

# Bioinformatics for Cross Species Extrapolation

Carlie A. LaLone, Ph.D.  
Research Bioinformaticist





Chemicals make up the world around us – necessary for our modern society







# Toxicity Testing to Understand Chemical Safety

- **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 21<sup>st</sup> 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**



**easy maintenance and good breeding capabilities**



**short lifespans and rapid life cycles**



# Surrogate



**ability to control diet and surroundings**

**requires least space and time-consuming care**





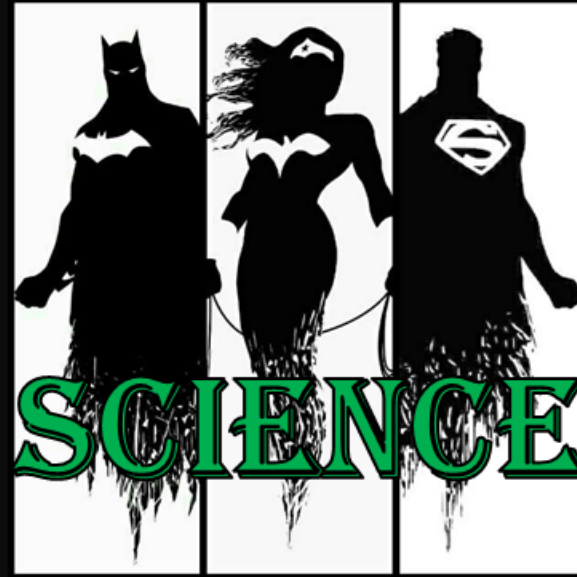
# Species Extrapolation

## What is it?

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

## Why is it important:

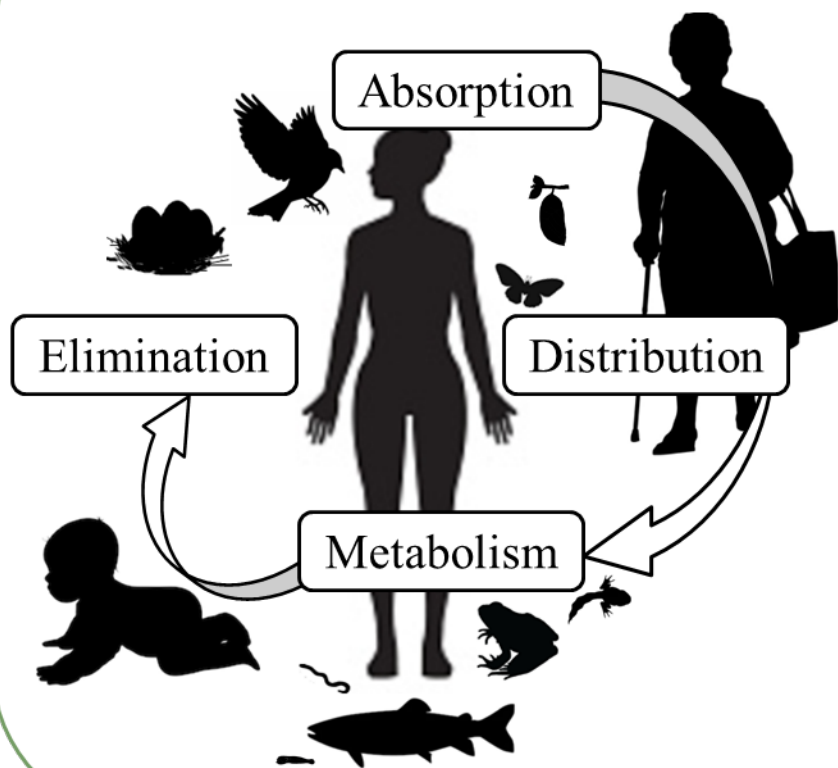
- **Limited or no toxicological data** for the animal or plant species of interest – reliance on surrogate (model organisms)
  - **Impractical to generate new data** for all species
- Testing **resources are limited**
  - International interest to **reduce animal use**
  - Ever-increasing demand to **evaluate more chemicals in a timely** and sometimes expedited manner





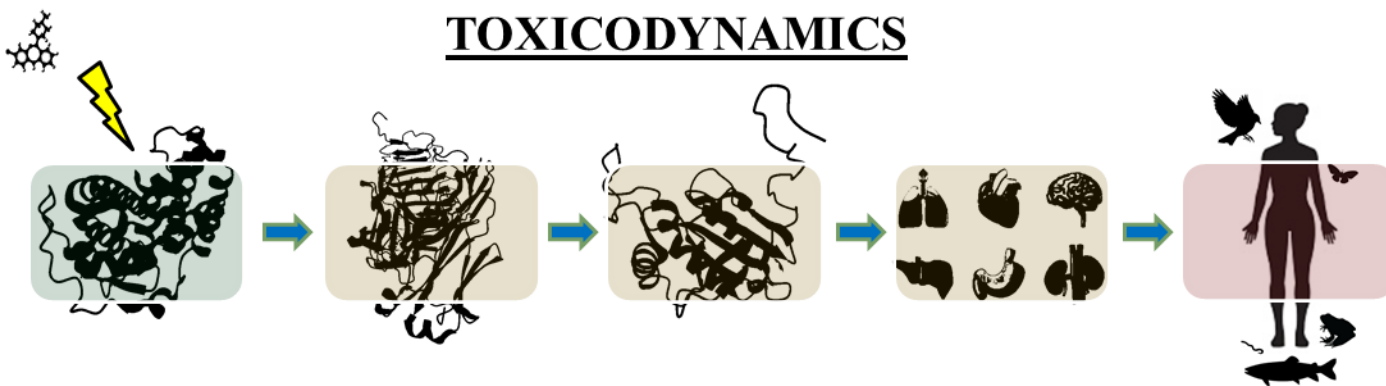
# Sensitivity to Chemical Perturbation

## TOXICOKINETICS



## Cross Species Extrapolation

### TOXICODYNAMICS





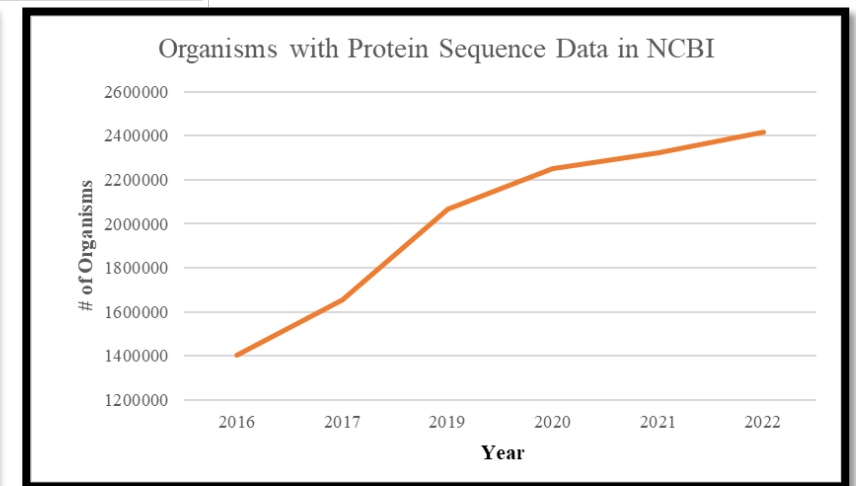
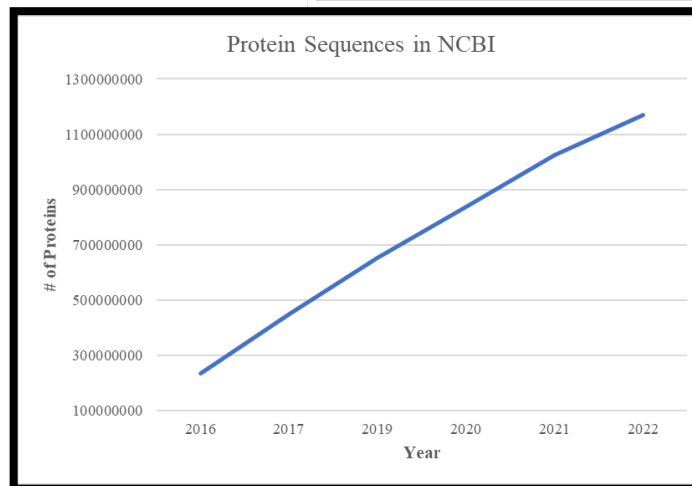
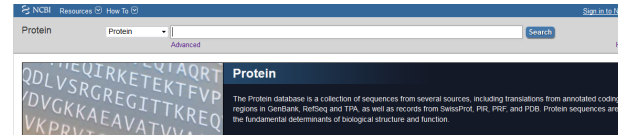
# Where could we begin in understanding species similarities and differences?

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

**Sequence and structural data: New tools and technologies have emerged**

- Improved sequencing technologies
- Large databases of sequence data

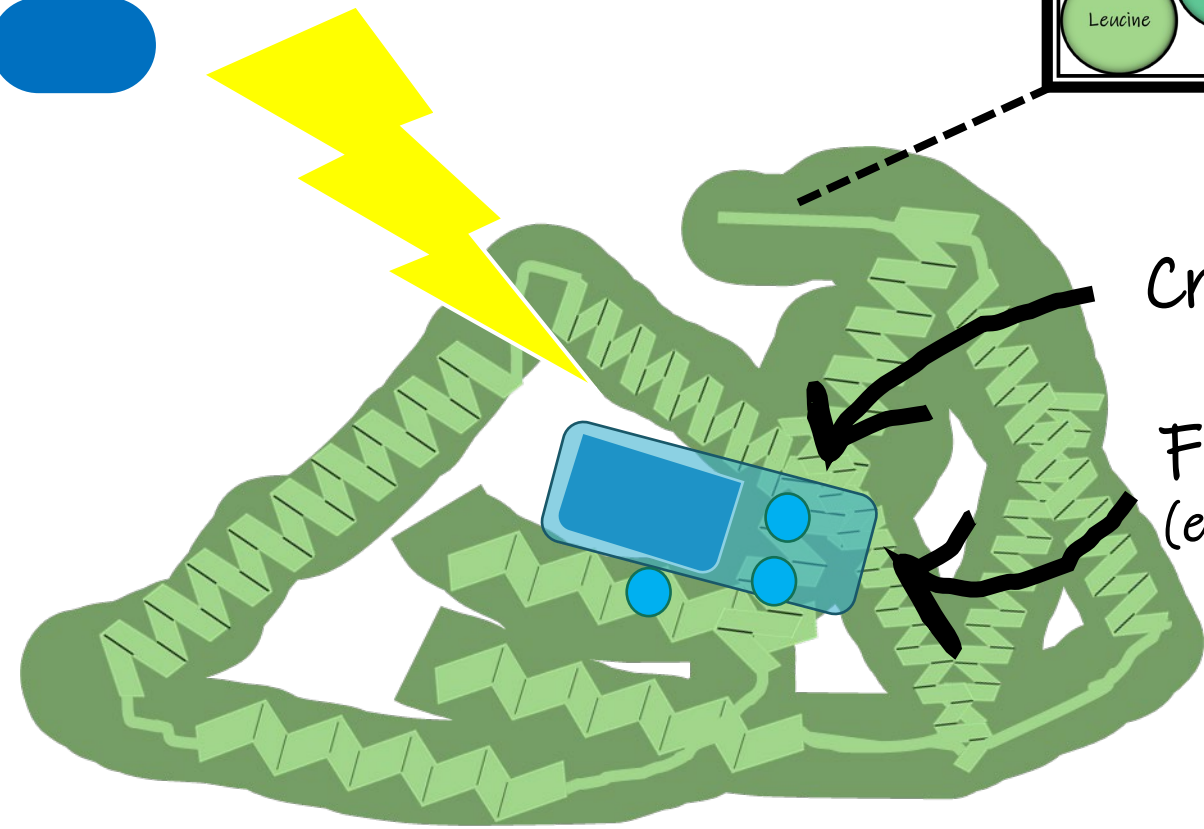
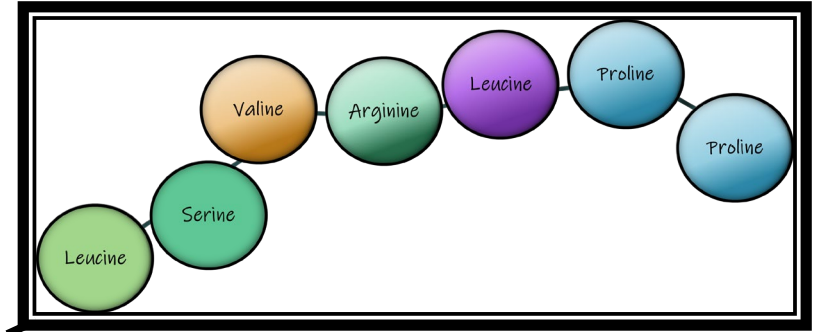
**NCBI: 1,170,638,058 Proteins representing 2,416,649 Organisms**







Chemical-Protein Interaction:



Critical amino acids

Functional domain  
(e.g. Ligand binding domain)



# 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

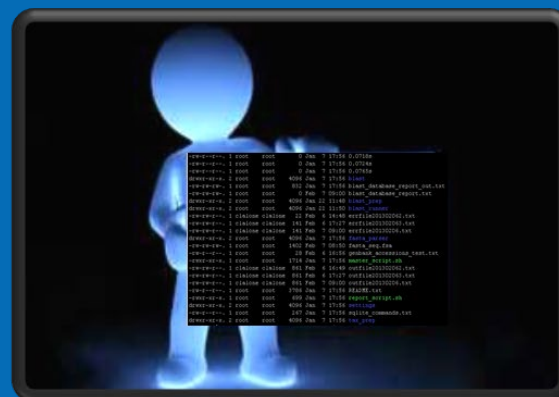
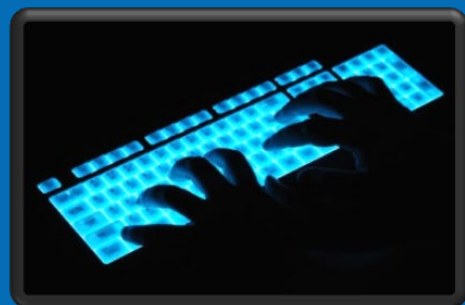
**Let the computers do the work**





<https://seqapass.epa.gov/seqapass/>

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)



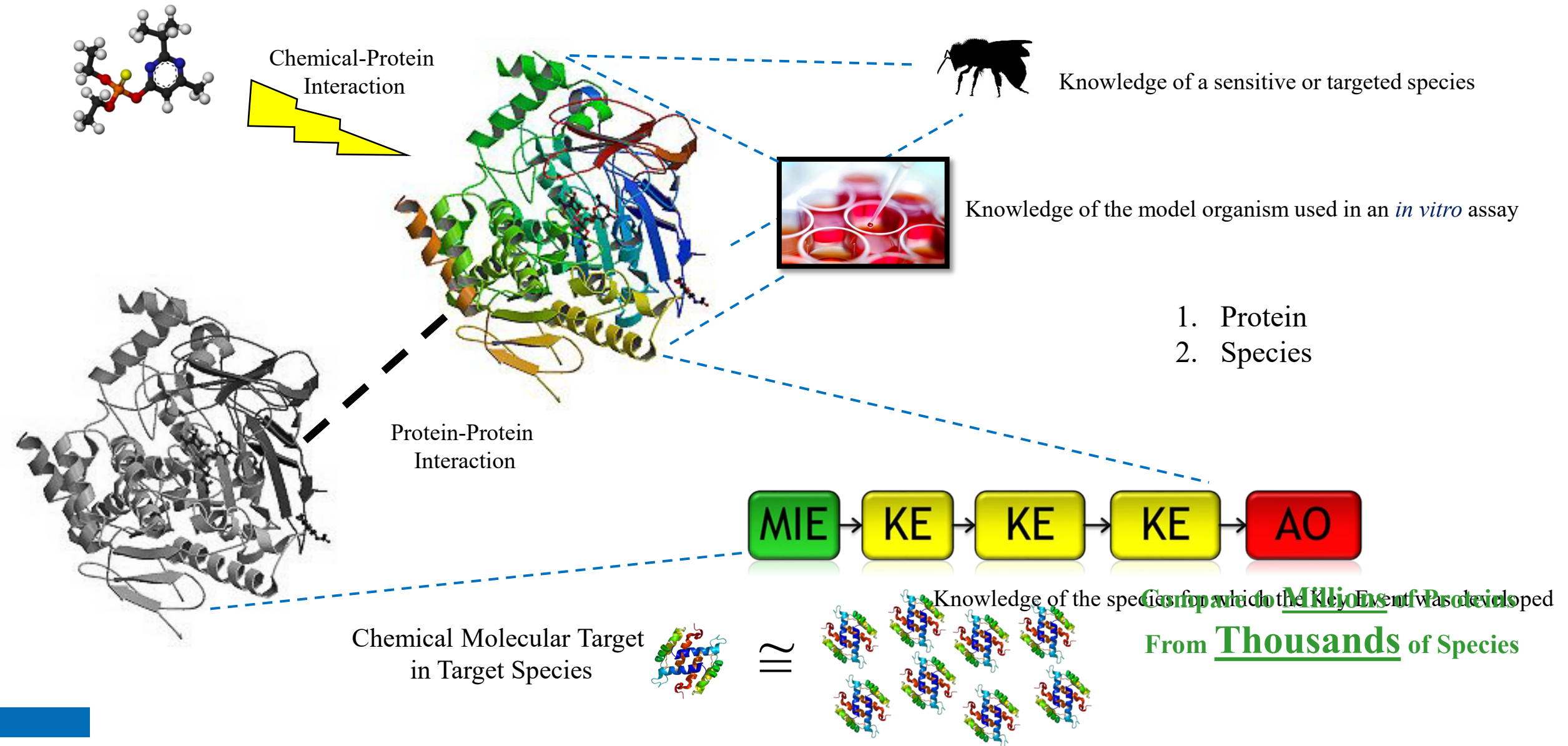
## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): A Web-Based Tool for Addressing the Challenges of Cross-Species Extrapolation of Chemical Toxicity

Charlie A. LaLone,<sup>\*,1</sup> Daniel L. Villeneuve,<sup>\*</sup> David Lyons,<sup>†</sup> Henry W. Helgen,<sup>‡</sup>  
Serina L. Robinson,<sup>§,2</sup> Joseph A. Swintek,<sup>¶</sup> Travis W. Saari,<sup>\*</sup> and  
Gerald T. Ankley<sup>\*</sup>



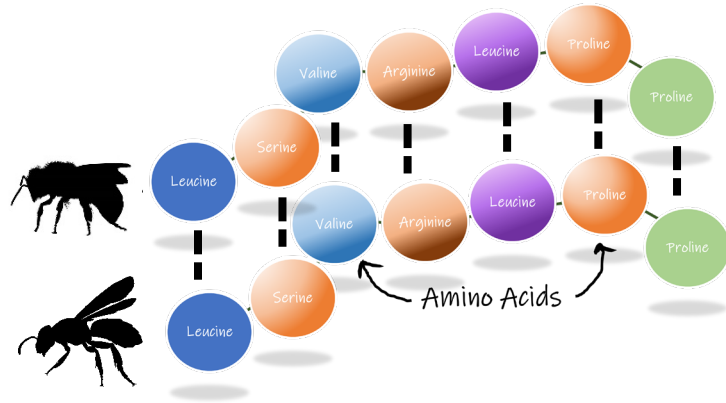


# What information is required for a SeqAPASS query?



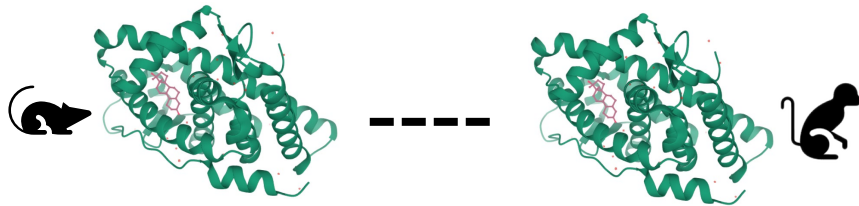
## Flexible Analysis Based On Available Data

- Level 1** Primary Amino Acid Sequence Alignments
- Level 2** Conserved Functional Domain Alignments
- Level 3** Critical (Close Contact) Amino Acid Conservation

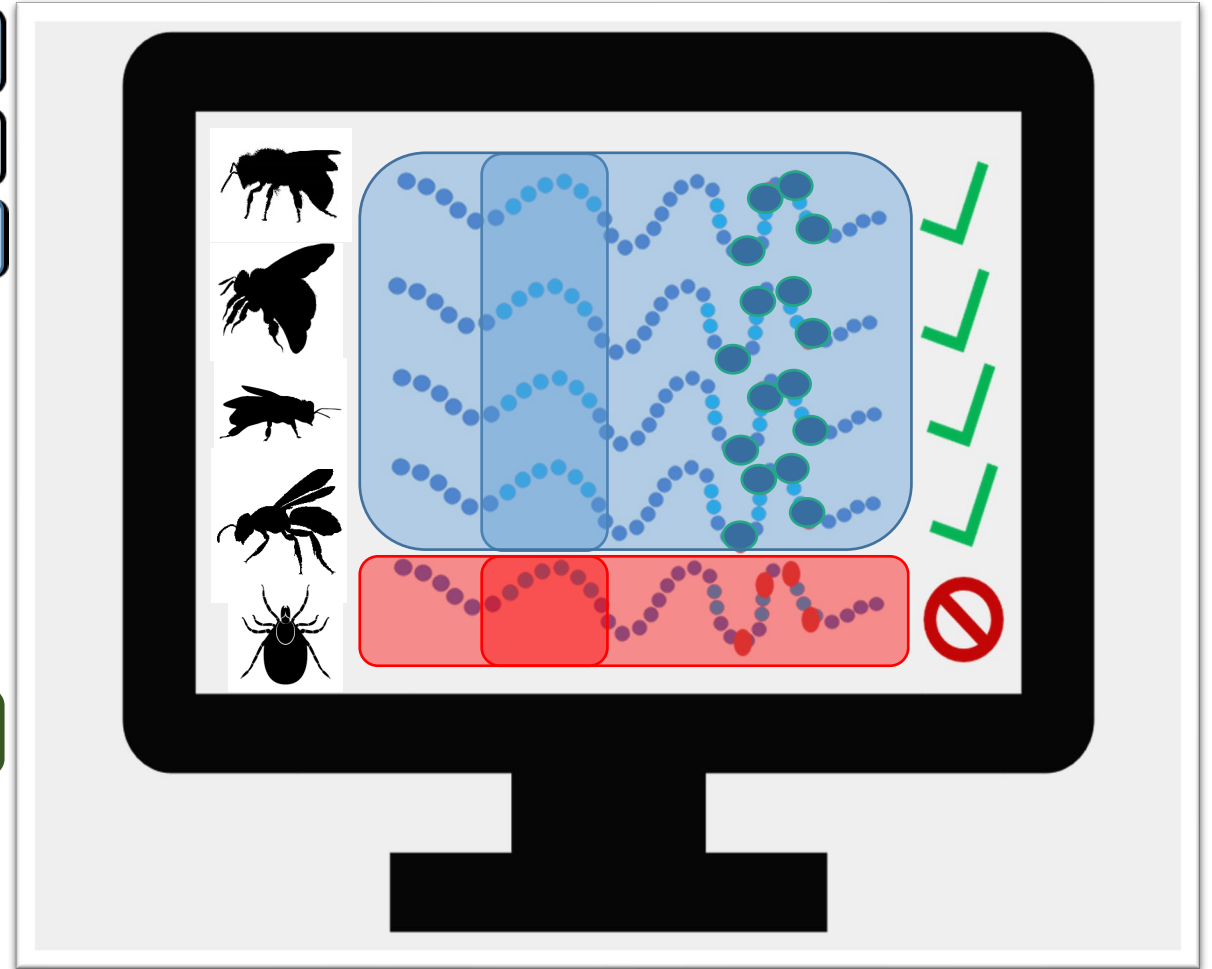
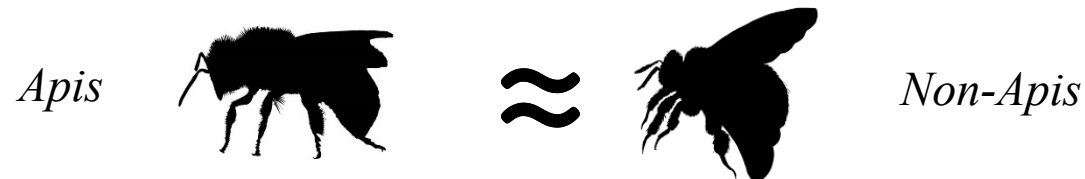


SeqAPASS v7.0  
(Coming 2023)

- Level 4 (expert users only)** Structure Alignments



# Gather Lines of Evidence Toward Protein Conservation



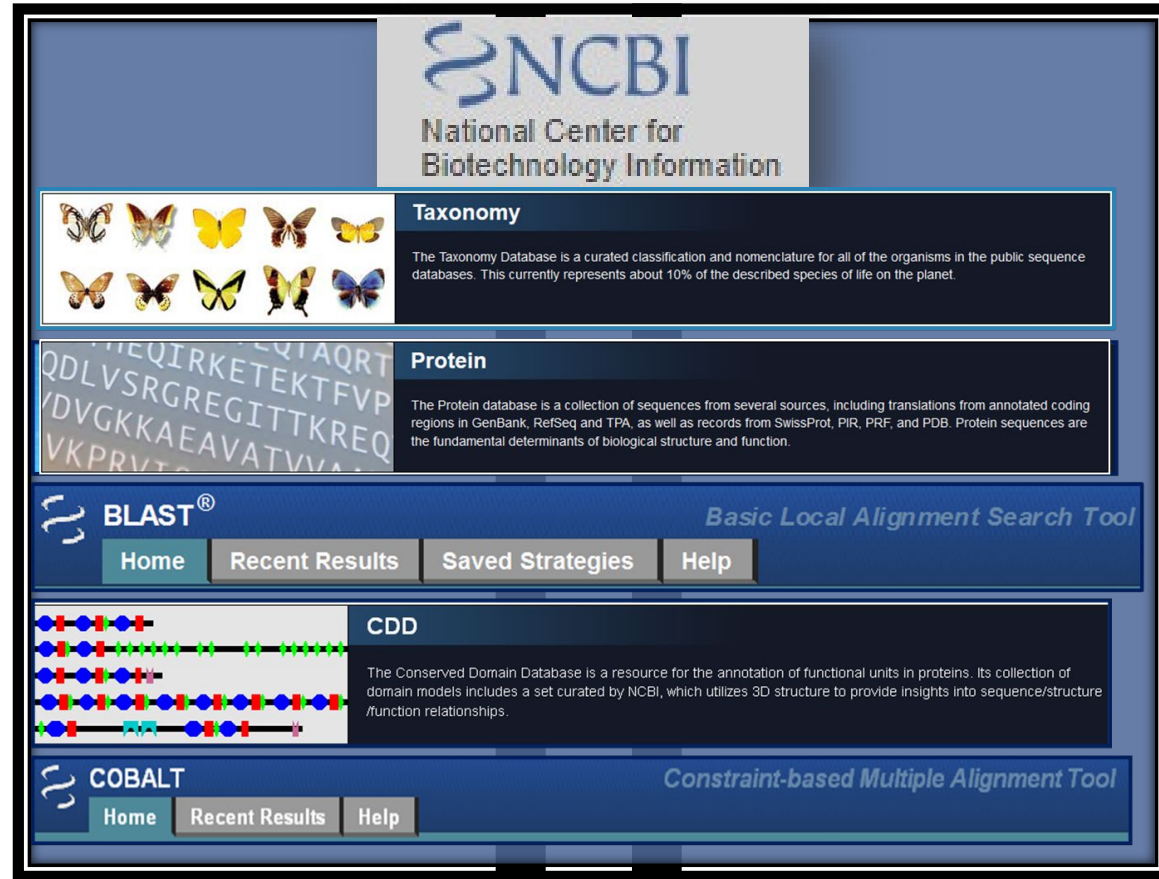


## SeqAPASS



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

# SeqAPASS

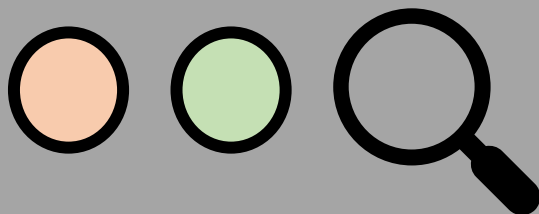
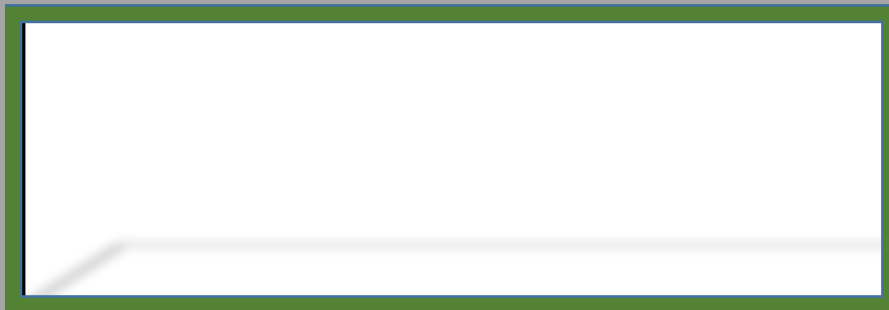


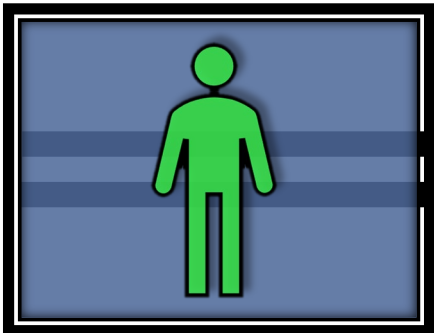
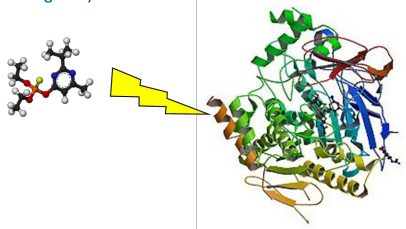
# SeqAPASS



# SeqAPASS

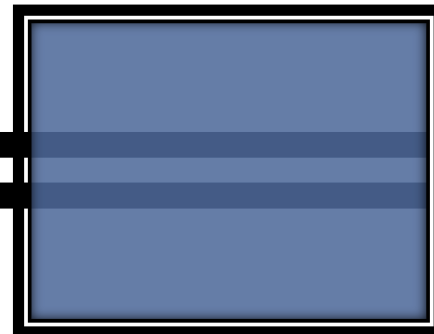
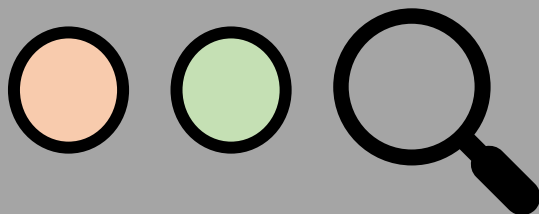
## Level 1



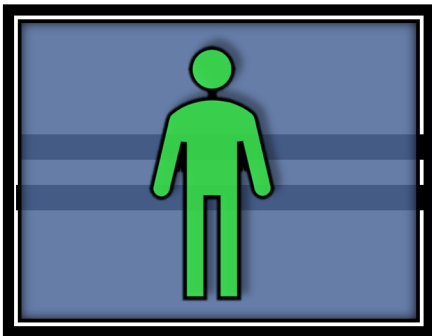
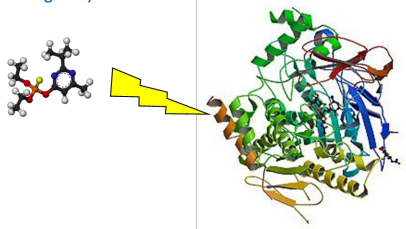


# SeqAPASS

## Level 1

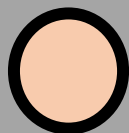
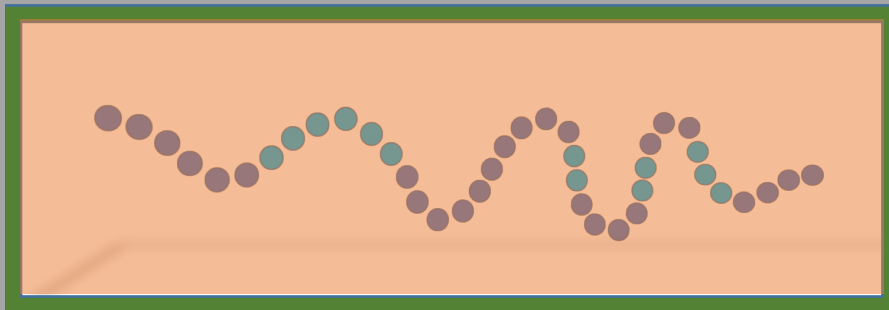






# SeqAPASS

## Level 1



Yes

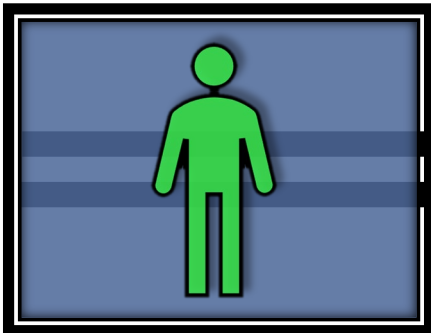
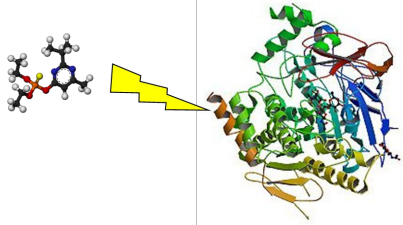
Line of Evidence:

Primary amino acid sequence

Conserved

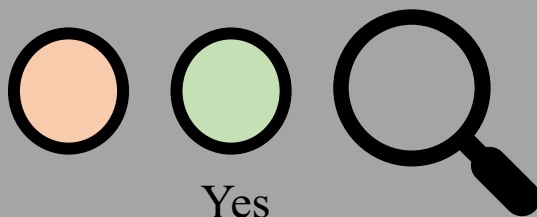
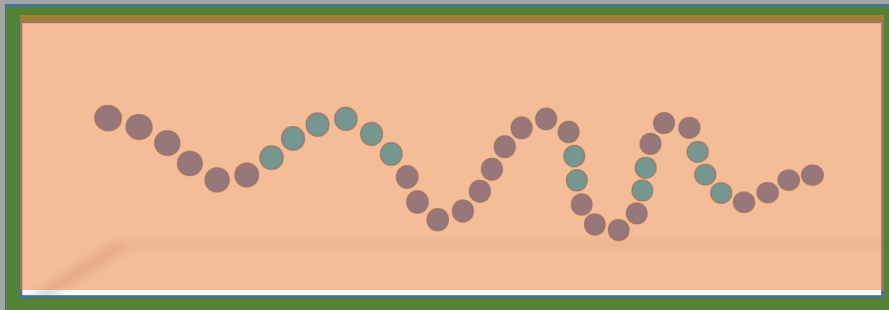


Percent similarity

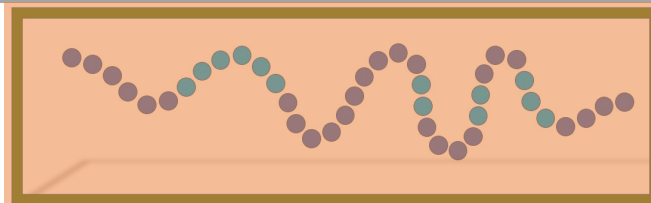
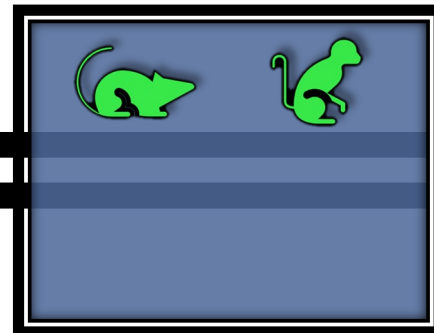


# SeqAPASS

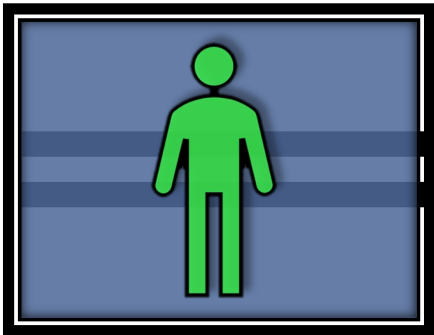
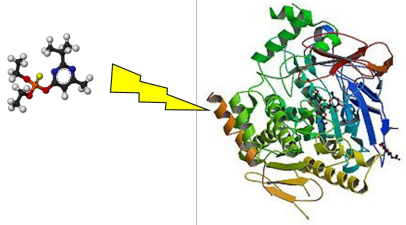
## Level 1



Line of Evidence:  
Primary amino acid sequence  
Conserved

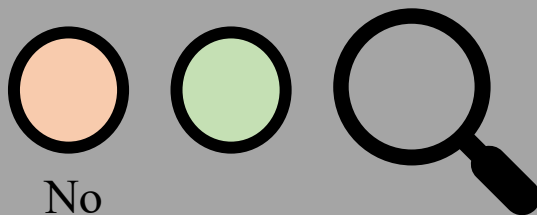
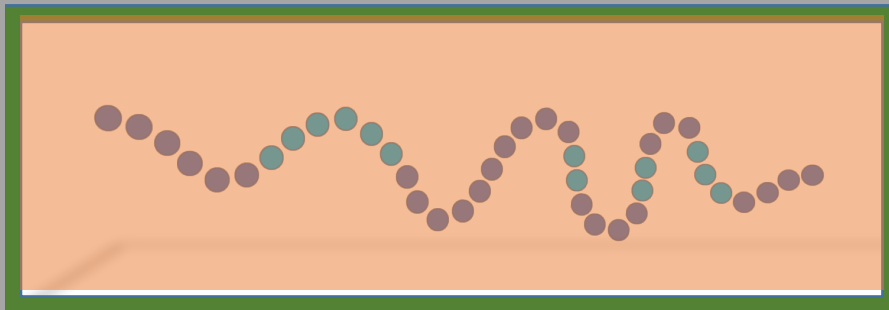




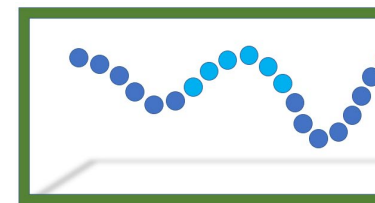
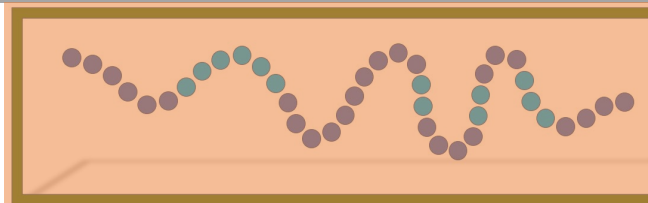
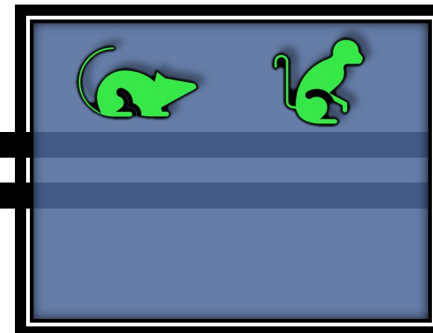


# SeqAPASS

## Level 1

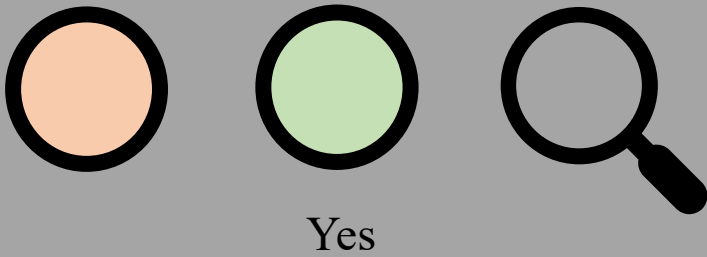


Line of Evidence:  
Primary amino acid sequence  
Conserved



# SeqAPASS Level 1

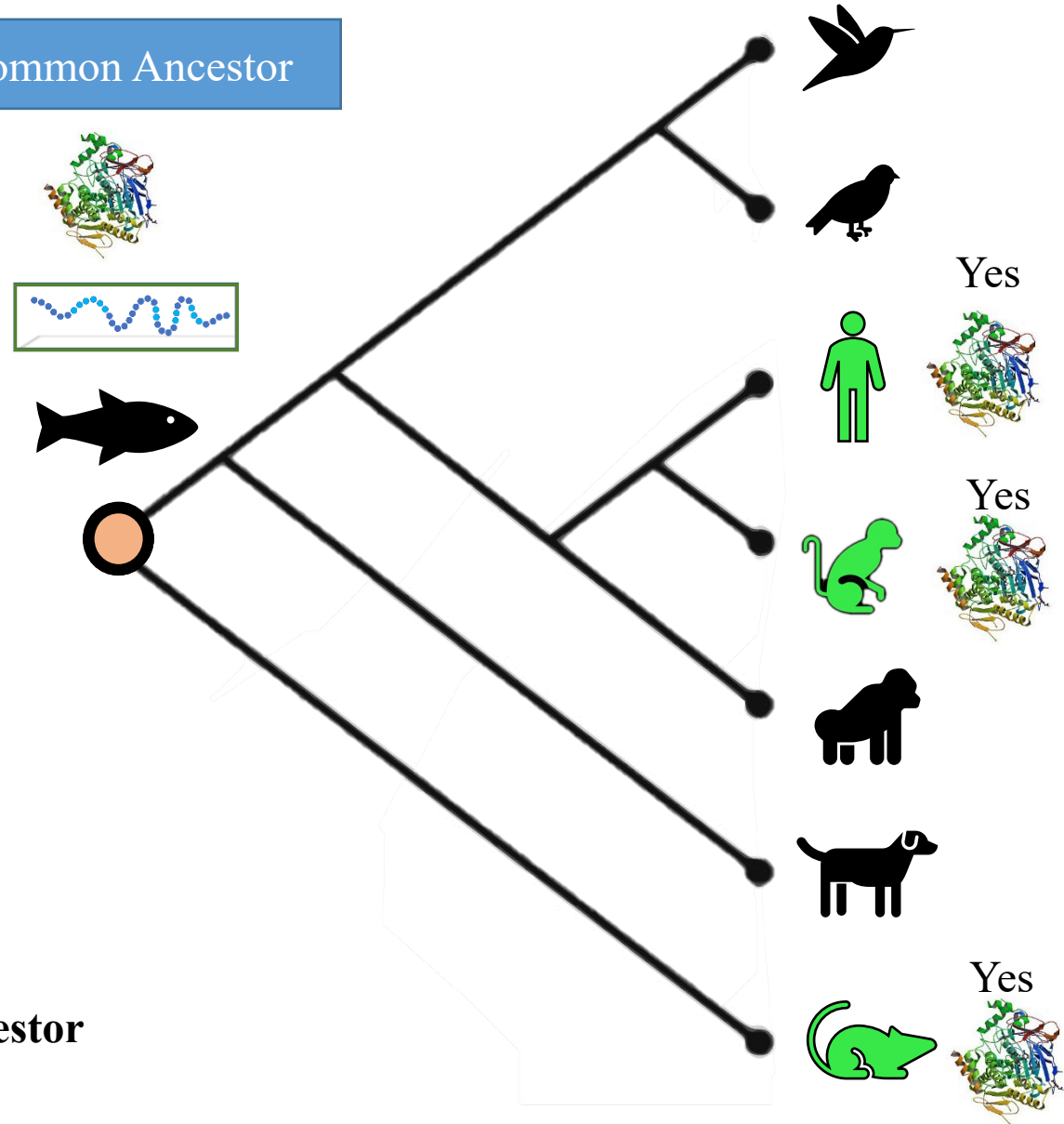
## Ortholog Candidate Identification



Proteins in different species that evolved from a common ancestor

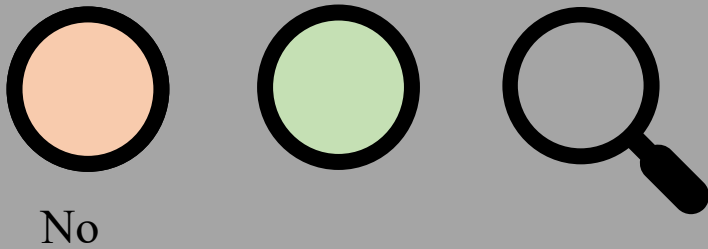
Typically maintain similar function

Common Ancestor

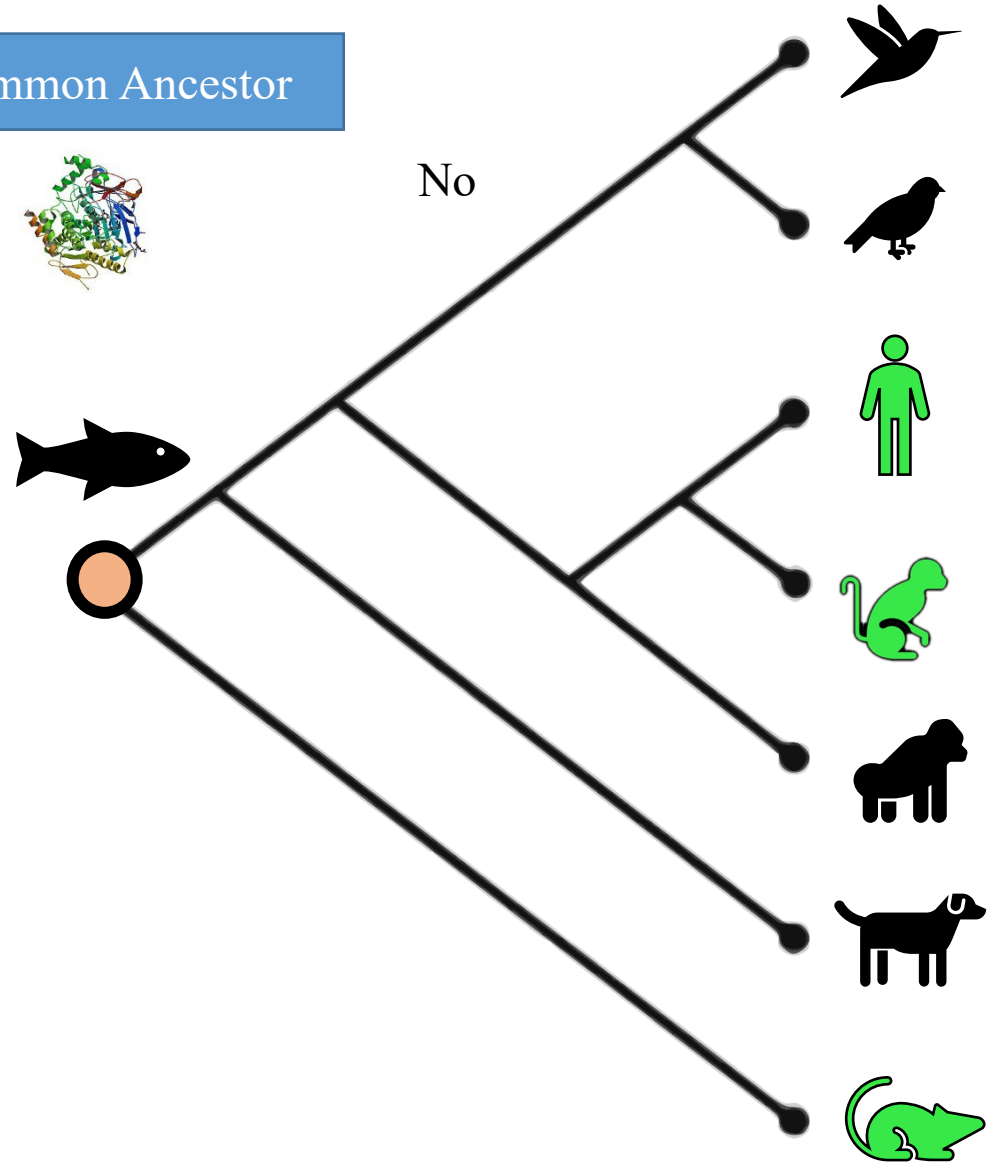


# SeqAPASS Level 1

## Ortholog Candidate Identification



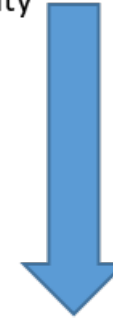
Common Ancestor





# 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

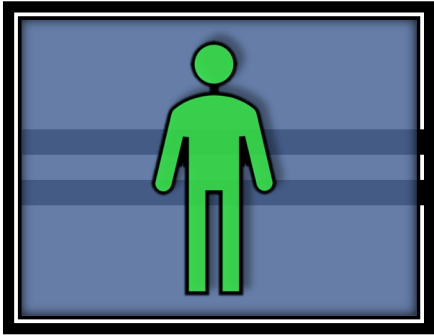
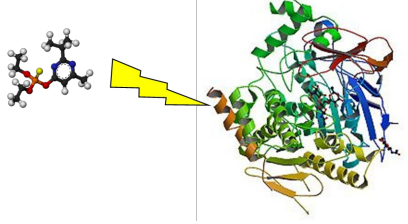


Lowest % Similarity that is still an ortholog

## Example:

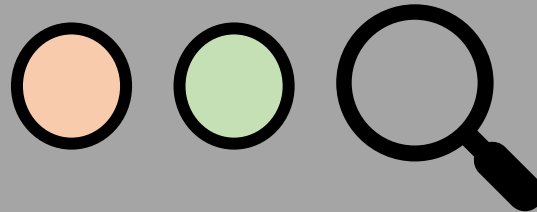
Susceptibility Cut-off: Set at 33.15

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



# SeqAPASS

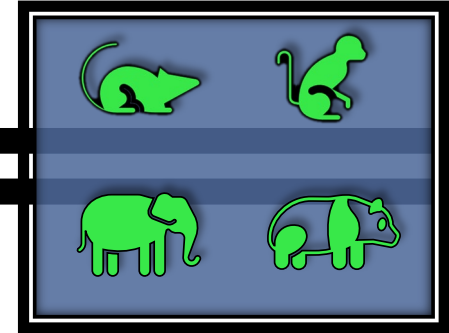
## Level 1



Line of Evidence:

Primary amino acid sequence

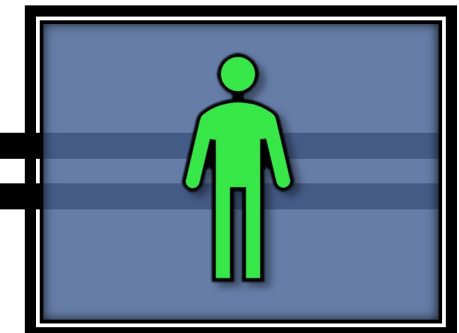
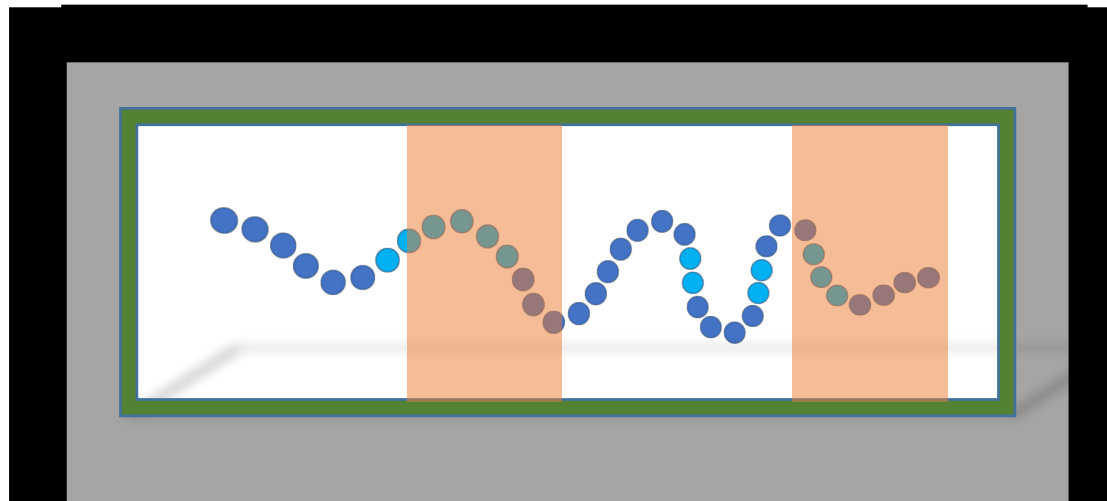
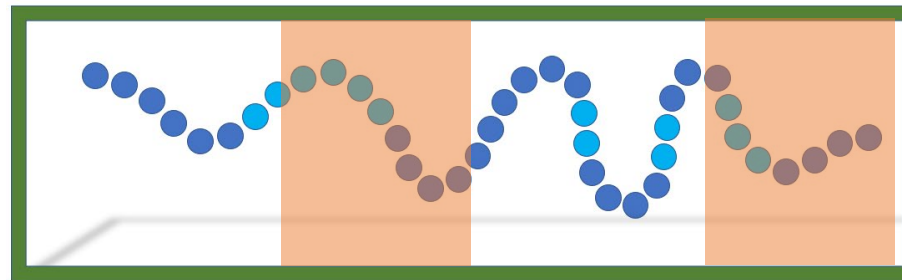
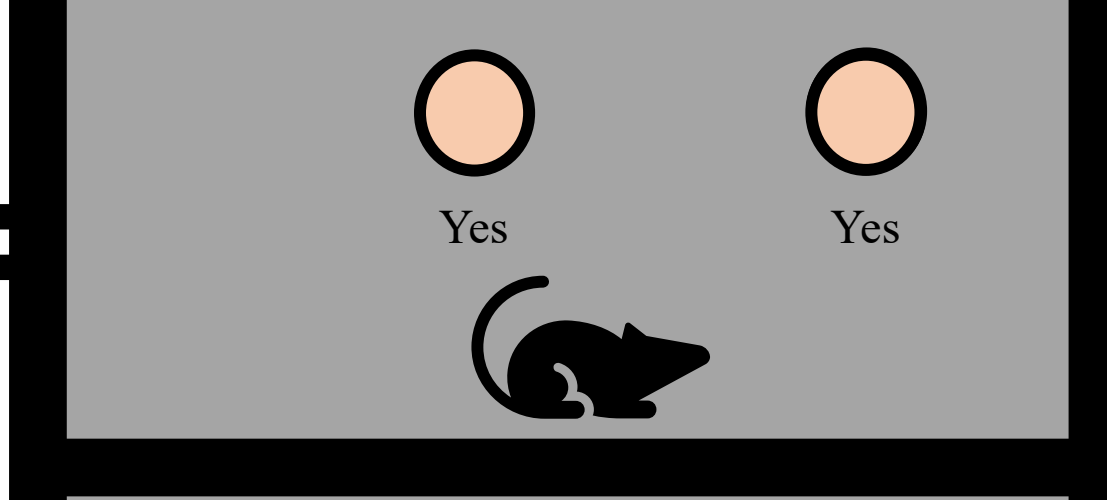
Conserved



Hundreds to Thousands of Species

# Level 2

Line of Evidence:  
Domain  
Conserved

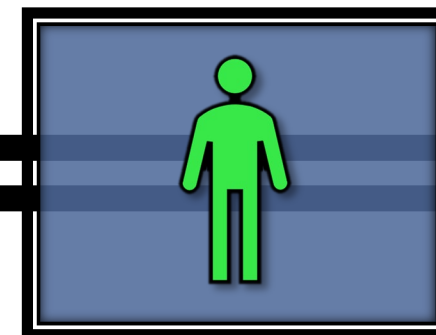
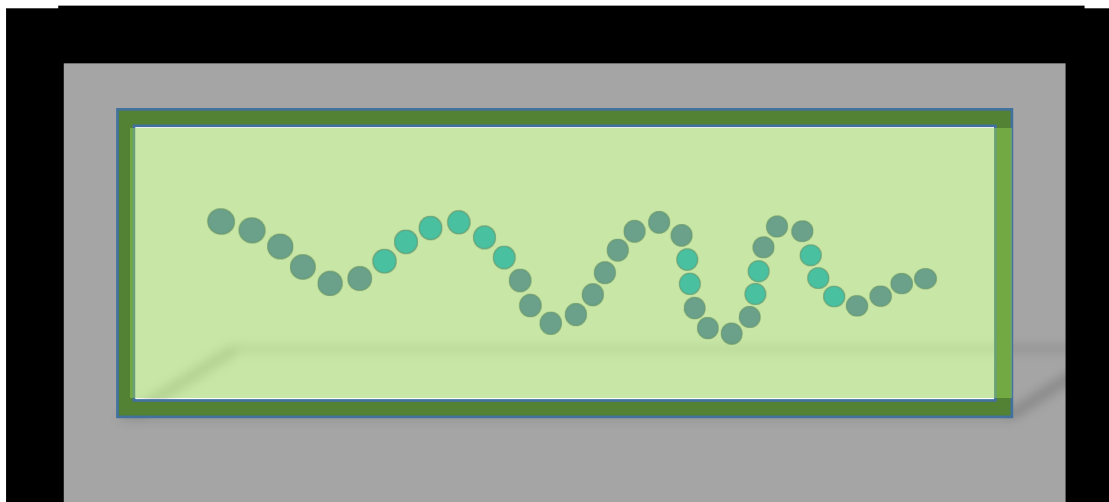
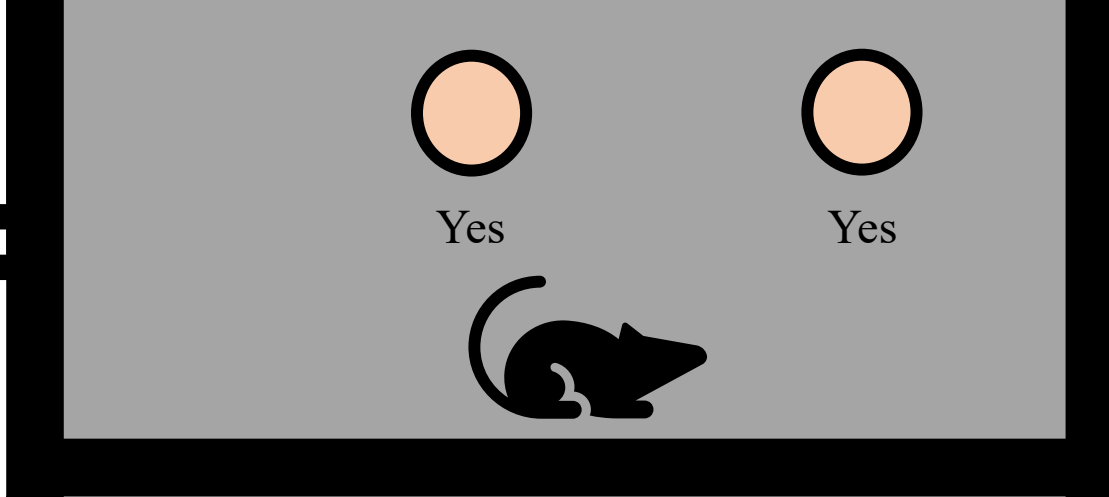


Human Functional Domain(s)





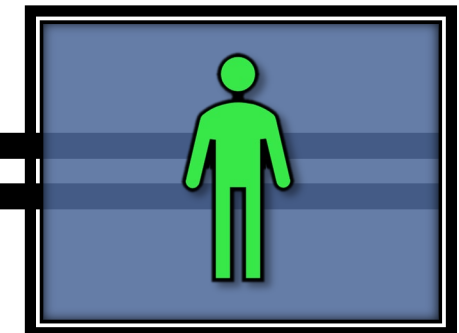
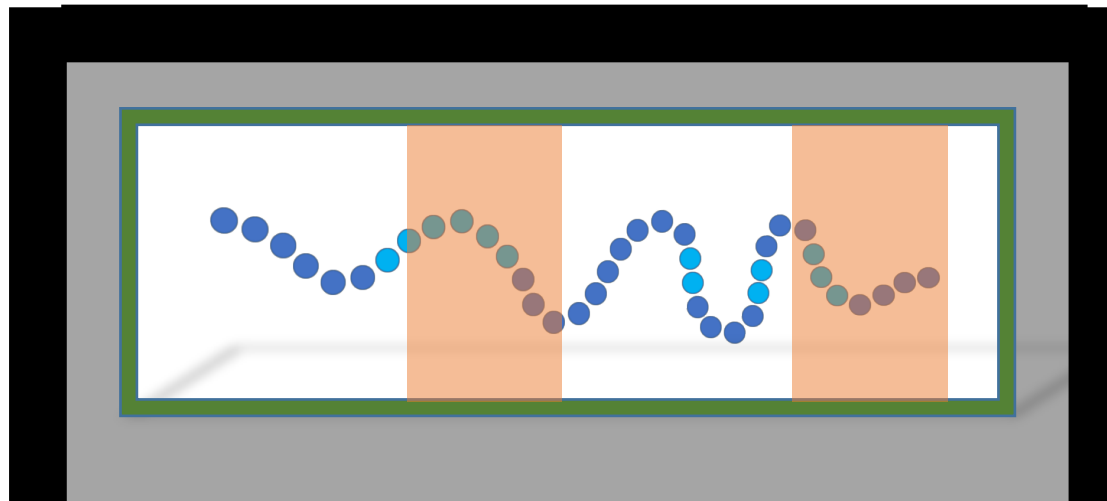
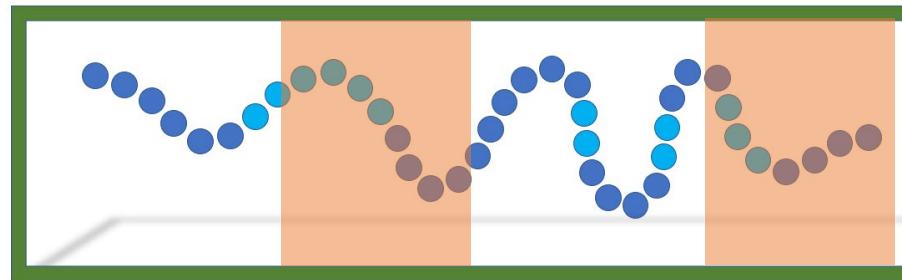
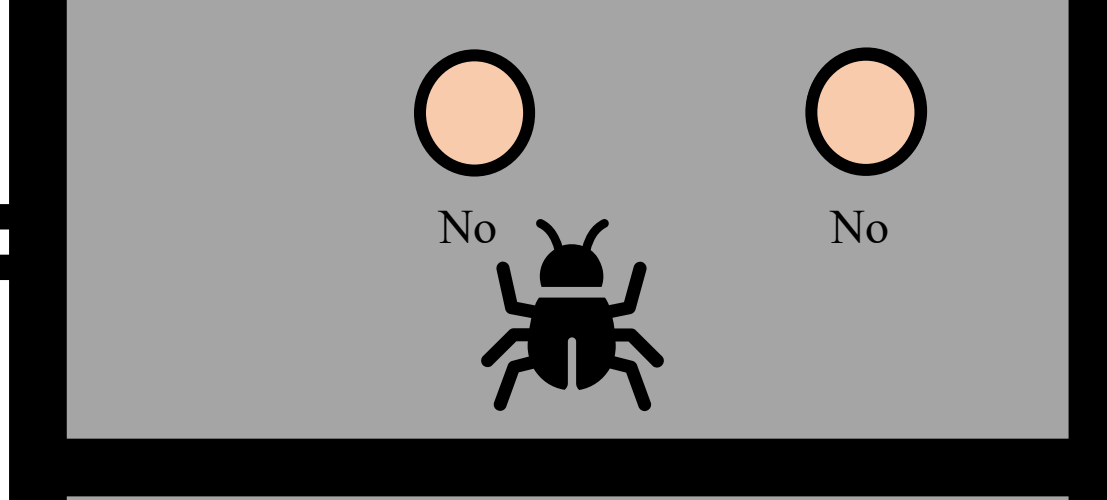
Line of Evidence:  
Domain  
Conserved



Human Functional Domain(s)

# Level 2

Line of Evidence:  
Domain  
Conserved



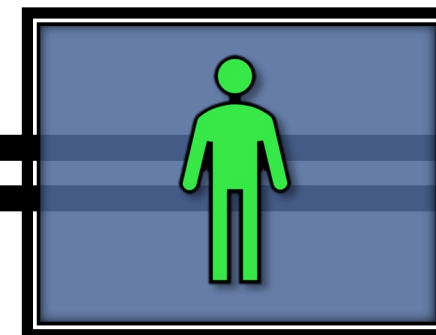
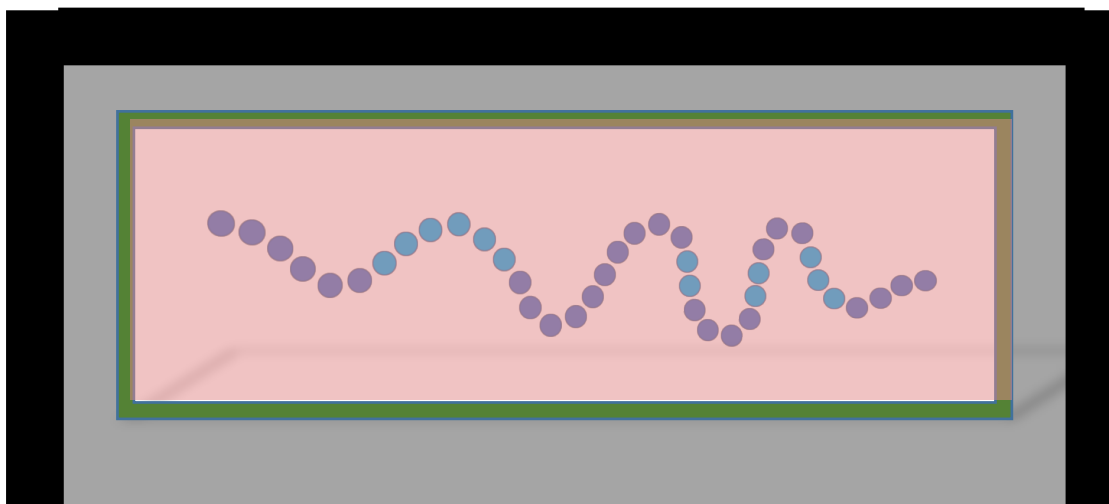
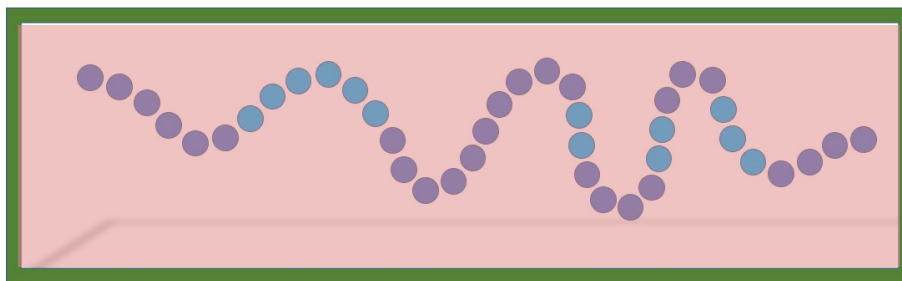
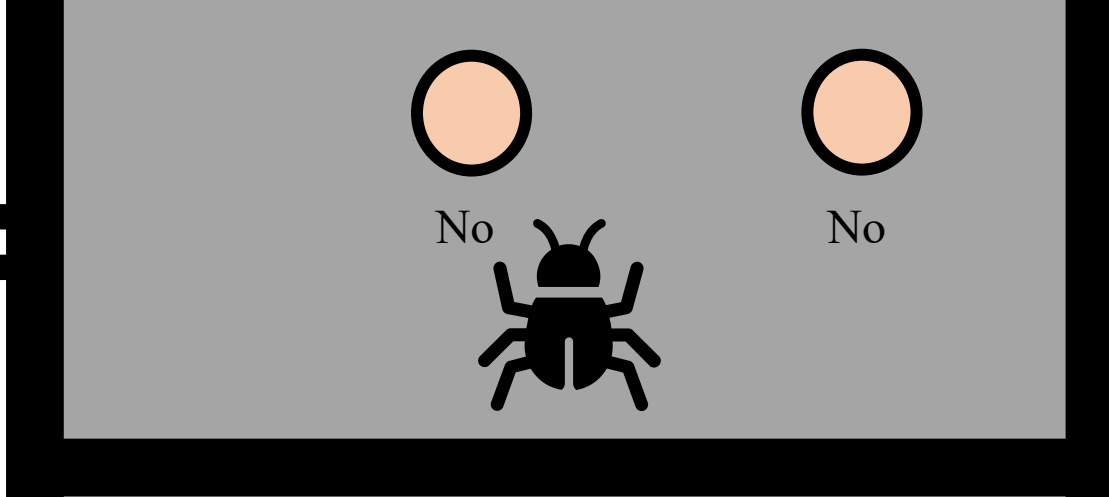
Human Functional Domain(s)

# Level 2

Line of Evidence:

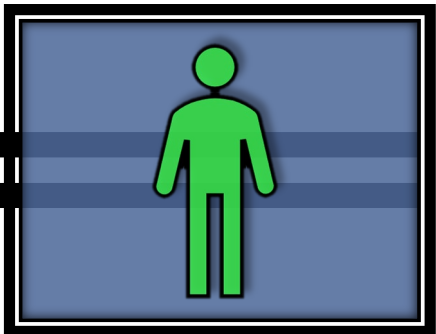
Domain

Not Conserved

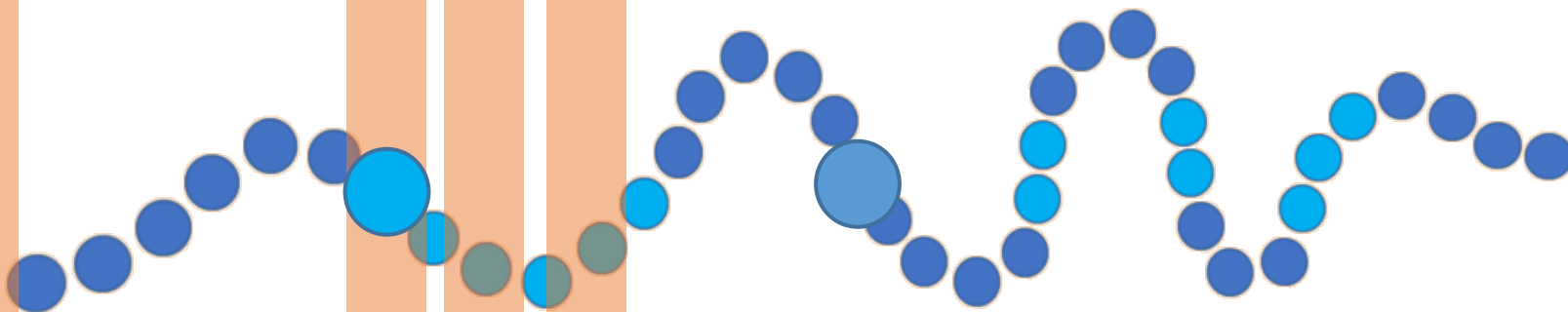
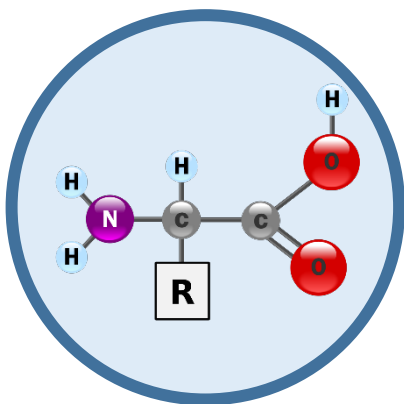
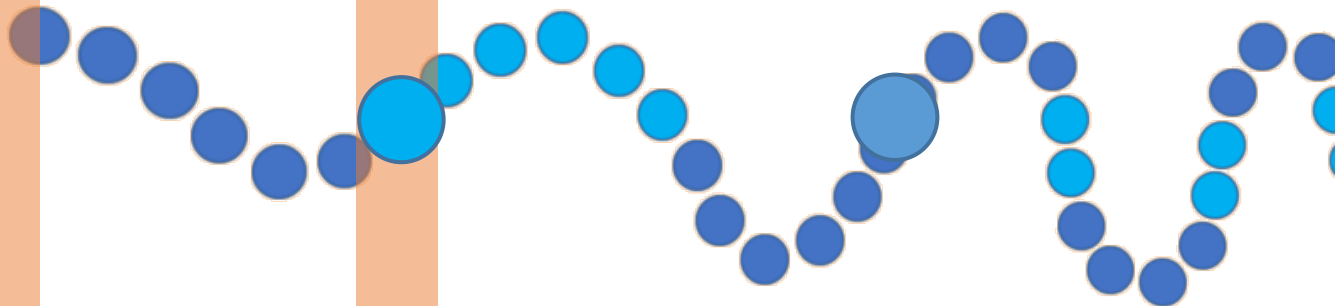
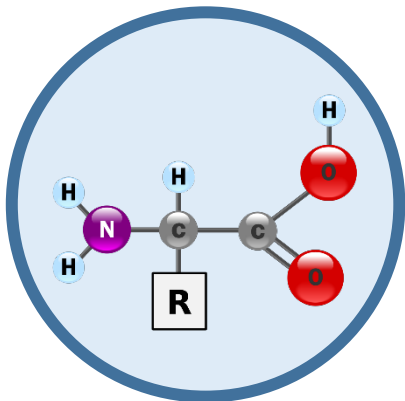


Human Functional Domain(s)

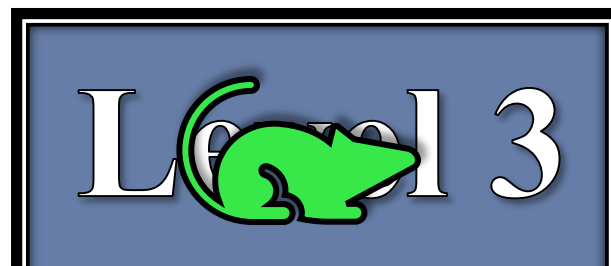


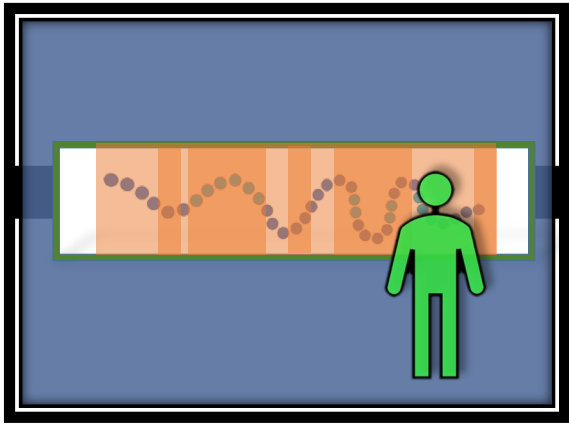


Human Critical Amino Acids



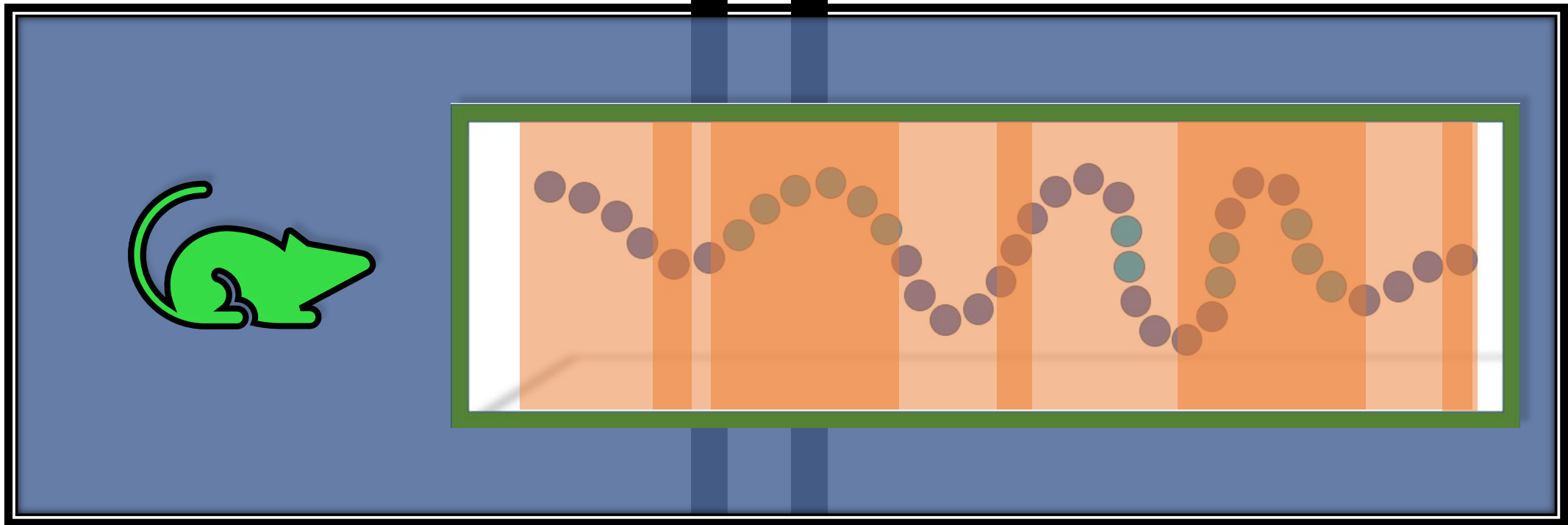
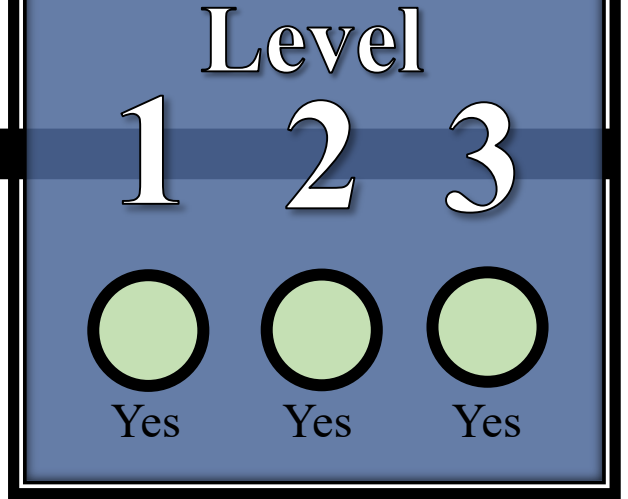
Line of Evidence: Conserved





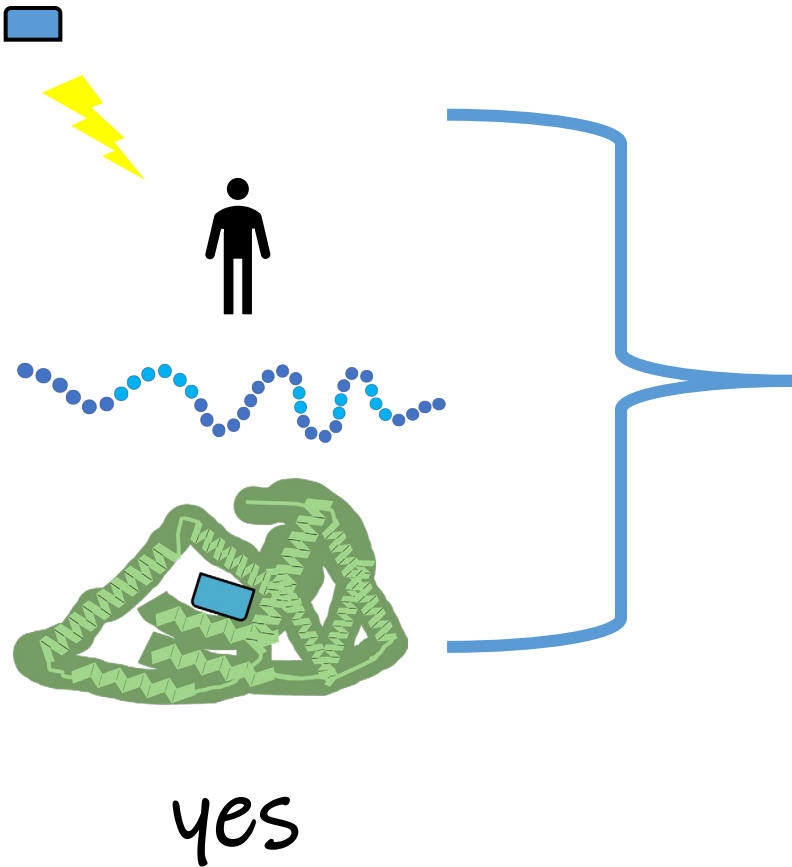
# SeqAPASS











## Summary



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

# SeqAPASS Predicts Likelihood of Similar Susceptibility based on Sequence Conservation:



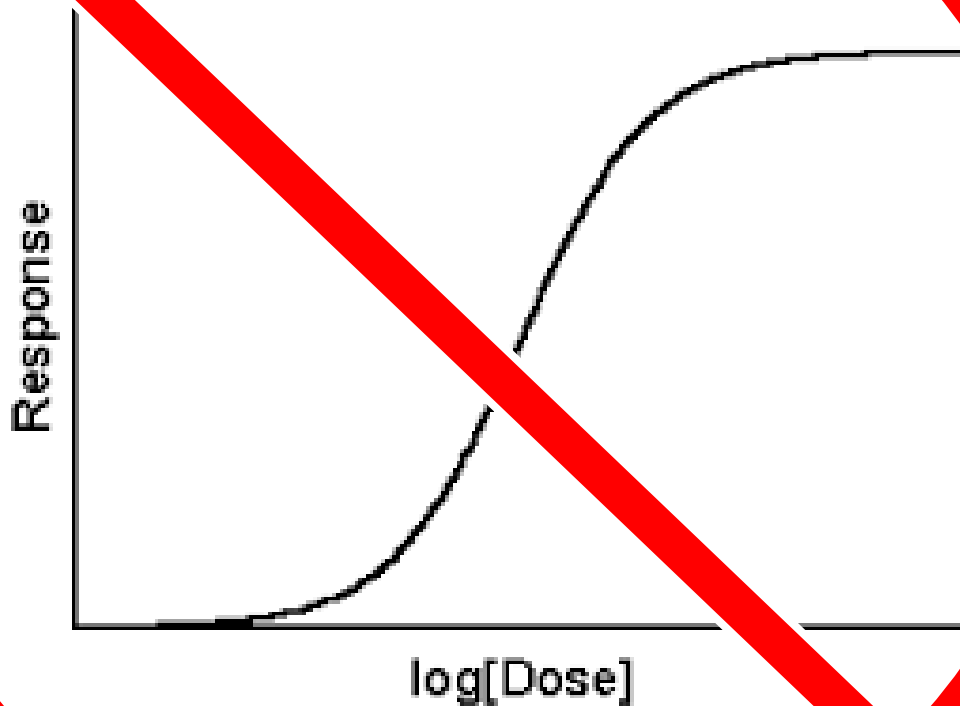
	yes
	yes
	yes
	yes
	yes
	yes
	yes
	no
	yes
	no

Line(s) of evidence indicate

- The protein is conserved
- The protein is NOT conserved



# SeqAPASS DOES NOT predict the degree of sensitivity/susceptibility:



## Factors that make a species sensitive

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





# Strengths of SeqAPASS

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





# Application of SeqAPASS

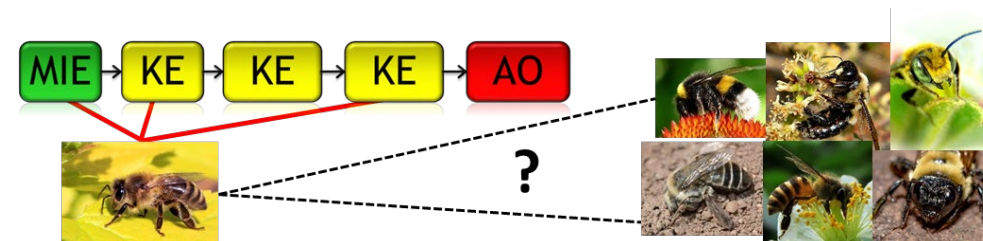




# Applications of Bioinformatics: Case Studies

- **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

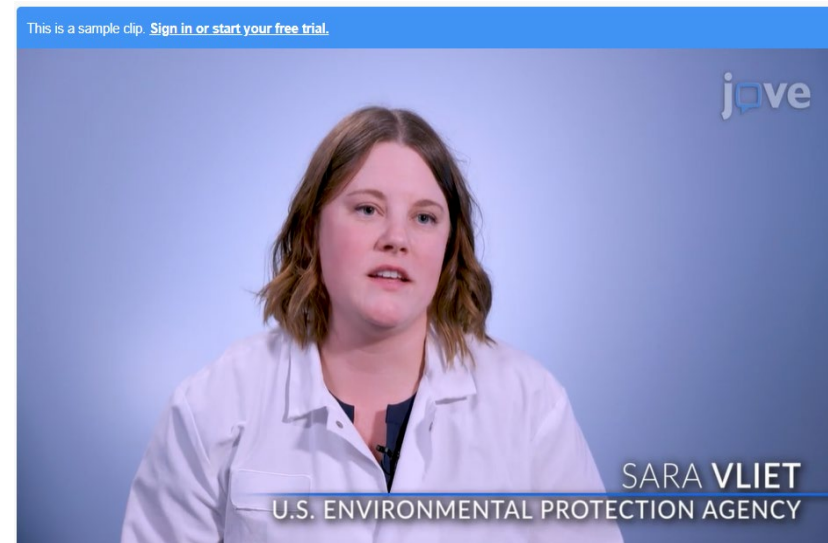


Genetics

# Demonstration of the Sequence Alignment to Predict Across Species Susceptibility Tool for Rapid Assessment of Protein Conservation

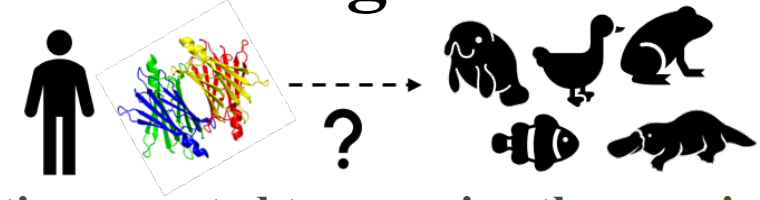
Published: February 10, 2023 DOI: [10.3791/63970](https://doi.org/10.3791/63970)

Sara M. F. Vliet<sup>1</sup>, Monique Hazemi<sup>2</sup>, Donovan Blatz<sup>2</sup>, Marissa Jensen<sup>3</sup>, Sally Mayasich<sup>4</sup>, Thomas R. Transue<sup>5</sup>, Cody Simmons<sup>5</sup>, Audrey Wilkinson<sup>5</sup>, Carlie A. LaLone<sup>6</sup>



# Coupling SeqAPASS with the ECOTOXicology Knowledgebase

Environmental contaminants (e.g., PCBs, PBDEs, PFAS) can competitively bind to the thyroid hormone distribution protein transthyretin (TTR) in mammals and disrupt normal thyroid processes



→ Is this interaction expected to occur in other species?

## Strategy: SeqAPASS Analysis

Taxonomic Group	Number of Species	Majority Similar Susceptibility?	Number Y	Number N
Mammals	135	Y	129	6
Birds	79	Y	79	0
Fish	53	Y	49	4
Reptiles	23	Y	23	0
Amphibians	4	Y	3	1

- **Most species are predicted similarly susceptible, except for 11 aquatic species**

Common Name	Similar Susceptibility	Amino Acid 1	Amino Acid 2	Amino Acid 3	Amino Acid 4	Amino Acid 5	Amino Acid 6
Human	Y	35K	128A	129A	130L	137S	139T
Sperm whale	N	35K	128M	129A	130L	137S	139T
Florida manatee	N	35K	128T	129A	130L	137S	139T
Vaquita	N	35K	127M	128A	129L	136S	138T
Yangtze finless porpoise	N	35K	127M	128A	129L	136S	138T
Narwhal	N	35K	126M	127A	128L	135S	137T
Platypus	N	37K	130A	131T	132L	139S	141T
Pike-perch	N	39K	132A	133M	134L	141S	143T
Blunt-snouted clingfish	N	32H	125P	126L	127L	134S	136Y
Black rockcod	N	10H	103P	104L	105L	112S	114Y
Turquoise killifish	N	11H	104P	105L	106L	113S	115Y
Gabon caecilian	N	40K	133A	134L	135F	142I	144T

→ Can we connect these predictions to empirical data?

## Strategy: Mine the ECOTOXicology Knowledgebase

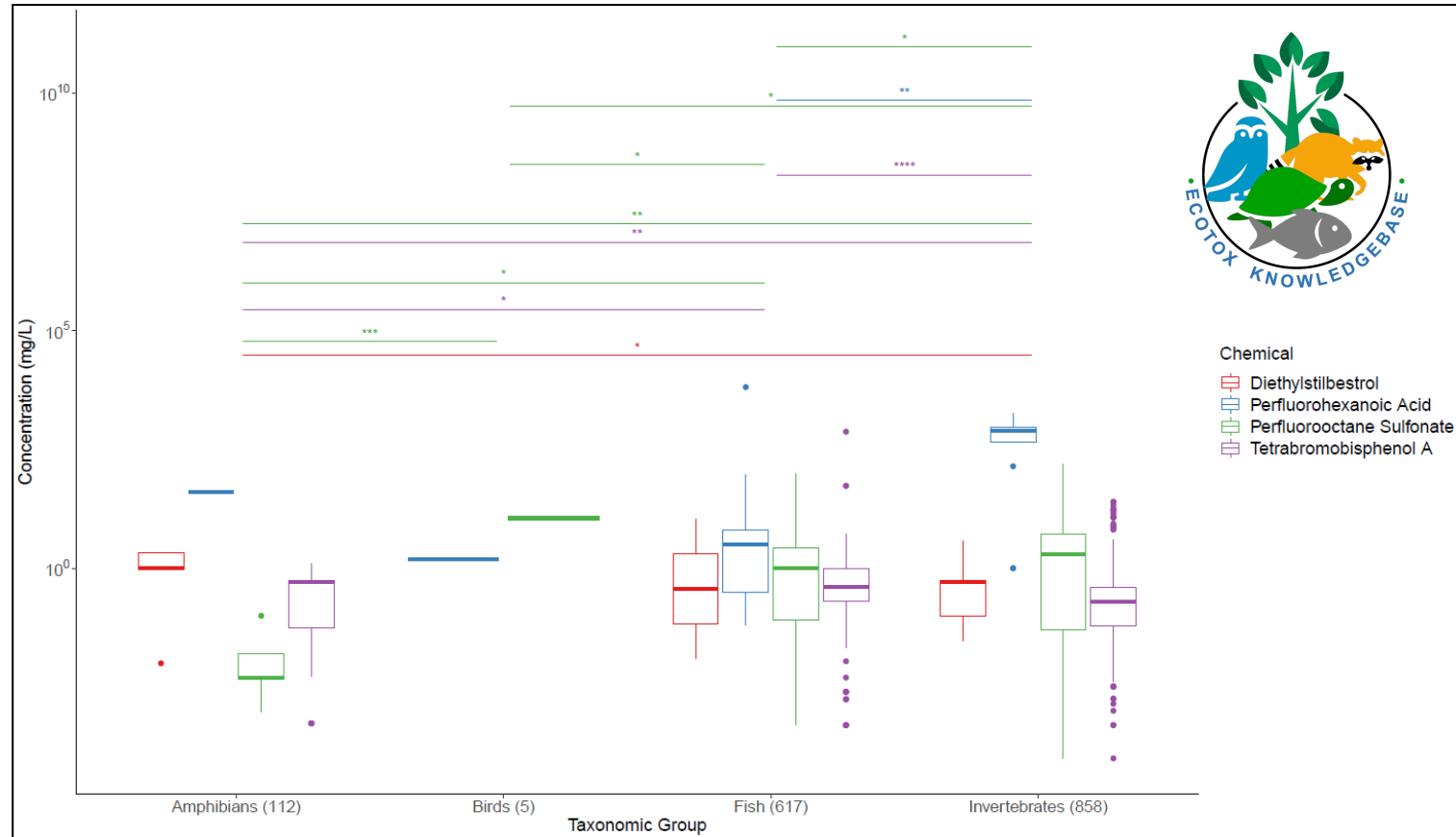
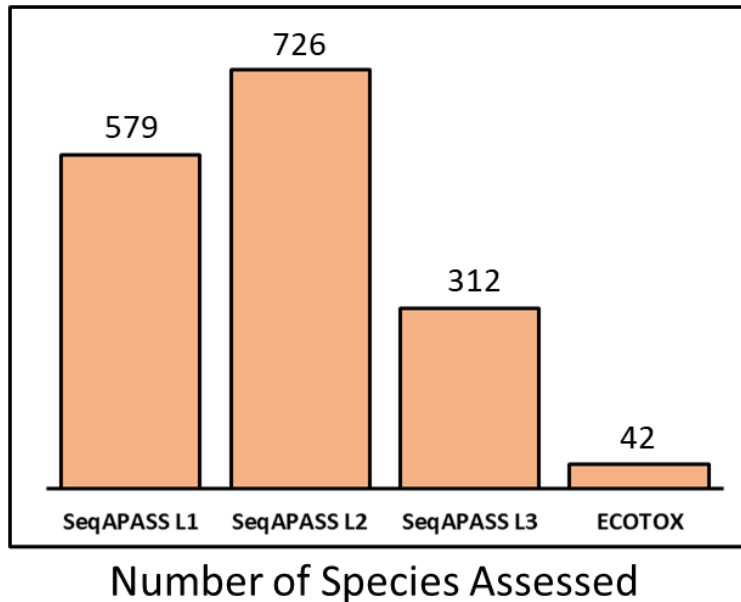
1. Choose chemicals of interest known to act on TTR at the chosen residues
2. Query ECOTOX for Aquatic data by CAS number
3. Filter ECOTOX data to species groups of interest
4. Calculate average effect concentrations
5. Statistically compare the mean effect concentrations of different taxonomic groups



# Coupling SeqAPASS with the ECOTOXicology Knowledgebase

Data suggest potential differences in sensitivity between taxonomic groups, potentially due to biological pathway differences (including TTR).

## → Implications for Cross-Species Extrapolation



- For all species assessed in SeqAPASS, a small number had corresponding ECOTOX data for the chemicals of interest
- For species lacking apical data, SeqAPASS predictions of susceptibility add additional lines of evidence that related species may behave similarly to those for which data are available





**SOT** | Society of  
Toxicology  
[academic.oup.com/toxsci](https://academic.oup.com/toxsci)



*Toxicological Sciences*, 2023, 1–15

<https://doi.org/10.1093/toxsci/kfad038>

Advance Access Publication Date: April 18, 2023

**Research article**

## Weight of evidence for cross-species conservation of androgen receptor-based biological activity

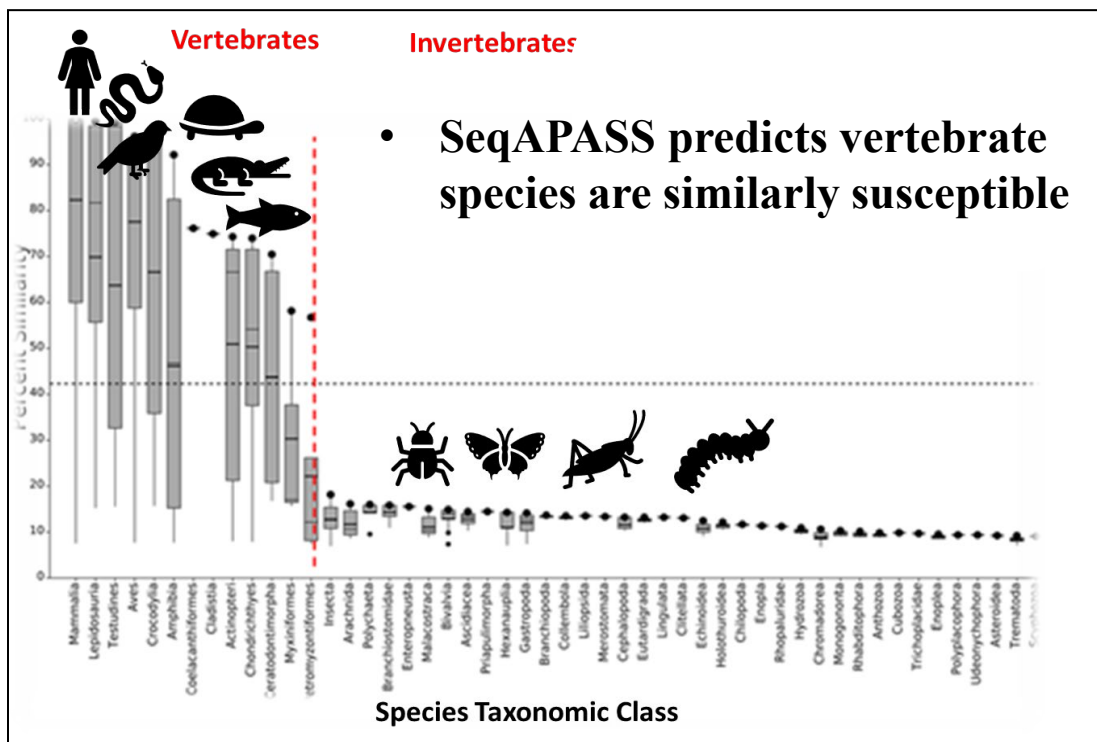
Sara M.F. Vliet <sup>1,\*</sup> Kristan J. Markey <sup>2</sup> Scott G. Lynn,<sup>2</sup> Anna Adetona,<sup>3</sup> Dawn Fallacara,<sup>3</sup> Patricia Ceger,<sup>4,†</sup> Neepa Choksi,<sup>4,‡</sup> Agnes L. Karmaus,<sup>4</sup> AtLee Watson,<sup>4</sup> Andrew Ewans,<sup>4</sup> Amber B. Daniel,<sup>4</sup> Jonathan Hamm,<sup>4</sup> Kelsey Vitense,<sup>1</sup> Kaitlyn A. Wolf,<sup>5</sup> Amy Thomas,<sup>3</sup> Carlie A. LaLone<sup>6</sup>

# Coupling SeqAPASS with Systematic Literature Review

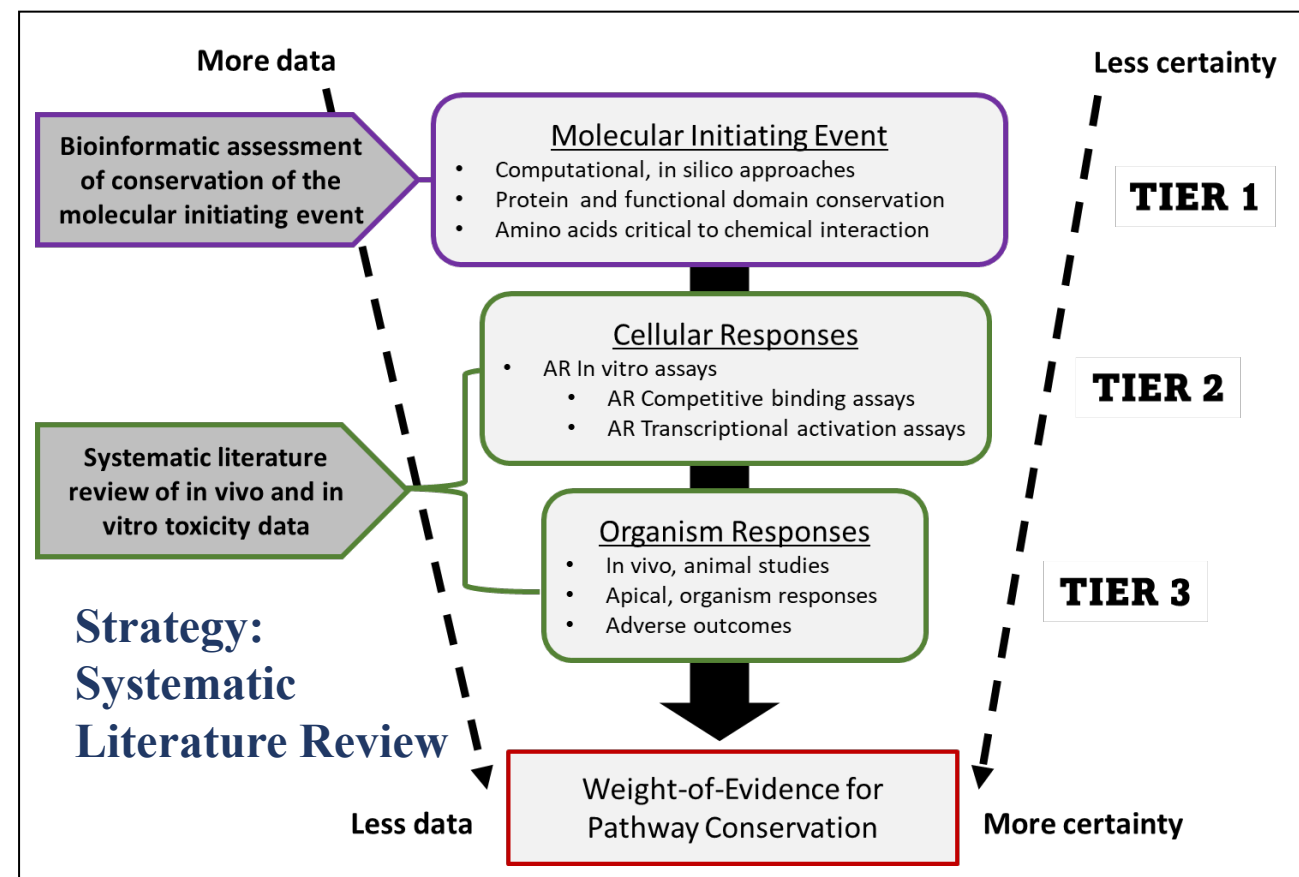
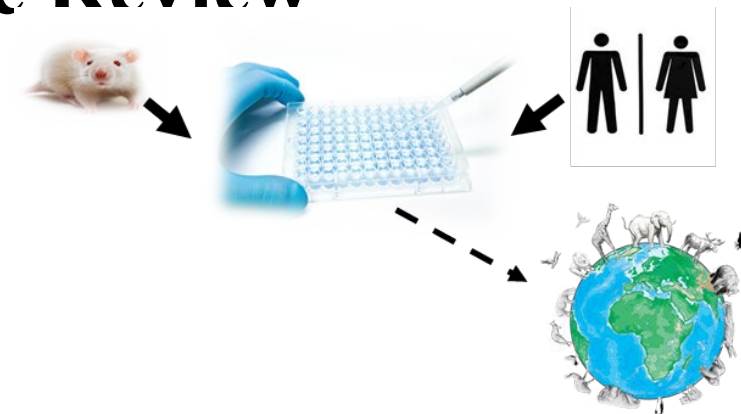
- Many high-throughput screening assays (e.g., ToxCast) rely on mammalian cell lines for determining bioactivity (e.g., androgen receptor agonism)
- However, the extent to which these results can be extrapolated across species and taxonomic groups remains unclear

→ Are HTS results reflective of those expected to occur in other species?

## Strategy: SeqAPASS Analysis



- However, to build weight-of-evidence for cross-species conservation, empirical data is needed







## Environmental Toxicology

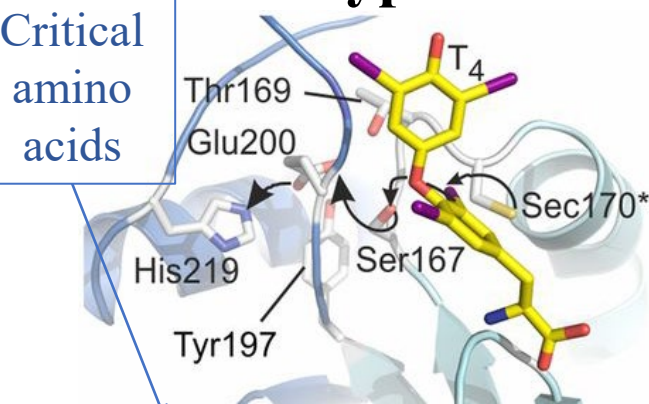
# Combining In Vitro and In Silico New Approach Methods to Investigate Type 3 Iodothyronine Deiodinase Chemical Inhibition Across Species

Sally A. Mayasich,<sup>a,b</sup> Michael R. Goldsmith,<sup>c,d</sup> Kali Z. Mattingly,<sup>e</sup> and Carlie A. LaLone<sup>b,\*</sup>



# Combining *in vitro* and *in silico* New Approach Methods to investigate type 3 iodothyronine deiodinase chemical inhibition across species

Critical amino acids



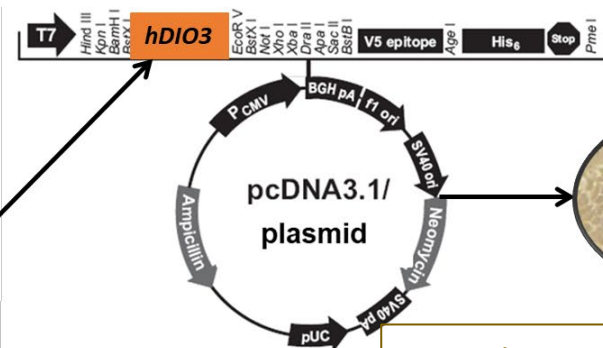
SeqAPASS Evaluation

	C168	T169	C239	A240	Y257
Common Name	Amino Acid 1	Amino Acid 2	Amino Acid 3	Amino Acid 4	Amino Acid 5
Human	168C	169T	239C	240A	257Y
Coelacanth	131C	132T	202C	203F	220Y
Atlantic herring	122C	123T	193C	194V	211Y
Atlantic cod	184C	185T	255C	256A	273Y
Goldfish	135C	136S	206S	207A	224Y
Clown anemonefish	123C	124T	194C	195L	212Y
Red drum	130C	131S	201S	202N	219Y
Blue tilapia	123C	124T	194C	195L	212Y
Zebra mbuna	130C	131S	201T	202N	219Y
Senegalese sole	130C	131S	201G	202N	219Y
Thornfishes	123C	124T	194C	195P	212Y
Antarctic dragonfishes	123C	124T	194C	195P	212Y
Striped sea-bass	123C	124T	194C	195M	212Y
Largemouth bass	123C	124T	194C	195L	212Y
Australian lungfish	133C	134T	204C	205L	222F
Puerto Rican coqui	130C	131T	201C	202R	219Y
Tropical clawed frog	130C	131T	201C	202R	219Y
African clawed frog	128C	129T	199C	200R	217Y
American bullfrog	131C	132T	202C	203R	220Y
Sea lamprey	143G	144S	215C	216P	233A

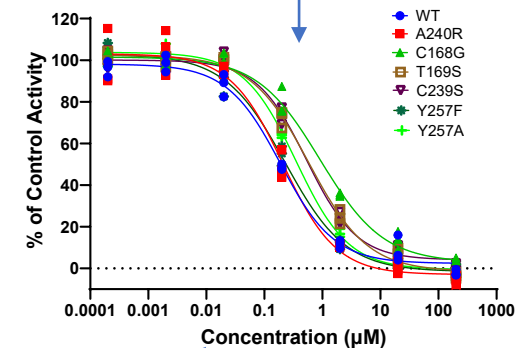
Transfect each mutated plasmid construct into HEK293 cells.

hDIO3 variants:

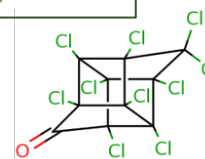
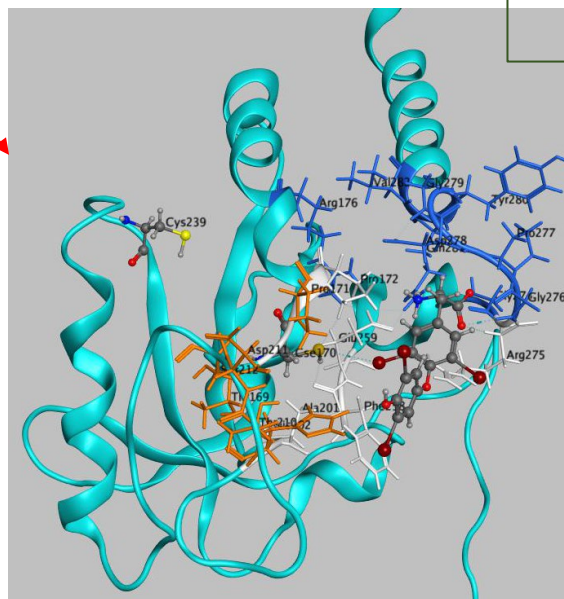
- C168G Lamprey
- T169S Fish
- C239S Fish
- A240R Frogs
- Y257A Lamprey
- Y257F Lungfish



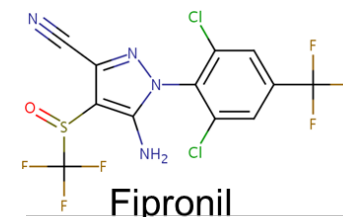
In vitro assays



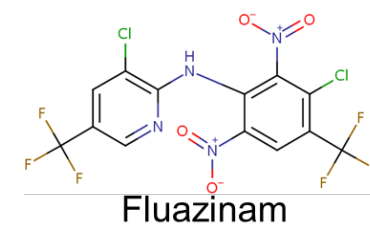
In silico molecular docking



Kepone



Fipronil



Fluazinam



## Environmental Toxicology

# Defining the Biologically Plausible Taxonomic Domain of Applicability of an Adverse Outcome Pathway: A Case Study Linking Nicotinic Acetylcholine Receptor Activation to Colony Death

Marissa A. Jensen,<sup>a,b</sup> Donovan J. Blatz,<sup>c</sup> and Carlie A. LaLone<sup>b,\*</sup>

<sup>a</sup>Department of Biology, Swenson College of Science and Engineering, University of Minnesota Duluth, Duluth, Minnesota, USA

<sup>b</sup>US Environmental Protection Agency, Center for Computational Toxicology and Exposure, Great Lakes Toxicology and Ecology Division, Duluth, Minnesota, USA

<sup>c</sup>Oak Ridge Institute for Science and Education, Duluth, Minnesota, USA

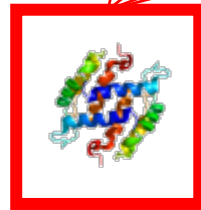
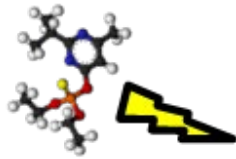
# Expanding taxonomic domain of applicability (tDOA) of AOPs



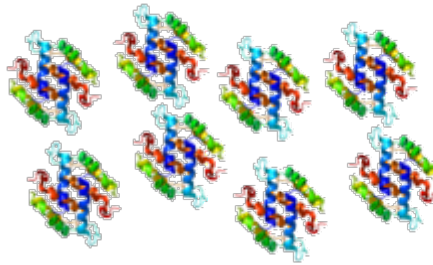
Nicotinic  
acetylcholine  
receptor activation



Colony  
death/failure



≡

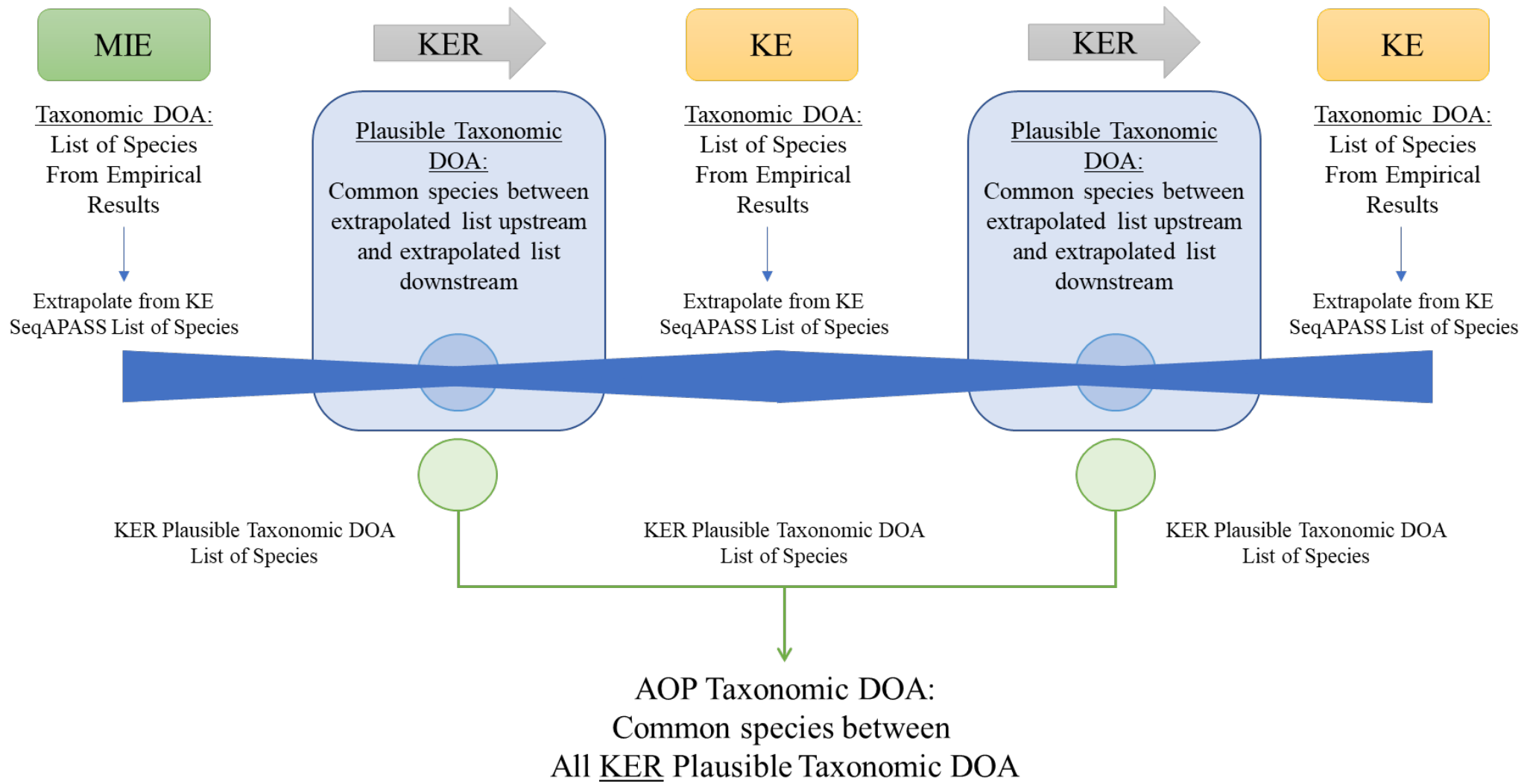


Compare to Millions of Proteins  
From Thousands of Species

Protein in MIE or KE for  
species used to develop the  
AOP

How broadly can we  
extrapolate this AOP across  
bee species?

# Expanding tDOA of AOP Using SeqAPASS

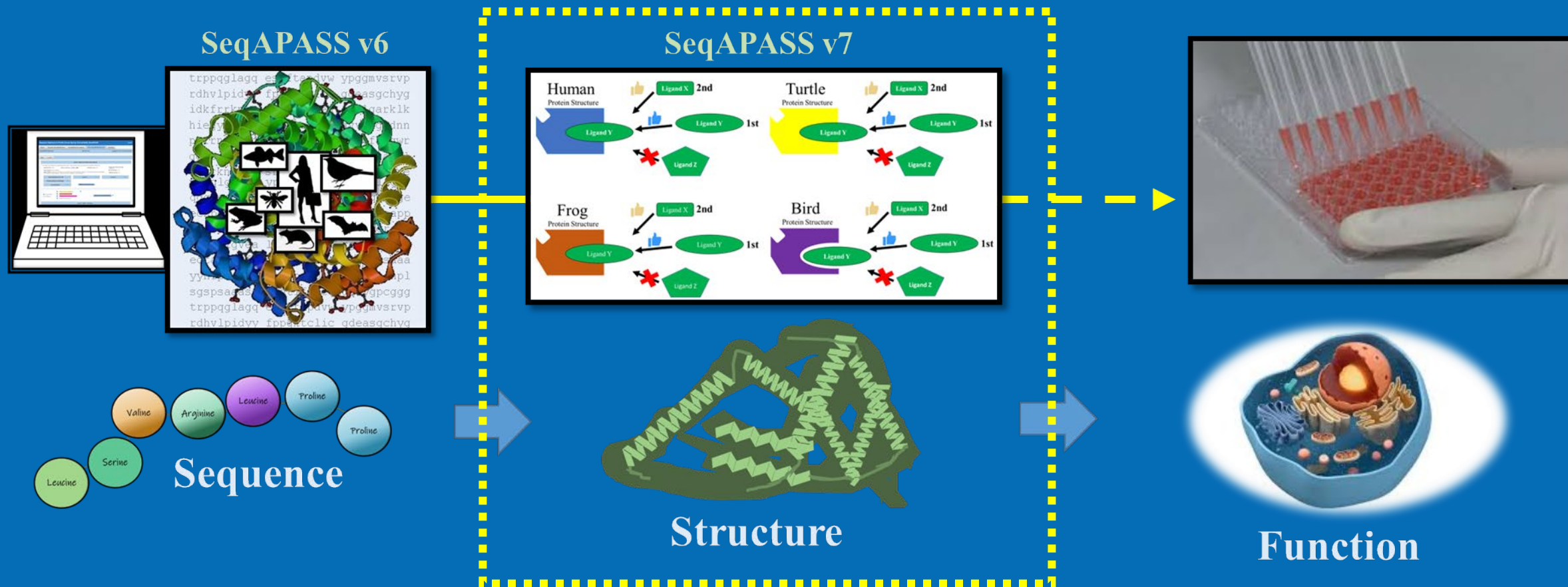


Empirical Evidence tDOA	SeqAPASS Evidence for tDOA
Honey bee ( <i>Apis mellifera</i> )	3-4 <i>Apis</i> species 13-14 non- <i>Apis</i> species

**SeqAPASS evaluations → Structural lines of evidence to extrapolate MIE and early KEs across 17 bee species**



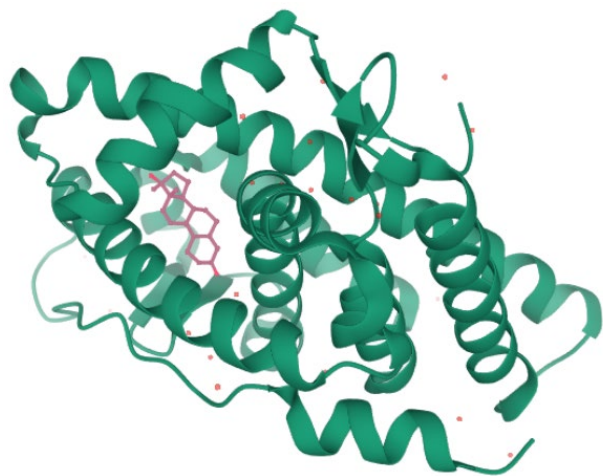
# Begin Simple and Advance as the Science Advances



Consider structural conservation across species and move toward predicted binding affinities

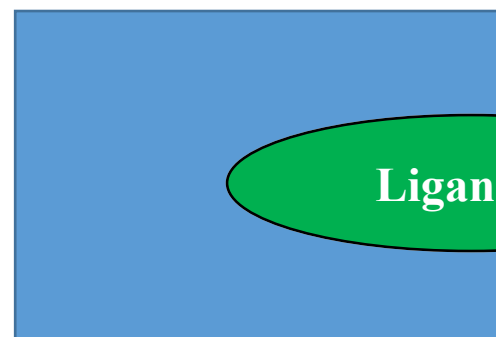
# Advances in Drug Discovery/Development

(COVID-19 has led to advances)



Structure derived  
from X-ray  
crystallography

Human  
Protein Structure



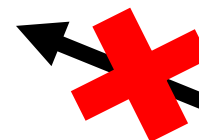
Ligand X

2nd



Ligand Y

1st



Ligand Z

## Bioinformatics Toolbox:

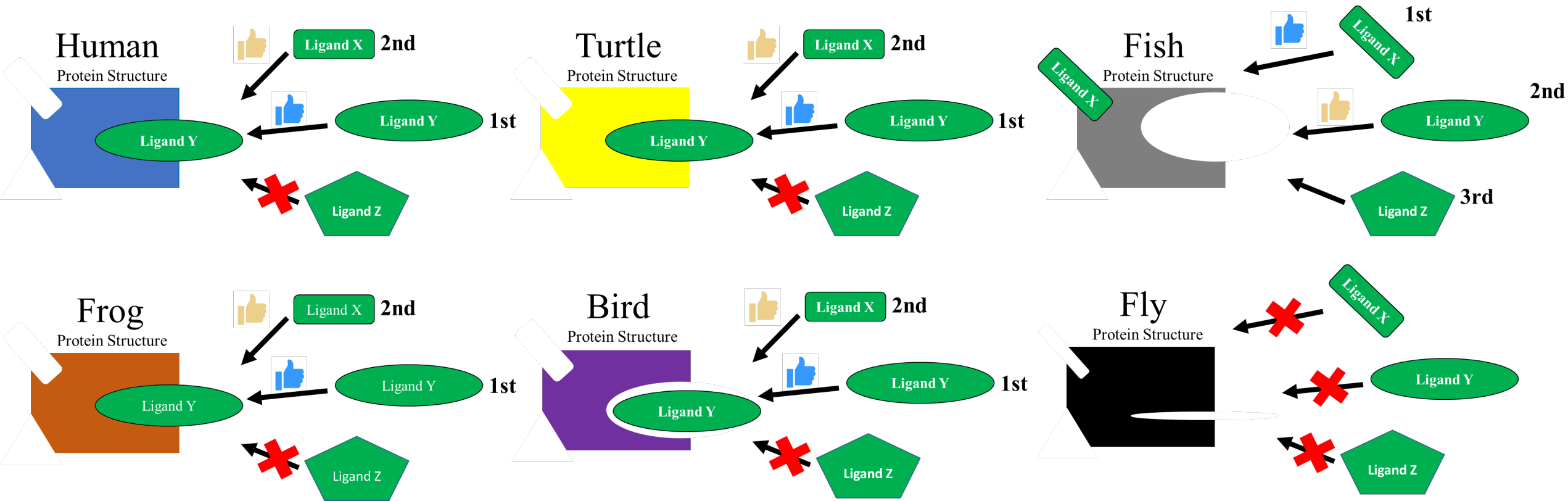
Molecular modeling

Molecular docking

Virtual screening

Molecular dynamic simulations

# Application to Species Extrapolation



## Bioinformatics Toolbox:

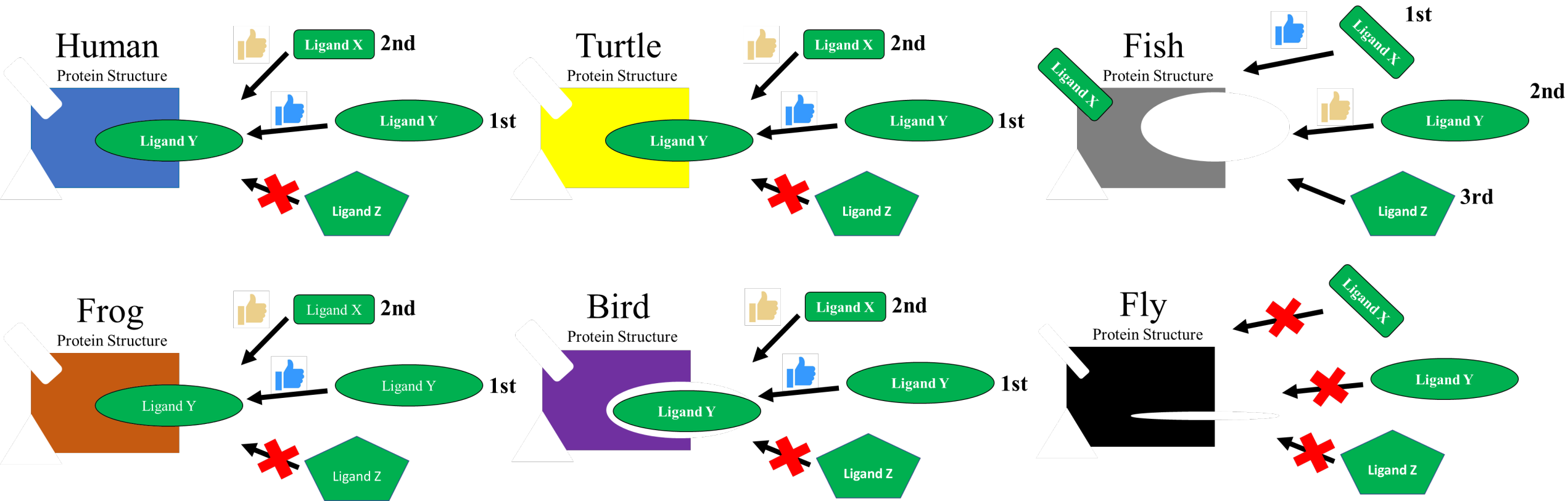
Molecular modeling

Molecular docking

Virtual screening

Molecular dynamic simulations

# Application to Species Extrapolation



**Bioinformatics Toolbox:**  
Molecular modeling  
Molecular docking  
Virtual screening  
Molecular dynamic simulations

*Thousands/Millions/Billions  
of  
Chemicals*

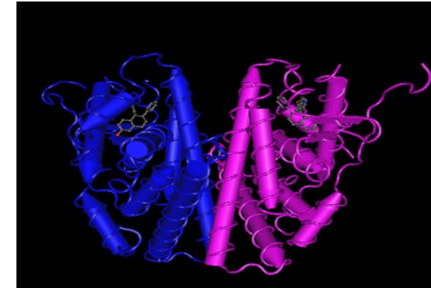


# Sequence

```
MTMTLHTKASGMALLHQIQGNELEPLNRPQLKIPLERPLGE
VYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYGQTLGPLYG
PGSEAAAFSGNSLGGFPPLNSVSPSPLMLLHPPQLSPFLQ
PHGQVQVPPYLENEPSGYTVREAGPPAFYRPNSDNRRQGGR
ERLASTNDKSGMAMESAKETRYCAVCNDYASGYHYGVWSC
EGCKAFFKRSIQGHNDYMCPTNQCTIDKNRRKSCQACRLR
KCYEVGMMKGGIRKDRRGGRLMKHRQRDDGEGRGEVG
SAGDMRAANLWPSPLMIKRSKKNLSLSTADQMVSALLA
EPPILYSEYDPTPRPFSEASMMGLLTNLADRELHMINWAKV
PGFVDLTLDQVHLLCAWLEILMIGLVWRSMHPGKLLFA
PNLLDRNGKQCEVGMVEIFDMLLATSSRFMMNQLQGEF
VCLKSILLNSGVYTLSTLSLEEKDHIHRVLDKITDTLIHLM
```



# Structure



## SeqAPASS Results from Level 1 Query Sequence FASTA + FASTA from 100s of Aligned Sequences Across Taxa

>NP\_001434.1 Protein X [Homo sapiens]  
MSFSGKYQLQSQENFEAFMKAIGLPELIQKGD  
KGVSEIVQNGKHFKFTITAGSKVIQNEFTVGEECE  
LETMTGEKVTVVQLGDNKLVTFKNIKSVTELN  
GDIITNTMTLGDIVFKRISKRI

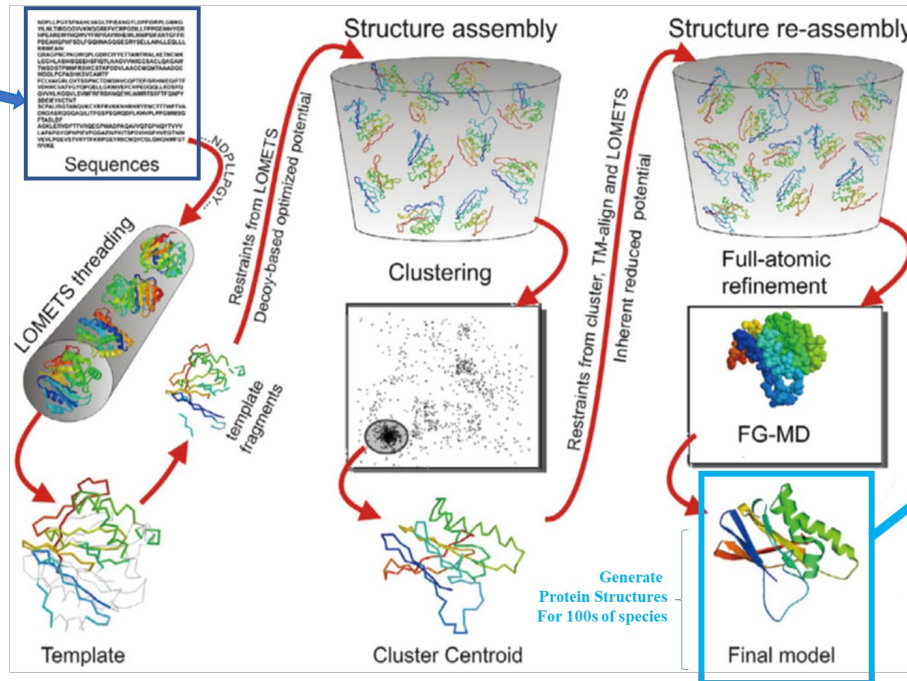
>NP\_787011.1 Protein X [Bos taurus]  
MNFSGKYVQVQENYEFMKAIGLPELIQKGD  
DIKGVSEIVQNGKHFKFTITAGSKVIQNEFTVGEECE  
MEFMTGEKIKAVVQEGDNKLVTFKNIKSVTEFN  
GDTVSTMTKGDVVKRISKRI

>KFQ76585.1 Protein X [Phoenixcopterus ruber  
ruber]  
MSFTGKYLQSQENFEAFMKAIGLPELIQKGD  
IKSISEIVQDGGKFKVTVTTSKVMQNEFTIGEECD  
IEMLTGEKVKAVVQMEGNRLVANLKGKLSVTEL  
NGDIITHTMTMGDLTYKRISKRI

>NP\_001116883.1 Protein X [Xenopus  
tropicalis]  
MAFAGKYELVHQENFEAFMKAIGLPELIQKGDV  
KSVTEIQNGKHFKFTITAGSKVLNFEFTIGEEAE  
LETPTGKVKSVKLEGDNLVQLKAITSTTELSG  
DTITHVLTNNLVFKRISKRV

100s of FASTA

## Iterative Threading ASSEMBly Refinement Develop Models for 100s of Species Based on Aligned Sequences (I-TASSER; <https://zhanglab.cmb.med.umich.edu/I-TASSER/>)

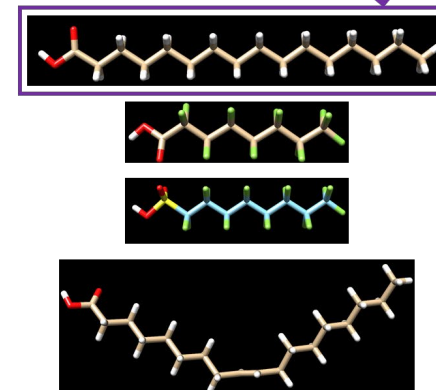


## UCSF Chimera DockPrep Structures and Minimize Ligands

### Protein Structure Models From 100s of Species



### Ligands of Interest for Docking



## AutoDock Vina Dock Multiple Ligands to Protein Structures



### Collect Predicted Binding Affinity

S	Score	RMSE Lib	RMSE u.b	HBonds (all)	HBond Ligand Atoms	HBond Receptor Atoms
V	-7.1	0.0	0.0	0	0	0
V	-7.0	1.212	2.436	0	0	0
V	-7.0	2.148	6.837	1	1	1
V	-6.9	1.128	2.04	0	0	0
V	-6.9	4.472	7.133	0	0	0
V	-6.7	3.27	7.552	0	0	0
V	-6.7	2.637	3.461	2	2	2
V	-6.6	1.572	3.516	0	0	0
V	-6.6	1.725	3.368	0	0	0

Chimera Model #3.1

REMARK VINA RESULT: -7.1 0.000 0.000

REMARK 15 active torsions:

REMARK status: 'A' for Active; 'I' for Inactive

REMARK 1 A between atoms: C2\_2 and C3\_3

REMARK 2 A between atoms: C3\_3 and C4\_4

REMARK 3 A between atoms: C4\_4 and C5\_5

REMARK 4 A between atoms: C5\_5 and C6\_6

REMARK 5 A between atoms: C6\_6 and C7\_7

REMARK 6 A between atoms: C7\_7 and C8\_8

REMARK 7 A between atoms: C8\_8 and C9\_9

REMARK 8 A between atoms: C10\_10 and C9\_9

REMARK 9 A between atoms: C10\_10 and C11\_11

REMARK 10 A between atoms: C11\_11 and C12\_12

REMARK 11 A between atoms: C12\_12 and C13\_13

REMARK 12 A between atoms: C13\_13 and C14\_14

REMARK 13 A between atoms: C14\_14 and C15\_15

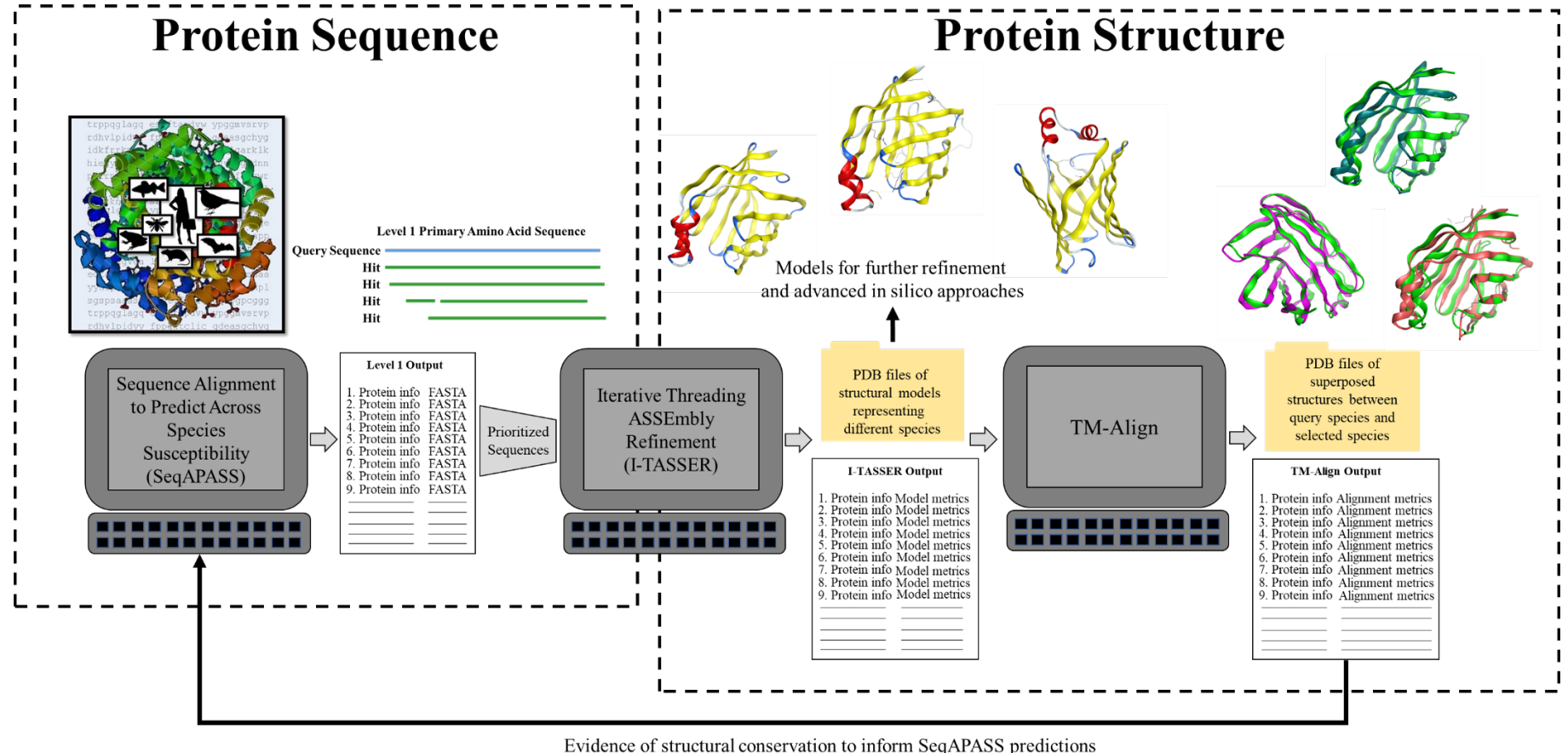
REMARK 14 A between atoms: C15\_15 and C16\_16

REMARK 15 A between atoms: C16\_16 and C17\_17

# Expanding the SeqAPASS Pipeline: I-TASSER and TM-Align



Maxwell Botz  
ORISE



Evidence of structural conservation to inform SeqAPASS predictions

**Query Species:**

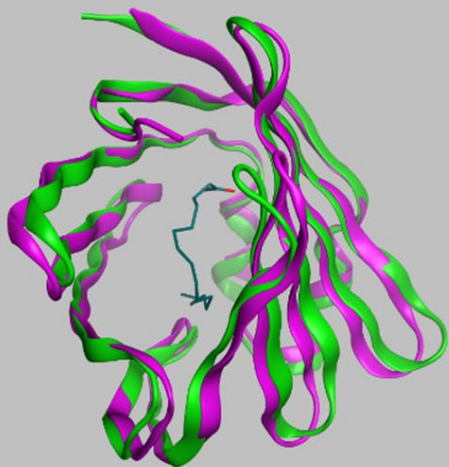
Human

Androgen receptor

Ligand Binding Domain

**A.**

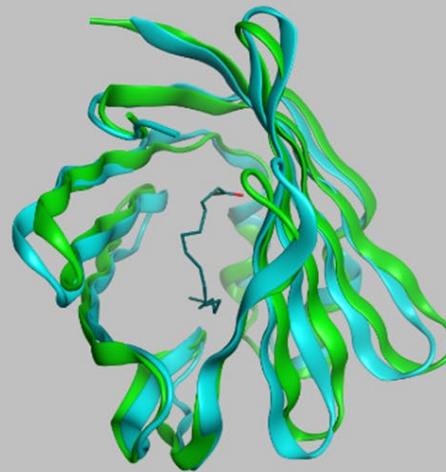
Average TM-align Score 0.96



Dog (*Canis lupus familiaris*)

**B.**

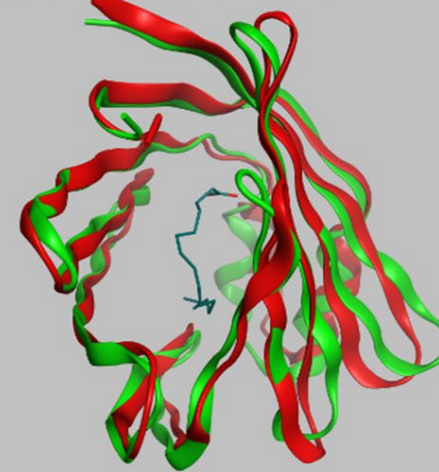
Average TM-align Score 0.95



Chicken (*Gallus gallus*)

**C.**

Average TM-align Score 0.92





African clawed frog (*Xenopus laevis*)

## Environmental Toxicology and Chemistry

Environmental Toxicology | [Full Access](#)

### From Protein Sequence to Structure: The Next Frontier in Cross-Species Extrapolation for Chemical Safety Evaluations

Carlie A. LaLone , Donovan J. Blatz, Marissa A. Jensen, Sara M. F. Vliet, Sally Mayasich, Kali Z. Mattingly, Thomas R. Transue, Wilson Melendez, Audrey Wilkinson, Cody W. Simmons, Carla Ng ... [See all authors](#) 

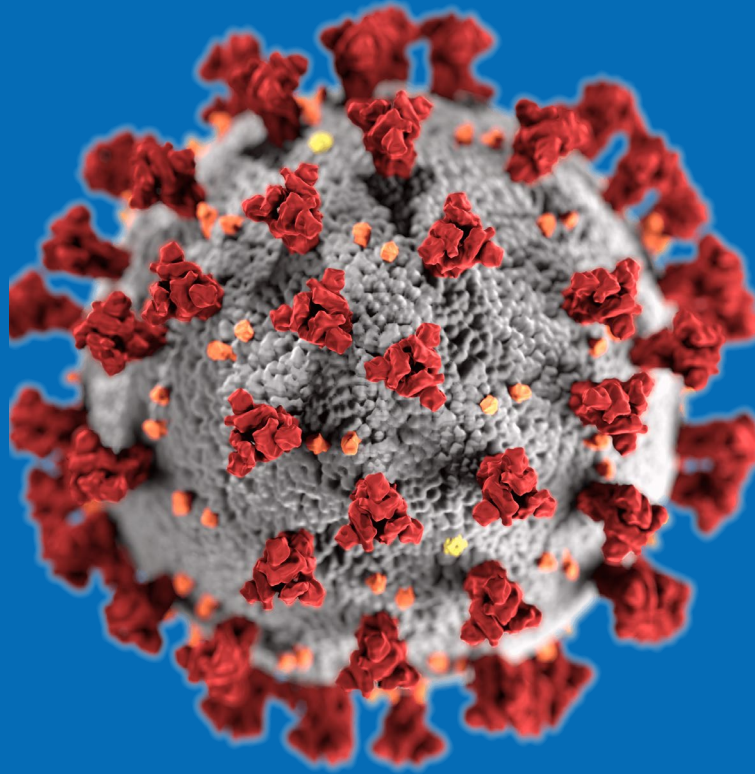
First published: 16 December 2022 | <https://doi.org/10.1002/etc.5537>



What's  
Cooking?!







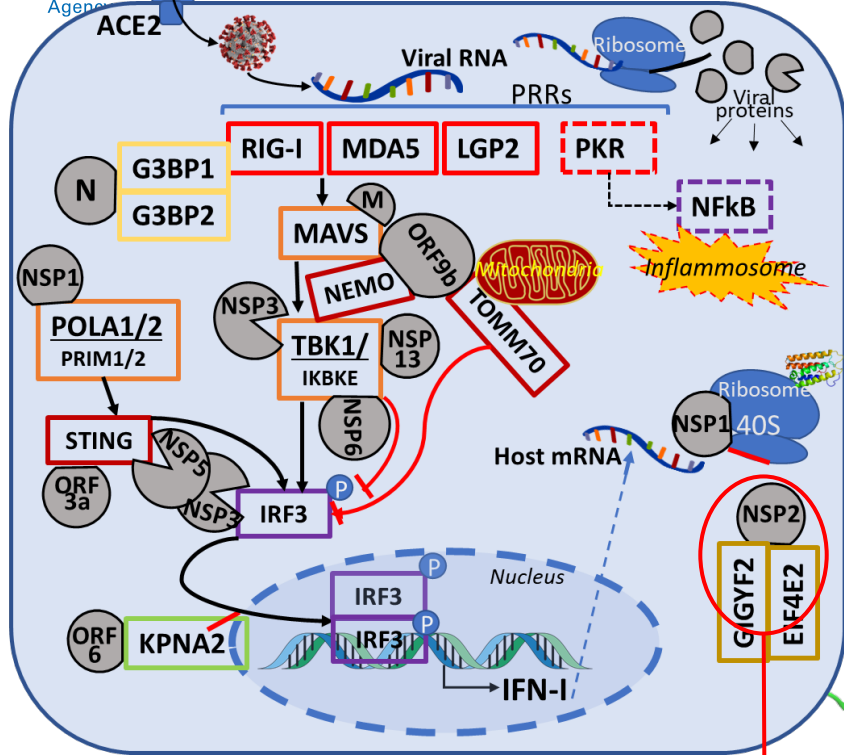
# Identification of potential SARS-CoV-2 intermediate host species using SeqAPASS

Sally Mayasich

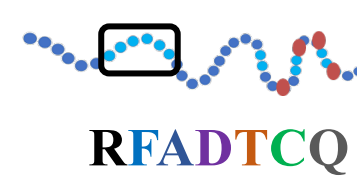
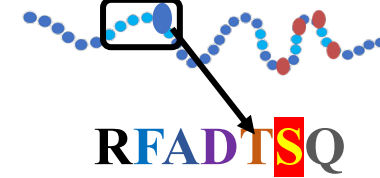
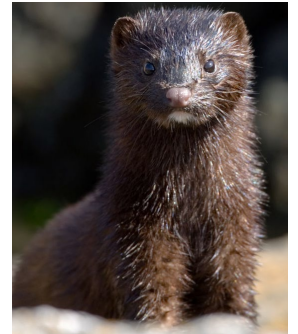
US EPA -UW-Madison Cooperative  
training Agreement post-doctoral fellow



Host antiviral interferon (IFN-I)  
response antagonized

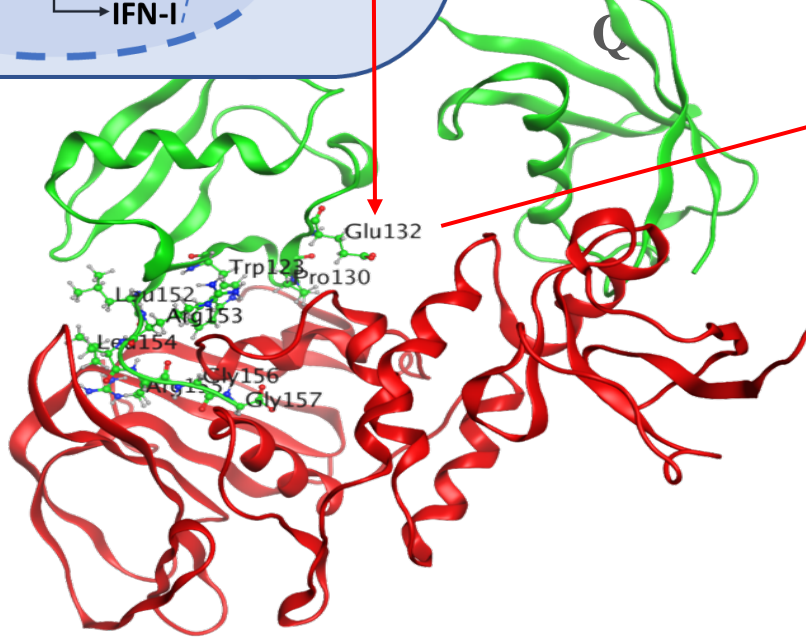


# Investigation of Interferon-I antiviral response protein sequence variation across species to predict non-human host susceptibility to SARS-CoV-2



SeqAPASS Level 3 Critical  
amino acid comparison

Human	RFADTCQ
Pangolin	RFADTCQ
Dog	RFADTSQ
Civet	RFADTCQ
Mink	RFADTCQ





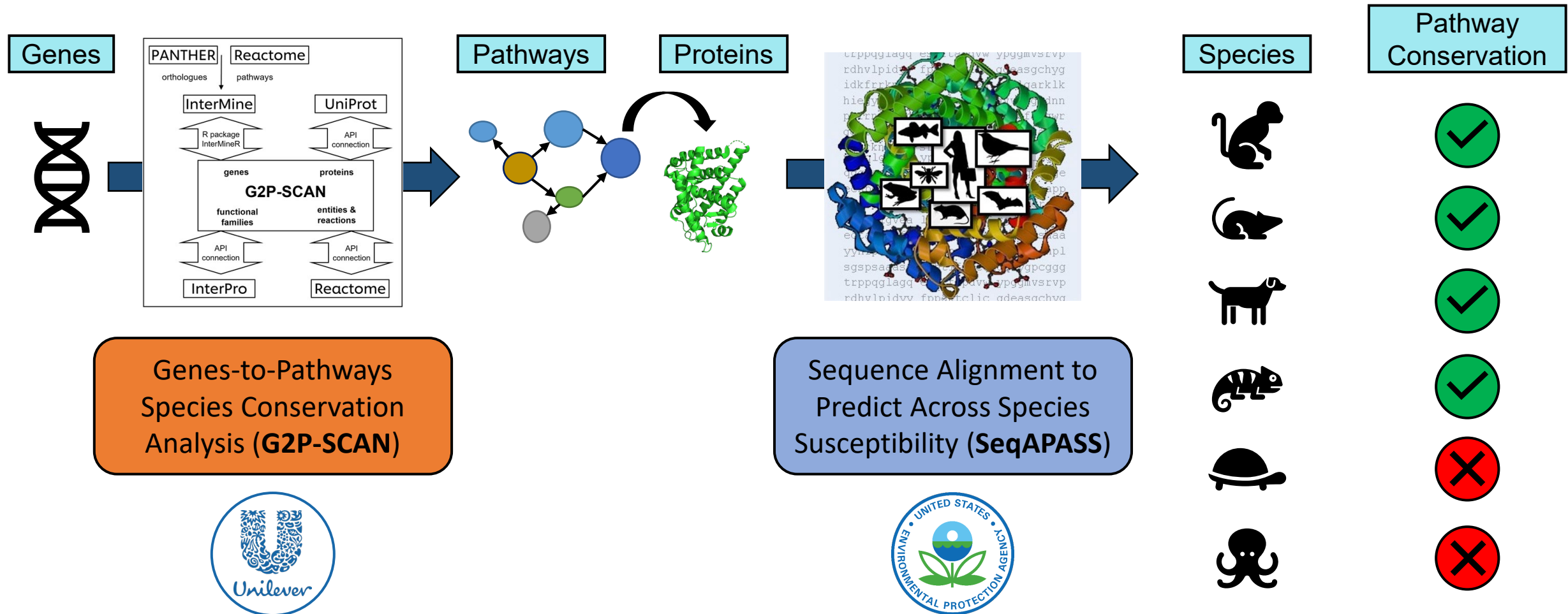
# CRADA(2) Project - Predictive toxicology to evaluate cross species differences in chemical susceptibility



**Peter Schumann, MS**  
(ORISE Research Fellow)

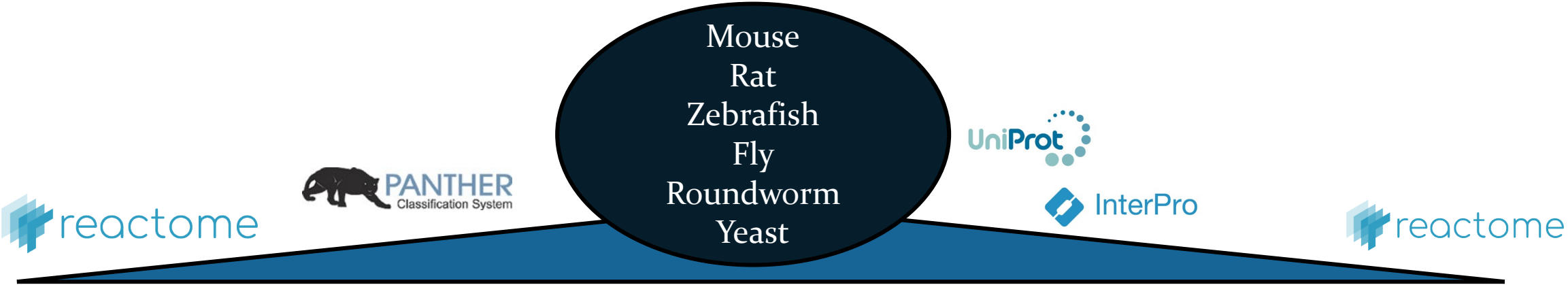


# Enhancing Species Extrapolation with Combined Approaches





# Target evaluation using G2P-SCAN



Pathways	Orthologs	Proteins	Protein Families	Entities and Reactions
----------	-----------	----------	------------------	------------------------

Pathways and Gene Counts		Least Divergent Orthologs		UniProt IDs & Counts		InterPro IDs & Counts		Entities, Reactions & Counts		
Pathway1	45 genes	<i>Gene1</i>	2 species	Prot1	2	IPR1	2	Pathway1	7	1
Pathway2	23 genes	<i>Gene2</i>	3 species	Prot2	3	IPR2	2	Pathway2	8	2
Pathway3	114 genes	<i>Gene3</i>	1 species	Prot3	1	IPR3	1	Pathway3	3	1
Pathway4	12 genes	<i>Gene4</i>	0 species	Prot4	0	IPR4	0	Pathway4	0	0
Pathway5	67 genes	<i>Gene5</i>	1 species	Prot5	1	IPR5	0	Pathway5	2	0
Etc.		Etc.		Etc.		Etc.		Etc.	58	

# Can molecular docking simulations help support predictions of cross-species susceptibility?

Peter Schumann, MS  
(ORISE Research Fellow)



Investigating with use of the  
androgen receptor



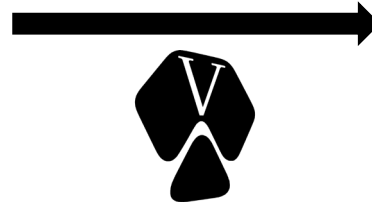
PDB ID: 2AMA



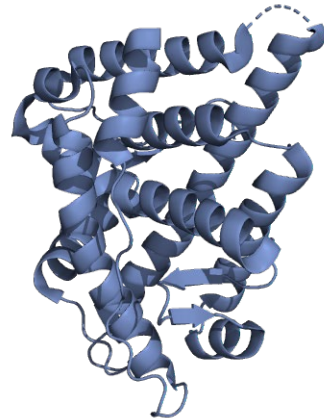
Species 1



Test binding



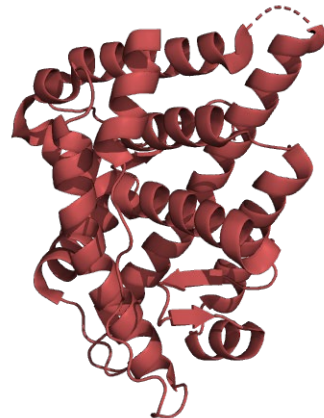
Species 2



Test binding



Species 3



Test binding



Exploring the use of these metrics for supporting species susceptibility calls

- PLIF similarity
- Ligand RMSD
- Binding pocket similarity
- Binding affinities

Susceptible species

Government

Industry

# Consortium to Advance Cross Species Extrapolation in Regulation

## Steering Committee:

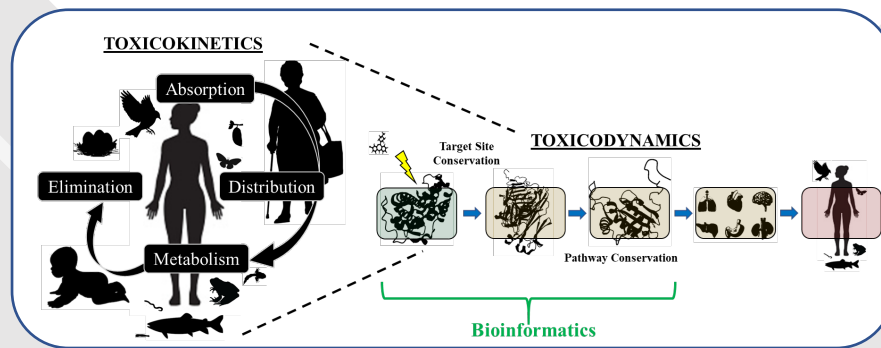
Carlie LaLone (US EPA)  
Geoff Hodges (Unilever)  
Nil Basu (McGill U)  
Steve Edwards (RTI)  
Fiona Sewell (NC3Rs)  
Michelle Embry (HESI)  
Patience Browne (OECD)

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

Interested in Learning more or Joining: Contact [LaLone.Carlie@epa.gov](mailto:LaLone.Carlie@epa.gov) or [Geoff.Hodges@unilever.com](mailto:Geoff.Hodges@unilever.com)

Academia

NGO





# Acknowledgements

## U.S. EPA, ORD

Maxwell Botz (ORISE)

Peter Schumann (ORISE)

Marissa Jensen (University of Minnesota Duluth)

Sally Mayasich (University of Wisconsin)

Monique Hazemi (ORISE)

Sara Vliet (past ORISE 2021)

Donovan Blatz (past ORISE 2021)

Jon Doering (past NRC 2018)

Colin Finnegan (past ORISE 2018)

## GDIT

Cody Simmons

Audrey Wilkinson

Wilson Menendez

Thomas Transue (past GDIT 2022)

(SeqAPASS v7.0 Coming 2023)



[LaLone.Carlie@epa.gov](mailto:LaLone.Carlie@epa.gov)

<https://seqapass.epa.gov/seqapass/>