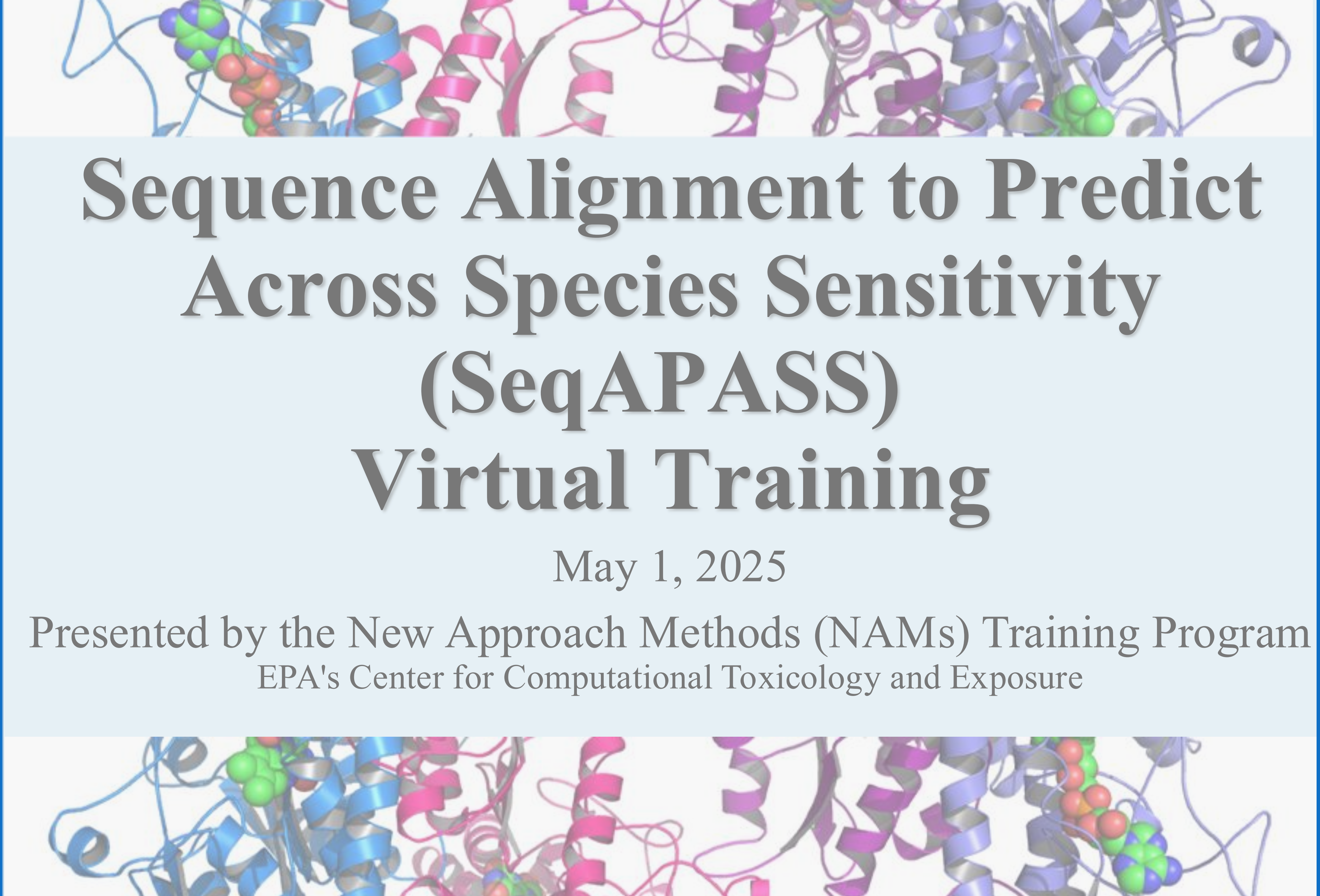


Sequence Alignment to Predict Across Species Sensitivity (SeqAPASS) Virtual Training

May 1, 2025

Presented by the New Approach Methods (NAMs) Training Program

EPA's Center for Computational Toxicology and Exposure

EPA NAMs Pilot Training Program

- New Approach Methodologies (NAMs) Training Program is a deliverable in the Agency's Work Plan, first released in 2019 and updated in 2021.
  - Previous trainings include ECOTOX, CompTox Chemicals Dashboard, GenRA, httk R Package, and the AOP Wiki.
- Goal: Develop, implement and maintain an engaging training program.
  - Interactive case studies to encourage active learning
  - Train the trainer
  - Obtain feedback
- The EPA NAMs training website includes existing training resources, including recordings and guidance documents.

Evaluate regulatory flexibility for accommodating NAMs

Develop baselines and metrics for assessing progress

Establish scientific confidence and demonstrate application

Develop NAMs that fill critical information gaps

Engage and communicate with stakeholders

EPA NAMs Training: www.epa.gov/chemical-research/new-approach-methods-nams-training

United States
Environmental Protection
Agency

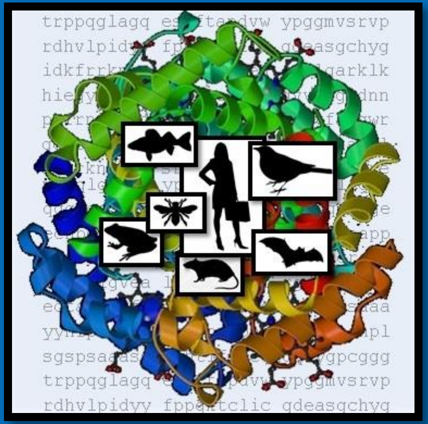
EPA New Approach Methods (NAMs) Virtual Training: SeqAPASS Tool

<u>Instructors:</u>

Ryan Staub

Peter Schumann, MS

Carlie A. LaLone, Ph.D.



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

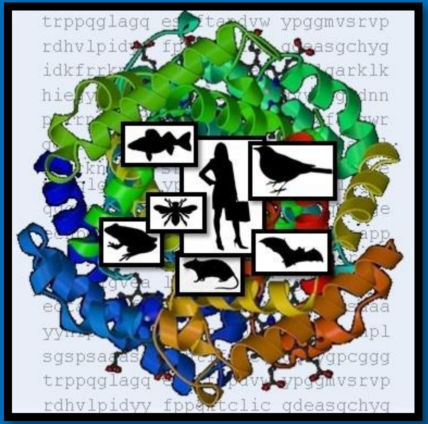
United States
Environmental Protection
Agency

Introduction – Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS):

Instructors:

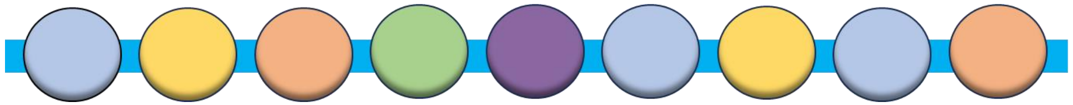
Peter Schumann

Carlie A. LaLone

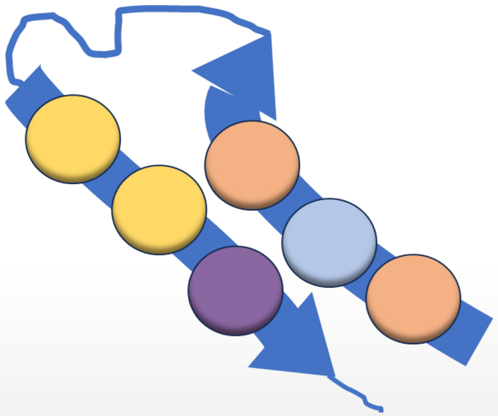


Protein 101

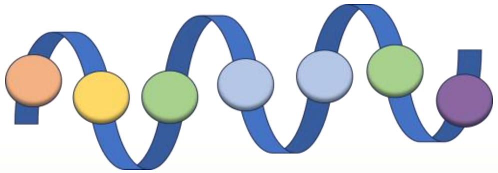
Primary Amino Acid Sequence



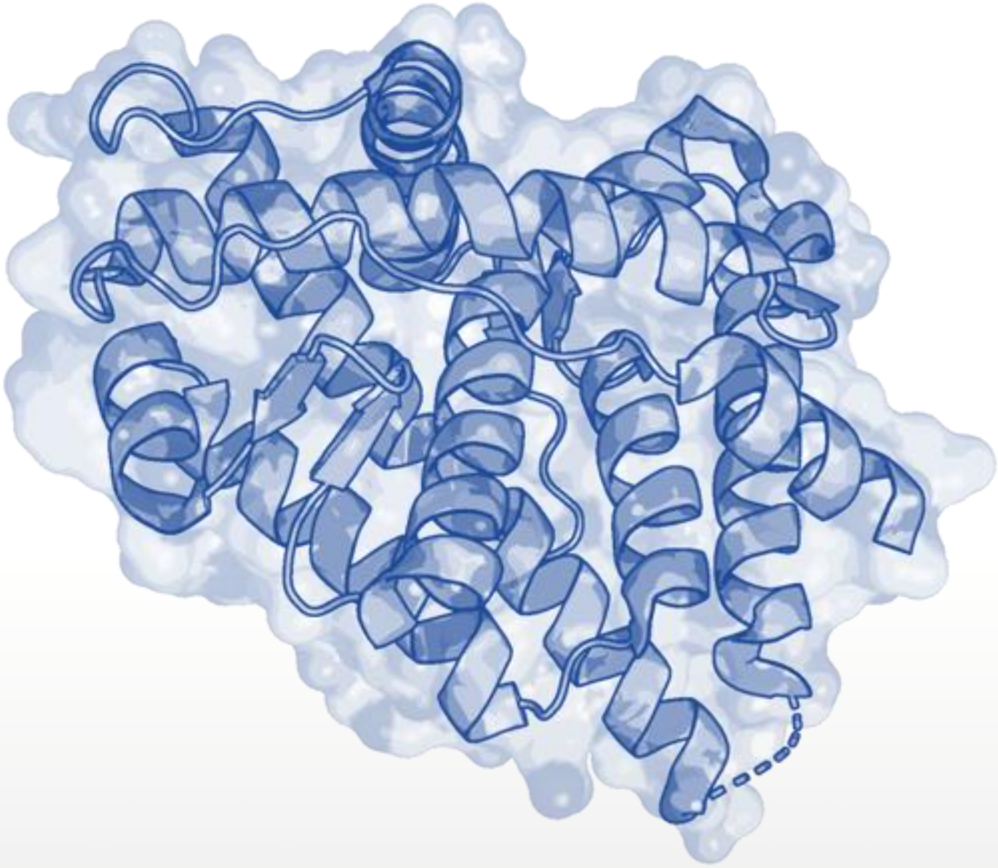
Beta Sheet



Alpha Helix



Tertiary Structure



Protein Structure = Function

Image from RCSB PDB

Regulate gene expression

Structural support

Biochemical catalysts

Transporting molecules

Repair and maintain muscle mass

Hormones

Enzymes

Regulate immune response

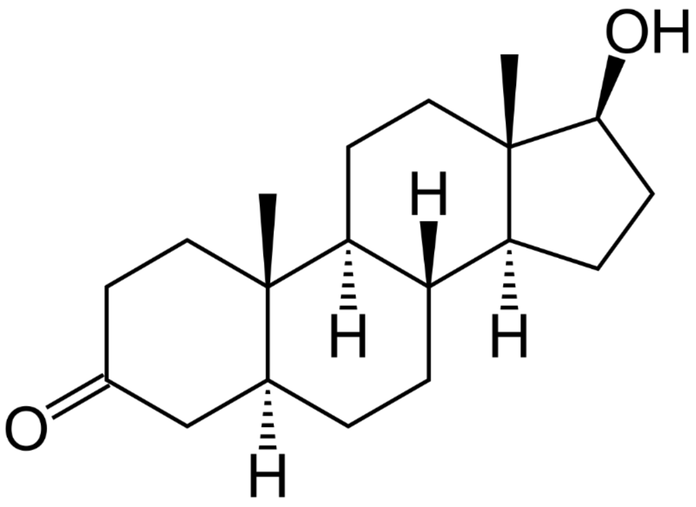
What’s all the fuss about proteins??

Our interest is in protecting human health and the environment

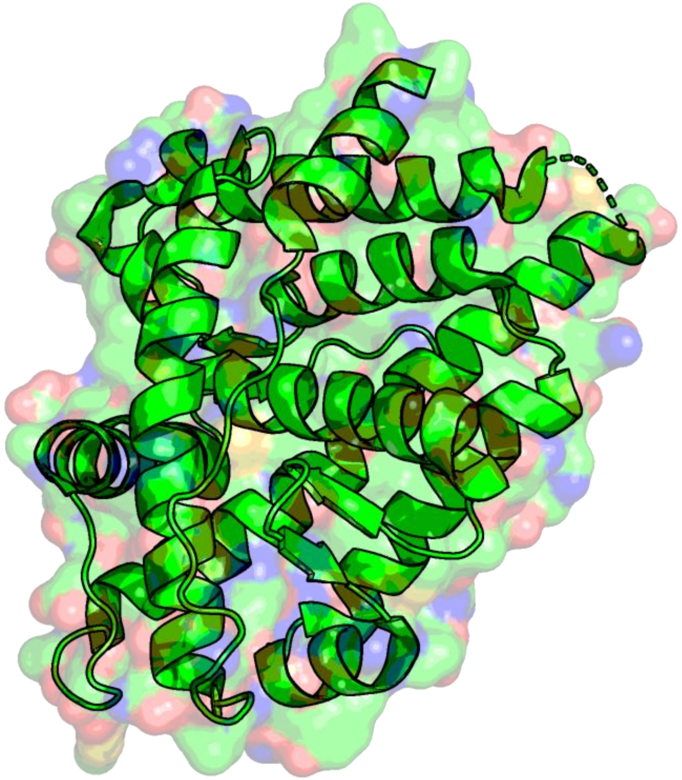

200
150
100

Chemical-Protein Interaction

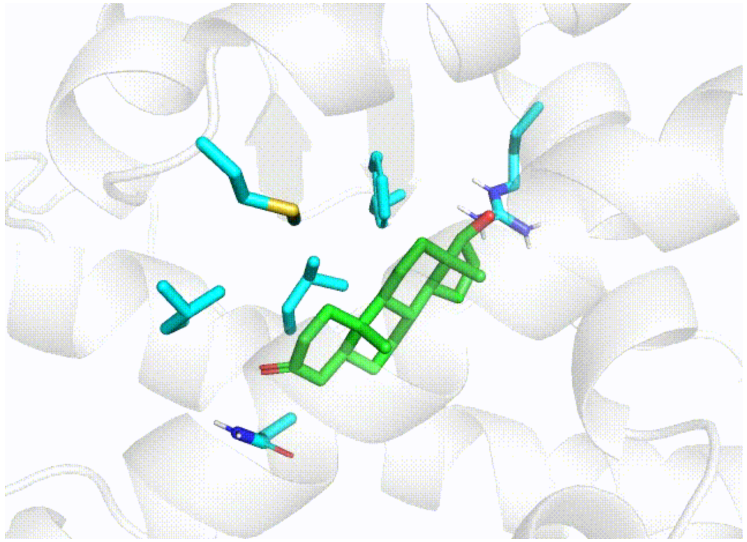
Chemical



e.g., Pharmaceutical



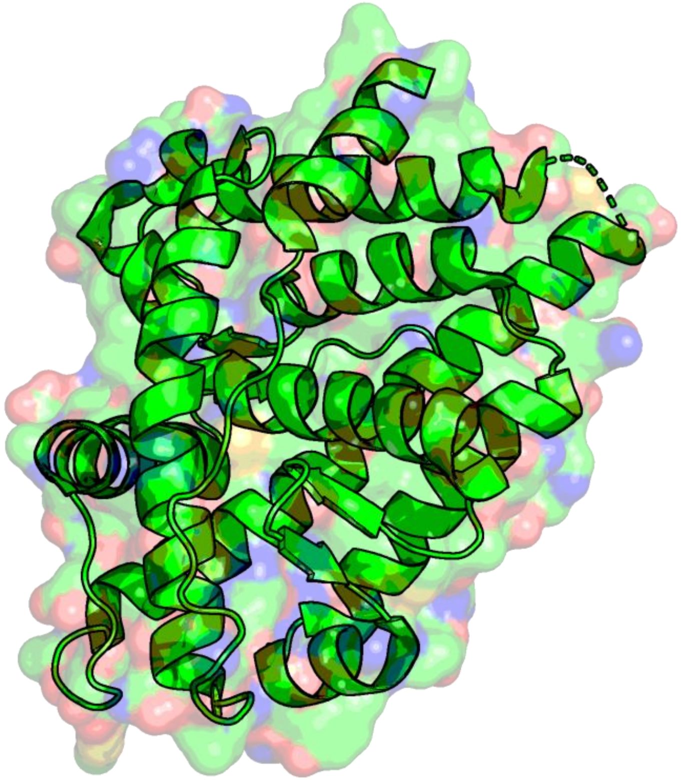
Beneficial effect



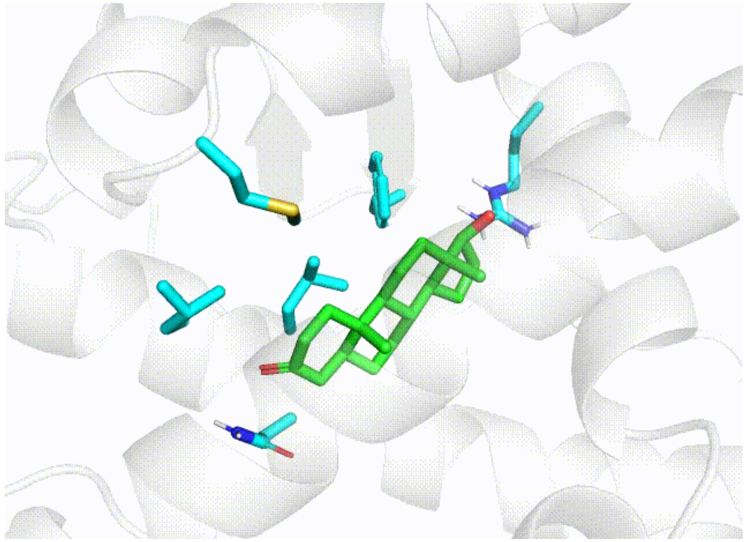
Chemical-Protein Interaction

Chemical(s)

Unintended Chemical Exposure



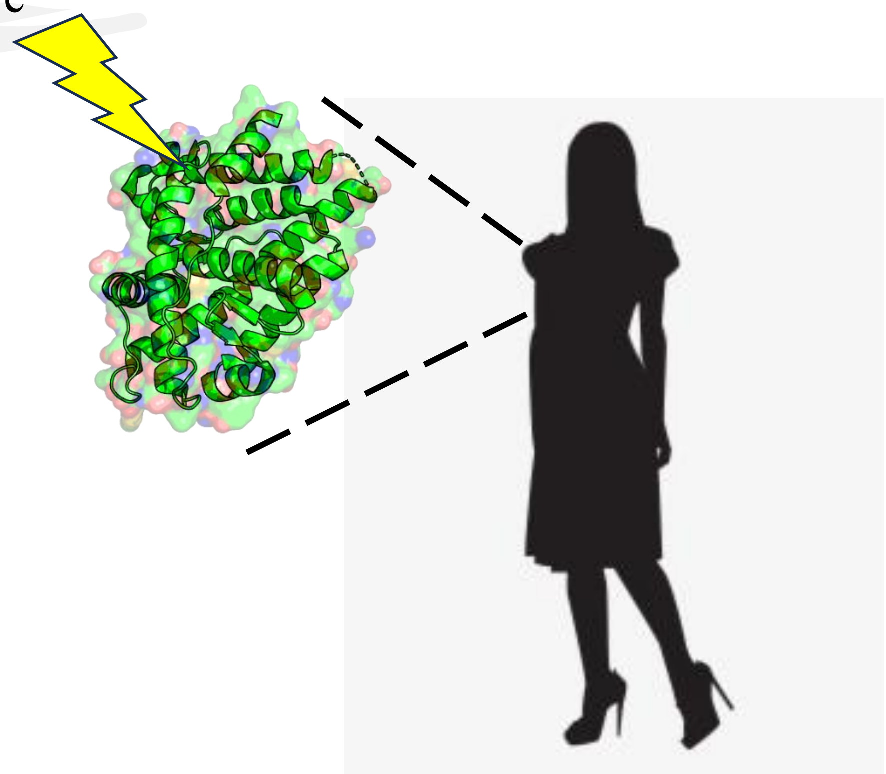
Adverse outcome

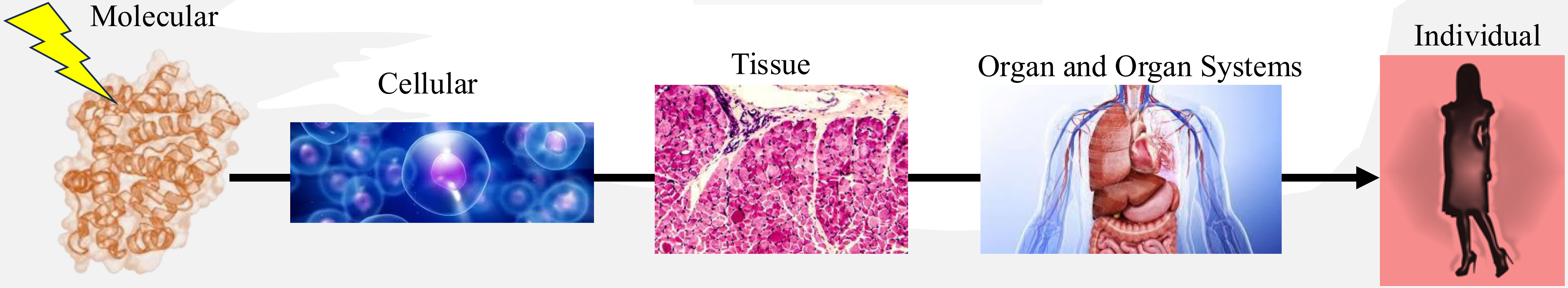


Disease
Reproduction
Growth
Mortality

Chemical-Protein Interactions are Important

Unintended Chemical Exposure





Molecular

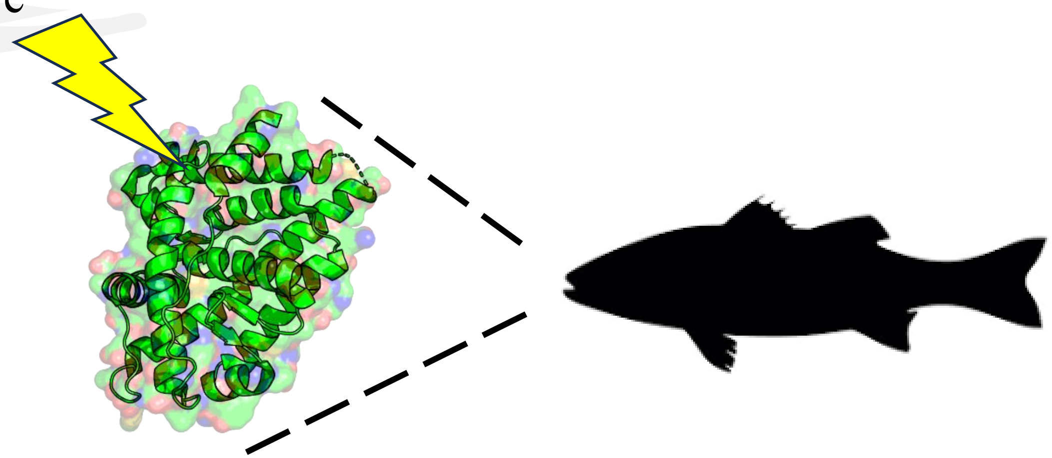
Cellular

Tissue

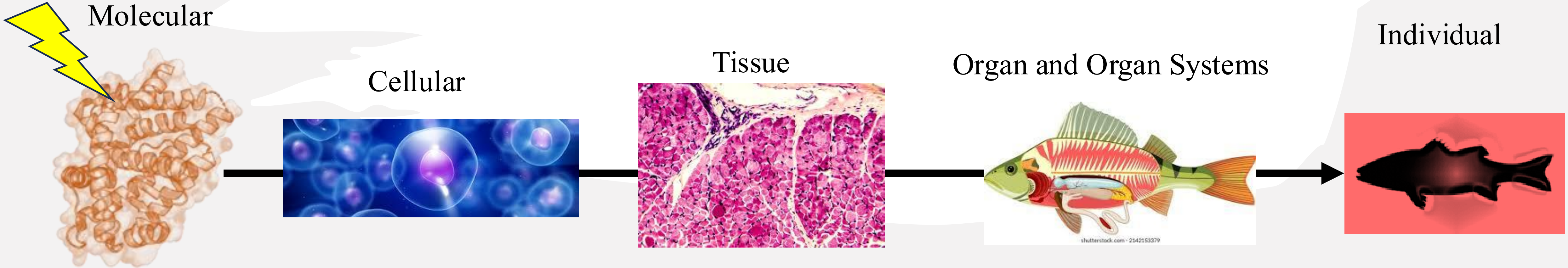
Organ and Organ Systems

Individual

Chemical Protein Interactions are Important



Unintended Chemical Exposure



Molecular

Cellular

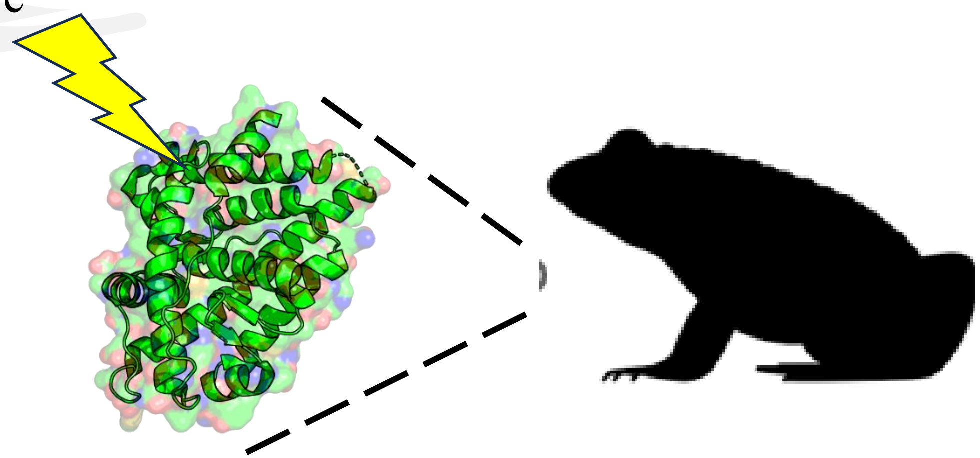
Tissue

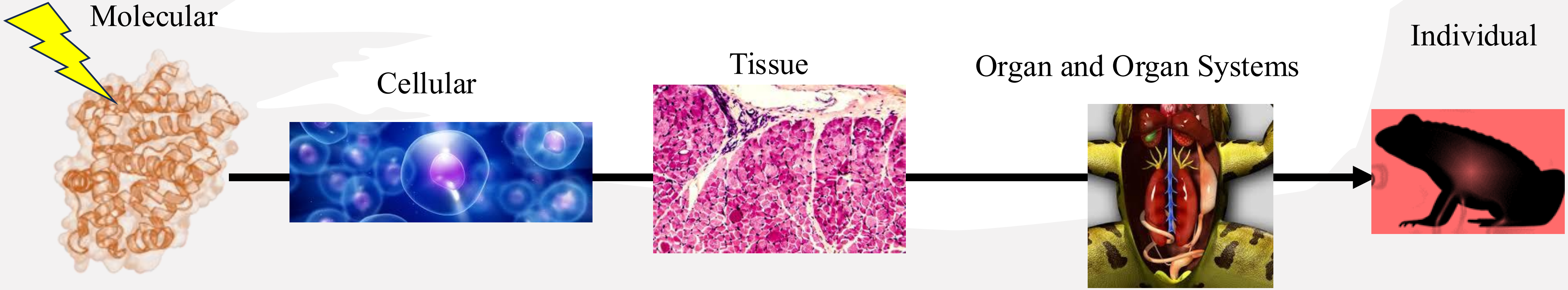
Organ and Organ Systems

Individual

Chemical Protein Interactions are Important

Unintended Chemical Exposure





Molecular

Cellular

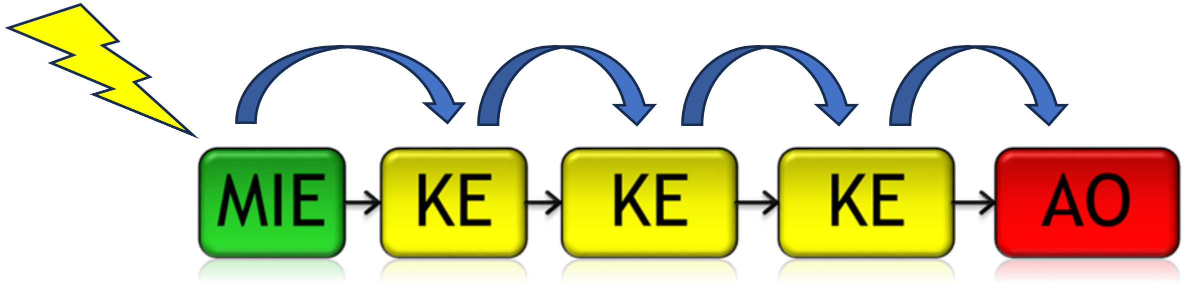
Tissue

Organ and Organ Systems

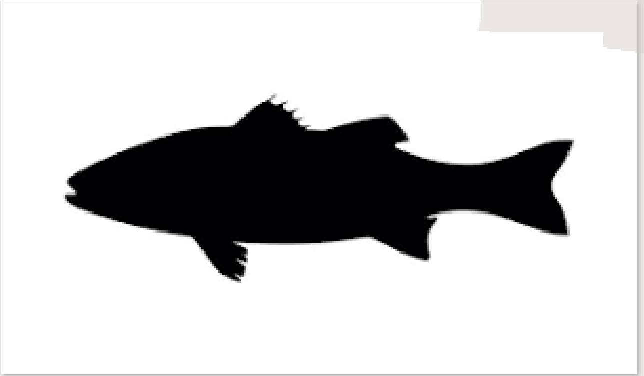
Individual

Chemical-Protein Interactions are Molecular Events that Initiate the Downstream Biological Events in Pathways

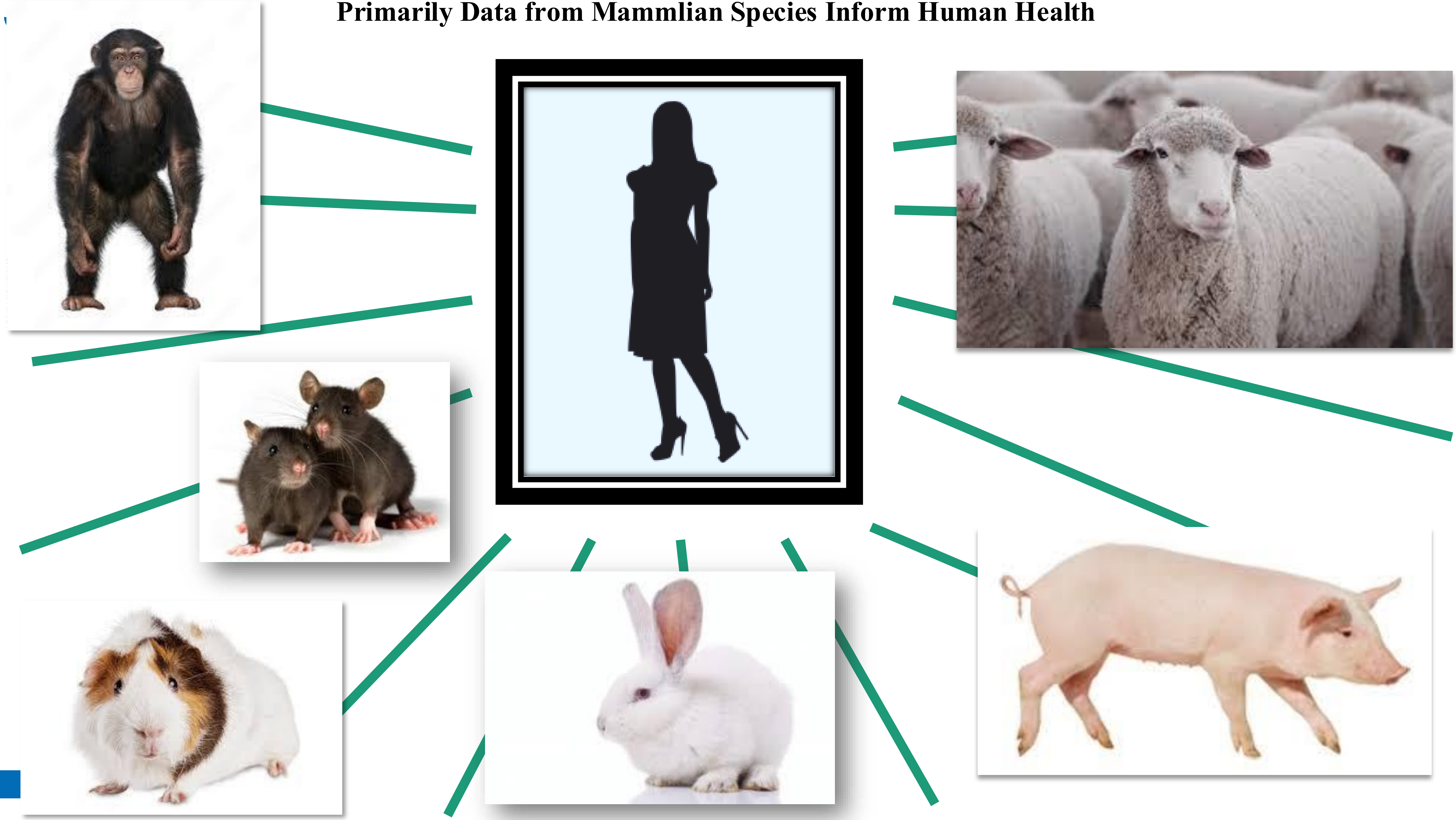
Why do we care about this knowledge across species?



Needed for Regulatory Decision-making



Primarily Data from Mammlian Species Inform Human Health



Cross Species Extrapolation: Decisions based on available data

Available Toxicity Data

Protection Goal:
Human

?

Ecotoxicology – Model Species

Cross Species Extrapolation: Decisions based on available data

Available Toxicity Data

Protection Goal:
Salmonids

?

Protection Goal:
Amphibia

Safety factors

Species Sensitivity Distributions

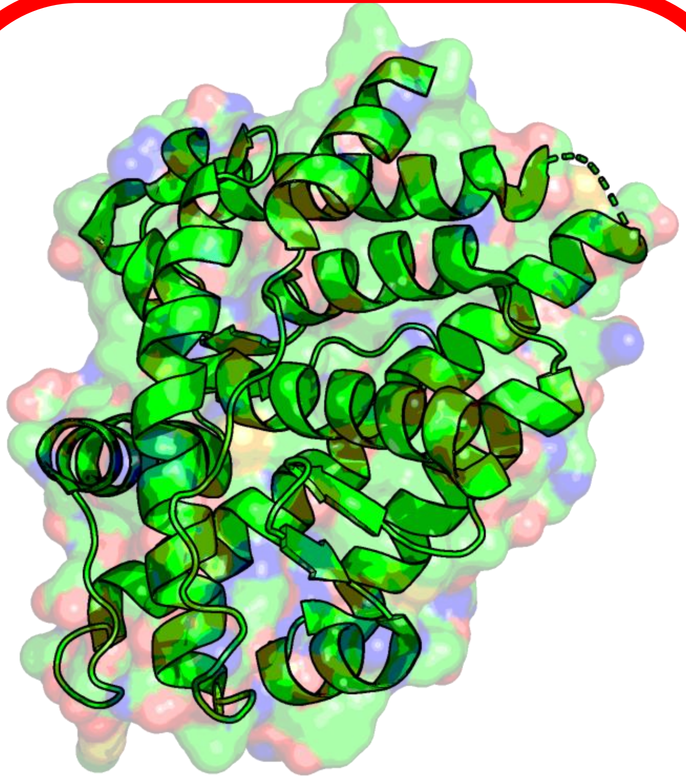
Species extrapolation is NOT a new challenge

Based on our knowledge of chemical-protein interactions, can we add evidence for extrapolating knowledge across species?

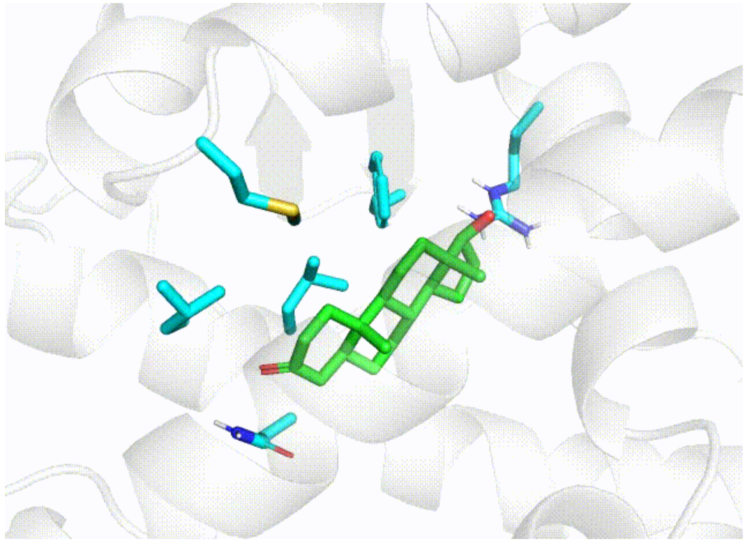
Chemical-Protein Interaction

Chemical(s)

Unintended Chemical Exposure

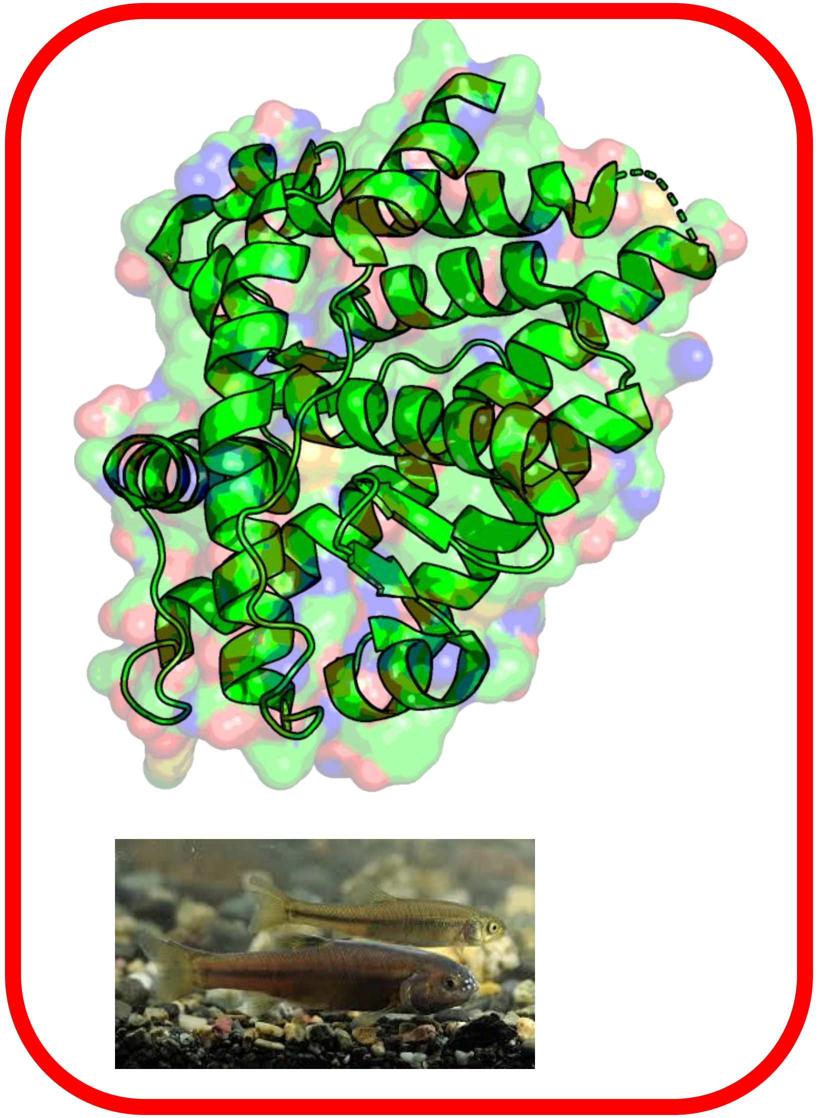


Adverse outcome

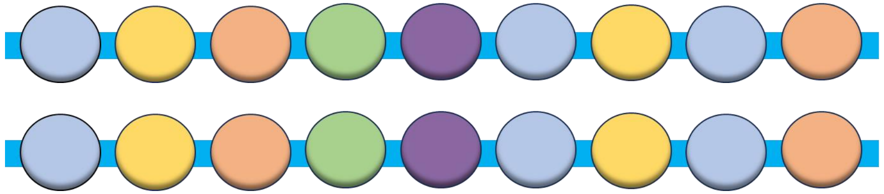


Impacts on reproduction

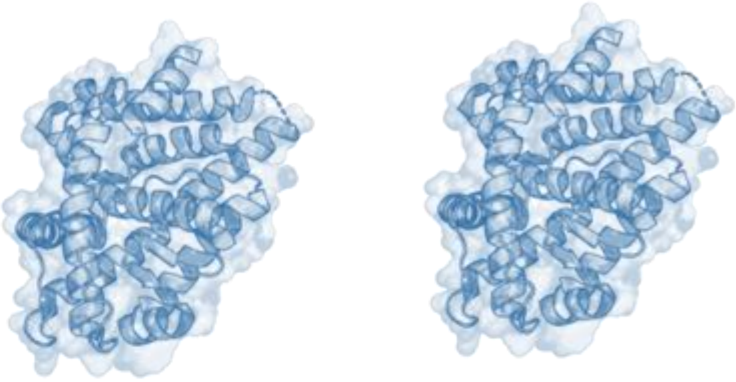
Evaluating Protein Conservation Across Species



Protein Sequence Conservation



Protein Structure Conservation



Chemical-Protein Interaction Conservation

Bioinformatics

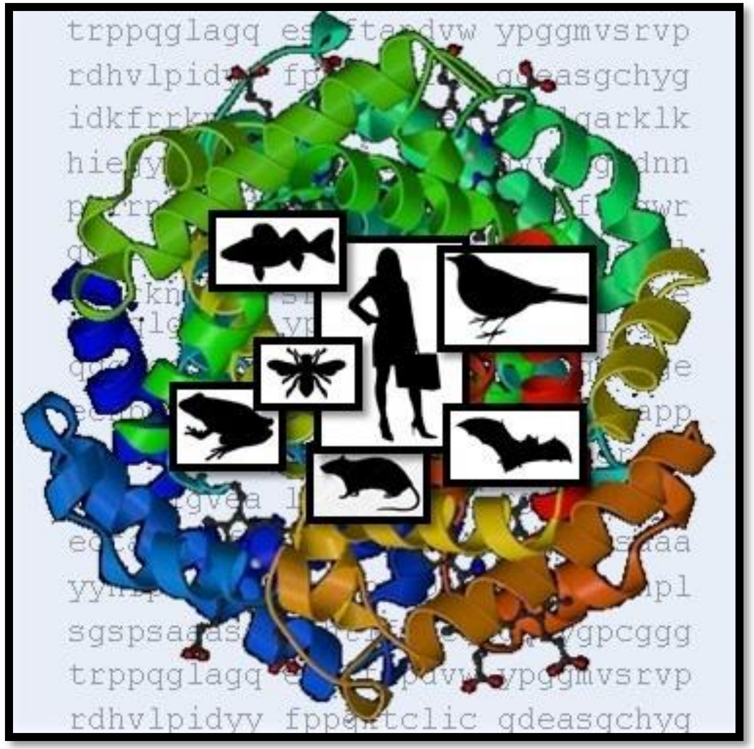
- **Combines mathematics, information science, and biology to answer biological questions**
- **Developing methodology and analysis tools to explore large volumes of biological data**
  - **Query, extract, store, organize, systematize, annotate, visualize, mine, and interpret complex data**
    - **Usually pertains to DNA and amino acid sequences**

Let the computers do the work

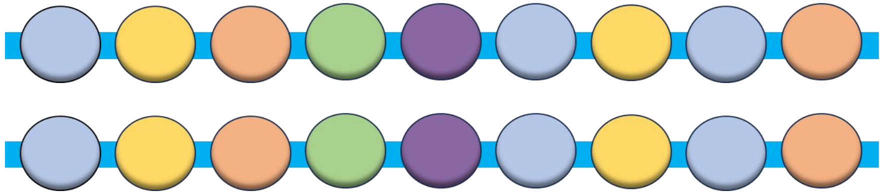
Evaluating Protein Conservation Across Species

<u>Seq</u>uence <u>A</u>lignment to <u>P</u>redict <u>A</u>cross <u>S</u>pecies <u>S</u>usceptibility

(SeqAPASS)



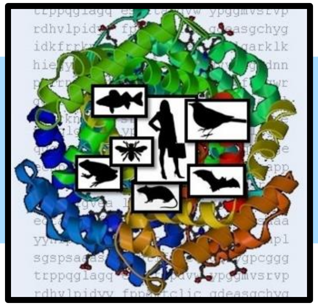
Protein Sequence Conservation



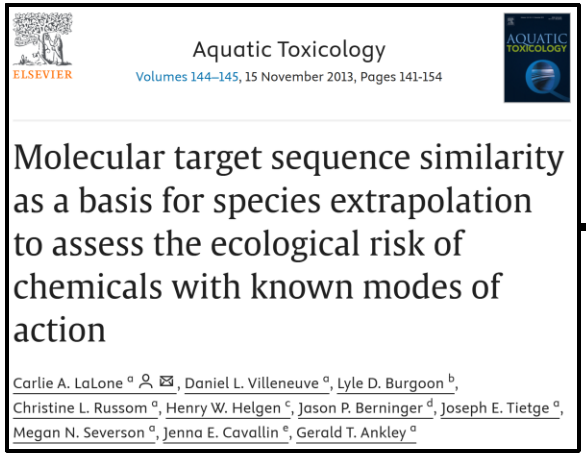
SeqAPASS

Evaluate protein conservation for predicting chemical susceptibility across species

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)



The concept of SeqAPASS was formalized in a 2013 publication that was later followed by a web-based tool release in 2016. It has continued to evolve to this day! (Last week we released v8.1)



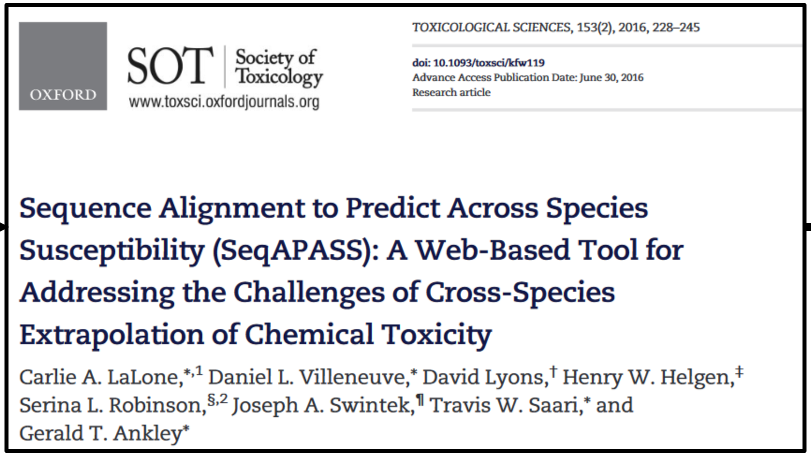
ELSEVIER

Aquatic Toxicology

Volumes 144–145, 15 November 2013, Pages 141-154

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]



OXFORD

SOT | Society of Toxicology

www.toxsci.oxfordjournals.org

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

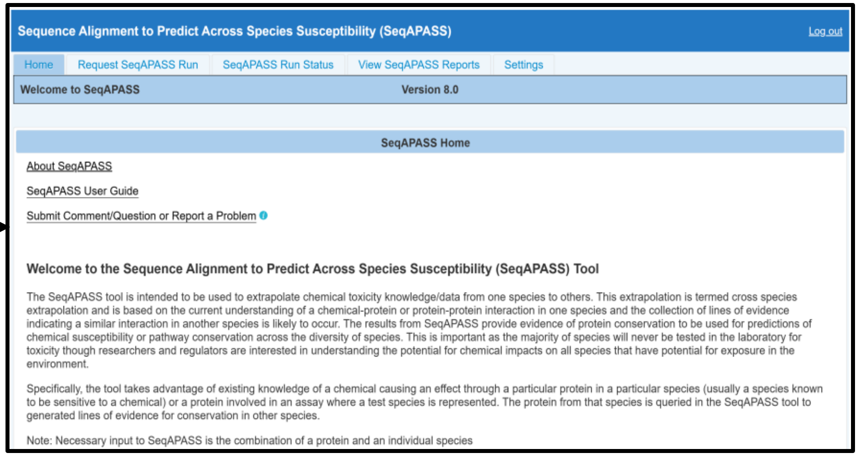
doi: 10.1093/toxsci/kfw119

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[*]



Welcome to the Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) Tool

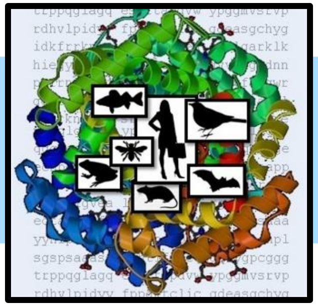
The SeqAPASS tool is intended to be used to extrapolate chemical toxicity knowledge/data from one species to others. This extrapolation is termed cross species extrapolation and is based on the current understanding of a chemical-protein or protein-protein interaction in one species and the collection of lines of evidence indicating a similar interaction in another species is likely to occur. The results from SeqAPASS provide evidence of protein conservation to be used for predictions of chemical susceptibility or pathway conservation across the diversity of species. This is important as the majority of species will never be tested in the laboratory for toxicity though researchers and regulators are interested in understanding the potential for chemical impacts on all species that have potential for exposure in the environment.

Specifically, the tool takes advantage of existing knowledge of a chemical causing an effect through a particular protein in a particular species (usually a species known to be sensitive to a chemical) or a protein involved in an assay where a test species is represented. The protein from that species is queried in the SeqAPASS tool to generated lines of evidence for conservation in other species.

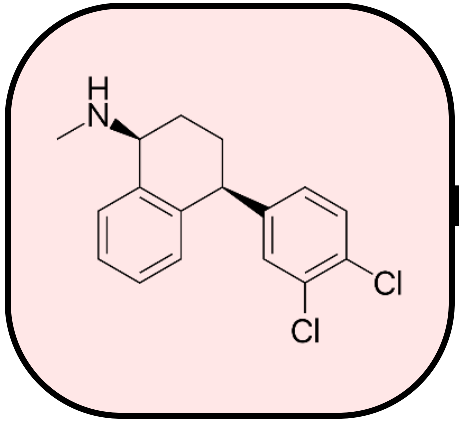
Note: Necessary input to SeqAPASS is the combination of a protein and an individual species

seqapass.epa.gov/seqapass/

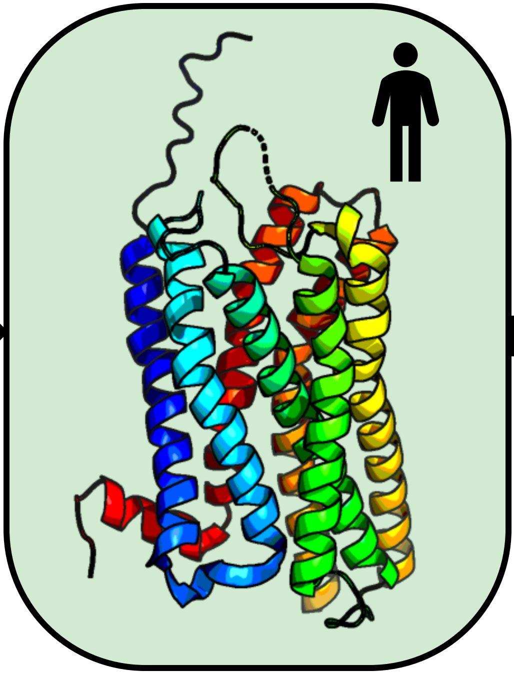
What is SeqAPASS?



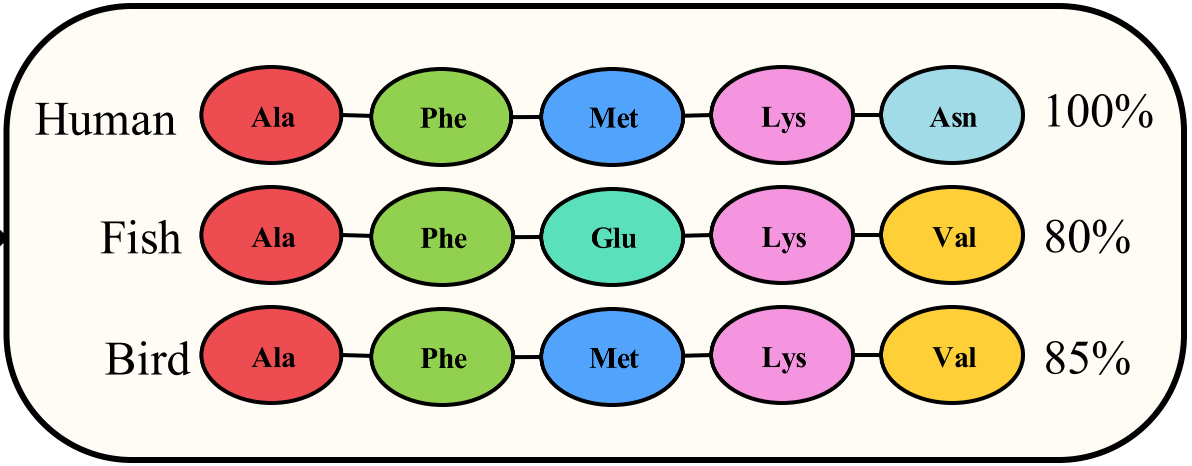
Chemical of Interest



Known Protein Target + Known Sensitive Species



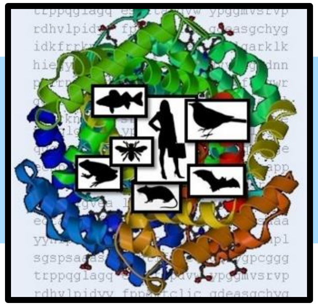
Cross-species Sequence Alignment



Human	Ala	Phe	Met	Lys	Asn	100%
Fish	Ala	Phe	Glu	Lys	Val	80%
Bird	Ala	Phe	Met	Lys	Val	85%

Protein sequence similarity (conservation) = shared likelihood of susceptibility to chemical!

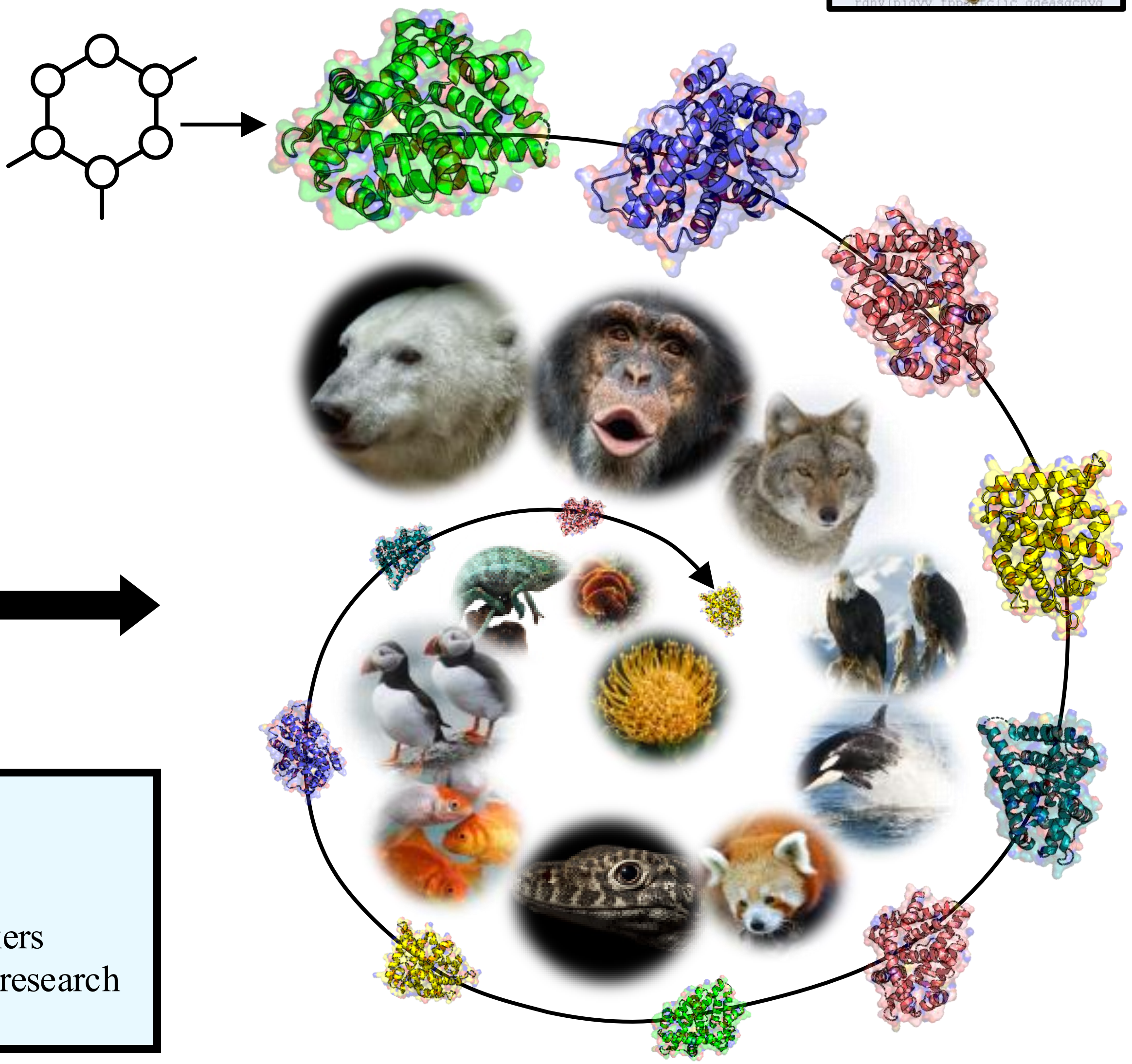
Why is SeqAPASS useful?



We can *rapidly* compare protein sequences across 100s to 1000s of species to generate predictions of chemical susceptibility, including in threatened or endangered species.

Cross-species Sequence Alignment Analysis

```
                      cov     pid 161        .         .         .         2         .         .         .         . 240
1 KAF1656078.1     100.0% 100.0%     TASIMHLCAISLDRYIAIRNPIHHSRFNSRTKAFAKIIAVWTISVGISMPIPVFGLQDDSKVFKKGSCLLVDDNFVLVGS
2 NP_766400.1       99.8%  76.3%     TASIMHLCAISLDRYVAIQNPIHHSRFNSRTKAFLKIIAVWTISVGISMPIPVFGLQDDSKVFKEGSCLLADDNFVLIGS
3 NP_001268288.1    99.8%  76.1%     TASIMHLCAISLDRYVAIQNPIHHSRFNSRTKAFLKIIAVWTISVGISMPIPVFGLQDDSKVFKEGSCLLADDNFVLIGS
4 NP_001005869.1    99.6%  78.4%     TASIMHLCAISLDRYVAIQNPIHHSRFNSRTKAFLKIIAVWTISVGISMPIPVFGLQDDSKVFKEGSCLLADDNFVLIGS
5 NP_999382.1       99.6%  77.8%     TASIMHLCAISLDRYVAIQNPIHHRRFNSRTKAFLKIIAVWTISVGISMPIPVFGLQDDSKVFKEGSCLLADDNFVLIGS
6 NP_001365853.1    99.8%  79.4%     TASIMHLCAISLDRYVAIQNPIHHSRFNSRTKAFLKIIAVWTISVGISMPIPVFGLQDDSKVFKEGSCLLADDNFVLIGS
7 NP_001028138.1    99.8%  79.4%     TASIMHLCAISLDRYVAIQNPIHHSRFNSRTKAFLKIIAVWTISVGISMPIPVFGLQDDSKVFKEGSCLLADDNFVLIGS
  consensus/100%                     TASIMHLCAISLDRYlAIpNPIHHpRFNSRTKAFhKIIAVWTISVGISMPIPVFGLQDDSKVFKcGSCLLsDDNFVLlGS
  consensus/90%                      TASIMHLCAISLDRYlAIpNPIHHpRFNSRTKAFhKIIAVWTISVGISMPIPVFGLQDDSKVFKcGSCLLsDDNFVLlGS
  consensus/80%                      TASIMHLCAISLDRYVAIQNPIHHSRFNSRTKAFLKIIAVWTISVGISMPIPVFGLQDDSKVFKEGSCLLADDNFVLIGS
  consensus/70%                      TASIMHLCAISLDRYVAIQNPIHHSRFNSRTKAFLKIIAVWTISVGISMPIPVFGLQDDSKVFKEGSCLLADDNFVLIGS
```

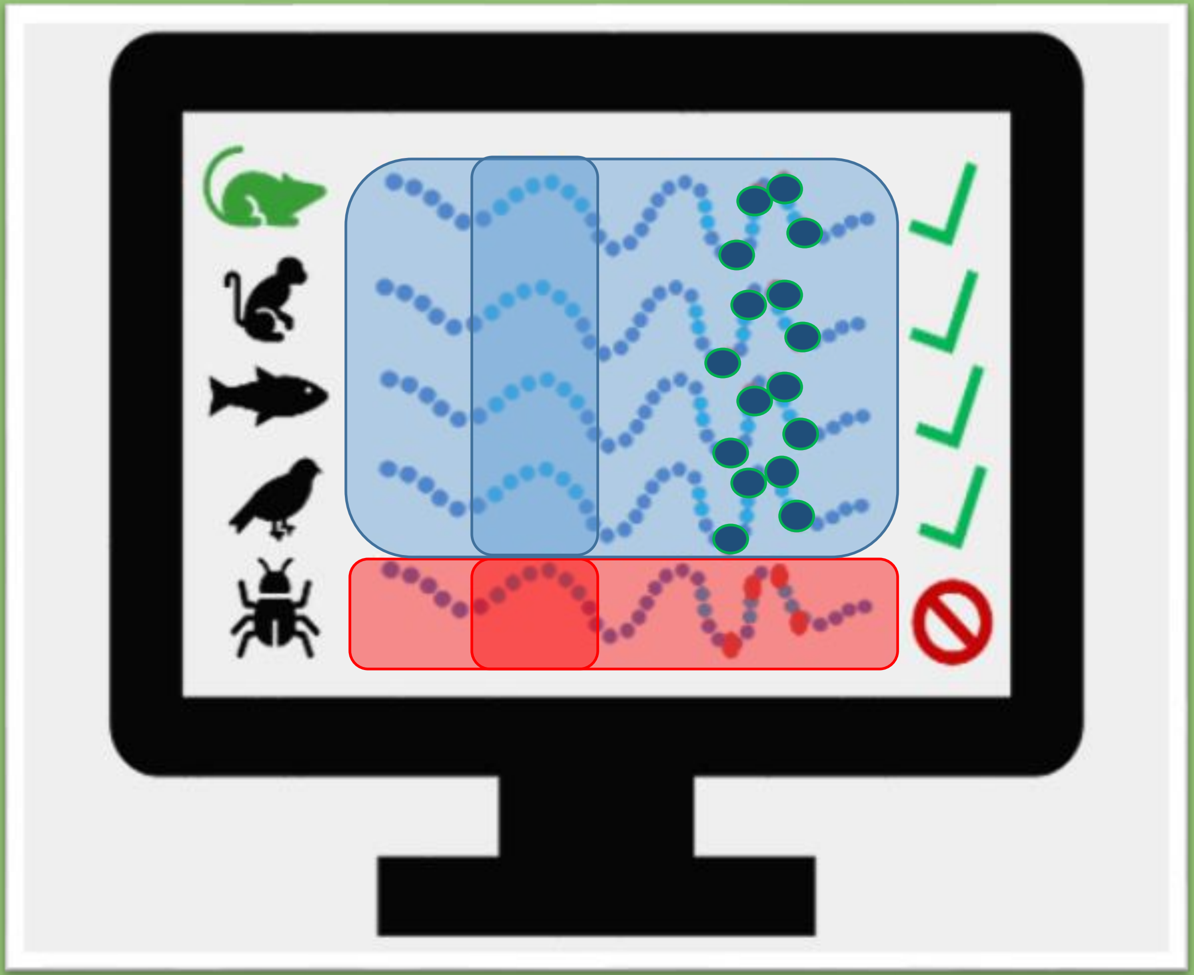


- Sequence data sourced from public databases
- SeqAPASS is publicly available
- Designed for a diversity of end users: researchers, regulators, decision makers
- Supports existing predictions and hypothesis generation for guiding future research

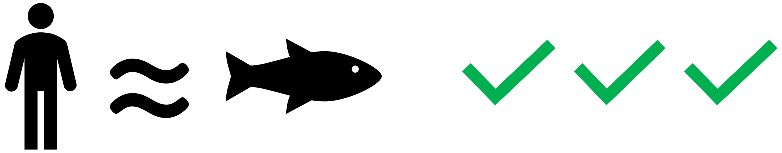
For this training

Level 1	Primary Amino Acid Sequence Alignments
Level 2	Conserved Functional Domain Alignments
Level 3	Critical Amino Acid Conservation
Level 4	Structural Conservation

seqapass.epa.gov/seqapass/

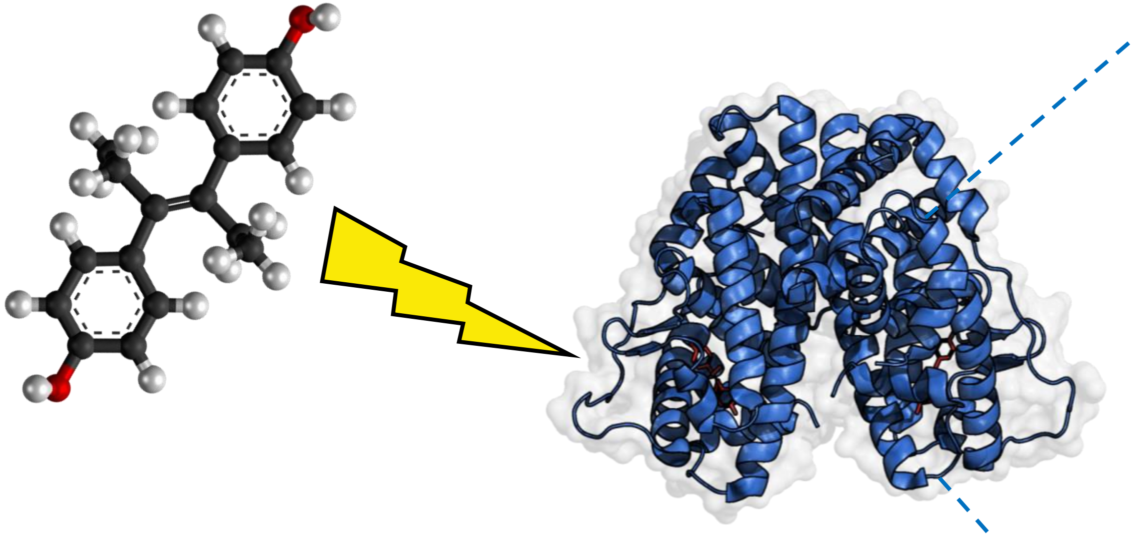


Gather Lines of Evidence Toward Protein Conservation



What information is required for a SeqAPASS query?

1) Protein Target



Knowledge of chemical-protein interaction

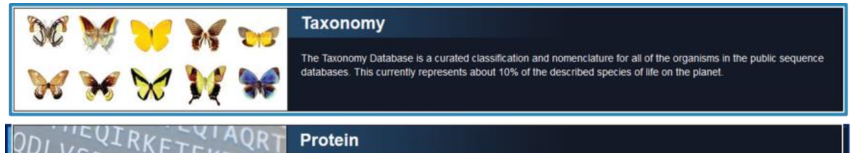
2) Sensitive Species

- Experimental evidence
- Model organism from *in vitro* assay

ToxCast

MIE → KE → KE → KE → AO

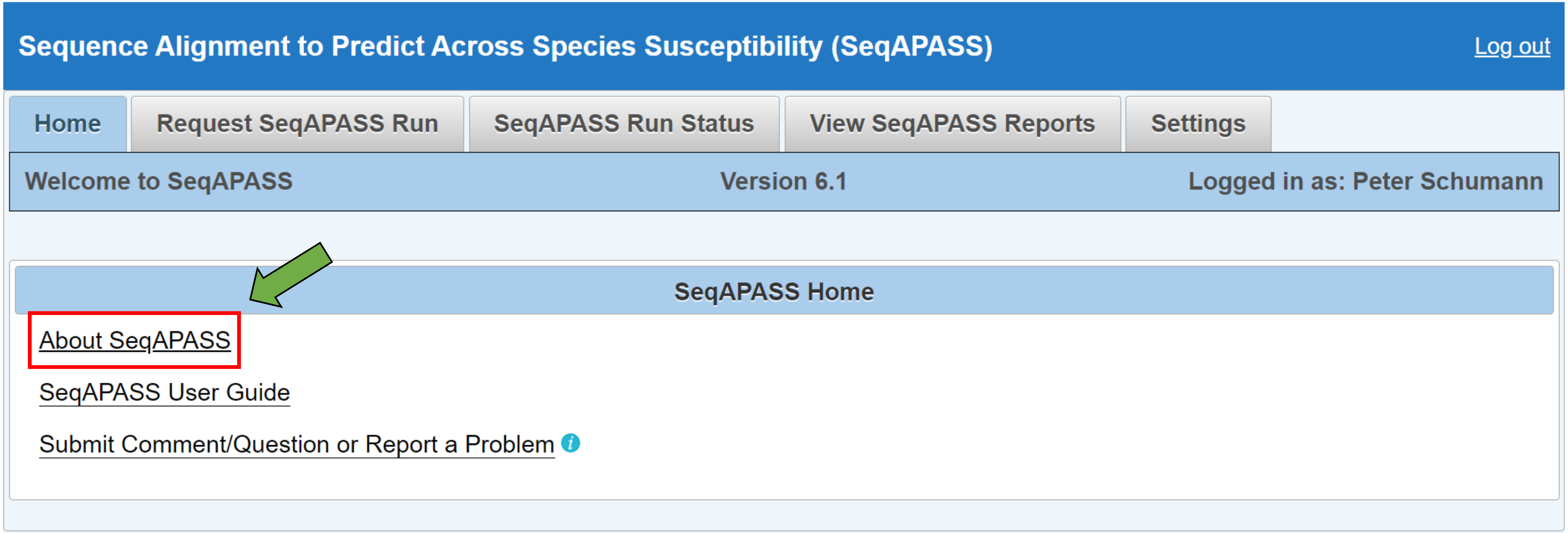
Ideally, with connections to an established AOP



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



Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)

Welcome to SeqAPASS

Version 6.1

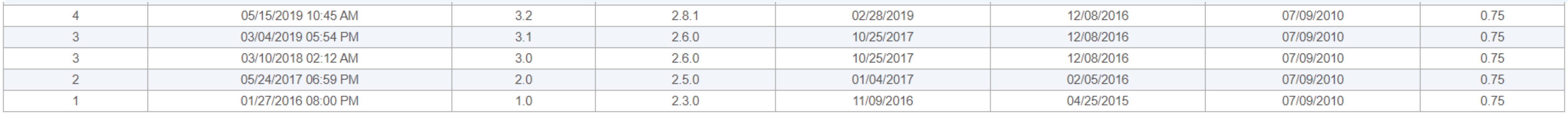
Logged in as: Peter Schumann

SeqAPASS Home

About SeqAPASS

SeqAPASS User Guide

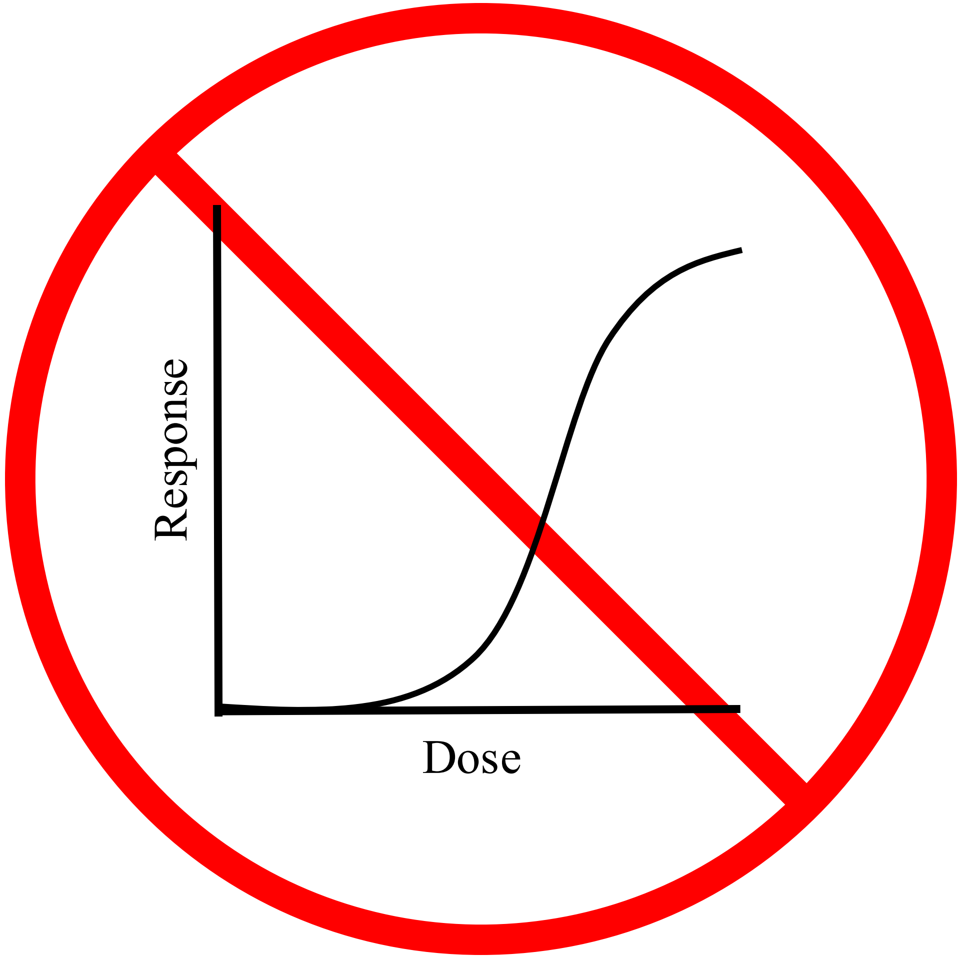
Submit Comment/Question or Report a Problem



4	05/15/2019 10:45 AM	3.2	2.8.1	02/28/2019	12/08/2016	07/09/2010	0.75
3	03/04/2019 05:54 PM	3.1	2.6.0	10/25/2017	12/08/2016	07/09/2010	0.75
3	03/10/2018 02:12 AM	3.0	2.6.0	10/25/2017	12/08/2016	07/09/2010	0.75
2	05/24/2017 06:59 PM	2.0	2.5.0	01/04/2017	02/05/2016	07/09/2010	0.75
1	01/27/2016 08:00 PM	1.0	2.3.0	11/09/2016	04/25/2015	07/09/2010	0.75

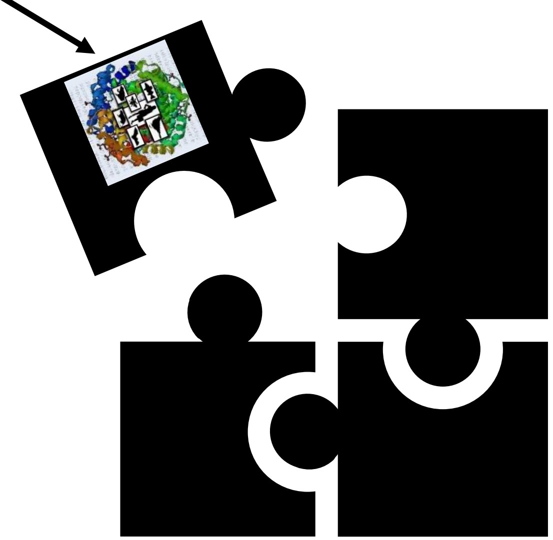
Threatened/Endangered Species Data obtained from EPA ECOTOX on Feb. 2, 2022

SeqAPASS ***does not*** predict the degree of sensitivity/susceptibility



Factors that make a species sensitive:

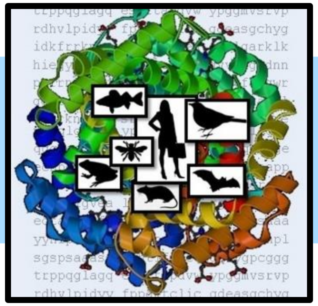
- Exposure
- Dose
- Toxicokinetics
- Target receptor availability
- Life stage
- Life history
- etc.
- etc.



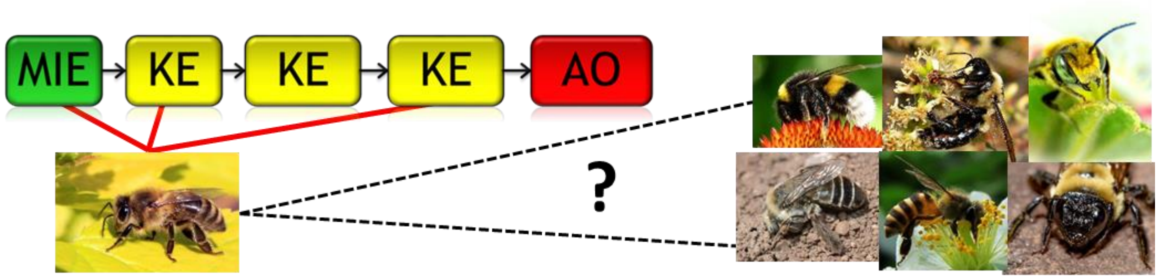
Applications of SeqAPASS

MIE → KE → KE → KE → AO

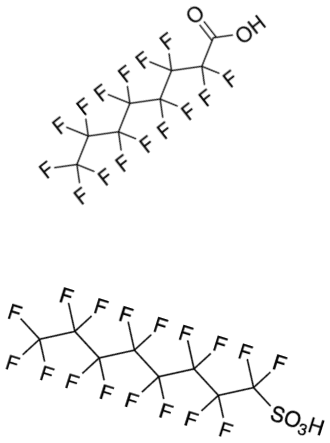
Case Studies



- **Extrapolate adverse outcome pathway knowledge across species**
  - Define the taxonomic domain of applicability
    - 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



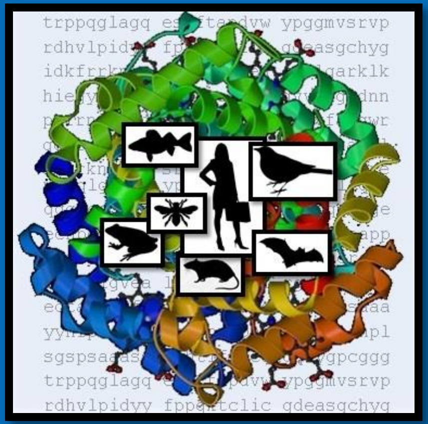
United States
Environmental Protection
Agency

Training Course – SeqAPASS Level 1: Primary sequence similarity and ortholog candidate identification

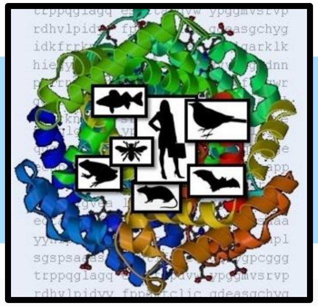
Instructors:

Peter Schumann, MS (US EPA)

Carlie A. LaLone, PhD (US EPA)

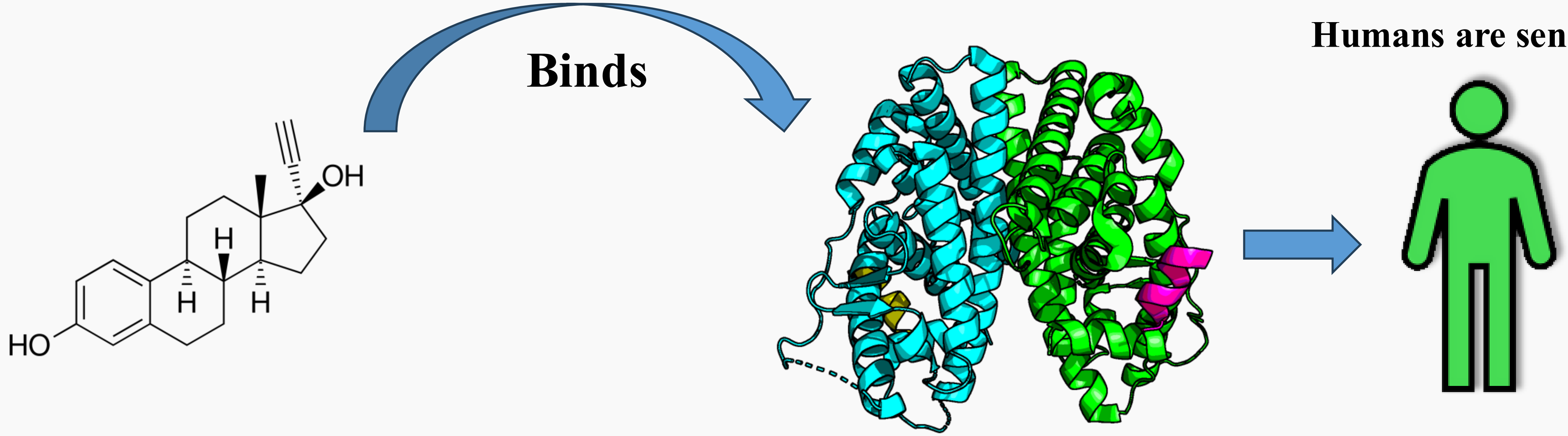


Query Formulation



1) **Protein Target**

2) **Sensitive species**



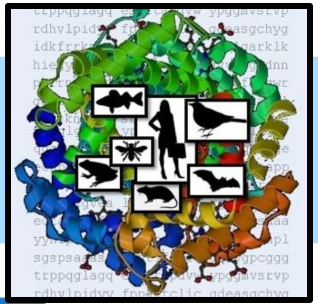
Humans are sensitive

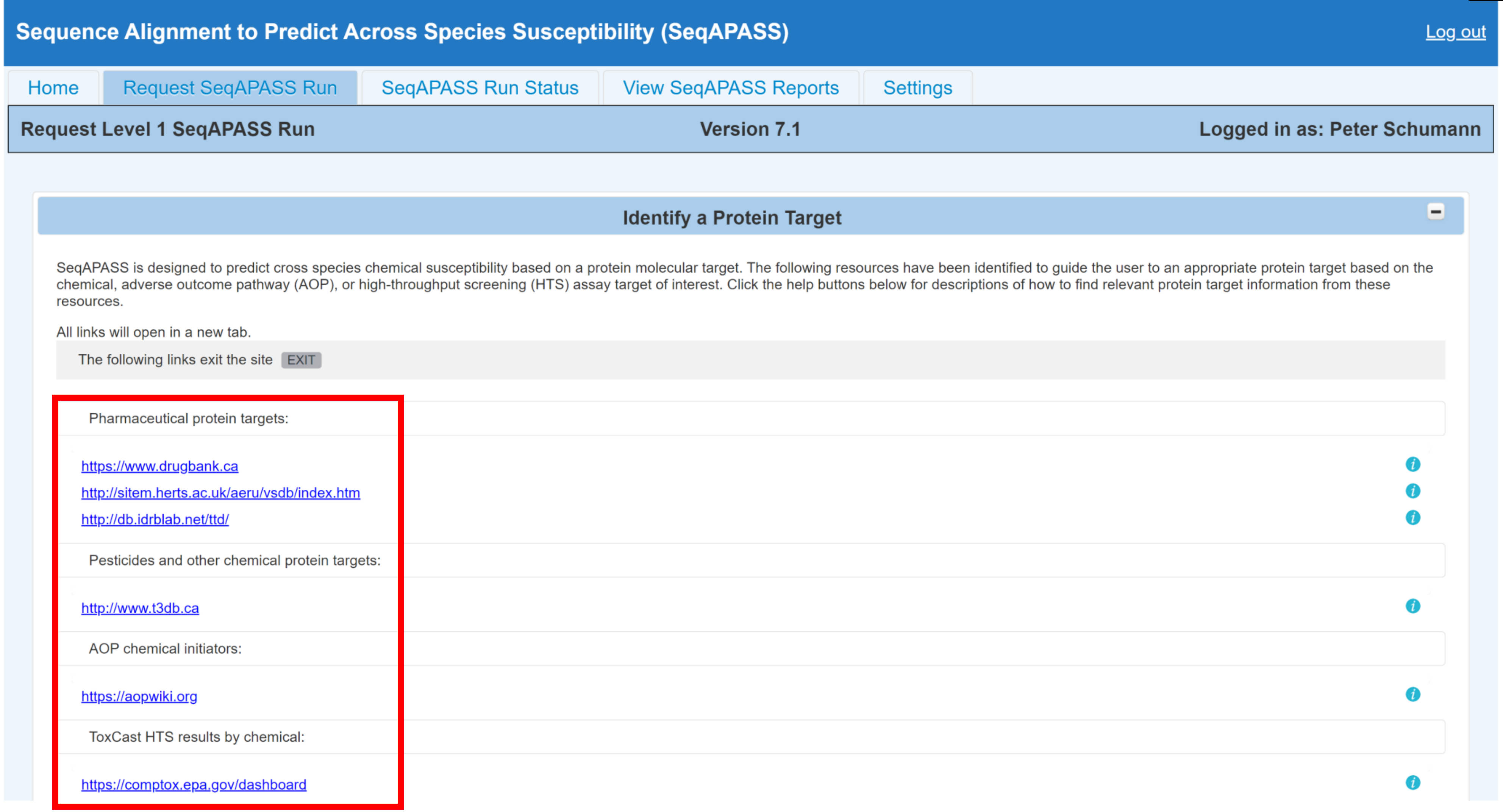
Ethinyl estradiol

Estrogen Receptor Alpha (ESR1;
NCBI Accession: NP_000116.2)

Question: What other species might be *susceptible* to ethinyl estradiol?

Target Protein Identification





Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)

Request Level 1 SeqAPASS Run **Version 7.1** **Logged in as: Peter Schumann**

Identify a Protein Target

SeqAPASS is designed to predict cross species chemical susceptibility based on a protein molecular target. The following resources have been identified to guide the user to an appropriate protein target based on the chemical, adverse outcome pathway (AOP), or high-throughput screening (HTS) assay target of interest. Click the help buttons below for descriptions of how to find relevant protein target information from these resources.

All links will open in a new tab.

The following links exit the site EXIT

Pharmaceutical protein targets:

https://www.drugbank.ca
http://sitem.herts.ac.uk/aeru/vsdb/index.htm
http://db.idrblab.net/ttd/

Pesticides and other chemical protein targets:

http://www.t3db.ca

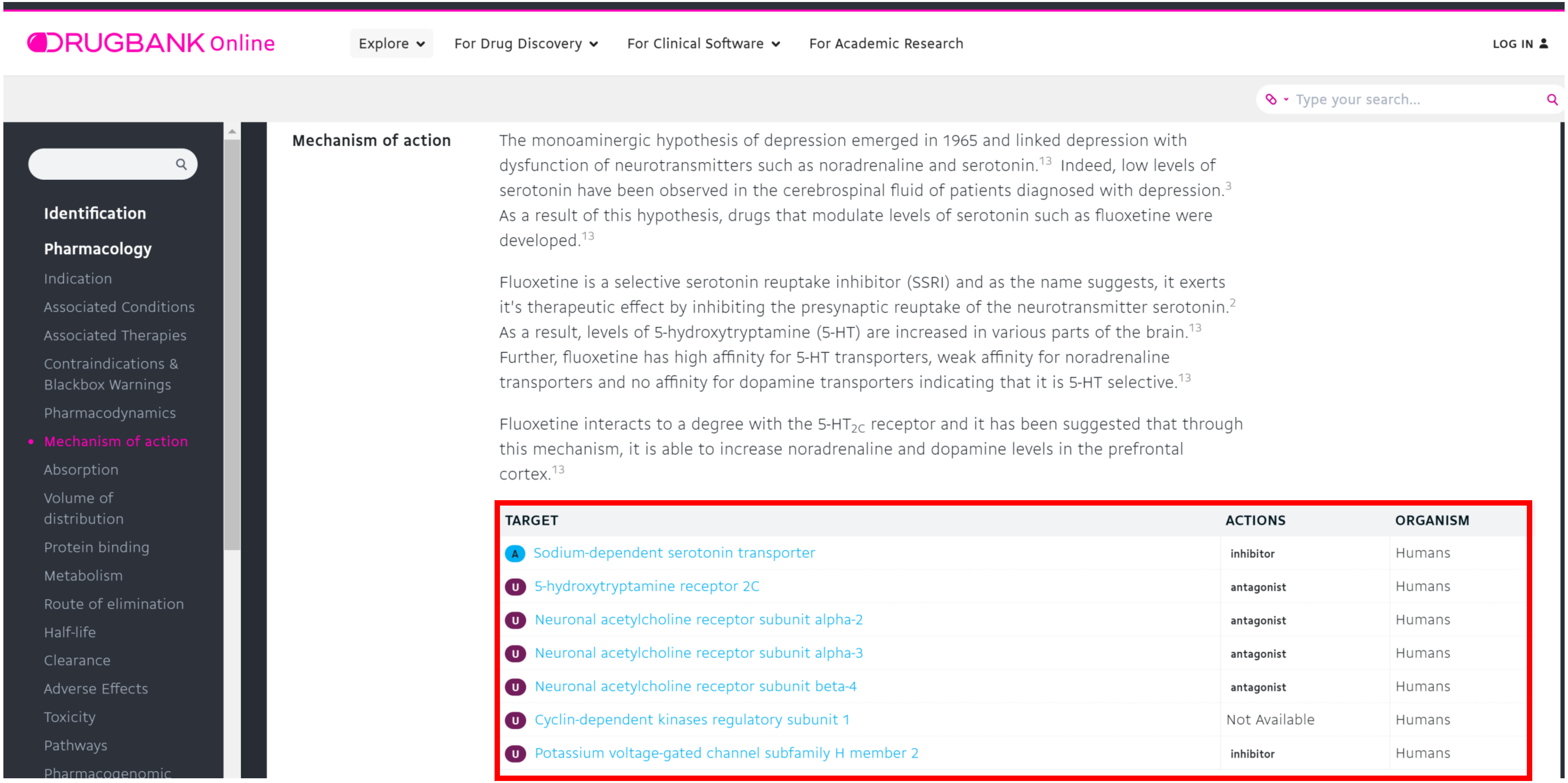
AOP chemical initiators:

https://aopwiki.org

ToxCast HTS results by chemical:

https://comptox.epa.gov/dashboard

Target Identification cont.



Mechanism of action

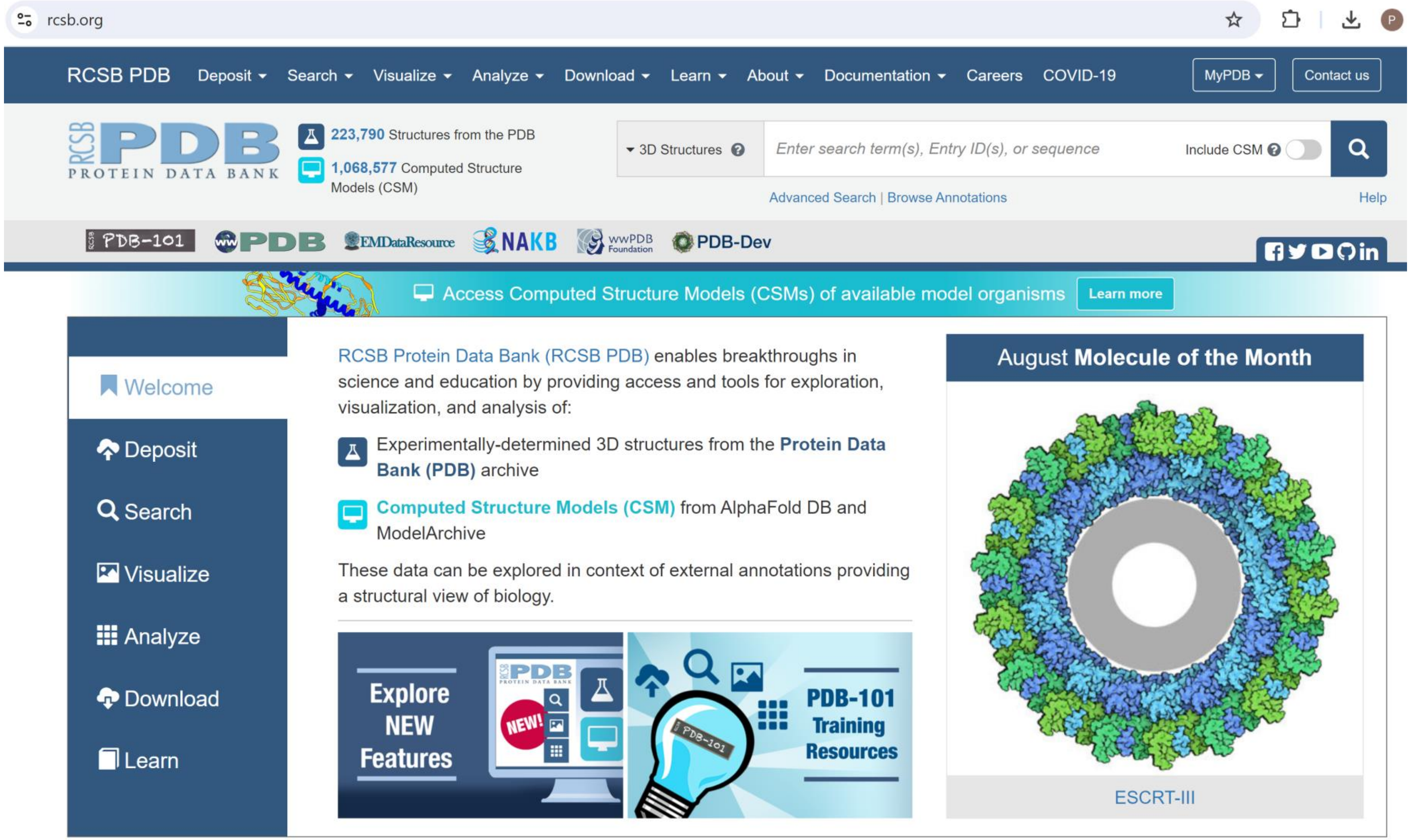
The monoaminergic hypothesis of depression emerged in 1965 and linked depression with dysfunction of neurotransmitters such as noradrenaline and serotonin.[13] Indeed, low levels of serotonin have been observed in the cerebrospinal fluid of patients diagnosed with depression.[3] As a result of this hypothesis, drugs that modulate levels of serotonin such as fluoxetine were developed.[13]

Fluoxetine is a selective serotonin reuptake inhibitor (SSRI) and as the name suggests, it exerts it's therapeutic effect by inhibiting the presynaptic reuptake of the neurotransmitter serotonin.[2] As a result, levels of 5-hydroxytryptamine (5-HT) are increased in various parts of the brain.[13] Further, fluoxetine has high affinity for 5-HT transporters, weak affinity for noradrenaline transporters and no affinity for dopamine transporters indicating that it is 5-HT selective.[13]

Fluoxetine interacts to a degree with the 5-HT_{2C} receptor and it has been suggested that through this mechanism, it is able to increase noradrenaline and dopamine levels in the prefrontal cortex.[13]

TARGET	ACTIONS	ORGANISM
A Sodium-dependent serotonin transporter	inhibitor	Humans
U 5-hydroxytryptamine receptor 2C	antagonist	Humans
U Neuronal acetylcholine receptor subunit alpha-2	antagonist	Humans
U Neuronal acetylcholine receptor subunit alpha-3	antagonist	Humans
U Neuronal acetylcholine receptor subunit beta-4	antagonist	Humans
U Cyclin-dependent kinases regulatory subunit 1	Not Available	Humans
U Potassium voltage-gated channel subfamily H member 2	inhibitor	Humans

Target Identification cont.



223,790 Structures from the PDB

1,068,577 Computed Structure Models (CSM)

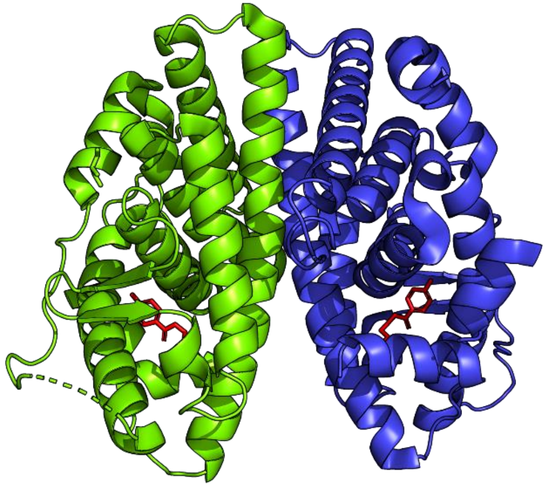
Access Computed Structure Models (CSMs) of available model organisms

RCSB Protein Data Bank (RCSB PDB) enables breakthroughs in science and education by providing access and tools for exploration, visualization, and analysis of:

- Experimentally-determined 3D structures from the **Protein Data Bank (PDB)** archive
- **Computed Structure Models (CSM)** from AlphaFold DB and ModelArchive

These data can be explored in context of external annotations providing a structural view of biology.

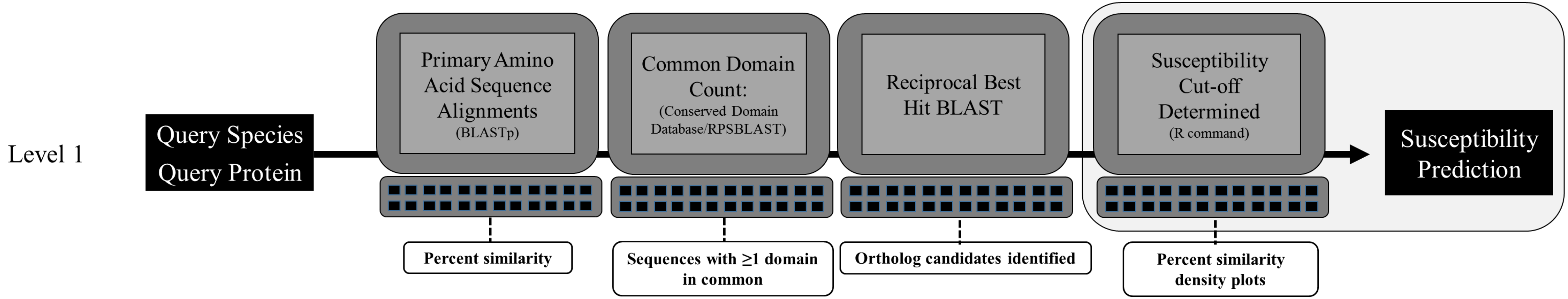
PDB ID: 4MG9



Human estrogen receptor alpha bound to butylparaben

EPA
United States
Environmental Protection
Agency

Level 1: Primary Amino Acid Sequence Comparisons



Level 1

Query Species
Query Protein

Primary Amino Acid Sequence Alignments
(BLASTp)

Percent similarity

Common Domain Count:
(Conserved Domain Database/RPSBLAST)

Sequences with ≥1 domain in common

Reciprocal Best Hit BLAST

Ortholog candidates identified

Susceptibility Cut-off Determined
(R command)

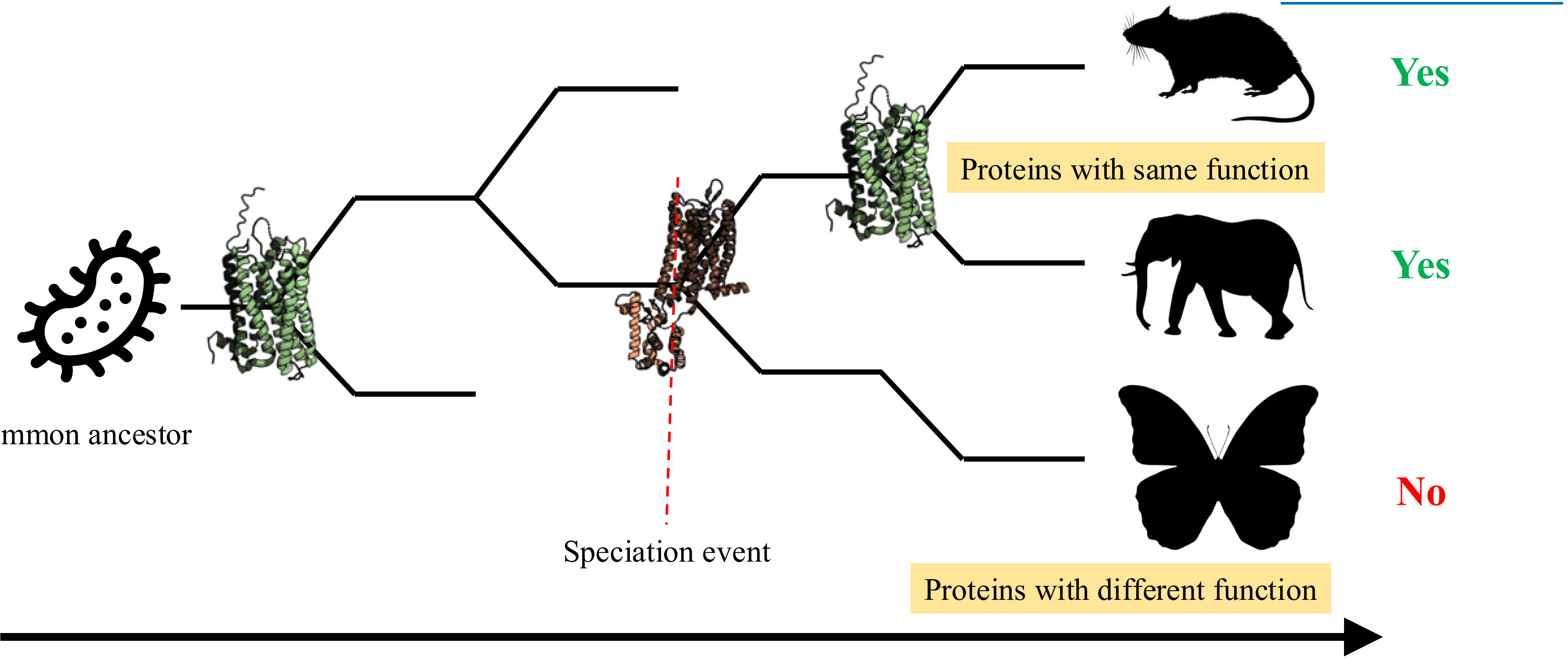
Percent similarity density plots

Susceptibility Prediction

	Align Full Primary Amino Acid Sequence	Bit Score	Percent Similarity
Query Sequence		1241.9	100
Hit		1229.5	99.0
Hit		1223.0	98.5
Hit		1111.3	89.5
Hit		862.4	69.4

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

What is a protein "ortholog"?



Evolutionary Time

Ortholog Candidates Set the Similarity Cutoff

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

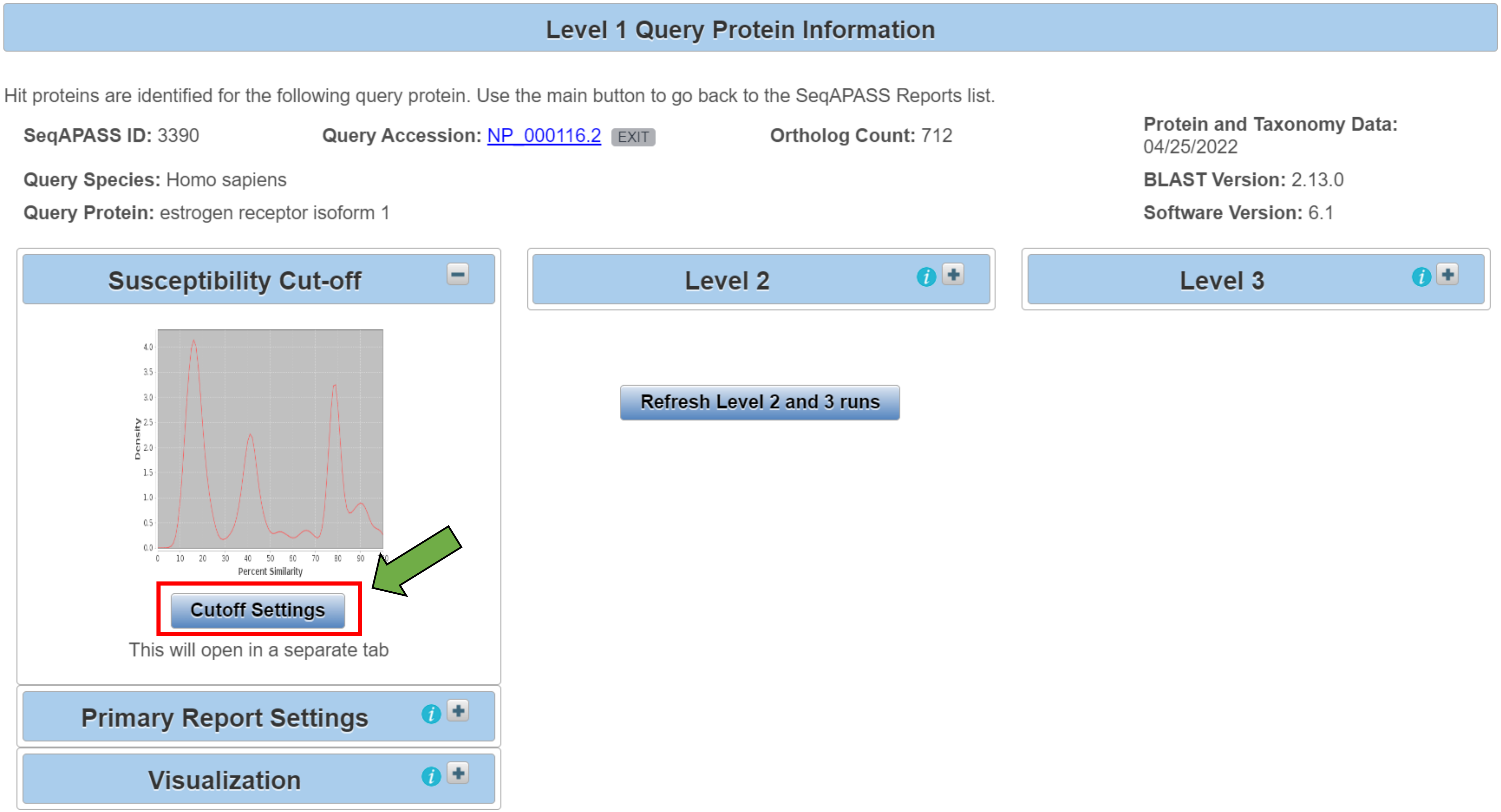
Susceptibility Prediction = Yes

Lowest % Similarity that is still an ortholog

Susceptibility Prediction = No

Even if the % sequence similarity is low, the protein *function* can still be conserved!

→ *Chemicals can disrupt or alter protein function*



Level 1 Query Protein Information

Hit proteins are identified for the following query protein. Use the main button to go back to the SeqAPASS Reports list.

SeqAPASS ID: 3390

Query Accession: NP_000116.2 EXIT

Ortholog Count: 712

Protein and Taxonomy Data: 04/25/2022

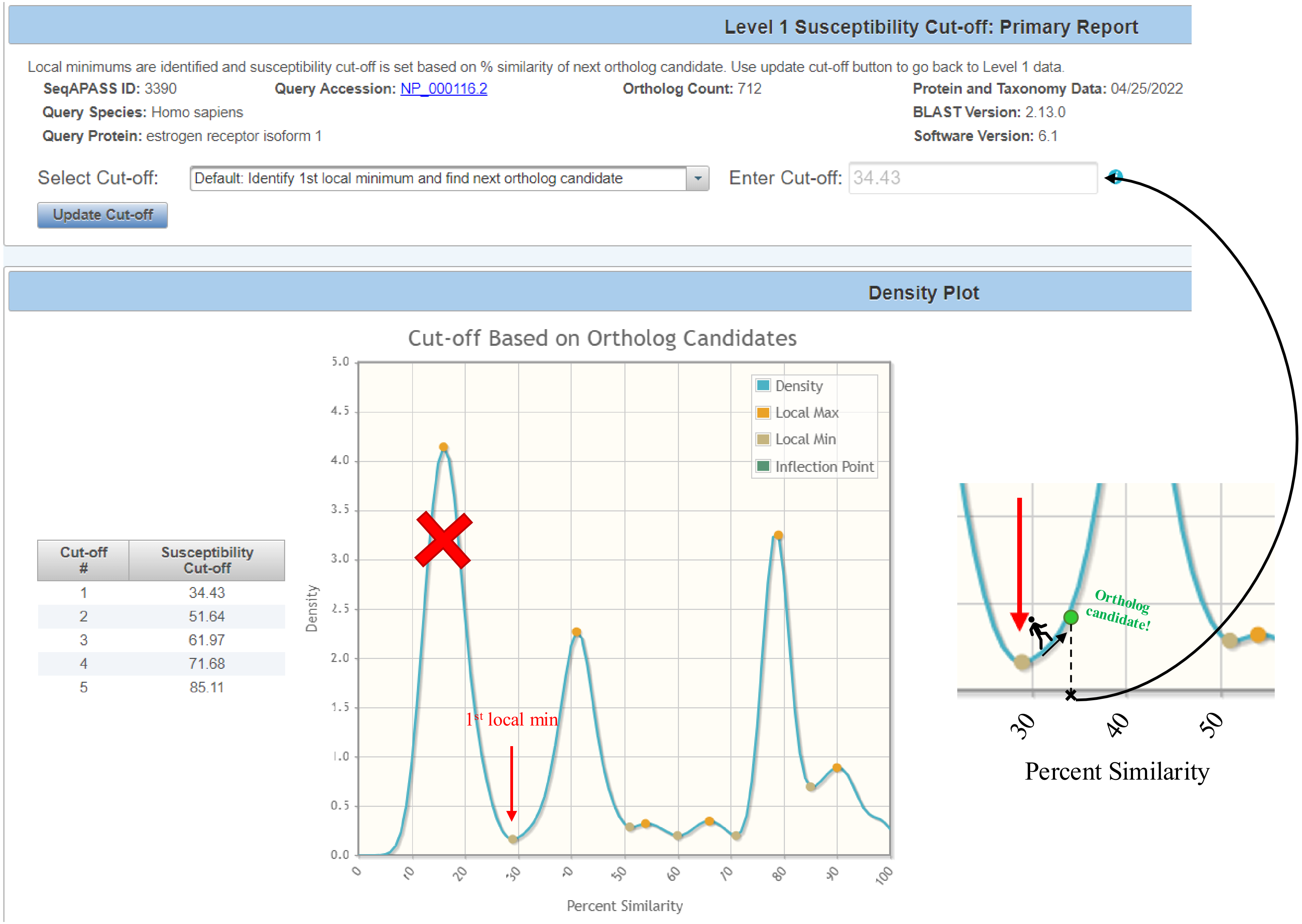
Query Species: Homo sapiens

BLAST Version: 2.13.0

Query Protein: estrogen receptor isoform 1

Software Version: 6.1

This will open in a separate tab



Level 1 Susceptibility Cut-off: Primary Report

Local minimums are identified and susceptibility cut-off is set based on % similarity of next ortholog candidate. Use update cut-off button to go back to Level 1 data.

SeqAPASS ID: 3390
Query Accession: NP_000116.2
Ortholog Count: 712
Protein and Taxonomy Data: 04/25/2022
Query Species: Homo sapiens
BLAST Version: 2.13.0
Query Protein: estrogen receptor isoform 1
Software Version: 6.1

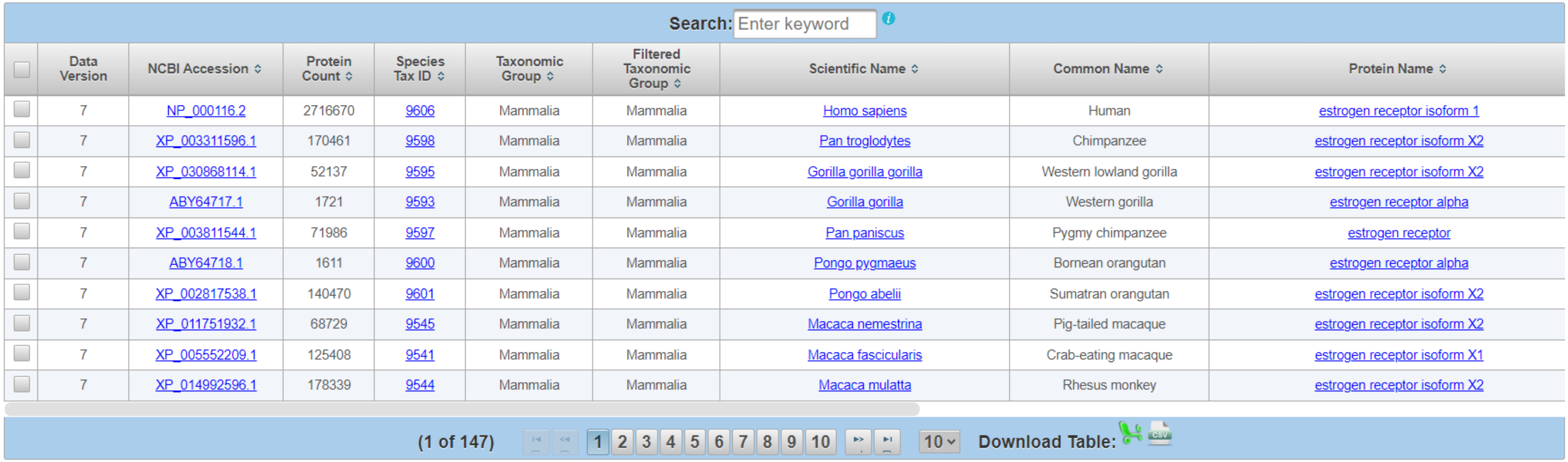
Select Cut-off: Default: Identify 1st local minimum and find next ortholog candidate
Enter Cut-off: 34.43

Density Plot

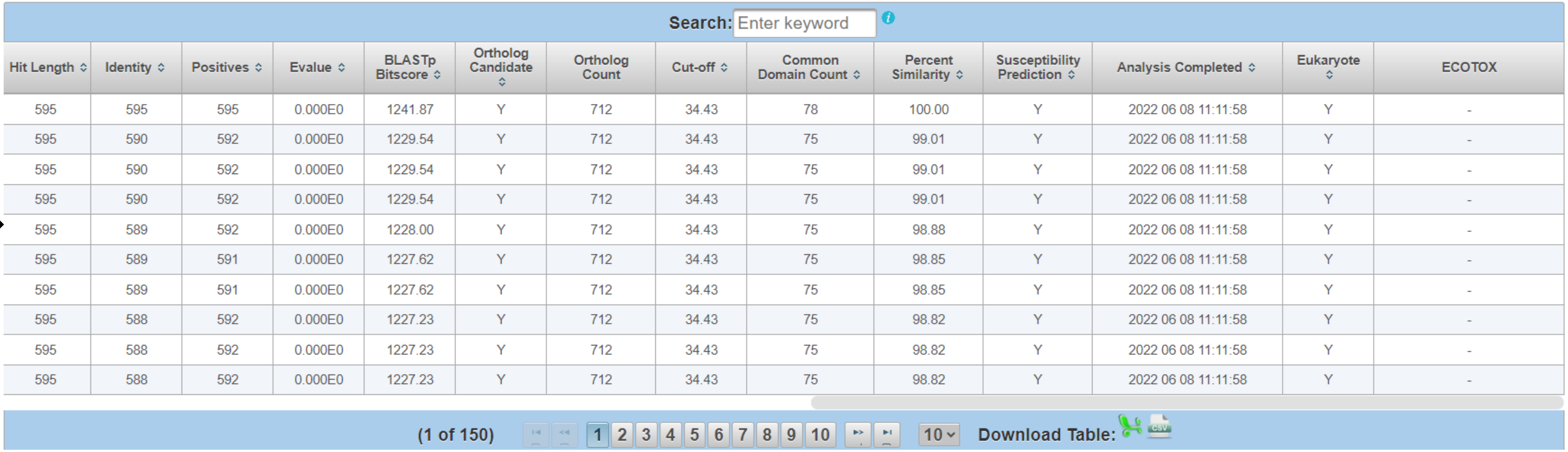
Cut-off #	Susceptibility Cut-off
1	34.43
2	51.64
3	61.97
4	71.68
5	85.11

Cut-off Based on Ortholog Candidates

Level 1 SeqAPASS Data

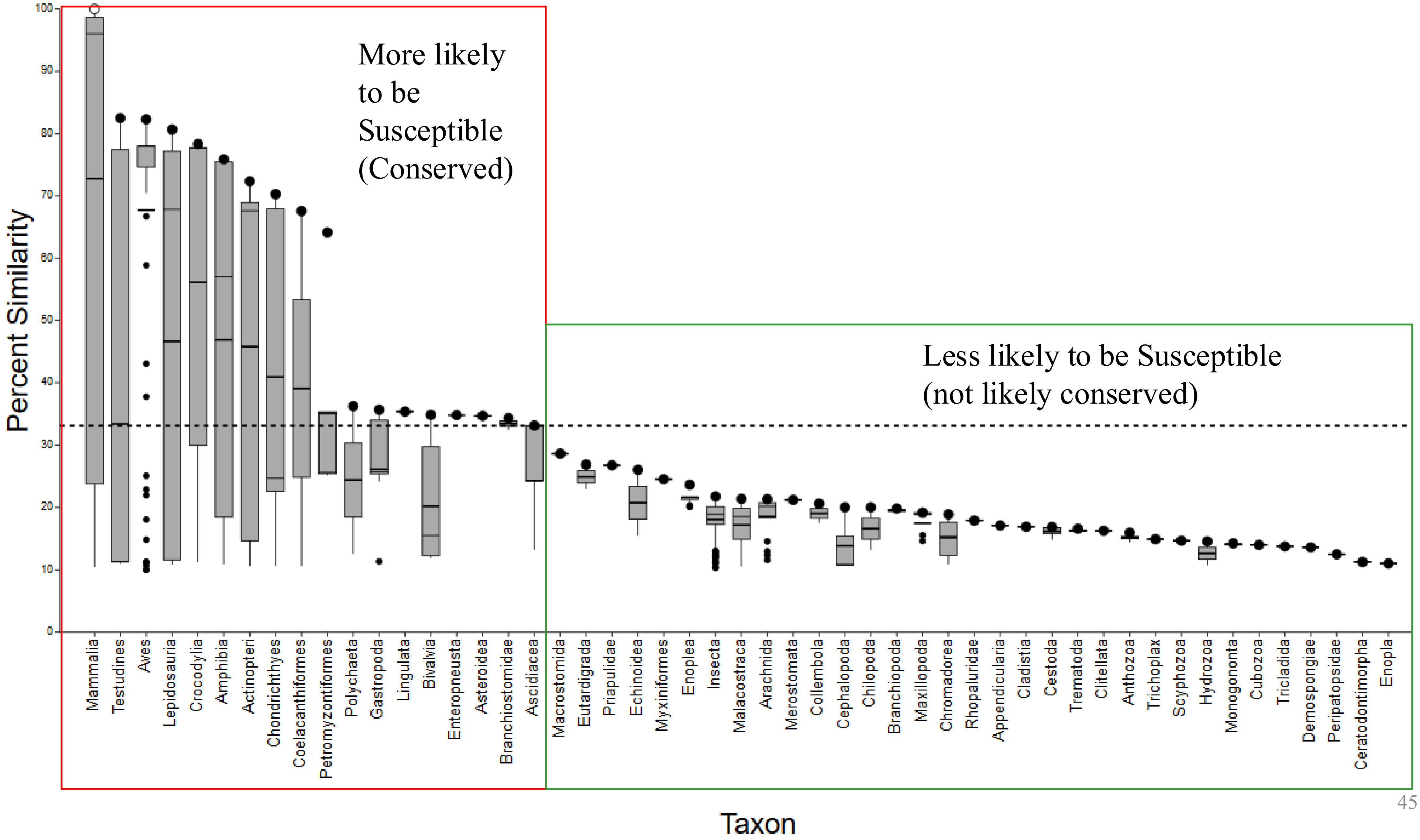


☐	Data Version	NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Filtered Taxonomic Group	Scientific Name	Common Name	Protein Name
☐	7	NP_000116.2	2716670	9606	Mammalia	Mammalia	Homo sapiens	Human	estrogen receptor isoform 1
☐	7	XP_003311596.1	170461	9598	Mammalia	Mammalia	Pan troglodytes	Chimpanzee	estrogen receptor isoform X2
☐	7	XP_030868114.1	52137	9595	Mammalia	Mammalia	Gorilla gorilla gorilla	Western lowland gorilla	estrogen receptor isoform X2
☐	7	ABY64717.1	1721	9593	Mammalia	Mammalia	Gorilla gorilla	Western gorilla	estrogen receptor alpha
☐	7	XP_003811544.1	71986	9597	Mammalia	Mammalia	Pan paniscus	Pygmy chimpanzee	estrogen receptor
☐	7	ABY64718.1	1611	9600	Mammalia	Mammalia	Pongo pygmaeus	Bornean orangutan	estrogen receptor alpha
☐	7	XP_002817538.1	140470	9601	Mammalia	Mammalia	Pongo abelii	Sumatran orangutan	estrogen receptor isoform X2
☐	7	XP_011751932.1	68729	9545	Mammalia	Mammalia	Macaca nemestrina	Pig-tailed macaque	estrogen receptor isoform X2
☐	7	XP_005552209.1	125408	9541	Mammalia	Mammalia	Macaca fascicularis	Crab-eating macaque	estrogen receptor isoform X1
☐	7	XP_014992596.1	178339	9544	Mammalia	Mammalia	Macaca mulatta	Rhesus monkey	estrogen receptor isoform X2



Hit Length	Identity	Positives	Evalue	BLASTp Bitscore	Ortholog Candidate	Ortholog Count	Cut-off	Common Domain Count	Percent Similarity	Susceptibility Prediction	Analysis Completed	Eukaryote	ECOTOX
595	595	595	0.000E0	1241.87	Y	712	34.43	78	100.00	Y	2022 06 08 11:11:58	Y	-
595	590	592	0.000E0	1229.54	Y	712	34.43	75	99.01	Y	2022 06 08 11:11:58	Y	-
595	590	592	0.000E0	1229.54	Y	712	34.43	75	99.01	Y	2022 06 08 11:11:58	Y	-
595	590	592	0.000E0	1229.54	Y	712	34.43	75	99.01	Y	2022 06 08 11:11:58	Y	-
595	589	592	0.000E0	1228.00	Y	712	34.43	75	98.88	Y	2022 06 08 11:11:58	Y	-
595	589	591	0.000E0	1227.62	Y	712	34.43	75	98.85	Y	2022 06 08 11:11:58	Y	-
595	589	591	0.000E0	1227.62	Y	712	34.43	75	98.85	Y	2022 06 08 11:11:58	Y	-
595	588	592	0.000E0	1227.23	Y	712	34.43	75	98.82	Y	2022 06 08 11:11:58	Y	-
595	588	592	0.000E0	1227.23	Y	712	34.43	75	98.82	Y	2022 06 08 11:11:58	Y	-
595	588	592	0.000E0	1227.23	Y	712	34.43	75	98.82	Y	2022 06 08 11:11:58	Y	-

Visualization of Level 1 Results



More likely to be Susceptible (Conserved)

Less likely to be Susceptible (not likely conserved)

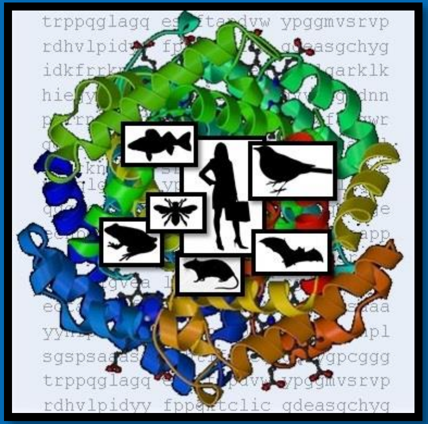
United States
Environmental Protection
Agency

Training Course – SeqAPASS Level 2: Conservation of Functional Domains

Instructors:

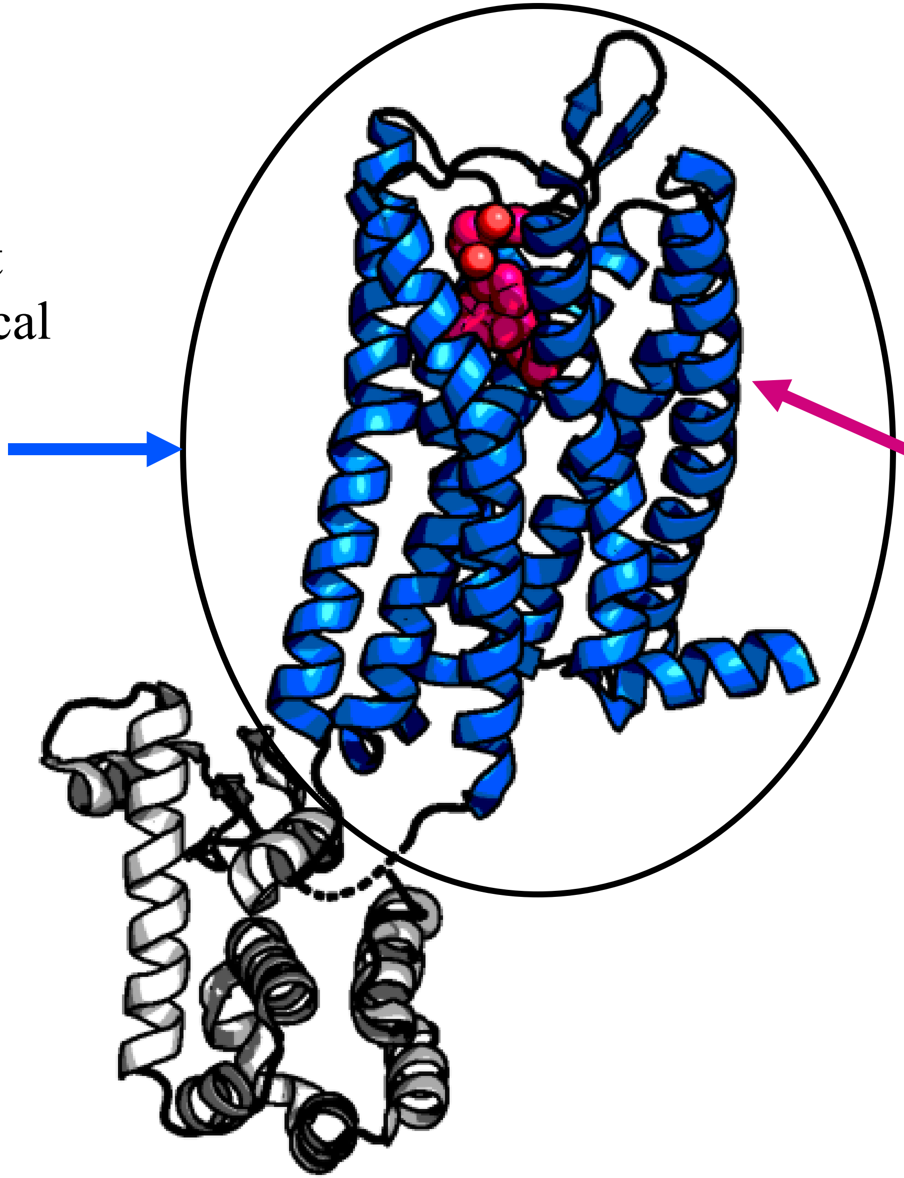
Peter Schumann, MS (US EPA)

Carlie A. LaLone, PhD (US EPA)



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

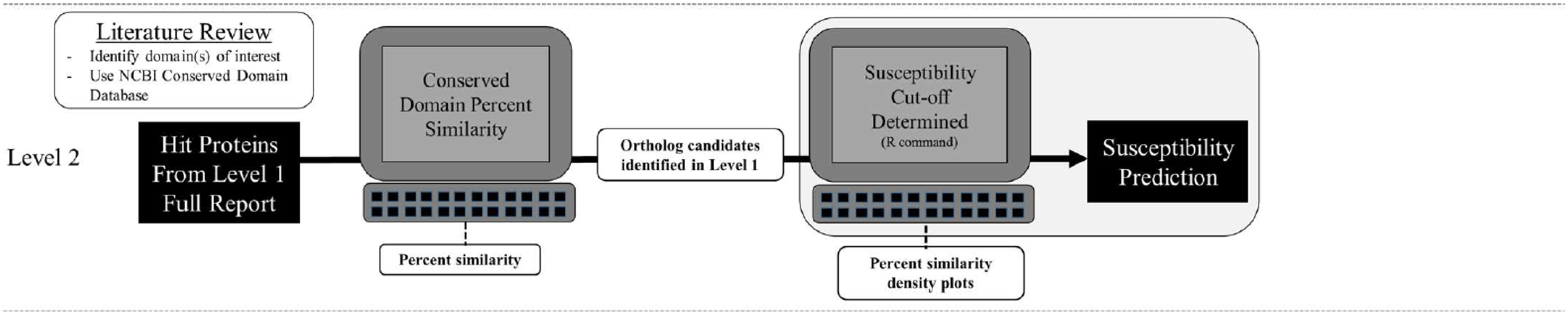
Protein function is structurally localized



Regions of a protein that perform specific biochemical functions are called "**functional domains**"

Chemicals often bind to a functional domain (e.g., "ligand binding domain")

Level 2: Functional domain comparisons



Level 2

Literature Review
- Identify domain(s) of interest
- Use NCBI Conserved Domain Database

Hit Proteins From Level 1 Full Report

Conserved Domain Percent Similarity

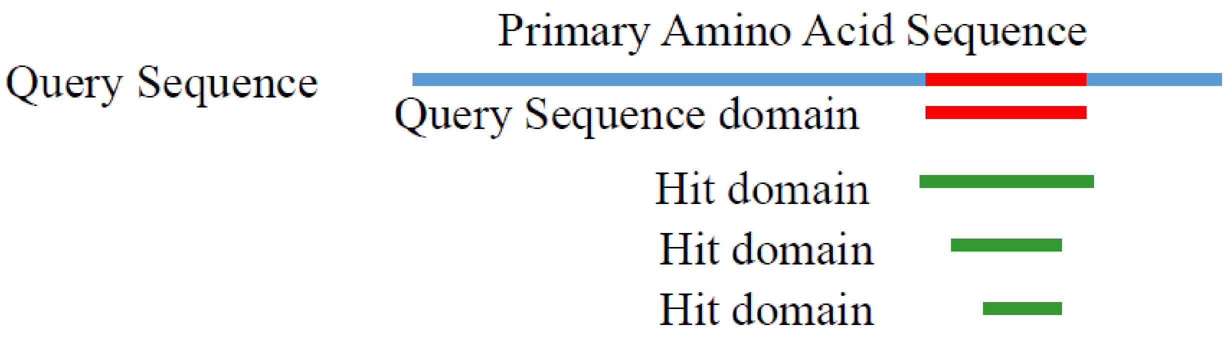
Percent similarity

Ortholog candidates identified in Level 1

Susceptibility Cut-off Determined (R command)

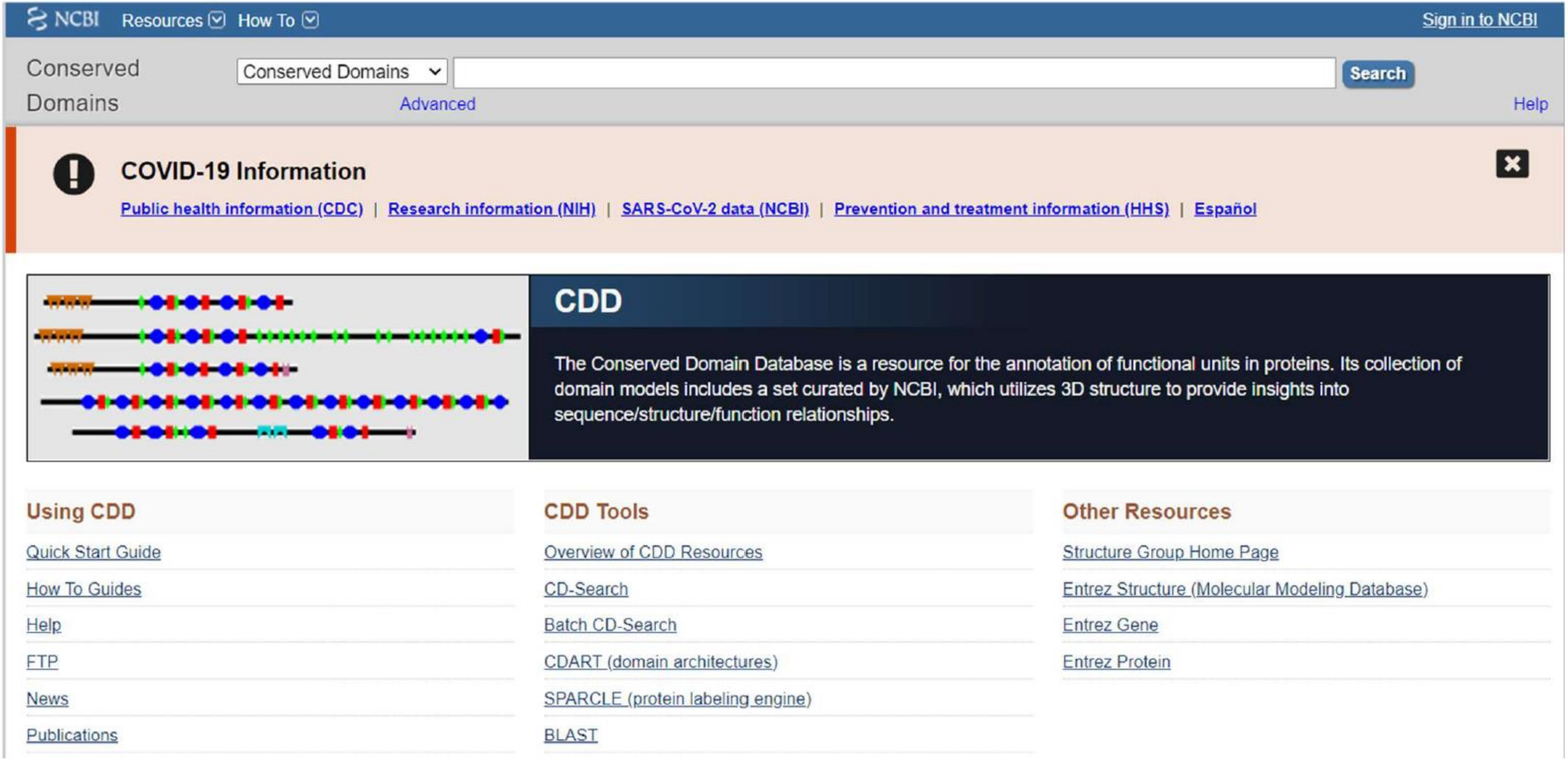
Percent similarity density plots

Susceptibility Prediction



	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

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



COVID-19 Information

Public health information (CDC) | Research information (NIH) | SARS-CoV-2 data (NCBI) | Prevention and treatment information (HHS) | Español

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.

Using CDD

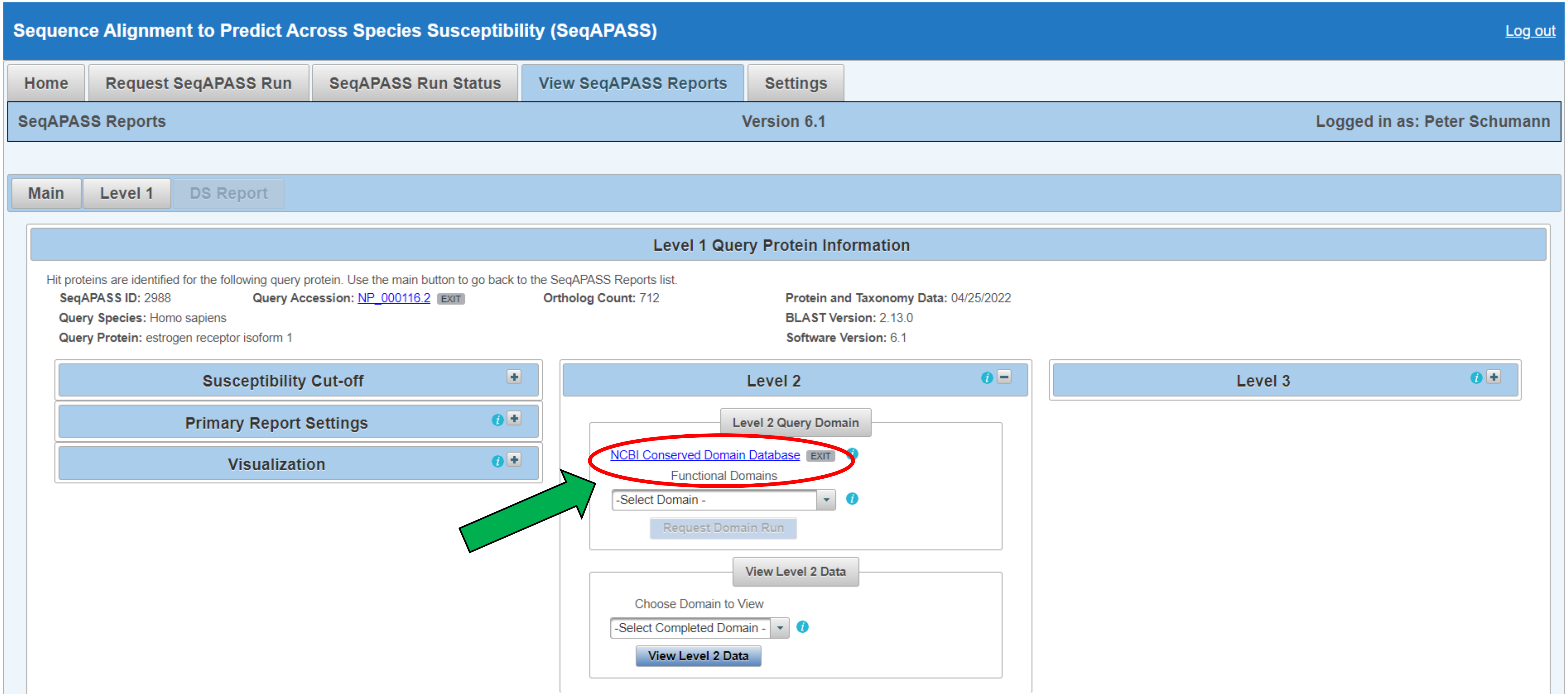
- Quick Start Guide
- How To Guides
- Help
- FTP
- News
- Publications

CDD Tools

- Overview of CDD Resources
- CD-Search
- Batch CD-Search
- CDART (domain architectures)
- SPARCLE (protein labeling engine)
- BLAST

Other Resources

- Structure Group Home Page
- Entrez Structure (Molecular Modeling Database)
- Entrez Gene
- Entrez Protein



Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)

SeqAPASS Reports

Version 6.1

Logged in as: Peter Schumann

Level 1 Query Protein Information

Hit proteins are identified for the following query protein. Use the main button to go back to the SeqAPASS Reports list.

SeqAPASS ID: 2988
Query Accession: NP_000116.2
Ortholog Count: 712
Protein and Taxonomy Data: 04/25/2022
Query Species: Homo sapiens
BLAST Version: 2.13.0
Query Protein: estrogen receptor isoform 1
Software Version: 6.1

Level 2

Level 2 Query Domain

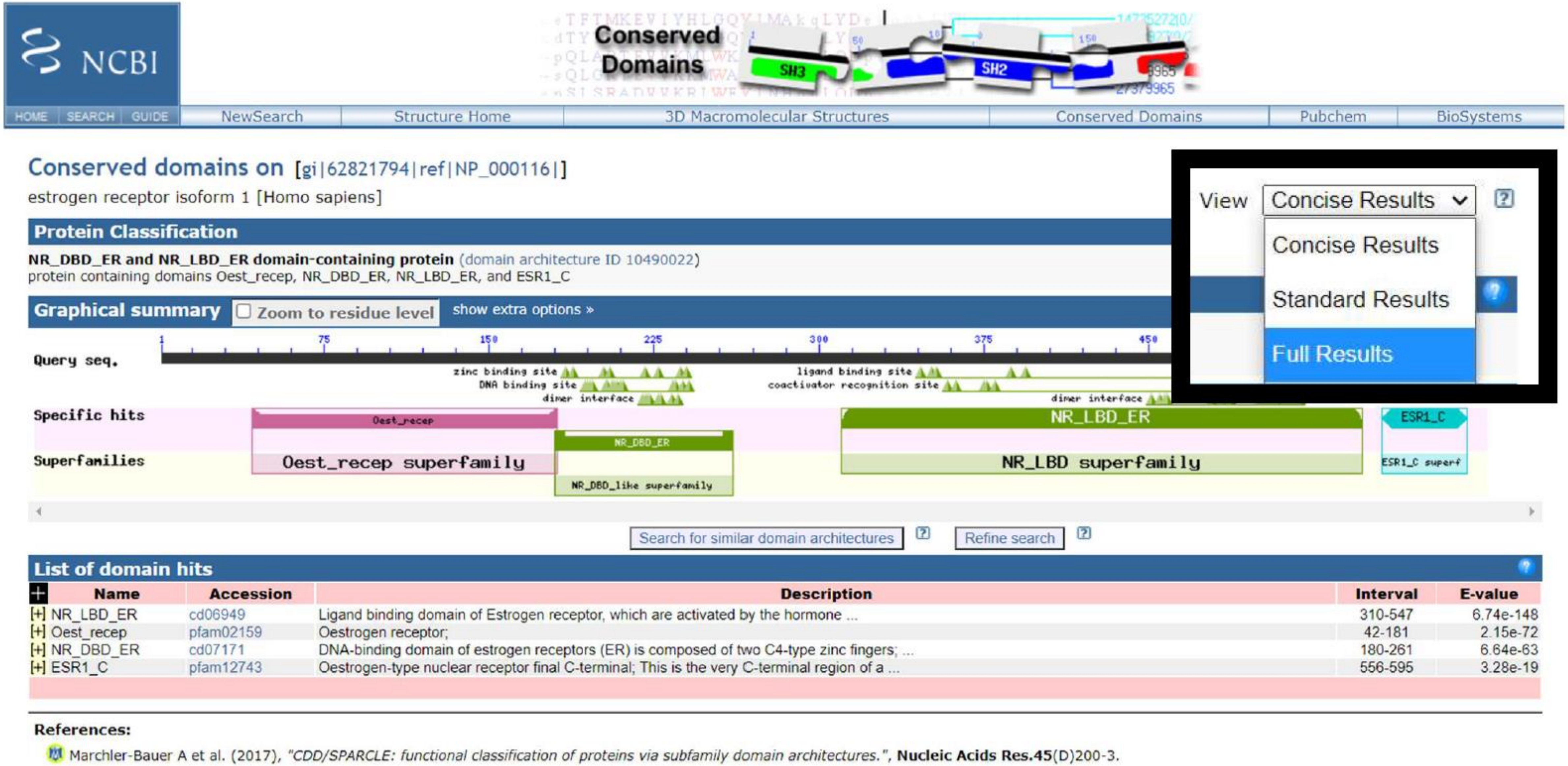
NCBI Conserved Domain Database

Functional Domains

View Level 2 Data

Choose Domain to View

EPA
United States
Environmental Protection
Agency



Conserved domains on [gi|62821794|ref|NP_000116|]

estrogen receptor isoform 1 [Homo sapiens]

Protein Classification

NR_DBD_ER and NR_LBD_ER domain-containing protein (domain architecture ID 10490022)
protein containing domains Oest_recep, NR_DBD_ER, NR_LBD_ER, and ESR1_C

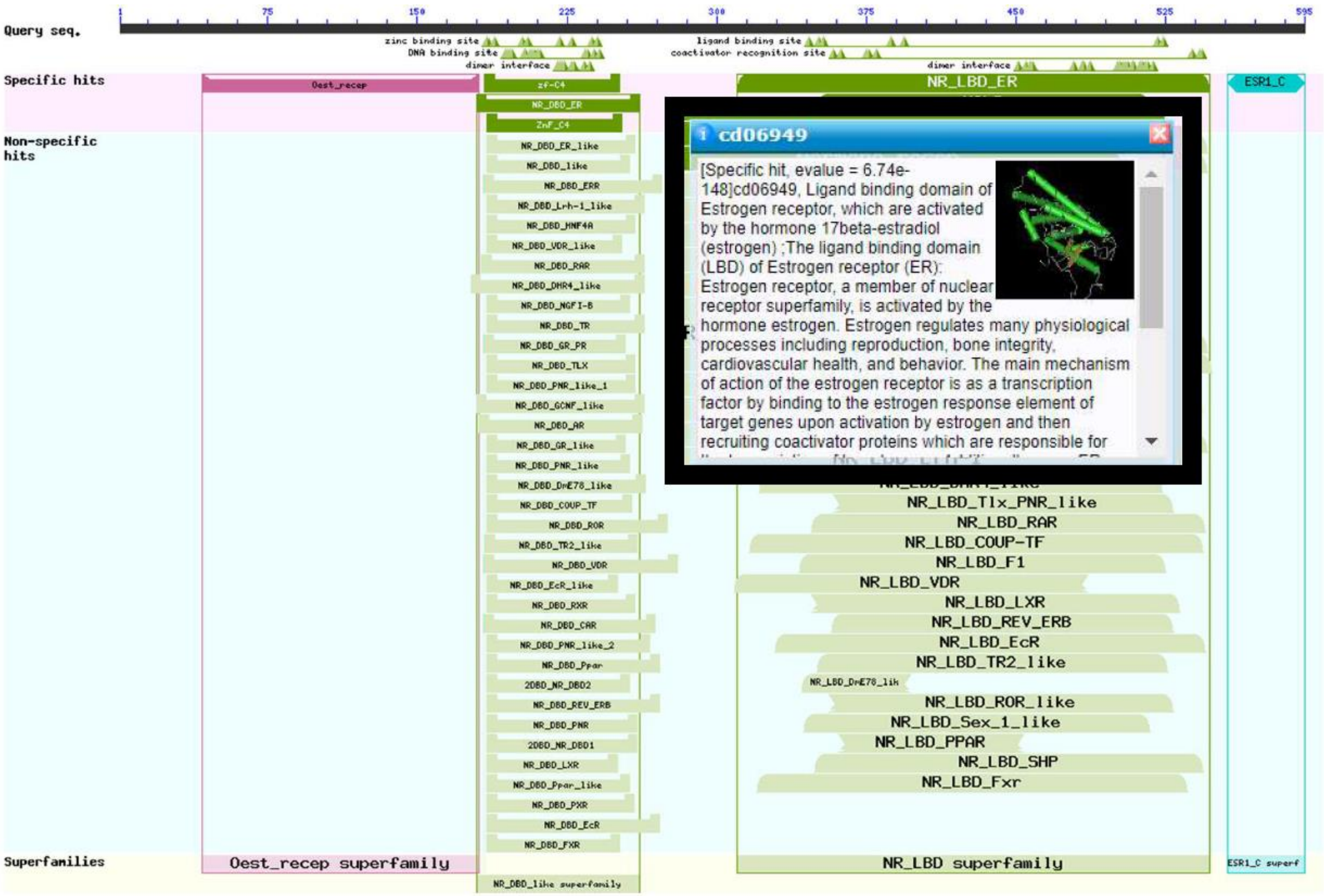
Graphical summary ☐ Zoom to residue level show extra options »

List of domain hits

Name	Accession	Description	Interval	E-value
[+] NR_LBD_ER	cd06949	Ligand binding domain of Estrogen receptor, which are activated by the hormone ...	310-547	6.74e-148
[+] Oest_recep	pfam02159	Oestrogen receptor;	42-181	2.15e-72
[+] NR_DBD_ER	cd07171	DNA-binding domain of estrogen receptors (ER) is composed of two C4-type zinc fingers; ...	180-261	6.64e-63
[+] ESR1_C	pfam12743	Oestrogen-type nuclear receptor final C-terminal; This is the very C-terminal region of a ...	556-595	3.28e-19

References:

Marchler-Bauer A et al. (2017), *"CDD/SPARCLE: functional classification of proteins via subfamily domain architectures."*, **Nucleic Acids Res.45**(D)200-3.



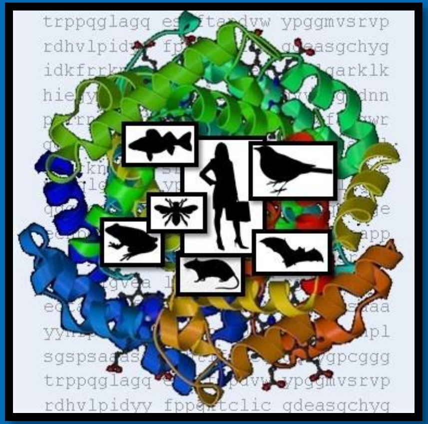
[Specific hit, evalue = 6.74e-148]cd06949, Ligand binding domain of Estrogen receptor, which are activated by the hormone 17beta-estradiol (estrogen) ;The ligand binding domain (LBD) of Estrogen receptor (ER): Estrogen receptor, a member of nuclear receptor superfamily, is activated by the hormone estrogen. Estrogen regulates many physiological processes including reproduction, bone integrity, cardiovascular health, and behavior. The main mechanism of action of the estrogen receptor is as a transcription factor by binding to the estrogen response element of target genes upon activation by estrogen and then recruiting coactivator proteins which are responsible for

Training Course – SeqAPASS Level 3: Key amino acid comparison

Instructors:

Peter Schumann, MS (US EPA)

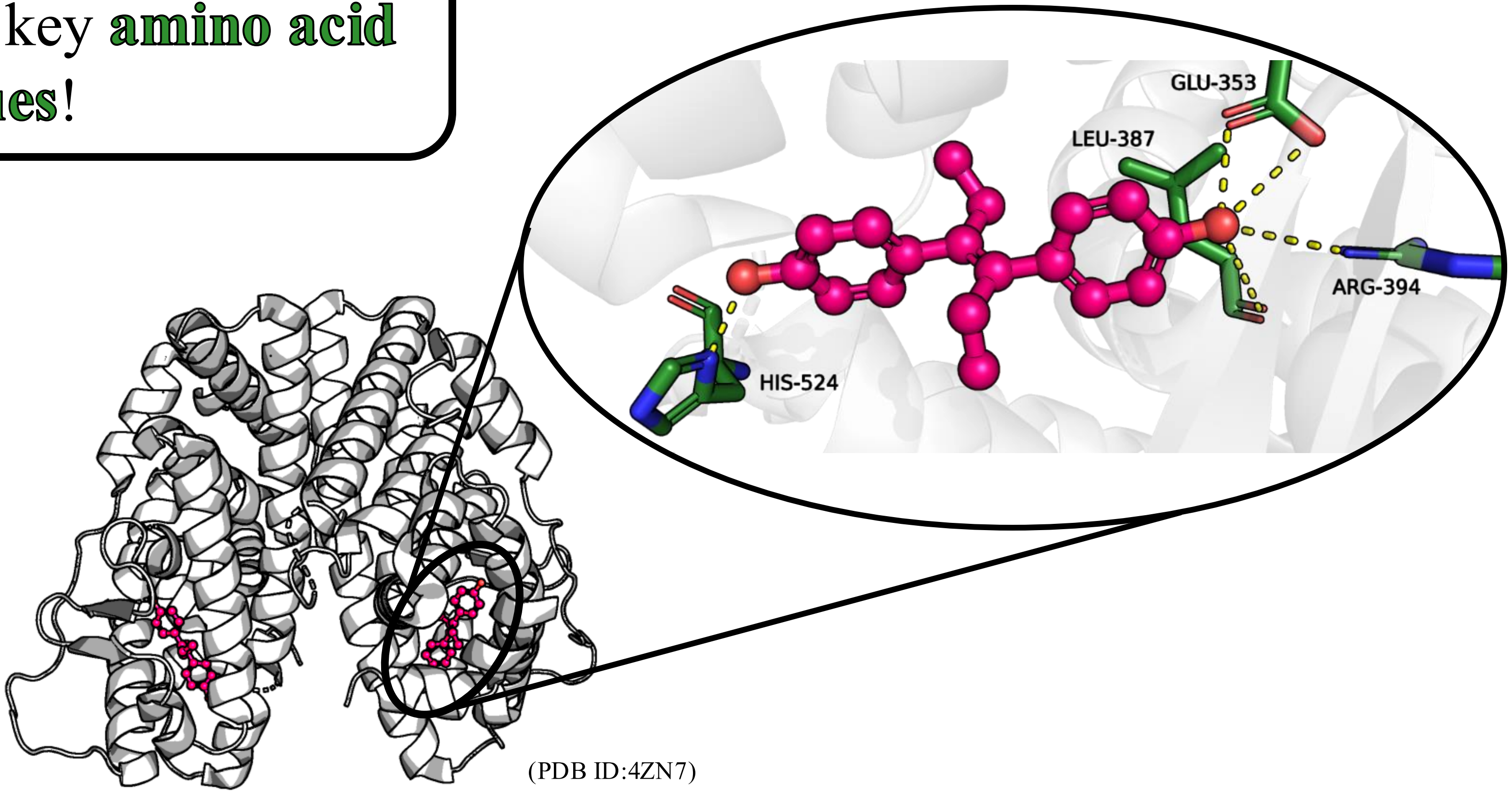
Carlie A. LaLone, PhD (US EPA)



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

Evaluating the Conservation of Key Protein Residues

Chemical-protein interactions are mediated by a few, key **amino acid residues**!



(PDB ID:4ZN7)

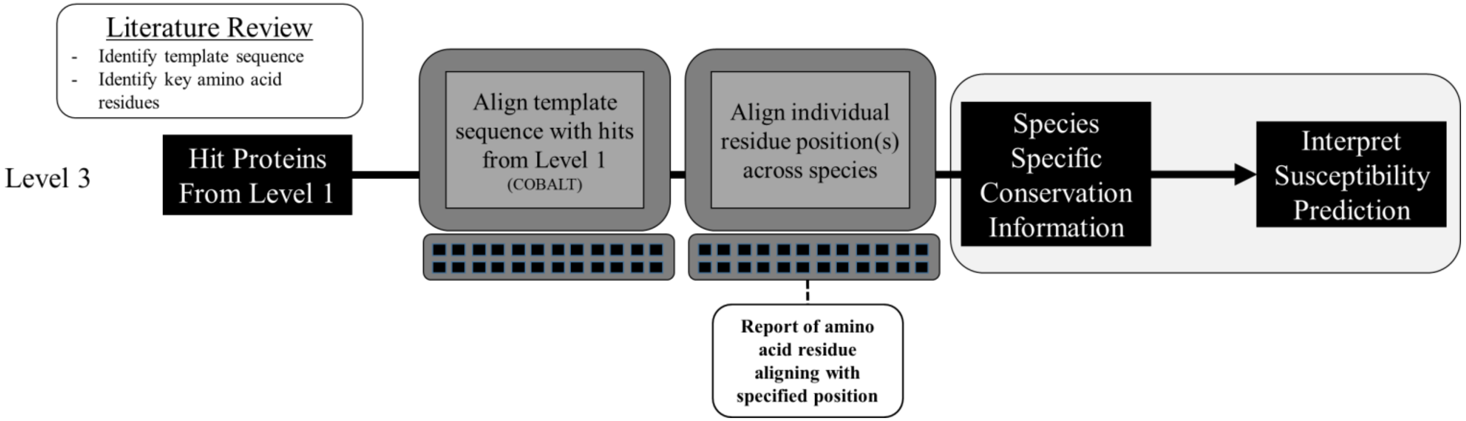
Level 3: Individual Amino Acid Residue Comparison

Information needed for Level 3 query

1. Template Species/Sequence
2. Identified Critical Amino Acids and positions of amino acids in template

Template Sequence
Critical amino acid residues
Hit aa position
Hit aa position
Hit aa position

Level 3: Individual Amino Acid Residue Comparison



Literature Review
- Identify template sequence
- Identify key amino acid residues

Level 3

Hit Proteins From Level 1

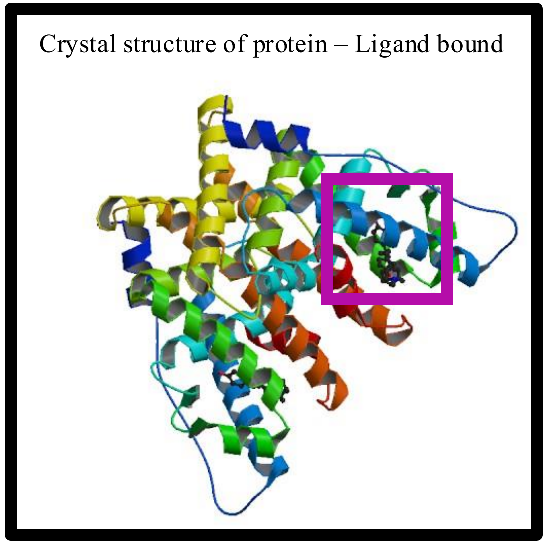
Align template sequence with hits from Level 1 (COBALT)

Align individual residue position(s) across species

Species Specific Conservation Information

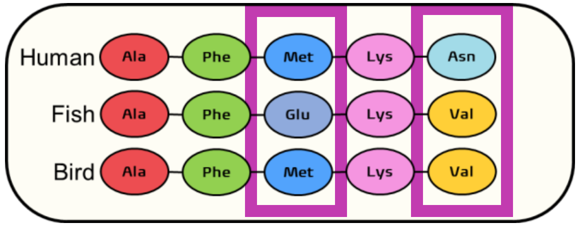
Interpret Susceptibility Prediction

Report of amino acid residue aligning with specified position



Crystal structure of protein – Ligand bound

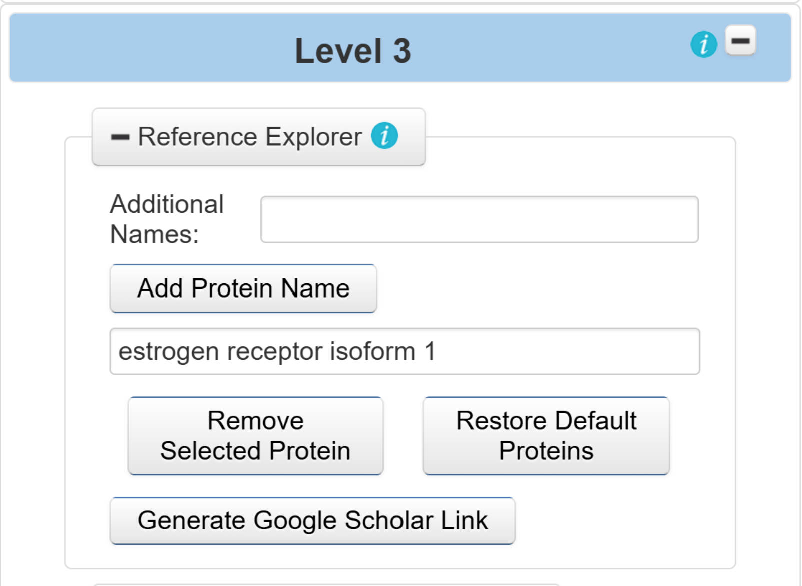
Amino acid residues that interact with the chemical



Human	Ala	Phe	Met	Lys	Asn
Fish	Ala	Phe	Glu	Lys	Val
Bird	Ala	Phe	Met	Lys	Val

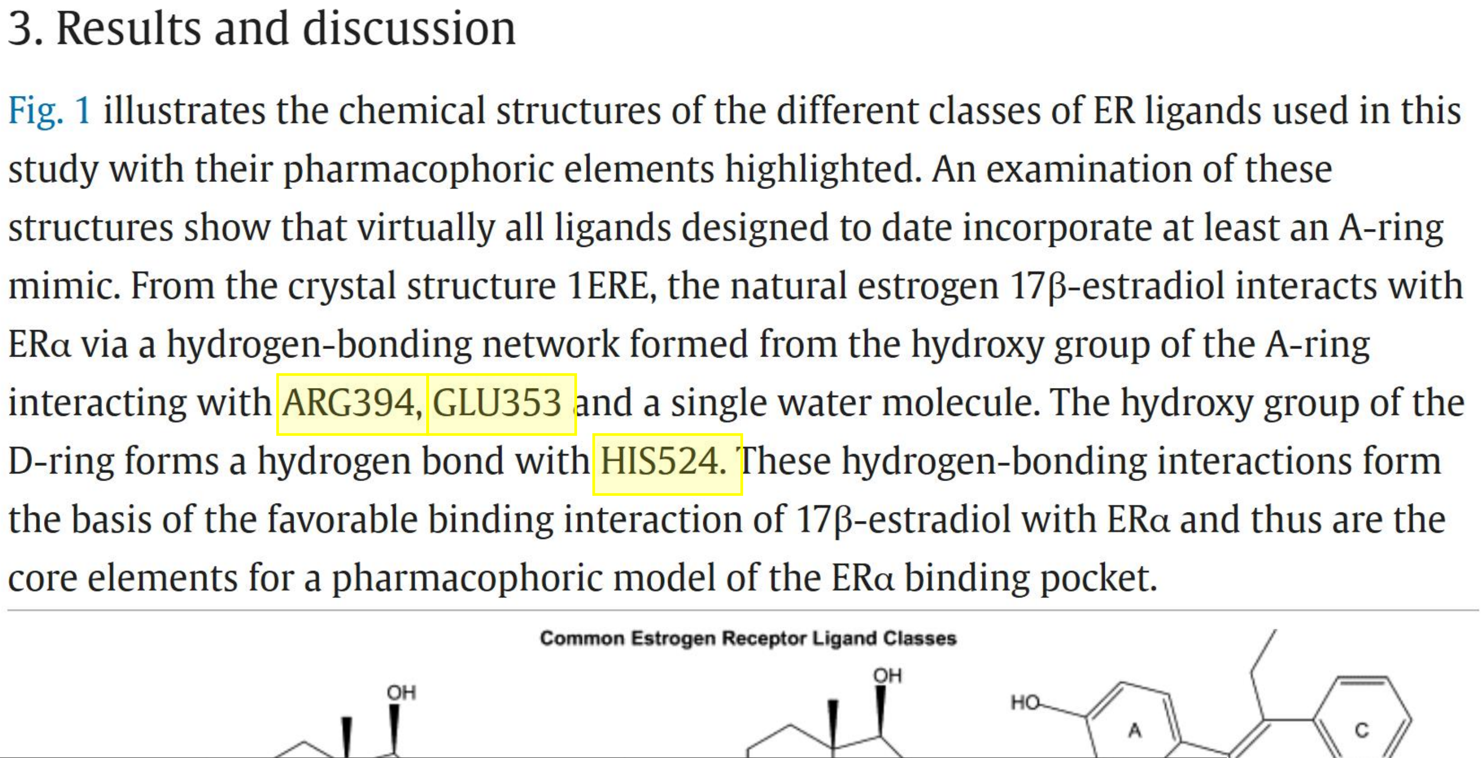
Where can I find this information?

- Literature Review
  - Types of studies:
    - Site-directed mutagenesis
    - Field resistance (pesticides)
    - Studies of x-ray crystallography
    - Homology modeling



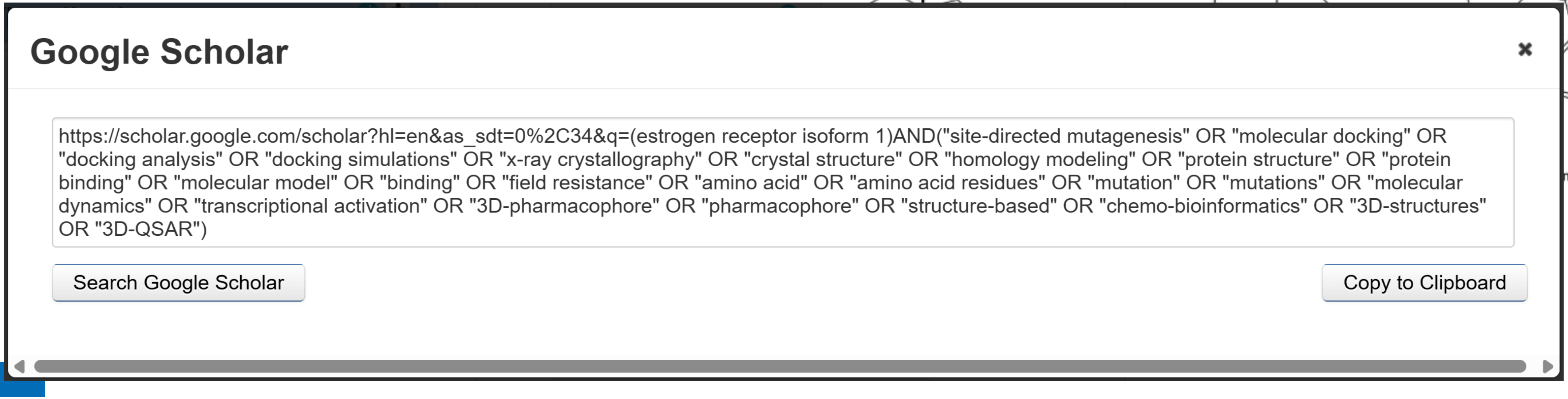
Additional Names:

estrogen receptor isoform 1



3. Results and discussion

Fig. 1 illustrates the chemical structures of the different classes of ER ligands used in this study with their pharmacophoric elements highlighted. An examination of these structures show that virtually all ligands designed to date incorporate at least an A-ring mimic. From the crystal structure 1ERE, the natural estrogen 17β-estradiol interacts with ERα via a hydrogen-bonding network formed from the hydroxy group of the A-ring interacting with ARG394, GLU353 and a single water molecule. The hydroxy group of the D-ring forms a hydrogen bond with HIS524. These hydrogen-bonding interactions form the basis of the favorable binding interaction of 17β-estradiol with ERα and thus are the core elements for a pharmacophoric model of the ERα binding pocket.



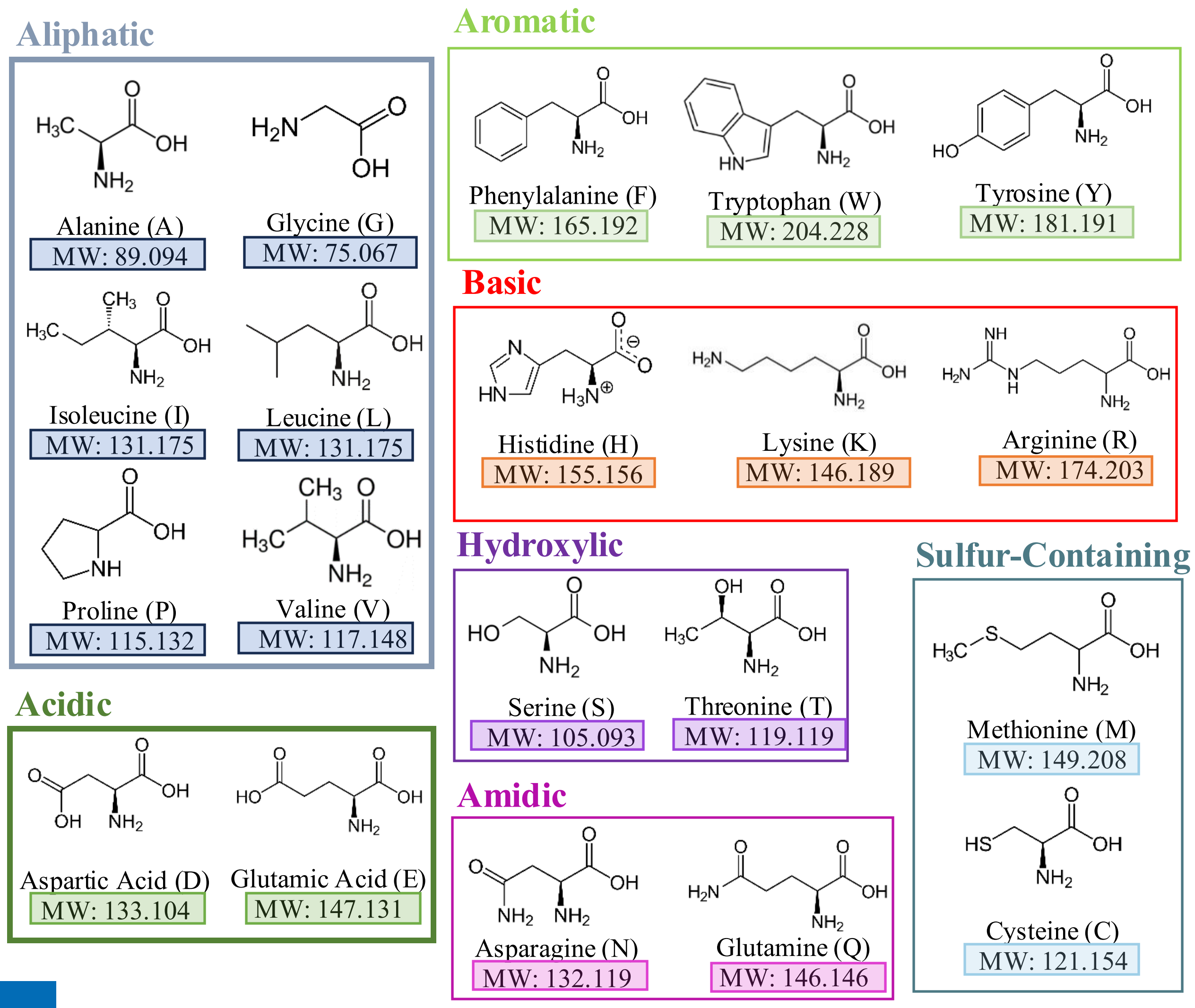
Google Scholar

https://scholar.google.com/scholar?hl=en&as_sdt=0%2C34&q=(estrogen receptor isoform 1)AND("site-directed mutagenesis" OR "molecular docking" OR "docking analysis" OR "docking simulations" OR "x-ray crystallography" OR "crystal structure" OR "homology modeling" OR "protein structure" OR "protein binding" OR "molecular model" OR "binding" OR "field resistance" OR "amino acid" OR "amino acid residues" OR "mutation" OR "mutations" OR "molecular dynamics" OR "transcriptional activation" OR "3D-pharmacophore" OR "pharmacophore" OR "structure-based" OR "chemo-bioinformatics" OR "3D-structures" OR "3D-QSAR")

EPA
United States
Environmental Protection
Agency

Level 3: Continued

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



- Automated Prediction

- 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: Continued

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.

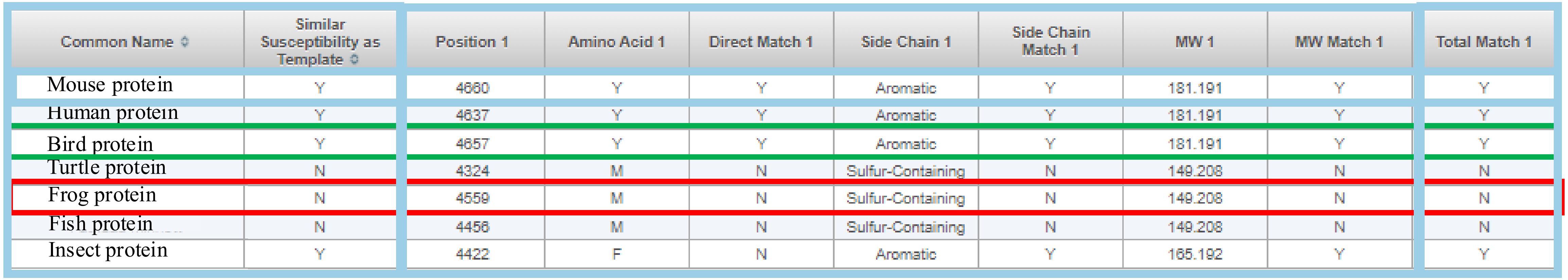
Tyr - Y

Yes/No

Yes/No

Yes/No

Yes/No

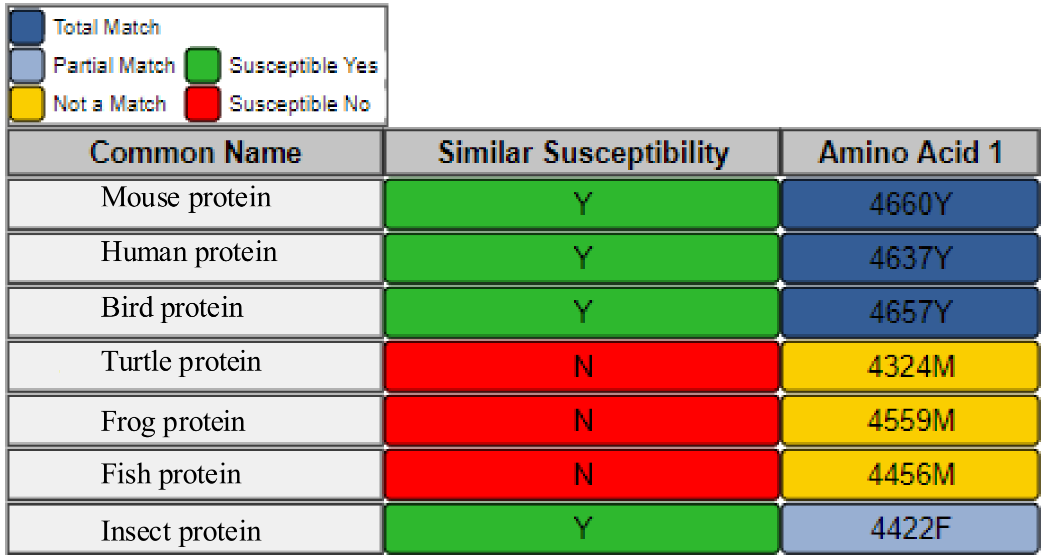


Common Name	Similar Susceptibility as Template	Position 1	Amino Acid 1	Direct Match 1	Side Chain 1	Side Chain Match 1	MW 1	MW Match 1	Total Match 1
Mouse protein	Y	4660	Y	Y	Aromatic	Y	181.191	Y	Y
Human protein	Y	4637	Y	Y	Aromatic	Y	181.191	Y	Y
Bird protein	Y	4657	Y	Y	Aromatic	Y	181.191	Y	Y
Turtle protein	N	4324	M	N	Sulfur-Containing	N	149.208	N	N
Frog protein	N	4559	M	N	Sulfur-Containing	N	149.208	N	N
Fish protein	N	4456	M	N	Sulfur-Containing	N	149.208	N	N
Insect protein	Y	4422	F	N	Aromatic	Y	165.192	Y	Y

Level 3: Visualization

Heat Map Visualization!

Simple report:



🟦 Total Match
▨ Partial Match 🟩 Susceptible Yes
🟨 Not a Match 🟥 Susceptible No

Common Name	Similar Susceptibility	Amino Acid 1
Mouse protein	Y	4660Y
Human protein	Y	4637Y
Bird protein	Y	4657Y
Turtle protein	N	4324M
Frog protein	N	4559M
Fish protein	N	4456M
Insect protein	Y	4422F

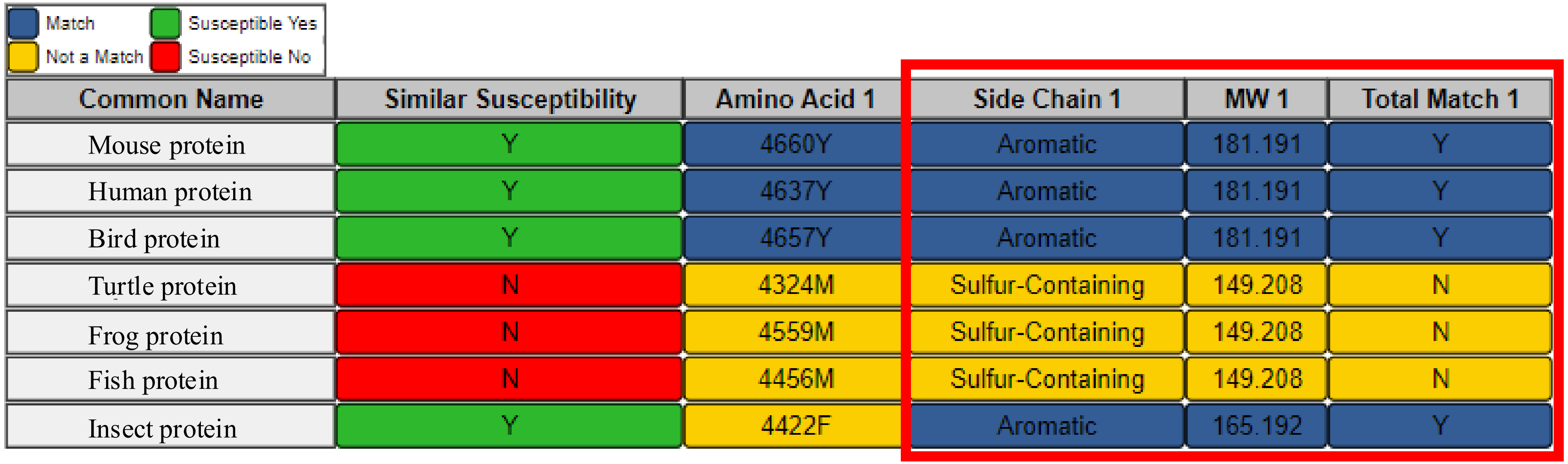
Customizable:

- Common/scientific name
- Ortholog candidates
- Endangered species
- Threatened species
- Common model organisms
- Amino acid information displayed

Level 3: Visualization

Heat Map Visualization!

Full report:



🟦 Match 🟩 Susceptible Yes
🟨 Not a Match 🟥 Susceptible No

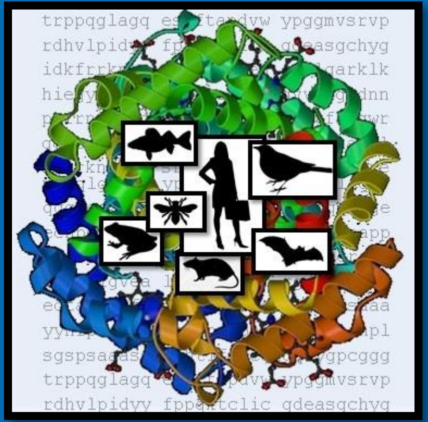
Common Name	Similar Susceptibility	Amino Acid 1	Side Chain 1	MW 1	Total Match 1
Mouse protein	Y	4660Y	Aromatic	181.191	Y
Human protein	Y	4637Y	Aromatic	181.191	Y
Bird protein	Y	4657Y	Aromatic	181.191	Y
Turtle protein	N	4324M	Sulfur-Containing	149.208	N
Frog protein	N	4559M	Sulfur-Containing	149.208	N
Fish protein	N	4456M	Sulfur-Containing	149.208	N
Insect protein	Y	4422F	Aromatic	165.192	Y

Training Course - SeqAPASS: Decision Summary Report

Instructors:

Peter Schumann, MS (US EPA)

Carlie A. LaLone, PhD (US EPA)



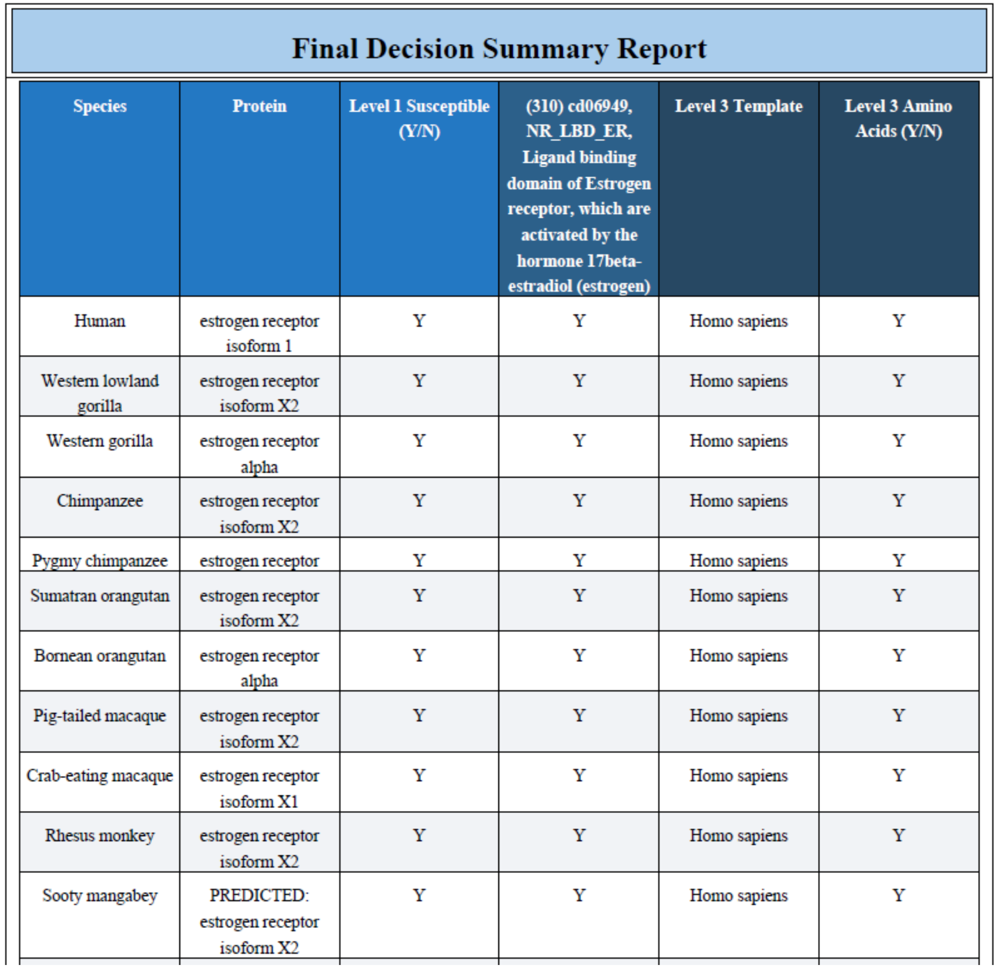
Office of Research and Development
Center for Computational Toxicology and Exposure, Great Lakes Toxicology and Ecology Division

Decision Summary (DS) Report

- Organize data
  - Can contain one Level 1, multiple Level 2, and one Level 3 run(s)

- Customized display for PDF

- Provides manuscript ready visualizations



Final Decision Summary Report

Species	Protein	Level 1 Susceptible (Y/N)	(310) cd06949, NR_LBD_ER, Ligand binding domain of Estrogen receptor, which are activated by the hormone 17beta-estradiol (estrogen)	Level 3 Template	Level 3 Amino Acids (Y/N)
Human	estrogen receptor isoform 1	Y	Y	Homo sapiens	Y
Western lowland gorilla	estrogen receptor isoform X2	Y	Y	Homo sapiens	Y
Western gorilla	estrogen receptor alpha	Y	Y	Homo sapiens	Y
Chimpanzee	estrogen receptor isoform X2	Y	Y	Homo sapiens	Y
Pygmy chimpanzee	estrogen receptor	Y	Y	Homo sapiens	Y
Sumatran orangutan	estrogen receptor isoform X2	Y	Y	Homo sapiens	Y
Bornean orangutan	estrogen receptor alpha	Y	Y	Homo sapiens	Y
Pig-tailed macaque	estrogen receptor isoform X2	Y	Y	Homo sapiens	Y
Crab-eating macaque	estrogen receptor isoform X1	Y	Y	Homo sapiens	Y
Rhesus monkey	estrogen receptor isoform X2	Y	Y	Homo sapiens	Y
Sooty mangabey	PREDICTED: estrogen receptor isoform X2	Y	Y	Homo sapiens	Y

United States
Environmental Protection
Agency

SeqAPASS v8.1 (Released Earth Day 2025)

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

Acknowledgements

U.S. EPA, ORD

Ryan Staub (ORISE)
Marissa Brickley (EPA/University of Minnesota Duluth)
Dylan Buglewicz (ORAU)
David Ryoo (URAU)
Maxwell Botz (past ORISE)
Monique Hazemi (past ORISE)
Donovan Blatz (past ORISE)
Colin Finnegan (past ORISE)
Sally Mayasich (past University of Wisconsin)

Choose Search Type: By Species

GDIT

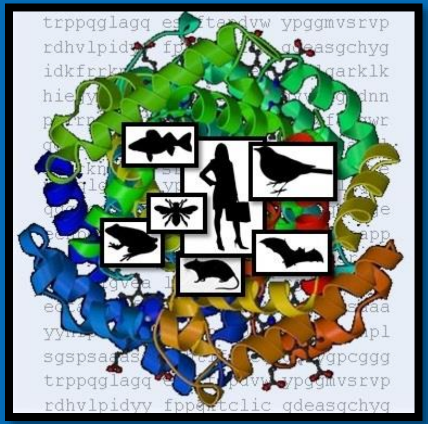
Cody Simmons
Audrey Wilkinson
Wilson Menendez

Schumann.Peter@epa.gov

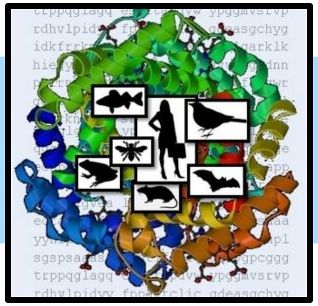
LaLone.Carlie@epa.gov

United States
Environmental Protection
Agency

SeqAPASS Concept Review: Levels 1 - 3

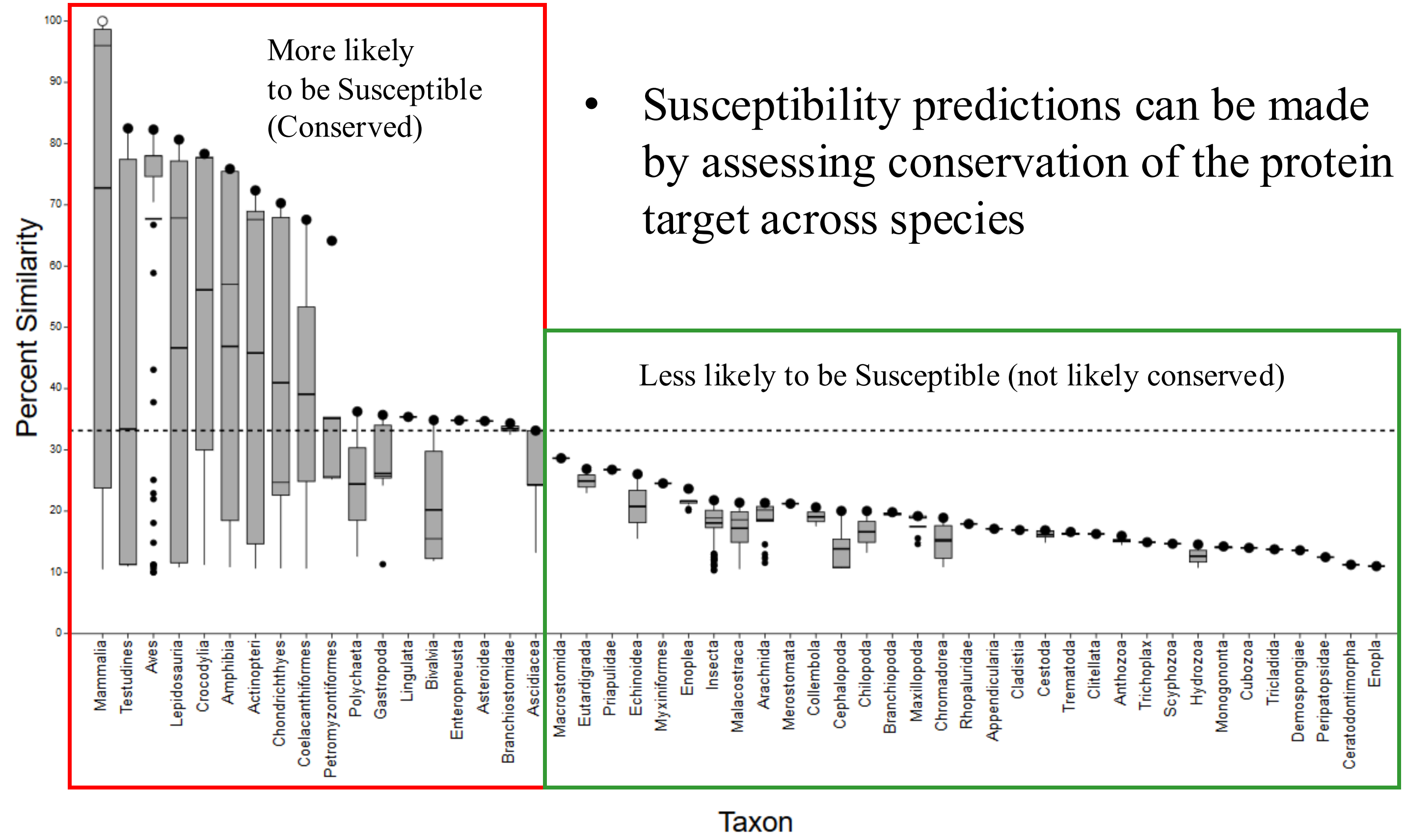


SeqAPASS Level 1 Evaluation: Primary Amino Acid Sequence Alignments



Proteins are chains of amino acids. The sequence of these amino acids can differ between species for each protein.

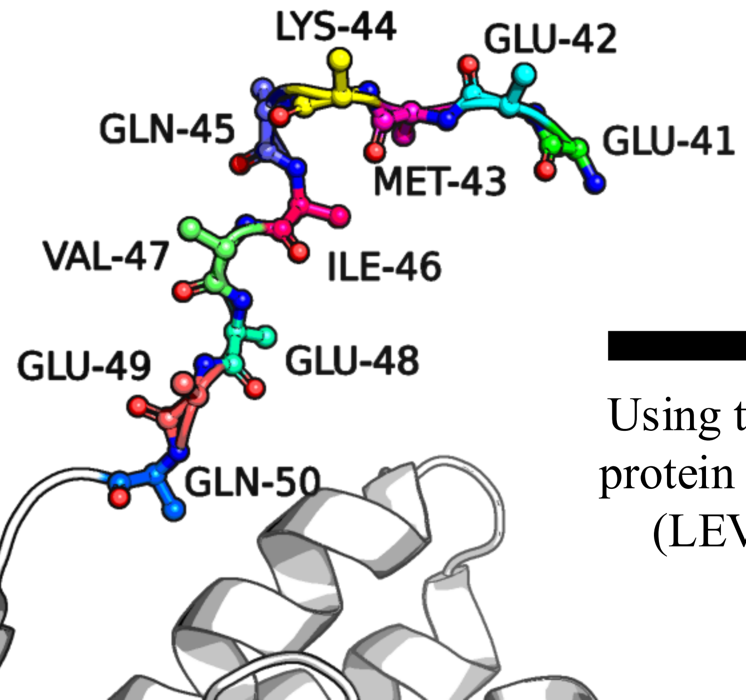
- Based on sequence alignments to a known sensitive species



More likely to be Susceptible (Conserved)

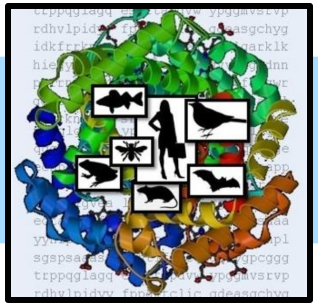
Less likely to be Susceptible (not likely conserved)

- Susceptibility predictions can be made by assessing conservation of the protein target across species

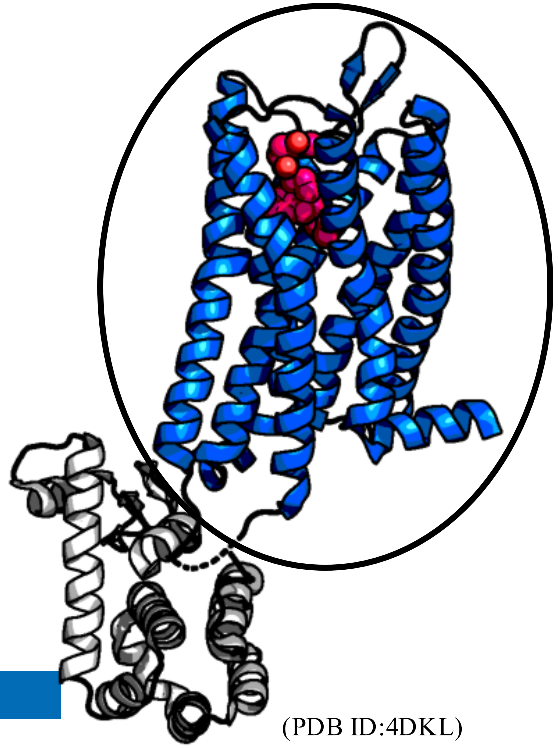


Using the *entire* protein sequence (LEVEL 1)

SeqAPASS Level 2 Evaluation: Conserved Functional Domain Sequence Alignments

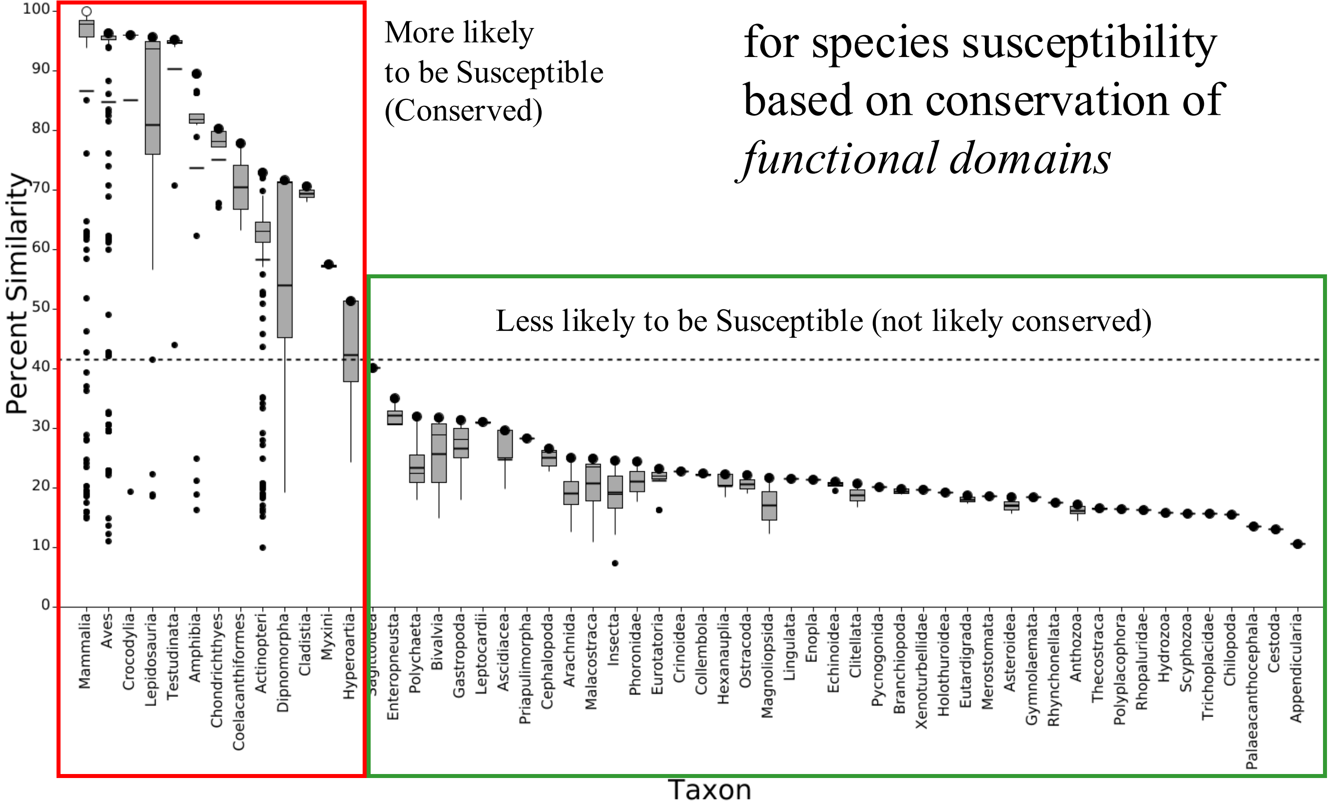


By focusing sequence comparisons to the conserved domains, estimates of functional conservation can be made.



(PDB ID:4DKL)

Using the *functional domain* sequence (LEVEL 2)

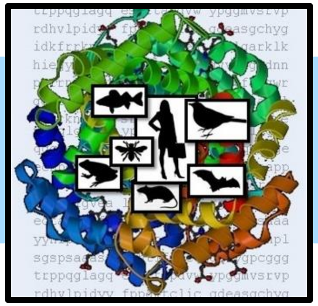


More likely to be Susceptible (Conserved)

Less likely to be Susceptible (not likely conserved)

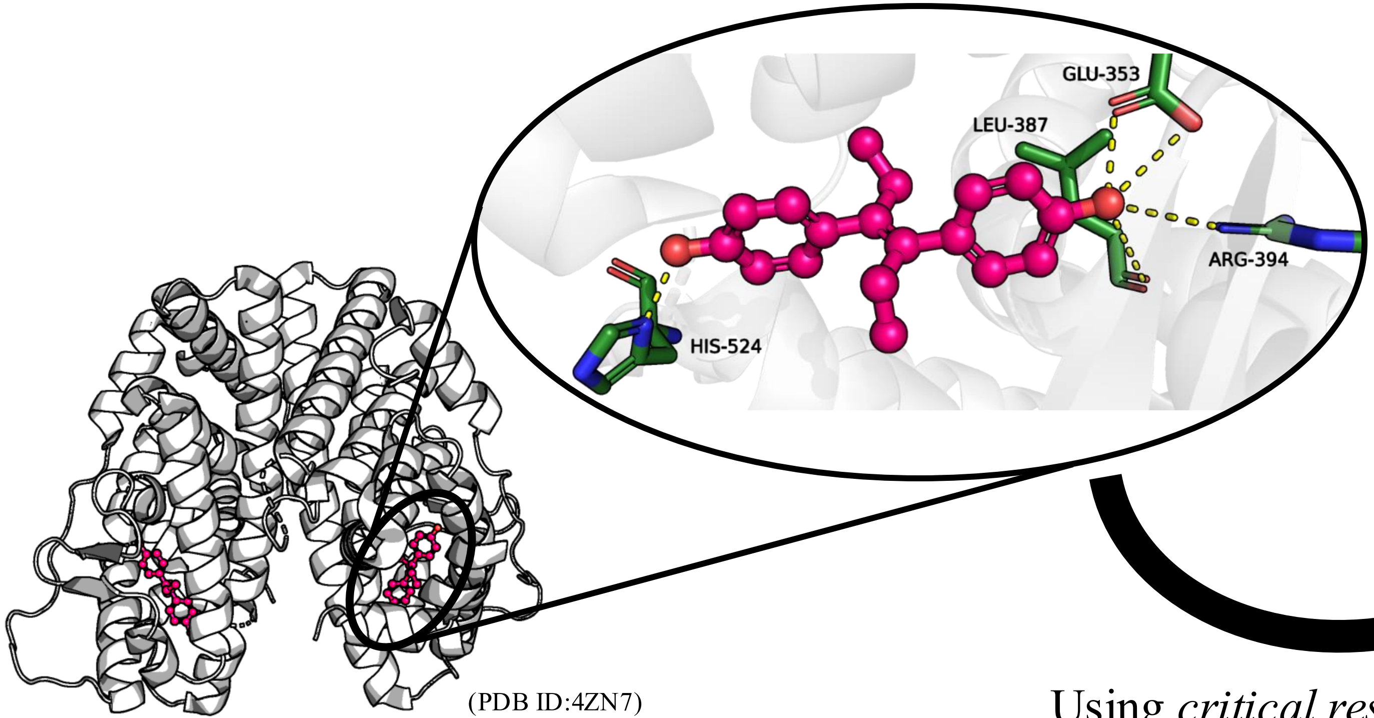
- Additional line of evidence for species susceptibility based on conservation of *functional domains*

SeqAPASS Level 3 Evaluation: Critical Amino Acid Residue Conservation

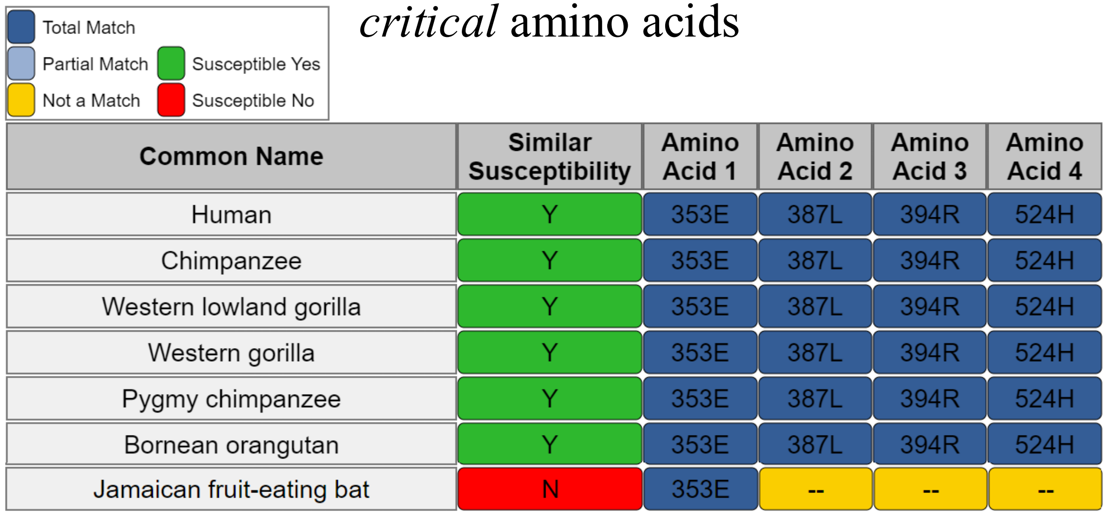


Protein-chemical interactions are often mediated by a few, critical amino acid residues.

- Additional line of evidence for species susceptibility based on conservation of *critical* amino acids



(PDB ID:4ZN7)



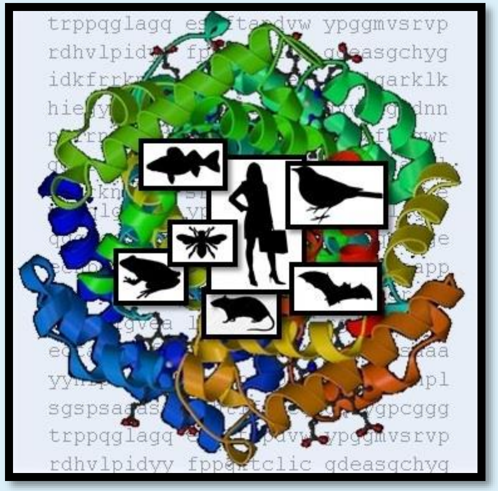
🟦 Total Match
🟦 Partial Match
🟨 Not a Match
🟩 Susceptible Yes
🟥 Susceptible No

Common Name	Similar Susceptibility	Amino Acid 1	Amino Acid 2	Amino Acid 3	Amino Acid 4
Human	Y	353E	387L	394R	524H
Chimpanzee	Y	353E	387L	394R	524H
Western lowland gorilla	Y	353E	387L	394R	524H
Western gorilla	Y	353E	387L	394R	524H
Pygmy chimpanzee	Y	353E	387L	394R	524H
Bornean orangutan	Y	353E	387L	394R	524H
Jamaican fruit-eating bat	N	353E	--	--	--

Using *critical residues* (LEVEL 3)

Determine critical amino acid residues from:
- Structural data
- Mutagenesis studies
- Field resistance studies
- Molecular dynamics
- Etc.

Q&A



We’ll see you in the next session!

Begins at 11:45 a.m. CT / 12:45 p.m. ET

You may stay on the call until Session 2 begins if you wish!

United States
Environmental Protection
Agency

SeqAPASS Demo