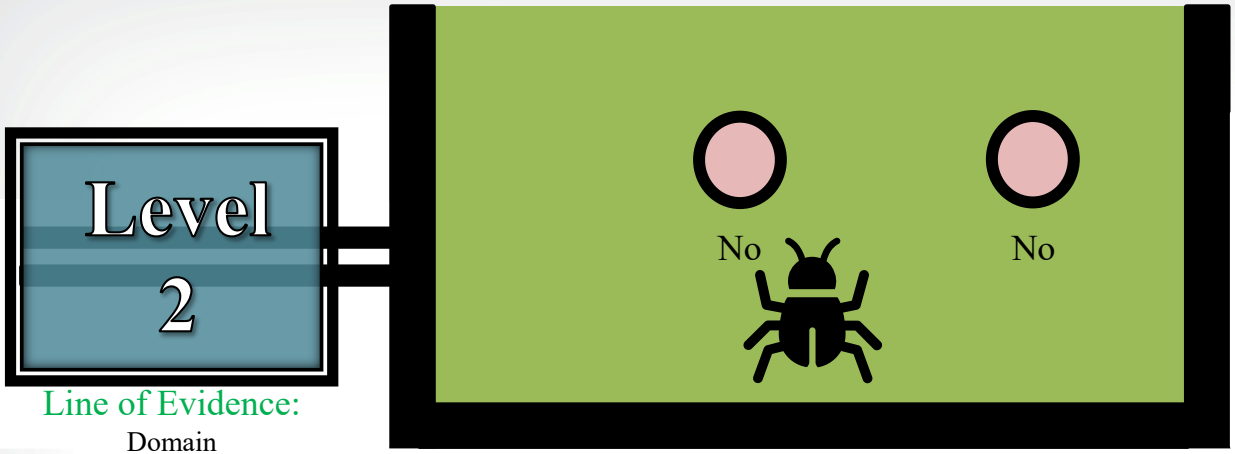
Conserved







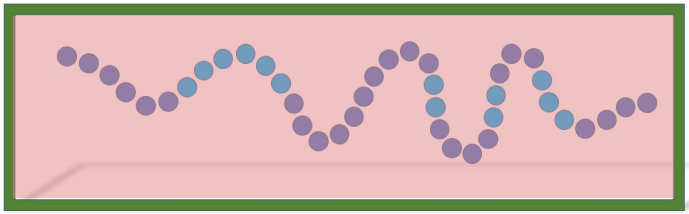
Human Functional Domain(s)

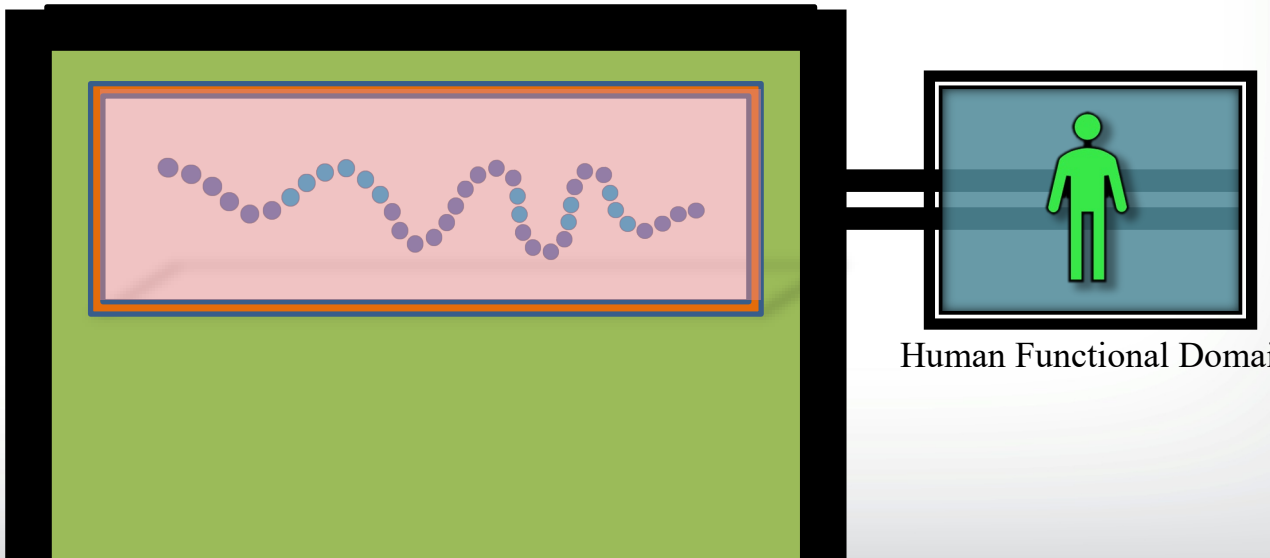


Line of Evidence:

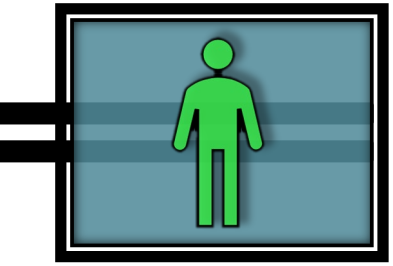
Domain

Not Conserved

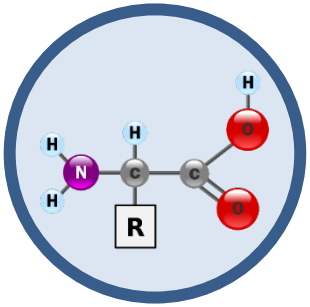


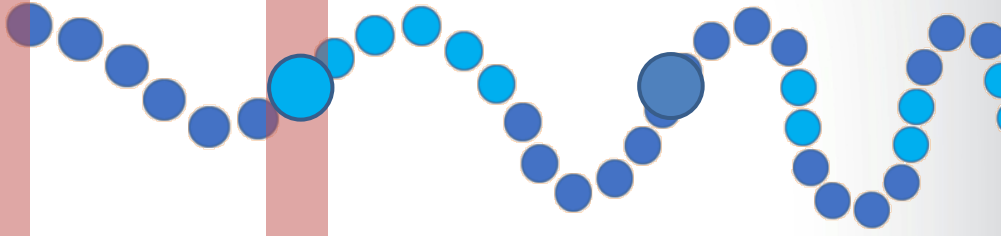


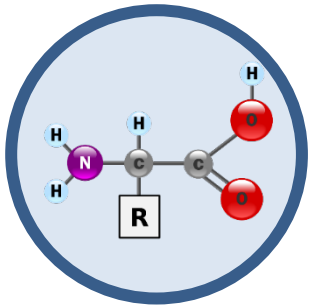
Human Functional Domain(s)

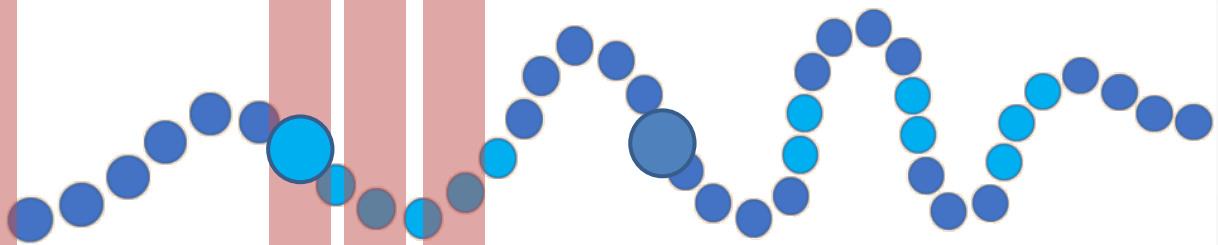


Human Critical Amino Acids

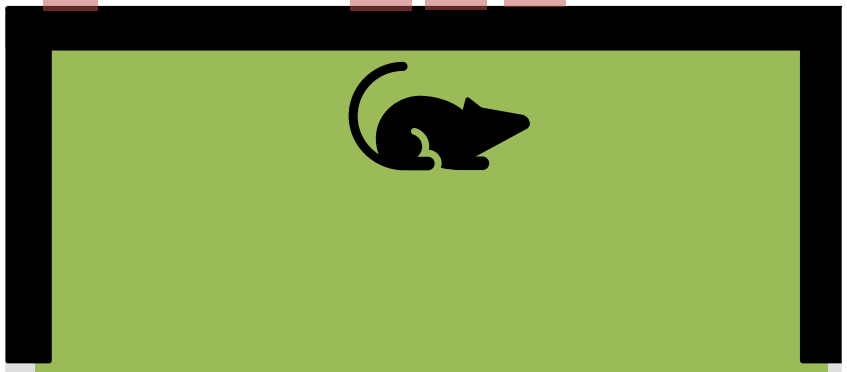


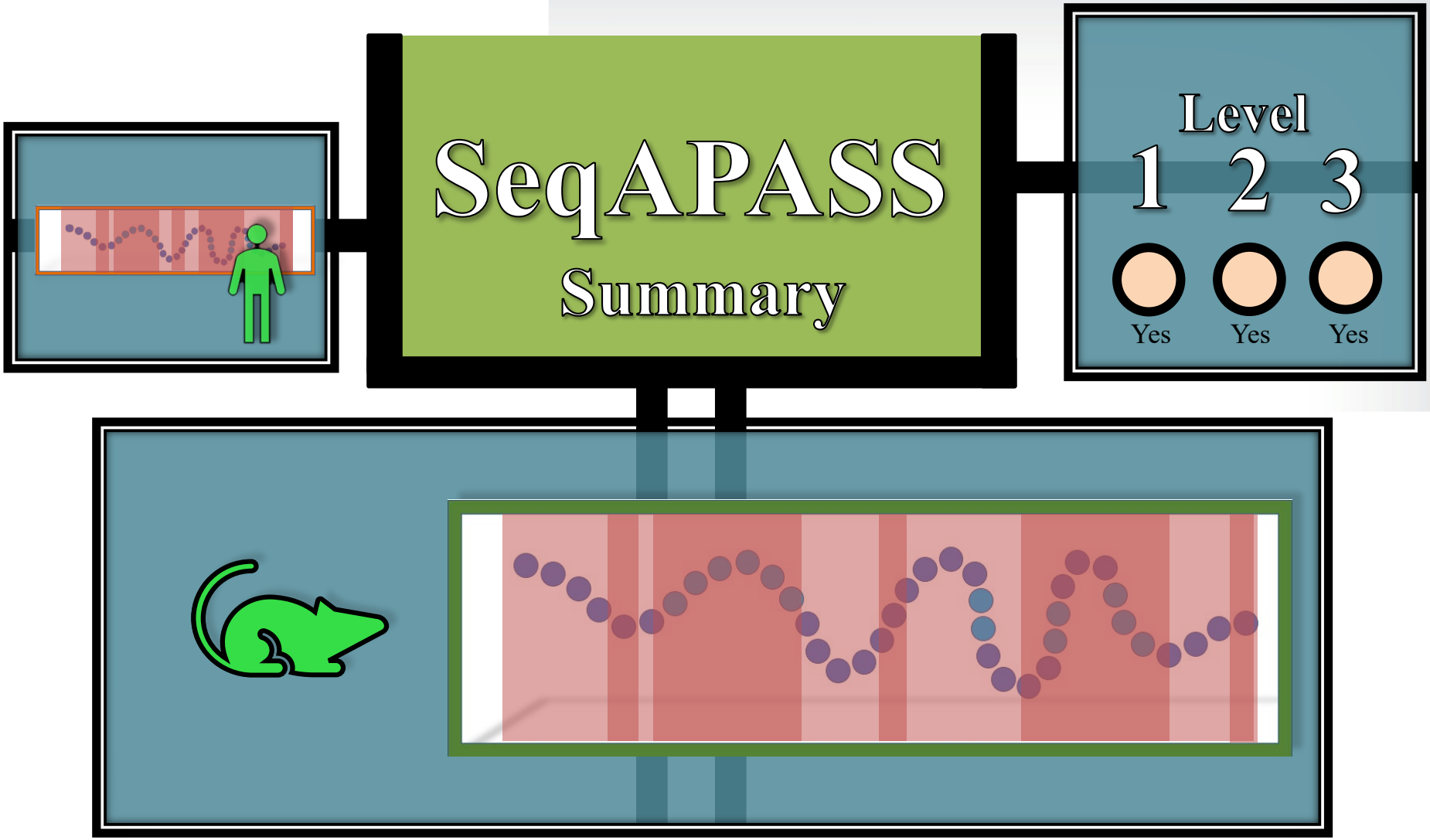






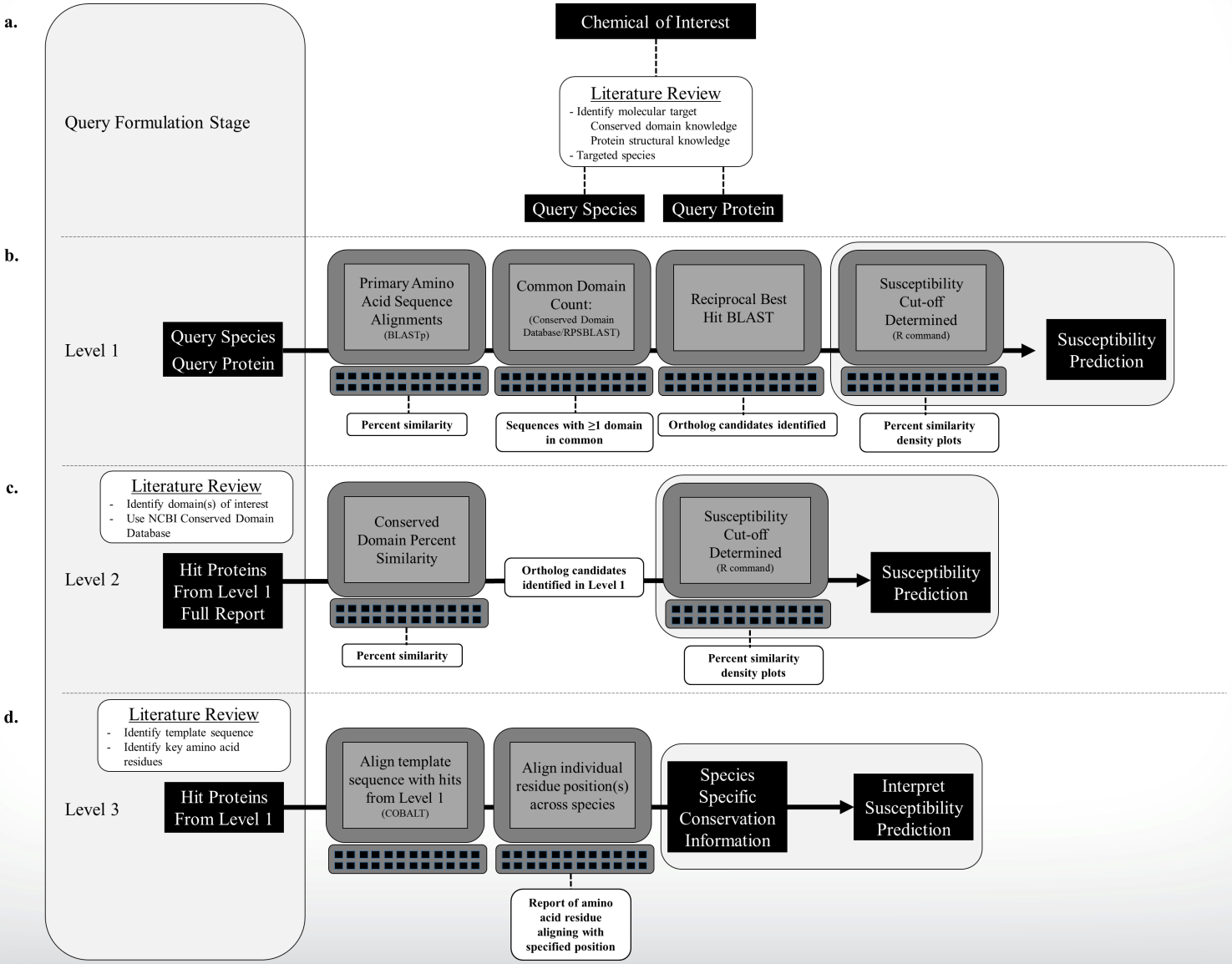
Line of Evidence: Conserved





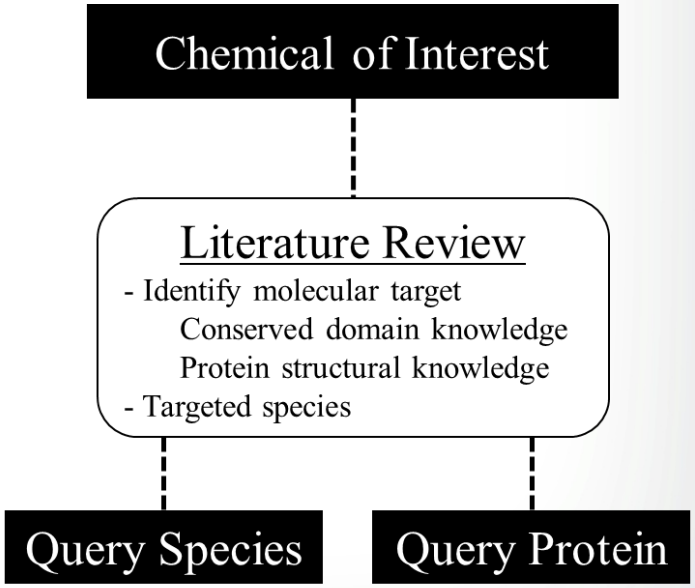
Gather Lines of Evidence for Conservation of Protein Target:
Susceptibility Prediction: Yes or No

SeqAPASS evaluation

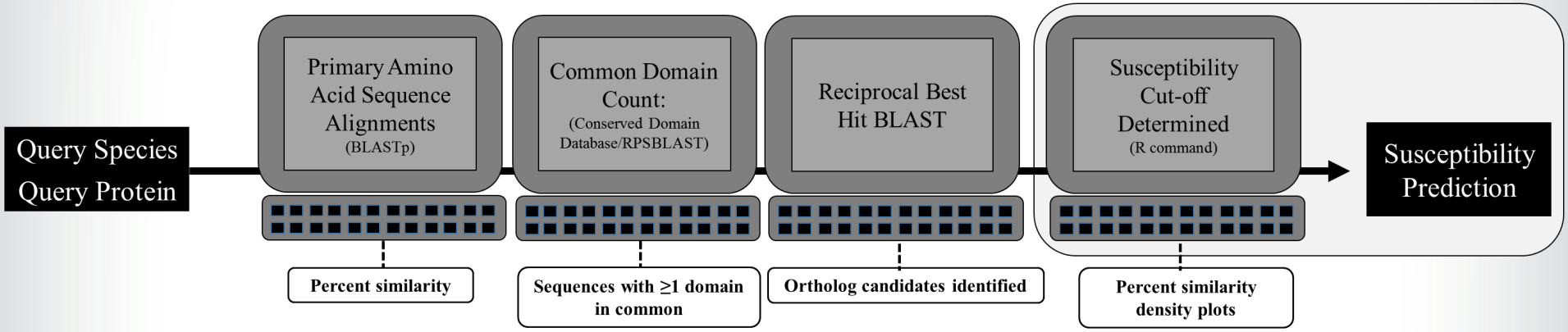


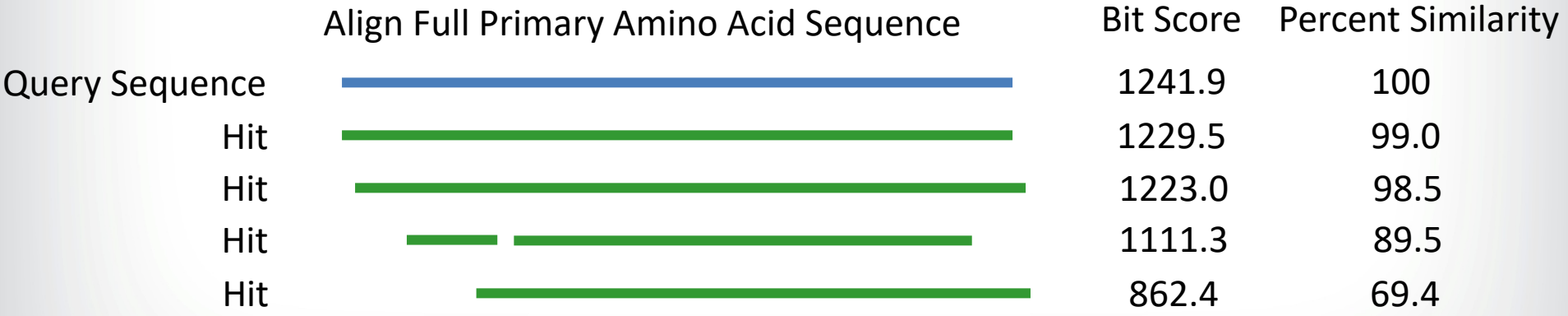
Query Formulation Stage

- **Needed information: Query Sequence**
  1. **The chemical's molecular (protein) target:**
     - **Pharmaceuticals**
       - » https://www.drugbank.ca/
       - » http://sitem.herts.ac.uk/aeru/vsdb/index.htm
       - » http://bidd.nus.edu.sg/group/cjttd/TTD_HOME.asp
     - **Pesticides and other chemicals**
       - » http://lilab.ecust.edu.cn/ptid/index.html
       - » http://www.t3db.ca/
     - **AOP initiators**
       - » https://aopwiki.org/
     - **EPA tested chemicals**
       - » https://comptox.epa.gov/dashboard
     - **Literature review**
  2. **The target species or sensitive species**
     - **Case Study:**
       - **https://cfpub.epa.gov/ecotox/**
       - **Literature review**



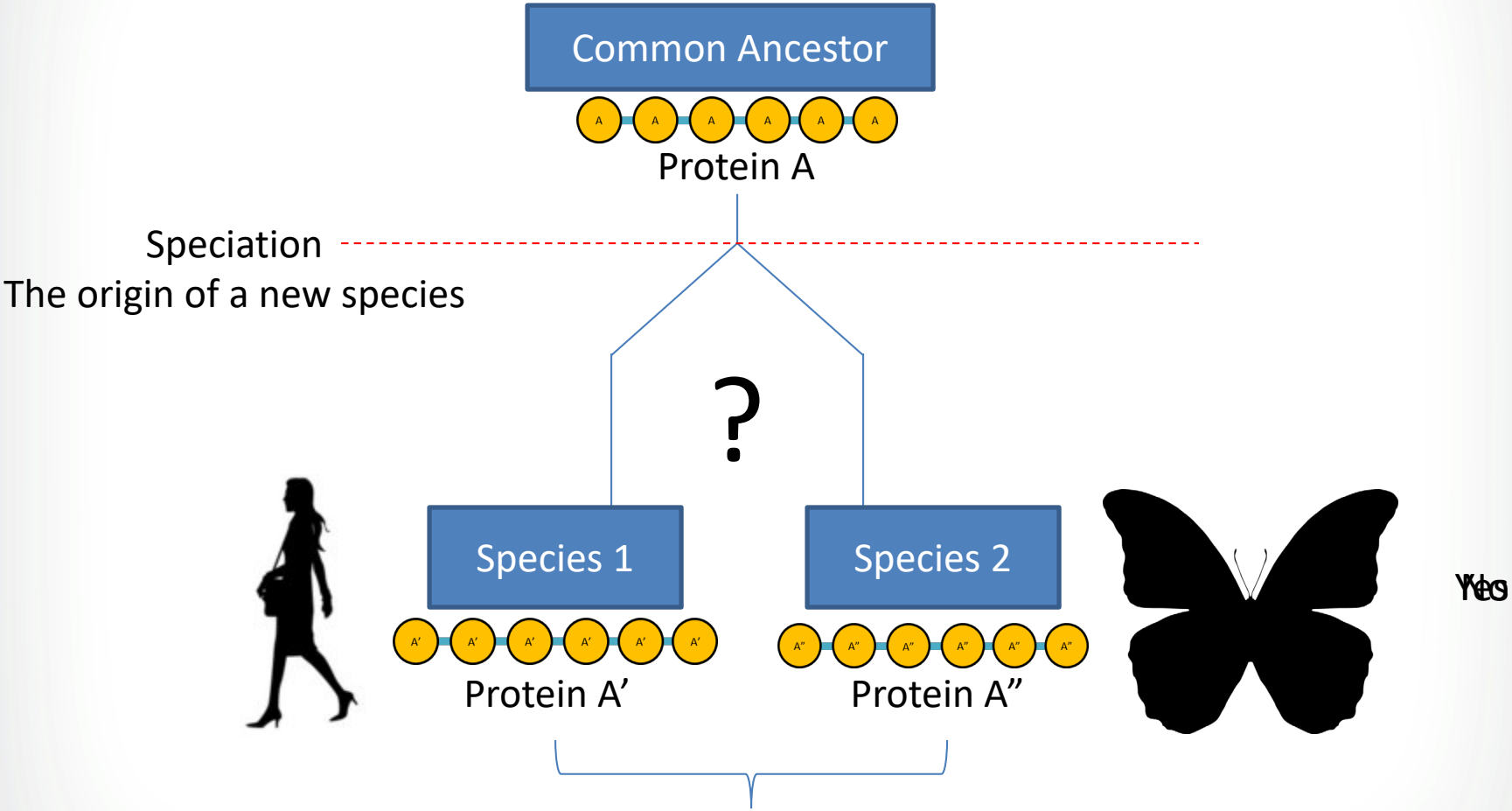
Level I: Primary Amino Acid Sequence Comparisons





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

Evolutionary Biology: Ortholog Candidate



Normally, orthologs retain the same **FUNCTION** in the course of evolution

Ortholog Candidates and Predicting Susceptibility with SeqAPASS

- **Ortholog = likely to maintain similar FUNCTION**

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

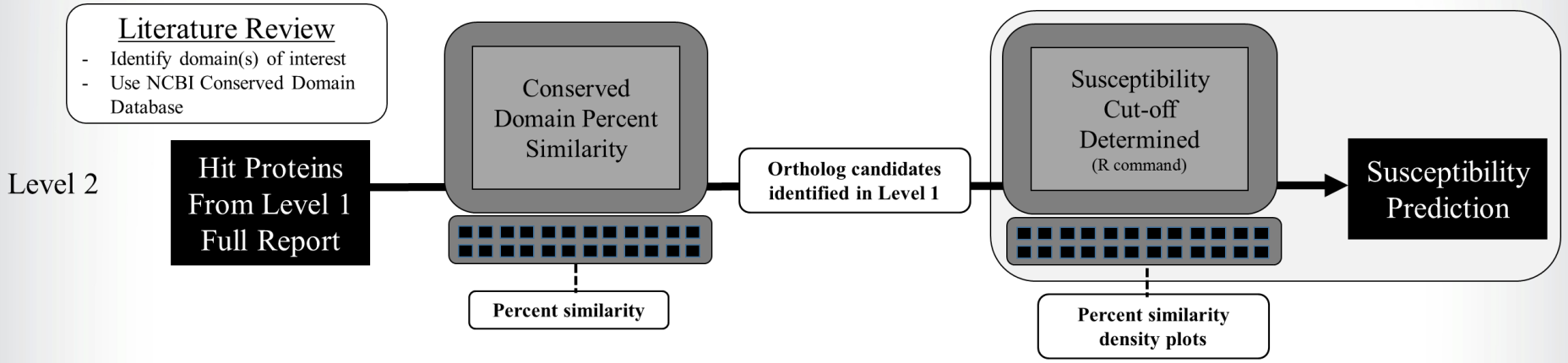
Lowest % Similarity that is still an ortholog

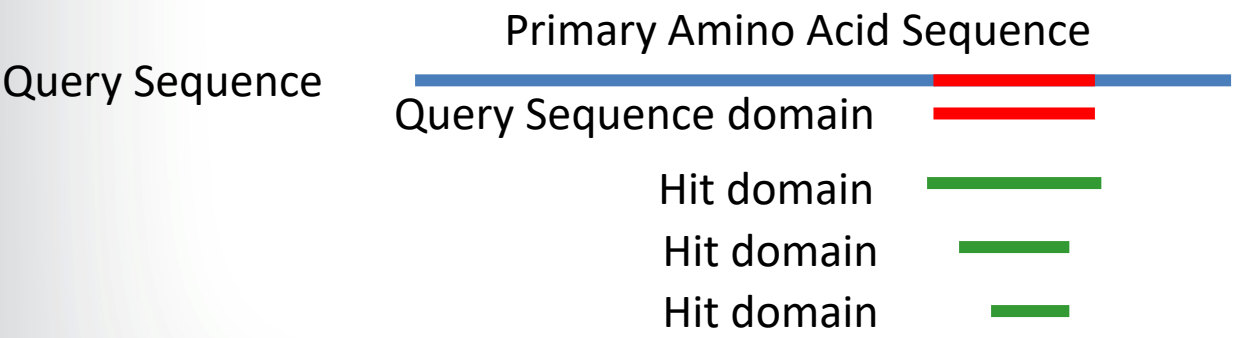
Example:

Susceptibility Cut-off: Set at 33.15

Above cut-off: More likely to be susceptible base on similar FUNCTION

Level 2: Functional domain comparisons





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

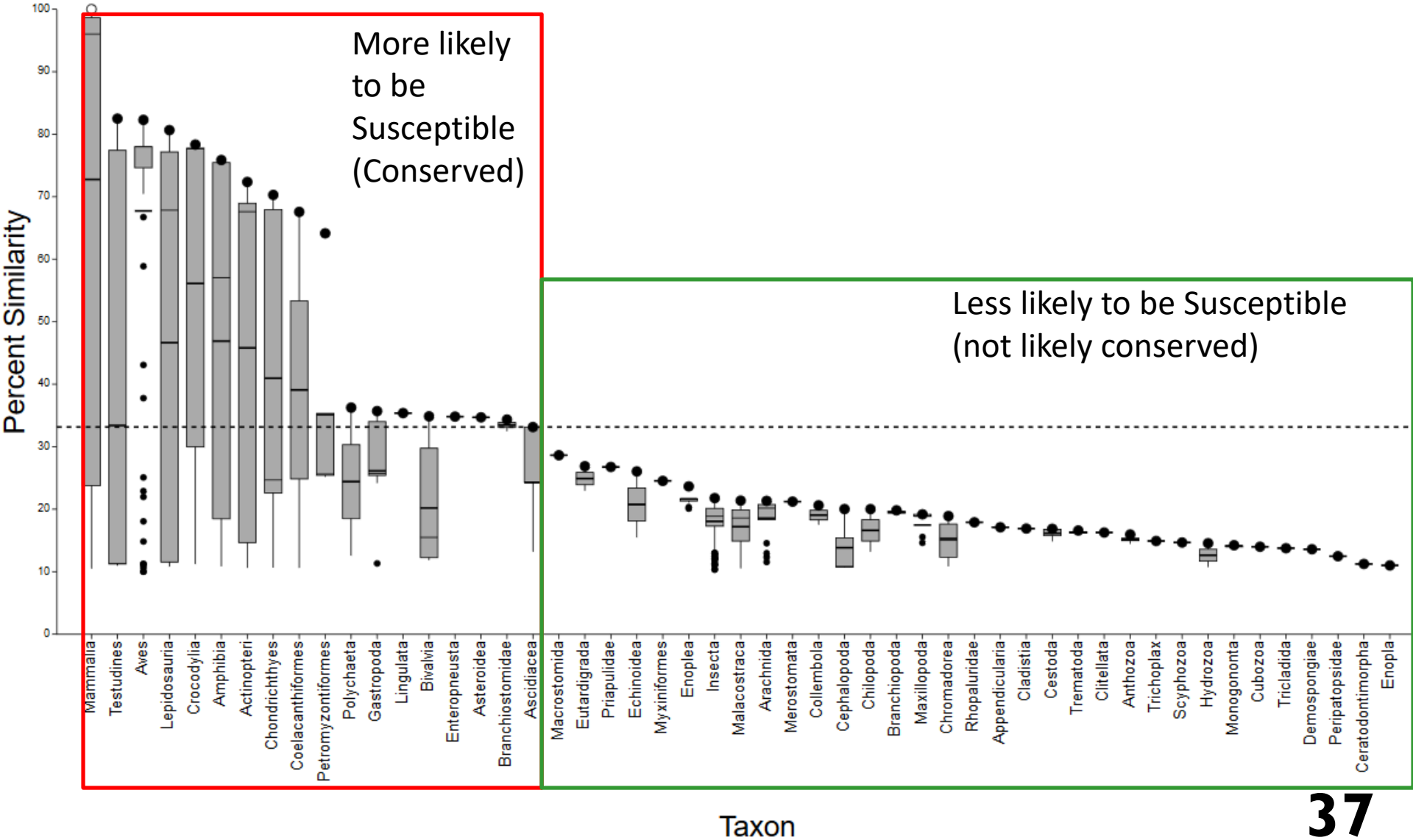
SeqAPASS output

- **Tables of data (csv file format)**

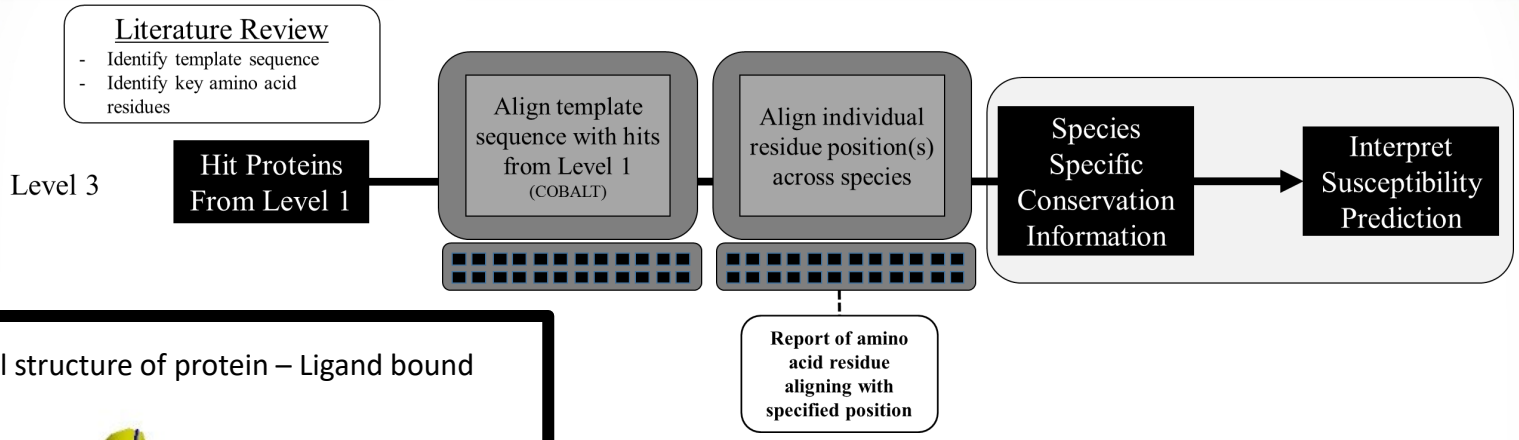
Data Versi	NCBI Acce	Protein Co	Species Ta	Taxonomi	Filtered T	Scientific	Common	Protein Na	BLASTp Bi	Ortholog (	Ortholog (	Cut-off	Percent Si	Susceptib	Analysis C	Eukaryote
3	P10827.1	1151619	9606	Mammalia	Mammalia	Homo sap	Human	RecName	1014.99	Y	26	59.28	100	Y	2018 03 19	Y
3	XP_001088	87173	9544	Mammalia	Mammalia	Macaca m	Rhesus m	PREDICTEI	1013.06	Y	26	59.28	99.81	Y	2018 03 19	Y
3	XP_003414	32342	9785	Mammalia	Mammalia	Loxodonta	African sa	PREDICTEI	1000.35	Y	26	59.28	98.56	Y	2018 03 19	Y
3	XP_003510	121846	10029	Mammalia	Mammalia	Cricetulus	Chinese h	PREDICTEI	996.5	Y	26	59.28	98.18	Y	2018 03 19	Y
3	XP_016057	29924	291302	Mammalia	Mammalia	Miniopter	Common	PREDICTEI	989.56	Y	26	59.28	97.49	Y	2018 03 19	Y
3	NP_11239	150717	10116	Mammalia	Mammalia	Rattus nor	Norway ra	c-erbAalp	989.56	Y	26	59.28	97.49	Y	2018 03 19	Y
3	O97716.1	84581	9823	Mammalia	Mammalia	Sus scrofa	Pig	RecName	987.64	Y	26	59.28	97.31	Y	2018 03 19	Y
3	NP_00130	306159	10090	Mammalia	Mammalia	Mus musc	House mo	thyroid ho	986.87	Y	26	59.28	97.23	Y	2018 03 19	Y
3	XP_006971	45629	230844	Mammalia	Mammalia	Peromysc	Prairie de	PREDICTEI	985.33	Y	26	59.28	97.08	Y	2018 03 19	Y
3	XP_004684	29273	143302	Mammalia	Mammalia	Condylura	Star-nose	PREDICTEI	984.17	Y	26	59.28	96.96	Y	2018 03 19	Y
3	XP_004591	26203	9978	Mammalia	Mammalia	Ochotona	American	PREDICTEI	982.25	Y	26	59.28	96.77	Y	2018 03 19	Y

SeqAPASS data (csv format) for all ToxCast Assays available through
The CompTox Chemistry Dashboard
https://comptox.epa.gov/dashboard

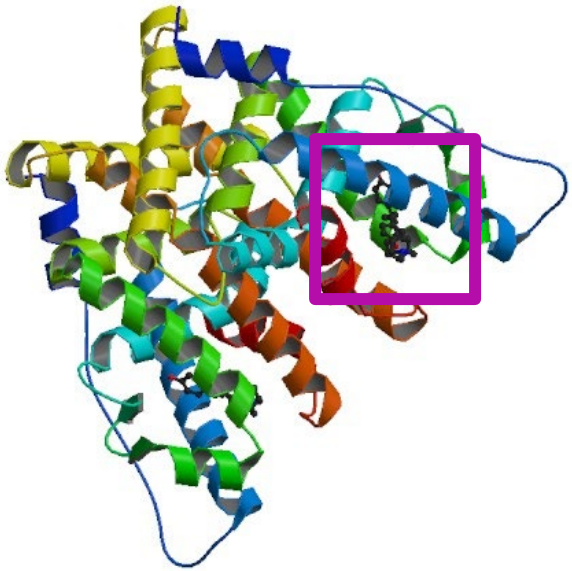
Box-Plot of SeqAPASS Data



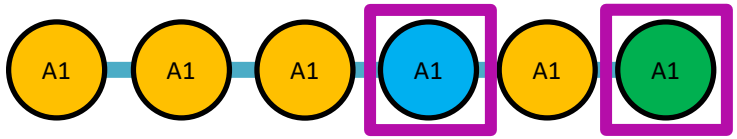
Level 3: Individual Amino Acid Residue Comparison



Crystal structure of protein – Ligand bound



Amino acid residues that interact with the chemical



Where can I find this information?

- Literature Review
  - Types of studies I'd be looking for:
    - Site-directed mutagenesis
    - Field resistance (pesticides)
    - Studies of x-ray crystallography
    - Homology modeling

This is not trivial….it can take some time to identify literature and formulate the Level 3 query

Level 3: Individual Amino Acid Residue Comparison

SeqAPASS can **AUTOMATICALLY** predict whether an amino acid difference is likely to change protein-chemical interaction.

Amino Acid	1-Letter	Side Chain Class	Molecular Weight (g/mol)
Aspartic Acid	D	Acidic	133.104
Glutamic Acid	E	Acidic	147.131
Alanine	A	Aliphatic	89.094
Glycine	G	Aliphatic	75.067
Isoleucine	I	Aliphatic	131.175
Leucine	L	Aliphatic	131.175
Proline	P	Aliphatic	115.132
Valine	V	Aliphatic	117.148
Asparagine	N	Amidic	132.119
Glutamine	Q	Amidic	146.146
Phenylalanine	F	Aromatic	165.192
Tryptophan	W	Aromatic	204.228
Tyrosine	Y	Aromatic	181.191
Histidine	H	Basic	155.156
Lysine	K	Basic	146.189
Arginine	R	Basic	174.203
Serine	S	Hydroxylic	105.093
Threonine	T	Hydroxylic	119.119
Methionine	M	Sulfur-Containing	149.208
Cysteine	C	Sulfur-Containing	121.154

- Automated Prediction
  - Each key amino acid is grouped into a “side chain class” and given a “size”.
  - Evaluated based on Rules:
    - Same side chain class as query (Y/N)?
    - Size 30g/mol or less from query (Y/N)?
  - If 2 “N” responses for ≥1 key amino acids then SeqAPASS predicts susceptibility of that species will differ from the query species.

Level 3: Individual Amino Acid Residue Comparison

Example of how Level 3 Susceptibility Predictions work (hypothetical case study):

Protein	Shared Susceptibility	Position 1	Residue	Side Chain Class	Match	Size (MW)	Match
Mouse protein	Yes	101	F	Aromatic	Yes	165.192	Yes
Human protein	Yes	101	F	Aromatic	Yes	165.192	Yes
Bird protein	Yes	101	Y	Aromatic	Yes	181.191	Yes
Turtle protein	Yes	101	R	Basic	No	174.203	Yes
Frog protein	No	101	D	Acidic	No	133.104	No
Fish protein	No	101	N	Amidic	No	132.119	No
Insect protein	No	101	A	Aliphatic	No	89.094	No

Level 3: Individual Amino Acid Residue Comparison

This example compares protein targets from 6 different species to Mouse as the query species.

Protein	Shared Susceptibility	Position 1	Residue	Side Chain Class	Match	Size (MW)	Match
Mouse protein	Yes	101	F	Aromatic	Yes	165.192	Yes
Human protein	Yes	101	F	Aromatic	Yes	165.192	Yes
Bird protein	Yes	101	Y	Aromatic	Yes	181.191	Yes
Turtle protein	Yes	101	R	Basic	No	174.203	Yes
Frog protein	No	101	D	Acidic	No	133.104	No
Fish protein	No	101	N	Amidic	No	132.119	No
Insect protein	No	101	A	Aliphatic	No	89.094	No

Level 3: Individual Amino Acid Residue Comparison

This example has 1 key amino acid position, but most proteins have 10 to 20 key positions that are directly involved in protein-chemical interactions.

Protein	Shared Susceptibility	Position 1	Residue	Side Chain Class	Match	Size (MW)	Match
Mouse protein	Yes	101	F	Aromatic	Yes	165.192	Yes
Human protein	Yes	101	F	Aromatic	Yes	165.192	Yes
Bird protein	Yes	101	Y	Aromatic	Yes	181.191	Yes
Turtle protein	Yes	101	R	Basic	No	174.203	Yes
Frog protein	No	101	D	Acidic	No	133.104	No
Fish protein	No	101	N	Amidic	No	132.119	No
Insect protein	No	101	A	Aliphatic	No	89.094	No

Level 3: Individual Amino Acid Residue Comparison

SeqAPASS compares the side-chain class of each amino acid from each species and whether it is a match to the query (mouse).

Protein	Shared Susceptibility	Position 1	Residue	Side Chain Class	Match	Size (MW)	Match
Mouse protein	Yes	101	F	Aromatic	Yes	165.192	Yes
Human protein	Yes	101	F	Aromatic	Yes	165.192	Yes
Bird protein	Yes	101	Y	Aromatic	Yes	181.191	Yes
Turtle protein	Yes	101	R	Basic	No	174.203	Yes
Frog protein	No	101	D	Acidic	No	133.104	No
Fish protein	No	101	N	Amidic	No	132.119	No
Insect protein	No	101	A	Aliphatic	No	89.094	No

Level 3: Individual Amino Acid Residue Comparison

SeqAPASS compares the molecular weight of each amino acid from each species and whether the difference is less than 30 g/mol from query (mouse).

Protein	Shared Susceptibility	Position 1	Residue	Side Chain Class	Match	Size (MW)	Match
Mouse protein	Yes	101	F	Aromatic	Yes	165.192	Yes
Human protein	Yes	101	F	Aromatic	Yes	165.192	Yes
Bird protein	Yes	101	Y	Aromatic	Yes	181.191	Yes
Turtle protein	Yes	101	R	Basic	No	174.203	Yes
Frog protein	No	101	D	Acidic	No	133.104	No
Fish protein	No	101	N	Amidic	No	132.119	No
Insect protein	No	101	A	Aliphatic	No	89.094	No

Level 3: Individual Amino Acid Residue Comparison

To provide conservative predictions, two "No" matches for one or more amino acids are required for SeqAPASS to predict that the species differ in susceptibility.

Protein	Shared Susceptibility	Position 1	Residue	Side Chain Class	Match	Size (MW)	Match
Mouse protein	Yes	101	F	Aromatic	Yes	165.192	Yes
Human protein	Yes	101	F	Aromatic	Yes	165.192	Yes
Bird protein	Yes	101	Y	Aromatic	Yes	181.191	Yes
Turtle protein	Yes	101	R	Basic	No	174.203	Yes
Frog protein	No	101	D	Acidic	No	133.104	No
Fish protein	No	101	N	Amidic	No	132.119	No
Insect protein	No	101	A	Aliphatic	No	89.094	No

APPLICATION OF SeqAPASS
MIE
KE
KE
KE
AO

Contents lists available at ScienceDirect

Aquatic Toxicology

journal homepage: www.elsevier.com/locate/aquatox

Molecular target sequence similarity as a basis for species extrapolation to assess the ecological risk of chemicals with known modes of action

Carlie A. LaLone[a,*], Daniel L. Villeneuve[a], Lyle D. Burgoon[b], Christine L. Russom[a], Henry W. Helgen[c], Jason P. Berninger[d], Joseph E. Tietge[a], Megan N. Severson[a], Jenna E. Cavallin[e], Gerald T. Ankley[a]

Case studies: 17alpha-ethinyl estradiol, human, estrogen receptor; Permethrin, mosquito, voltage-gated sodium channel; 17β-trenbolone, bovine, androgen receptor

Environmental Toxicology and Chemistry, Vol. 35, No. 11, pp. 2806–2816, 2016
Published 2016 SETAC
Printed in the USA

EVALUATION OF THE SCIENTIFIC UNDERPINNINGS FOR IDENTIFYING ESTROGENIC CHEMICALS IN NONMAMMALIAN TAXA USING MAMMALIAN TEST SYSTEMS

GERALD T. ANKLEY,*† CARLIE A. LALONE,† L. EARL GRAY,‡ DANIEL L. VILLENEUVE,† and MICHAEL W. HORNUNG†
†Mid-Continent Ecology Division, US Environmental Protection Agency, Duluth, Minnesota
‡Toxicity Assessment Division, US Environmental Protection Agency, Research Triangle Park, North Carolina

Case study: Human estrogen receptor alpha

Article
Cite This: Environ. Sci. Technol. 2018, 52, 13960−13971
pubs.acs.org/est

Evidence for Cross Species Extrapolation of Mammalian-Based High-Throughput Screening Assay Results

Carlie A. LaLone,*,† Daniel L. Villeneuve,† Jon A. Doering,‡ Brett R. Blackwell,† Thomas R. Transue,§ Cody W. Simmons,§ Joe Swintek,¶ Sigmund J. Degitz,† Antony J. Williams,∥ and Gerald T. Ankley†

Case studies: All mammalian-based ToxCast Targets (484)

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*

Case studies: Diacylhydrazines and bisacylhydrazine, tobacco budworm, ecdysone receptor; Neonicotinoids, honey bee, nicotinic acetylcholine receptor; Strobilurin fungicides, corn rust, cytochrome b

TOXICOLOGICAL SCIENCES, 166(1), 2018, 131–145
doi: 10.1093/toxsci/kfy186
Dryad Digital Repository DOI: https://doi.org/10.5061/dryad.2tg6967
Advance Access Publication Date: July 27, 2018
Research Article

In Silico Site-Directed Mutagenesis Informs Species-Specific Predictions of Chemical Susceptibility Derived From the Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) Tool

Jon A. Doering,*,† Sehan Lee,‡,§ Kurt Kristiansen,¶ Linn Evenseth,¶ Mace G. Barron,‡ Ingebrigt Sylte,¶ and Carlie A. LaLone*,[1]

Case studies: Mouse acetylcholinesterase; Tobacco budworm ecdysone receptor

SETAC SciCon

SETAC Europe 30th Annual Meeting

Open Science for Enhanced Global Environmental Protection

SETAC EU Posters: Applications and future directions for the SeqAPASS tool

ID	Title	Presenter	Type	Session	Session Room	Time
1.10P.4	A Weight-of-Evidence Approach for Androgen Receptor Conservation Across Vertebrate Species	Sara M Vliet	Poster	Novel Tools and Bioassays for the 21st Century Environmental Toxicology (P)	Exhibition Hall	Wednesday 8:30 AM
4.19P.2	Capitalizing on existing knowledge of pesticide-CYP9 enzyme interactions to understand chemical susceptibility across bee species using the SeqAPASS tool	Donovan Blatz	Poster	Environmental Risks of Neonicotinoid Insecticides: Are They the Outliers? (P)	Exhibition Hall	Wednesday 8:30 AM
1.10P.3	Protein structural similarity for extrapolation of toxicity knowledge across species	Carlie LaLone	Poster	Novel Tools and Bioassays for the 21st Century Environmental Toxicology (P)	Exhibition Hall	Wednesday 8:30 AM

Demo SeqAPASS

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

- **Getting started**
  - **Download user guide**

Terms

- **Query protein sequence/species** – the initial sequence/species submitted to SeqAPASS
- **Hit protein sequence/species** – all sequences/species that align with the query sequence/species
- **Protein Accession –** A unique identification tag given to every protein in the NCBI databases (e.g., ACD123211.2; NP_0000116.2)
- **Domain Accession –** A unique identification tag given to every protein domain in the NCBI Conserved Domain Databases (e.g., cd00891; pfam00067)

Thank you for attending

Introduction to SeqAPASS

Contact with any questions:

LaLone.Carlie@epa.gov; Blatz.Donovan@epa.gov; or Vliet.Sara@epa.gov