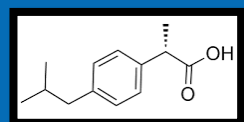
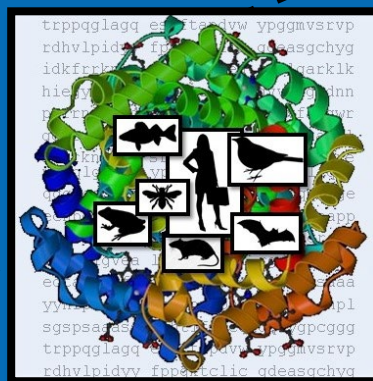


# A vision for the role of bioinformatics, specifically SeqAPASS in species extrapolation for chemical safety

Carlie A. LaLone, Ph.D.  
Research Bioinformaticist



Toxicity Data with  
Model Organism

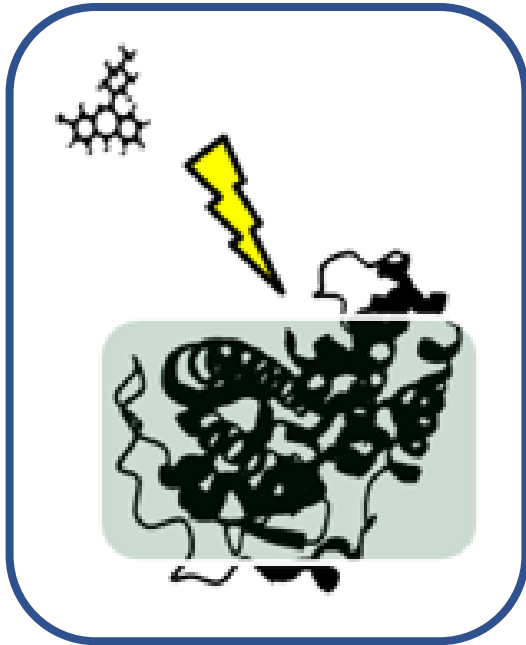


Predict Chemical Susceptibility Across All Species

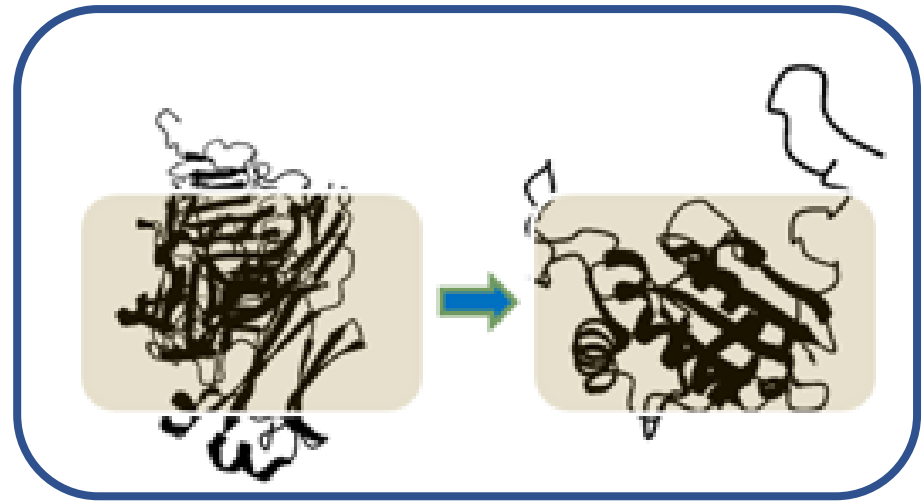


# Numerous bioinformatics approaches

- My interest and focus has been chemical-protein and protein-protein Interactions
  - Utility for cross species extrapolation



Chemical-Protein



Protein-Protein



# Overview Questions

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





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





# Toxicity Testing to Understand Chemical Safety

- Regulatory decision-making



**US EPA Examples:**

*Clean Air Act*

*Clean Water Act*

*Resource Recovery Act*

*Endangered Species Act*

*Food Quality Protection Act*

*Endocrine Disruptor Screening Program*

*Federal Insecticide, Fungicide, and Rodenticide Act*

*Frank R. Lautenberg Chemical Safety for the 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**



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



**ability to control diet and surroundings**





# Species Extrapolation



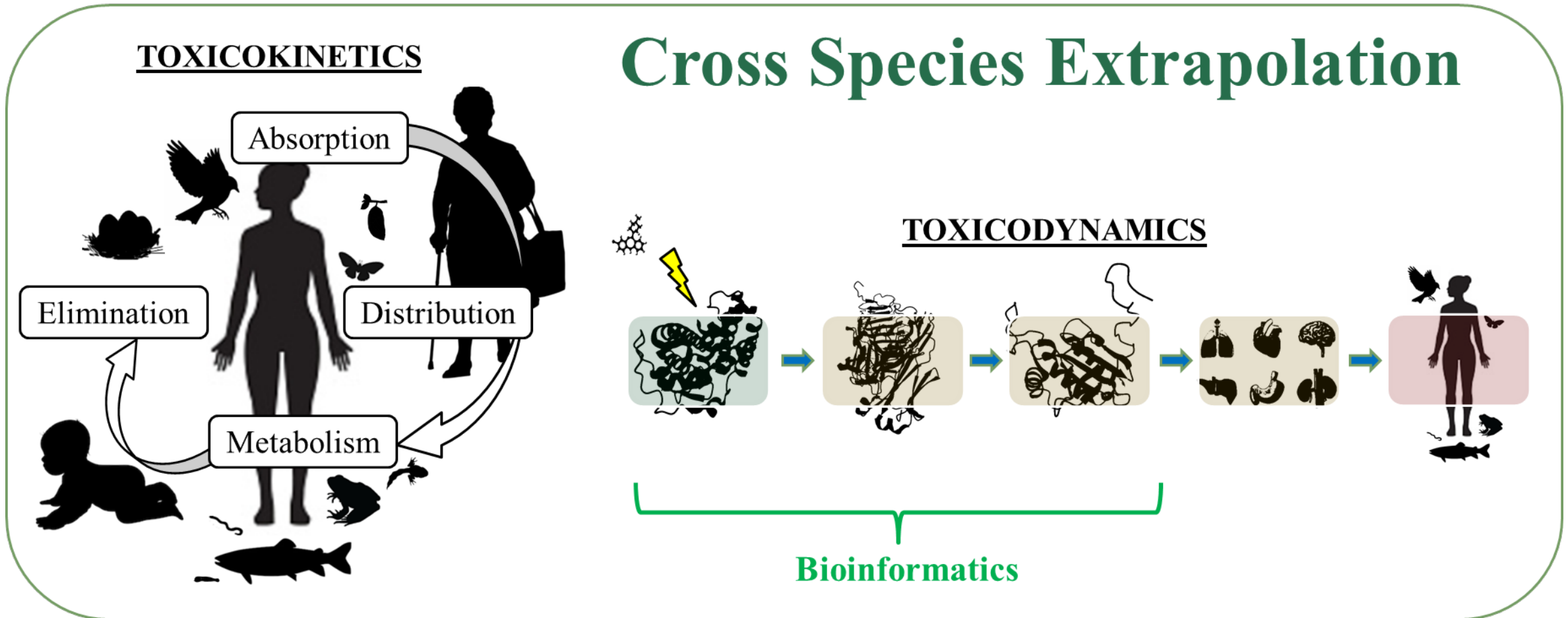
## What is it?

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

## Why is it important:

- Limited or no toxicological data for the animal or plant species of interest – reliance on surrogate (model organisms)
  - Impractical to generate new data for all species
- Testing resources are limited
  - International interest to reduce animal use
  - Ever-increasing demand to evaluate more chemicals in a timely and sometimes expedited manner
- Sensitivity of species must be estimated based on scientifically-sound methods of cross-species extrapolation
  - Immense diversity of species in the wild
  - Important challenge for species listed under the Endangered Species Act

# Sensitivity to Chemical Perturbation







# Bioinformatics

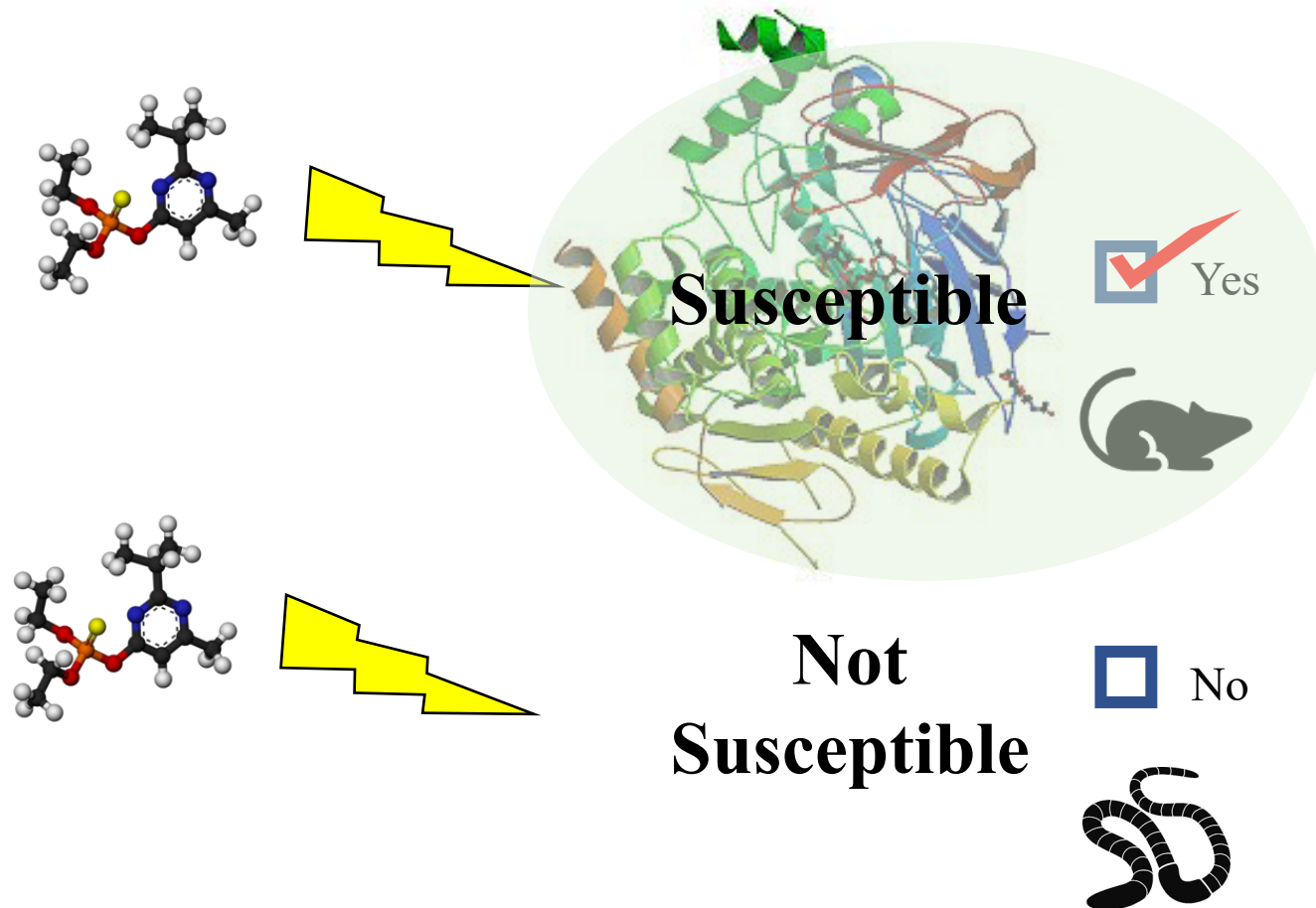
- Combines mathematics, information science, and biology to answer biological questions
- Developing methodology and analysis tools to explore large volumes of biological data
  - Query, extract, store, organize, systematize, annotate, visualize, mine, and interpret complex data
    - Usually pertains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

**Let the computers do the work**

# Considering chemical sensitivity?

## Factors that make a species sensitive

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



Simple question to address:

Is the known chemical target available in a species for a chemical to act upon?

Yes or No

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





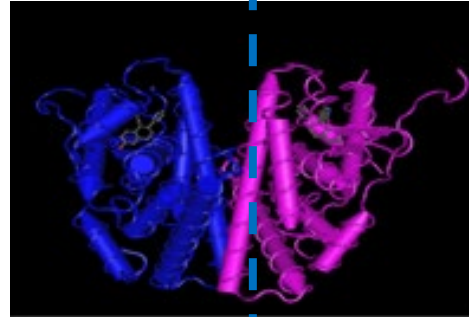
Start simple

## Sequence

```
MTMTLHTKASGMALLHQIQGNELEPLNRPQLKIPLERPLGE  
VYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYGQTGLPYG  
PGSEAAAFSGNGLGGFPPLNSVSPSPLMLLHPPQLSPFLQ  
PHGQQVPYYLENEPSGYTVREAGPPAFYRPNSDNRRQGGR  
ERLASTNDKGSMAVESAKETRYCAVCNDYASGYHYGVWSC  
EGCKAFFKRSIQGHNDYMCNATNQCTIDKNRRKSCQACRLR  
KCYEVGMMKGGIRKDRRGGRMLKHKRQRDDGEGRGEVG  
SAGDMRAANLWPSPLMIKRSKKNLALSLTADQMVSALLA  
EPPILYSEYDPTRPFSEASMMGLLTNLADRELHMINWAKV  
PGFVDLTLDQVHLLCAWLEILMIGLVWRSMHEHPGKLLFA  
PNLLDRNQGKCEGMVEIFDMLLATSSRFMMNLQGEF  
VCLKSIILLNSGVYFLSSTLKSLEEKDHIHRVLDKITDTLIHLM
```



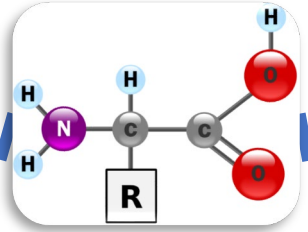
## Structure



## Function



# Bioinformatics



Valine

Arginine

Leucine

Proline

Pro

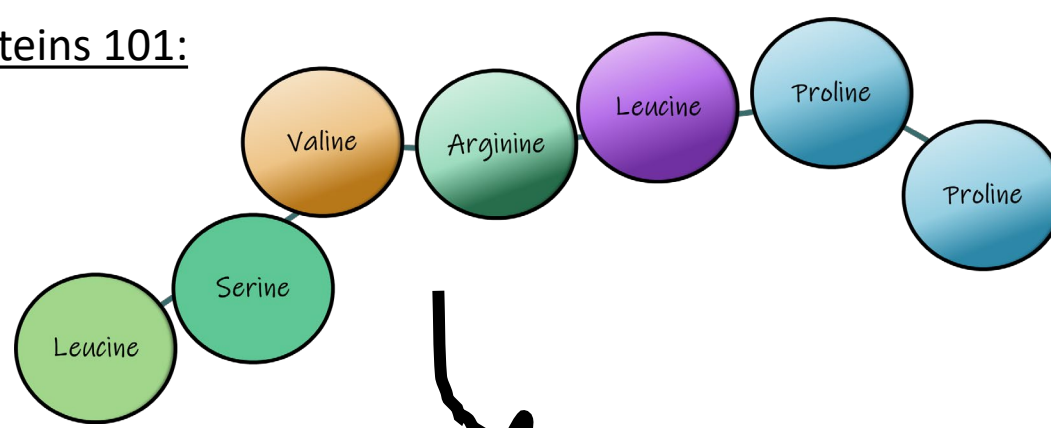
Serine

2

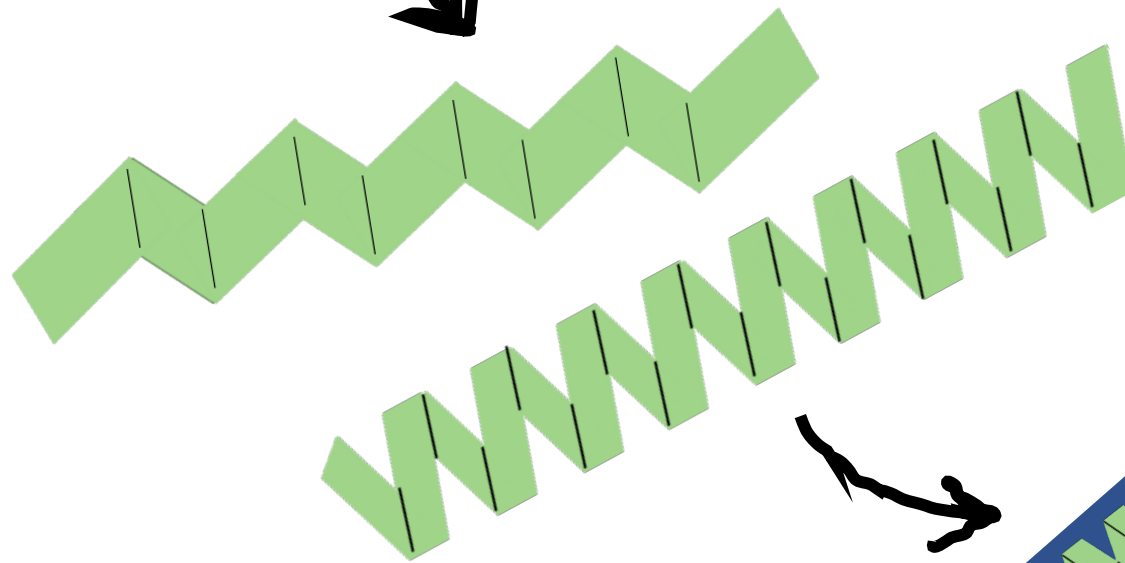
Amino Acids



## Proteins 101:



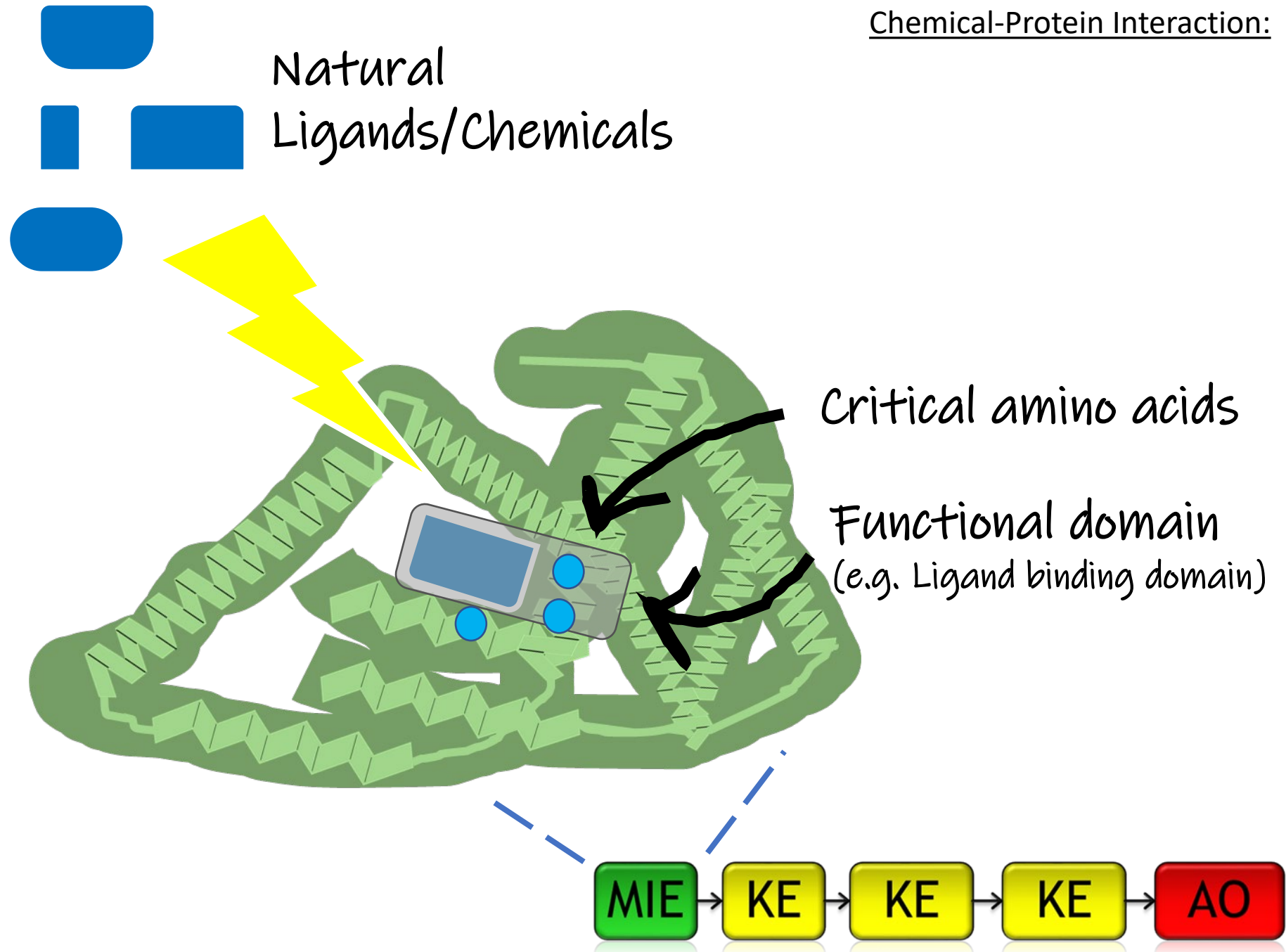
Primary amino acid sequence



Secondary Structure

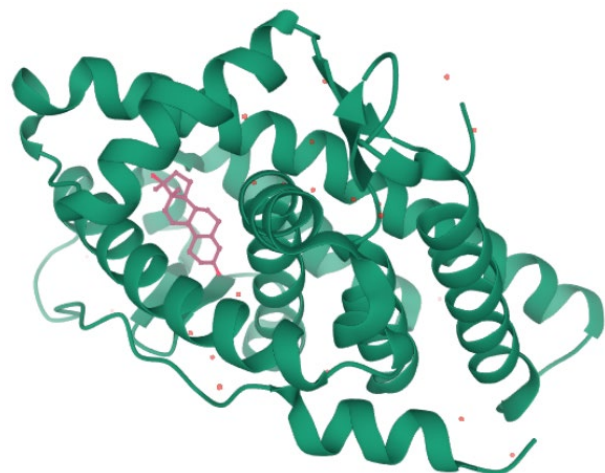
Tertiary Structure





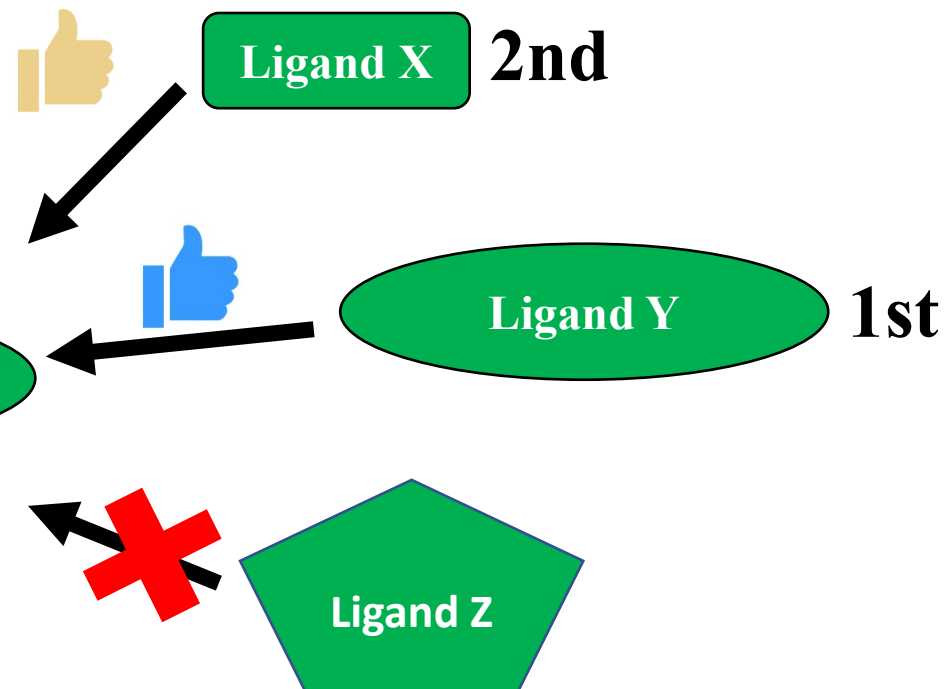
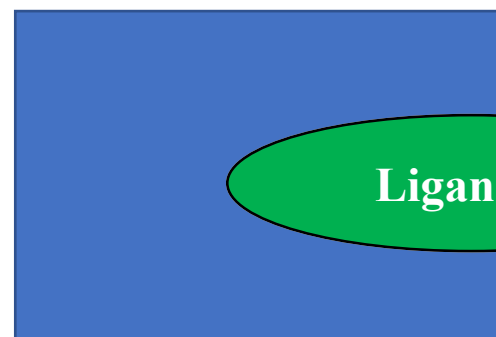


# Advances in Drug Discovery/Development



Structure derived  
from X-ray  
crystallography

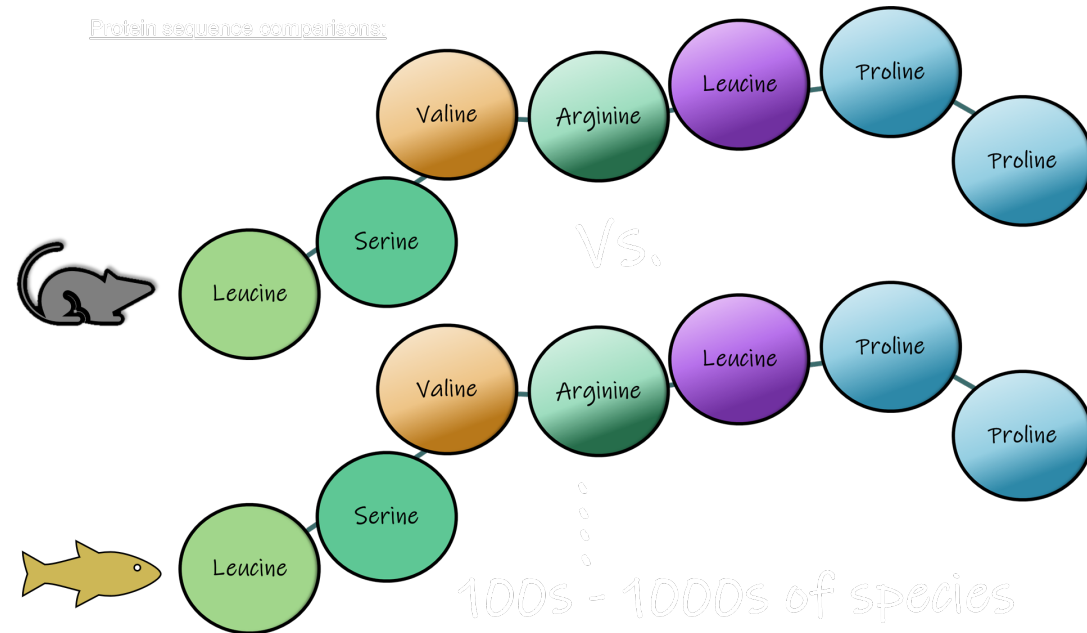
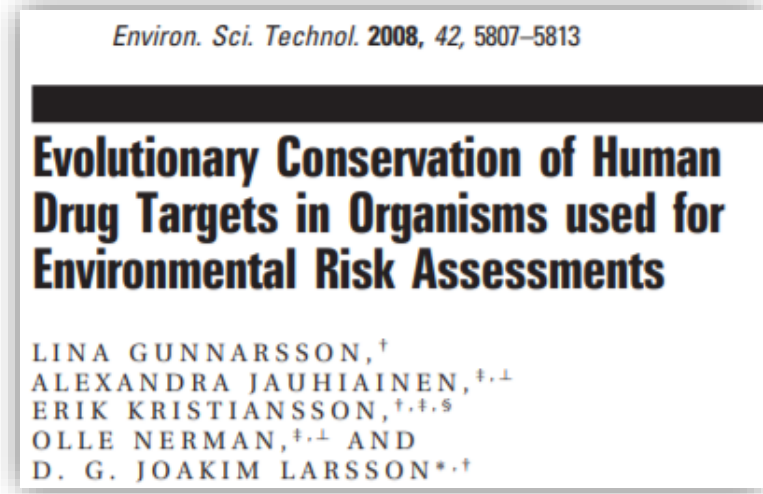
Human  
Protein Structure



## Bioinformatics Toolbox:

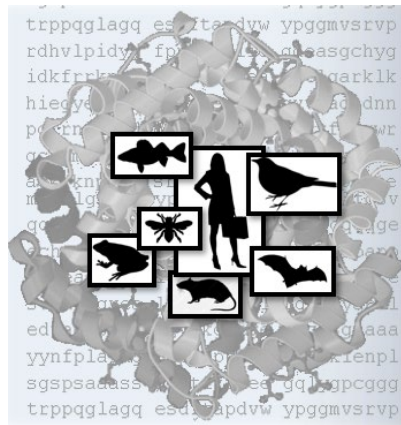
Molecular modeling  
Molecular docking  
Virtual screening  
Molecular dynamic simulations  
Functional prediction

# How to begin using bioinformatics for species extrapolation?



- Start simple, include target audiences and always look to future state of the science during development (prepare for what's to come)



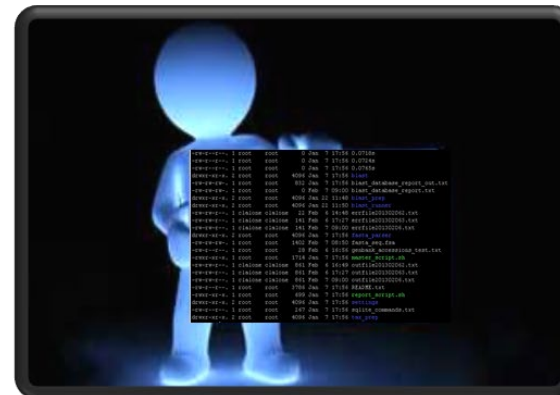
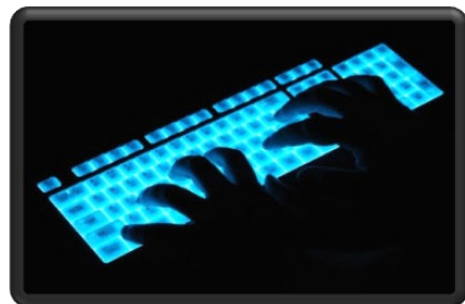


## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): A Web-Based Tool for Addressing the Challenges of Cross-Species Extrapolation of Chemical Toxicity

Carlie A. LaLone,<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>

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)

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



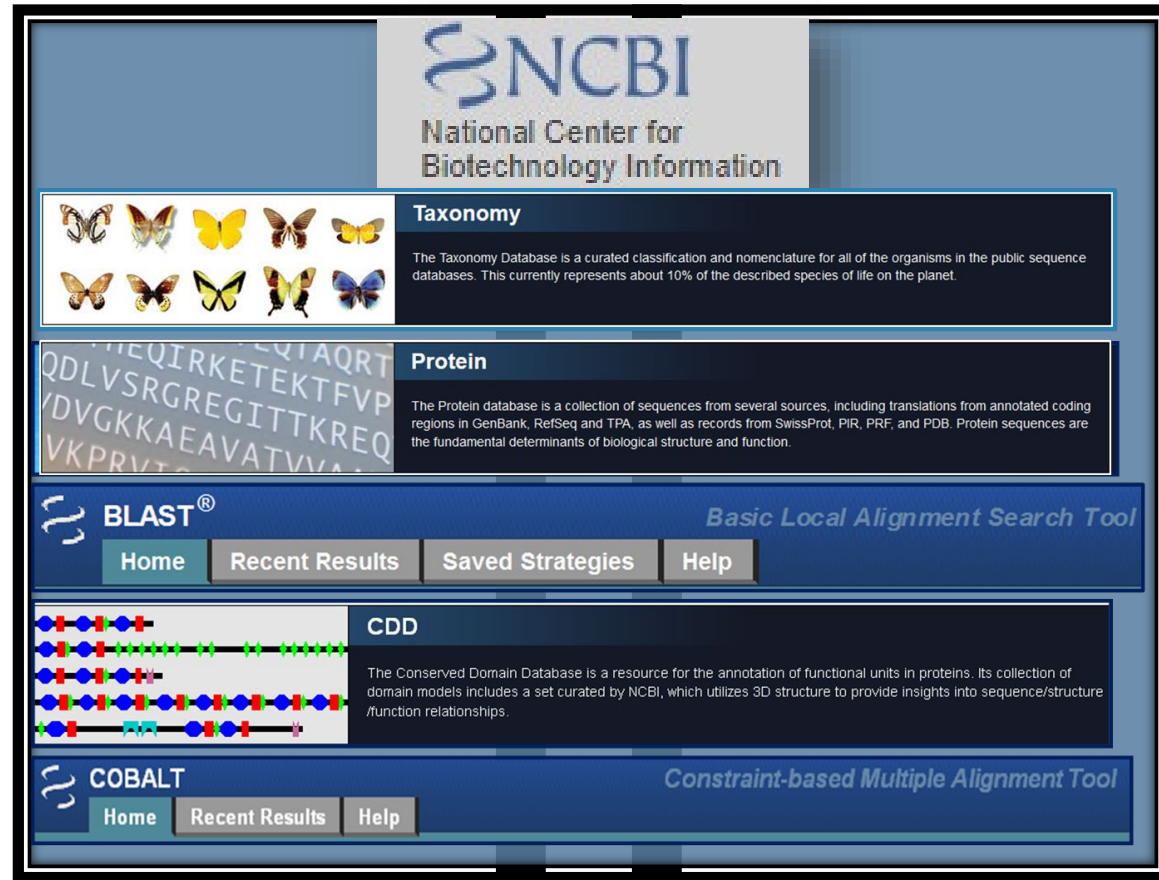


Animation by: Miguel Moravec (EPA CSS) & Andrew Patterson



SeqAPASS

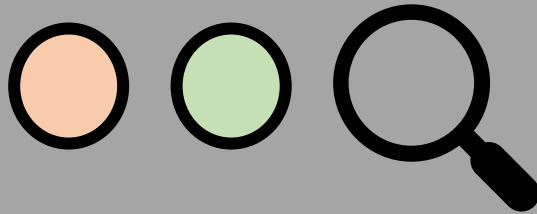
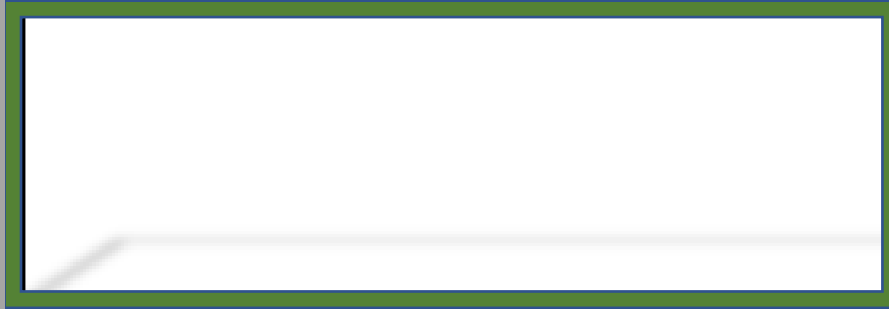




# SeqAPASS

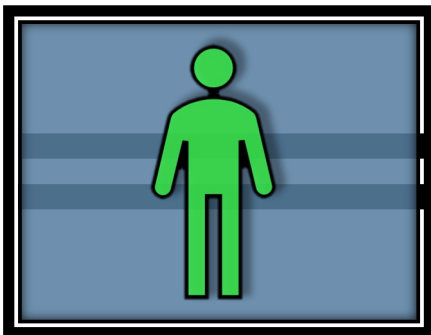
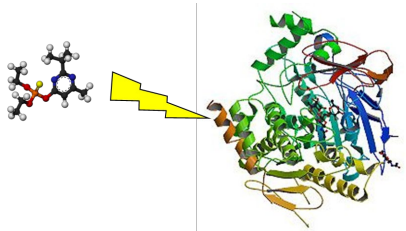
# SeqAPASS

## Level 1



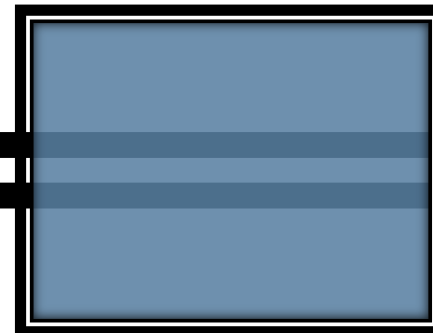
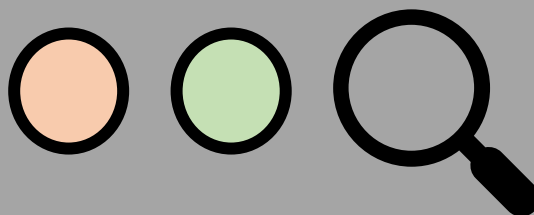


Human Protein Target

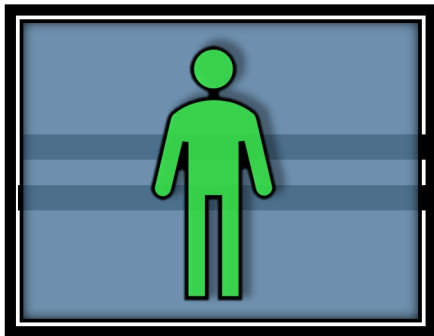
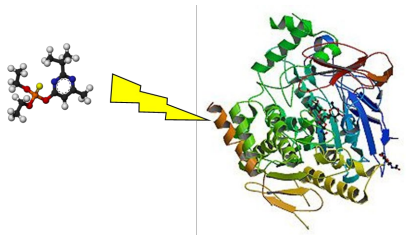


# SeqAPASS

## Level 1

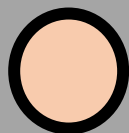
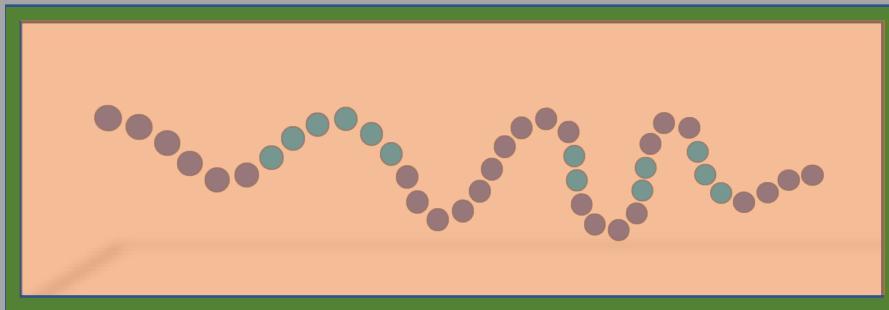


Human Protein Target



# SeqAPASS

## Level 1



Yes

Line of Evidence:

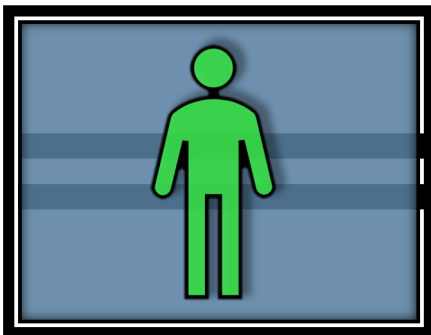
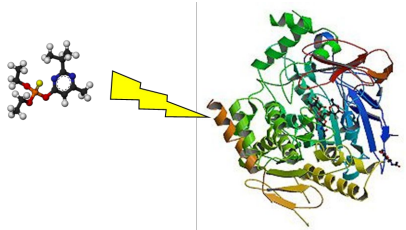
Primary amino acid sequence

Conserved



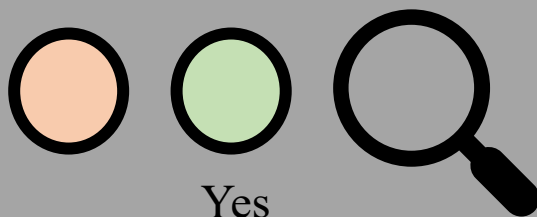
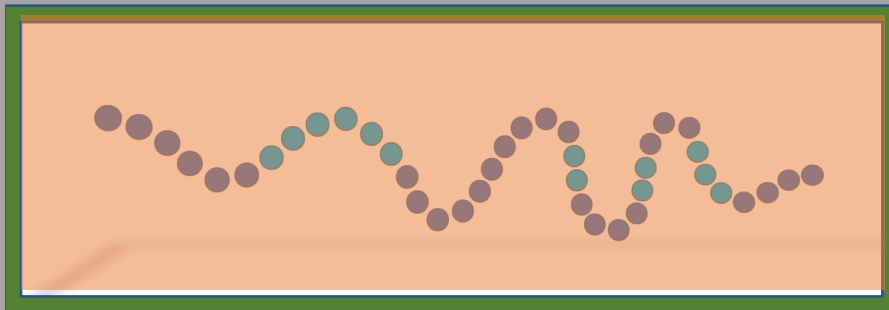
Percent similarity

Human Protein Target

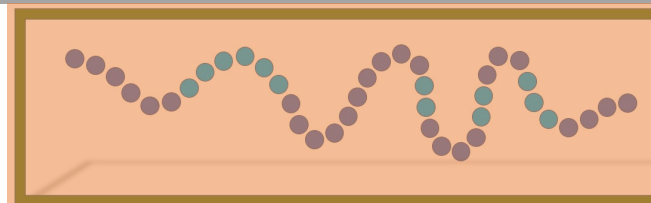
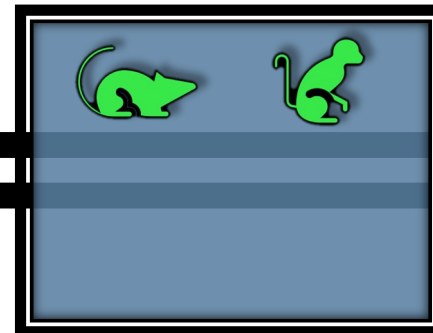


# SeqAPASS

## Level 1

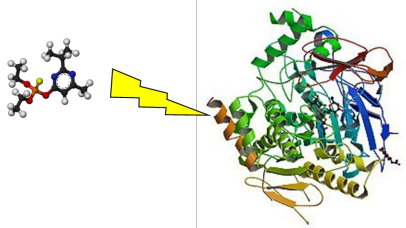


Line of Evidence:  
Primary amino acid sequence  
Conserved





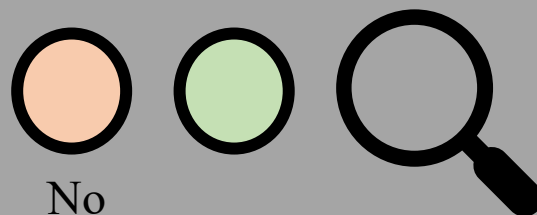
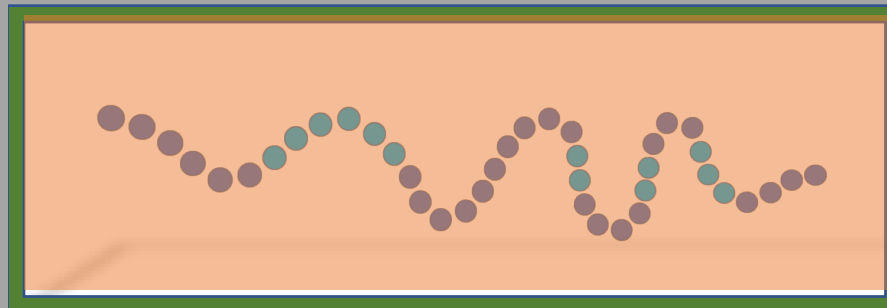
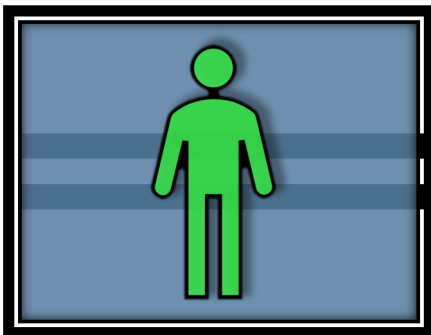
Human Protein Target



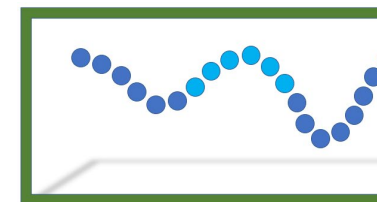
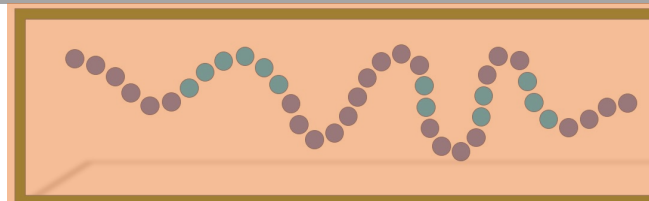
# SeqAPASS

## Level 1

Line of Evidence:  
Primary amino acid sequence  
Conserved

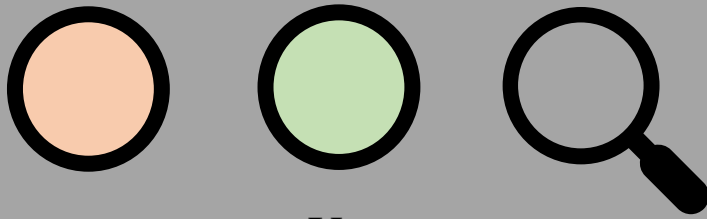


No



# SeqAPASS Level 1

## Ortholog Candidate Identification

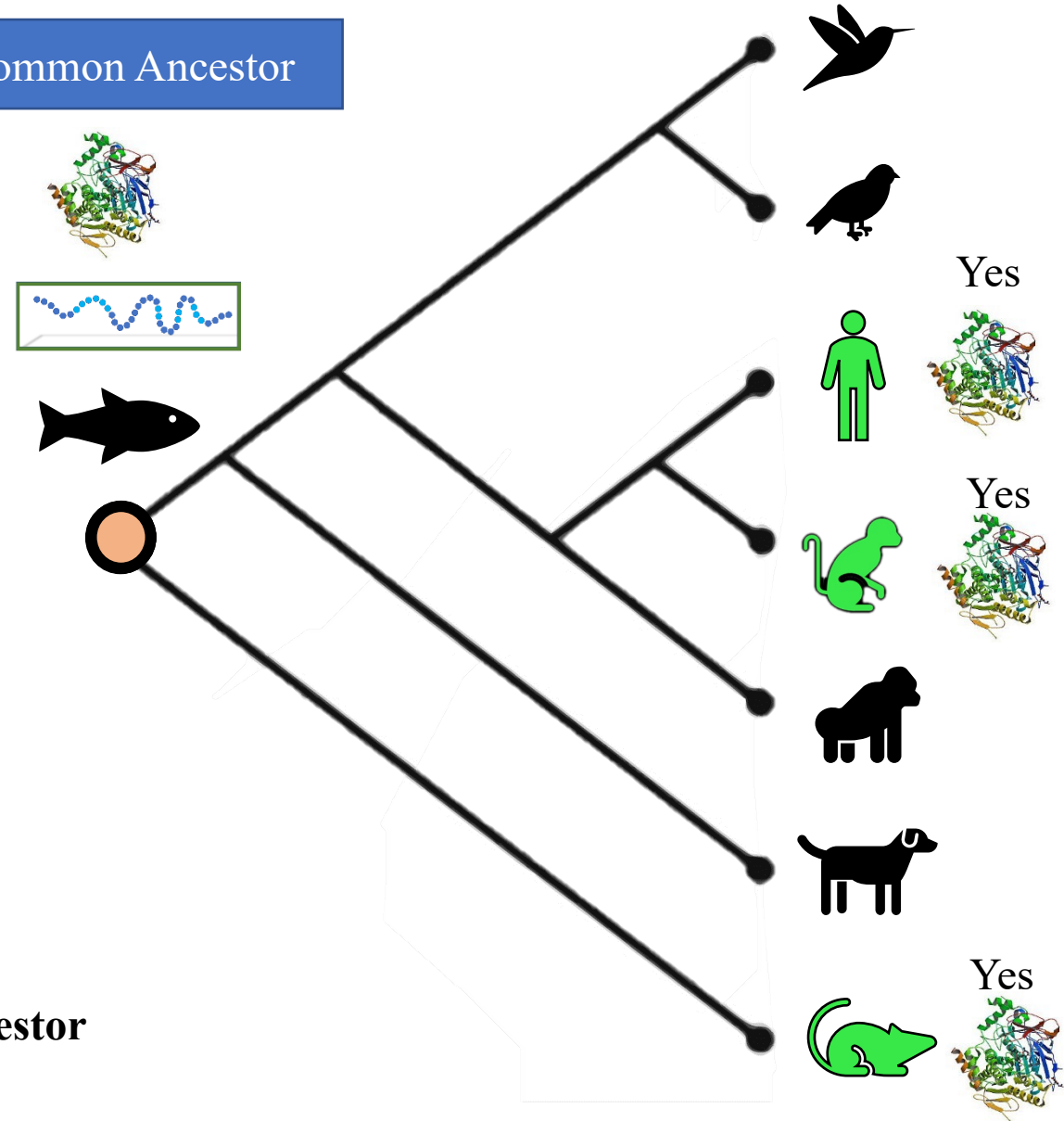


Yes

Proteins in different species that evolved from a common ancestor

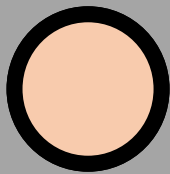
Typically maintain similar function

Common Ancestor

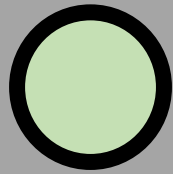


# SeqAPASS Level 1

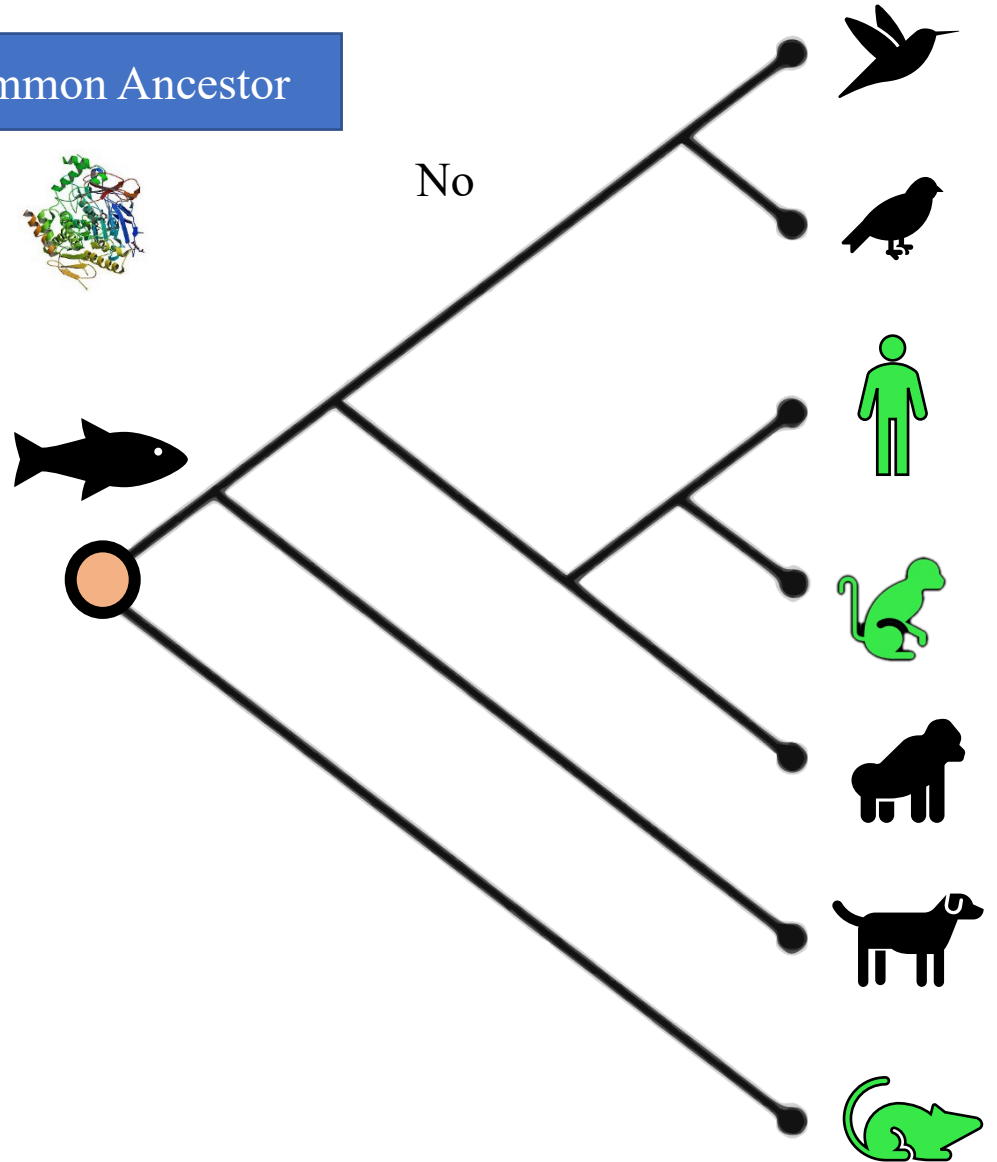
## Ortholog Candidate Identification



No




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



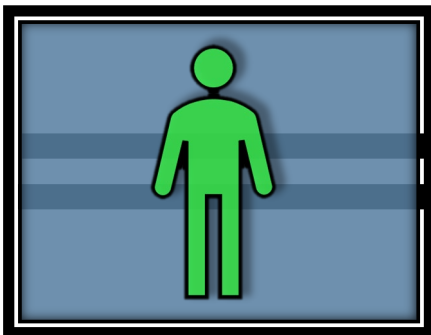
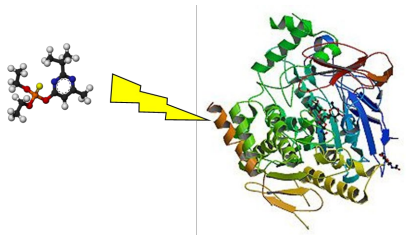
Lowest % Similarity that is still an ortholog

## Example:

Susceptibility Cut-off: Set at 33.15

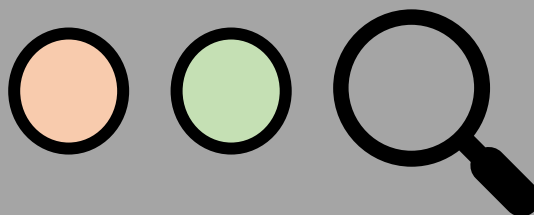
Above cut-off: More likely to be susceptible base on similar **FUNCTION**

Human Protein Target



# SeqAPASS

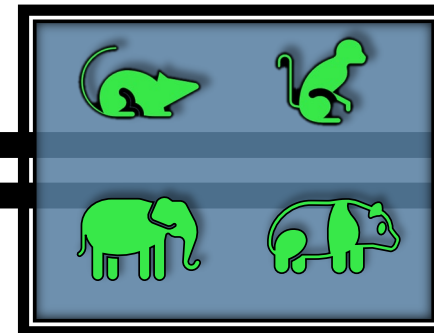
## Level 1



Line of Evidence:

Primary amino acid sequence

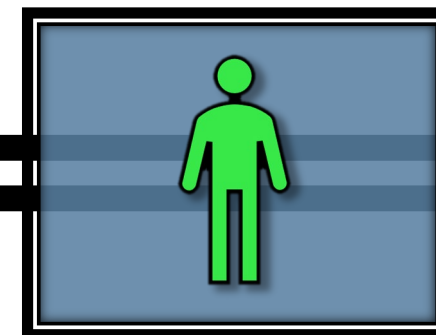
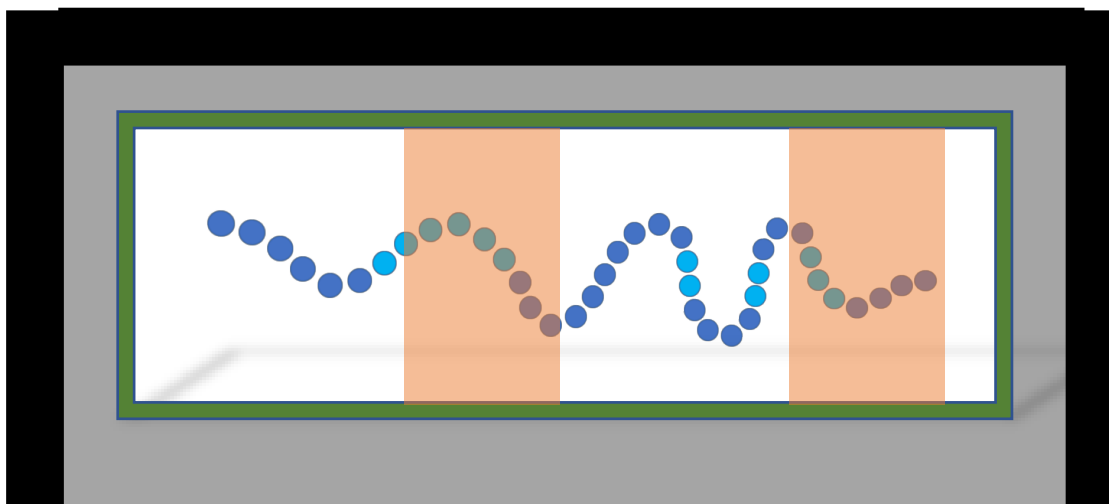
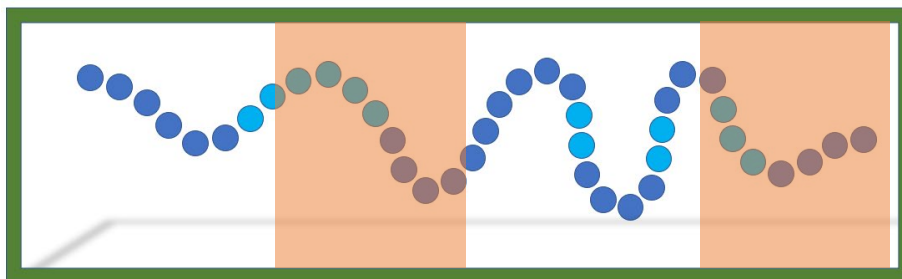
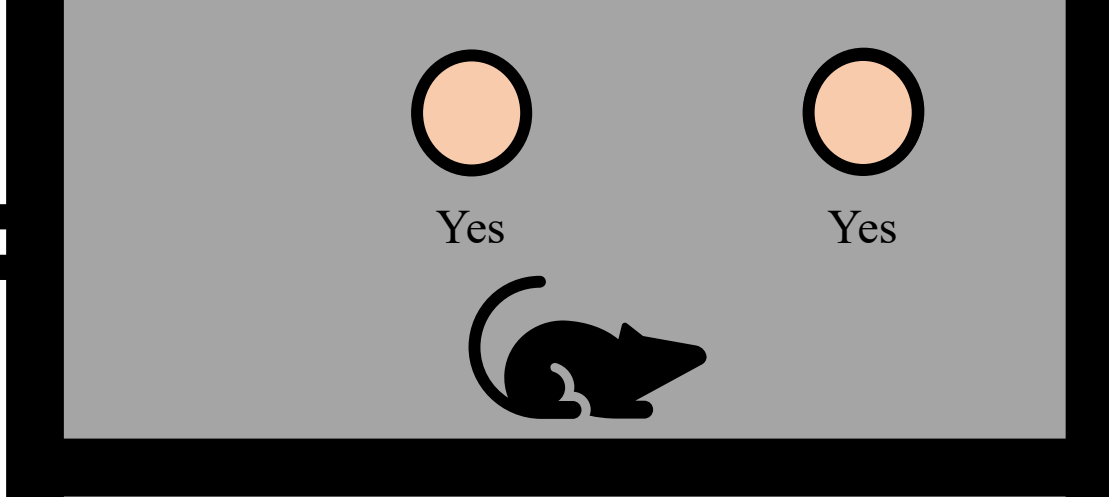
Conserved



Hundreds to Thousands of Species

# Level 2

Line of Evidence:  
Domain  
Conserved

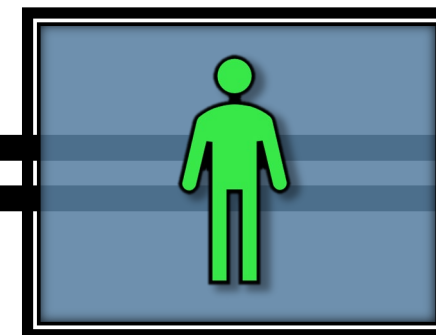
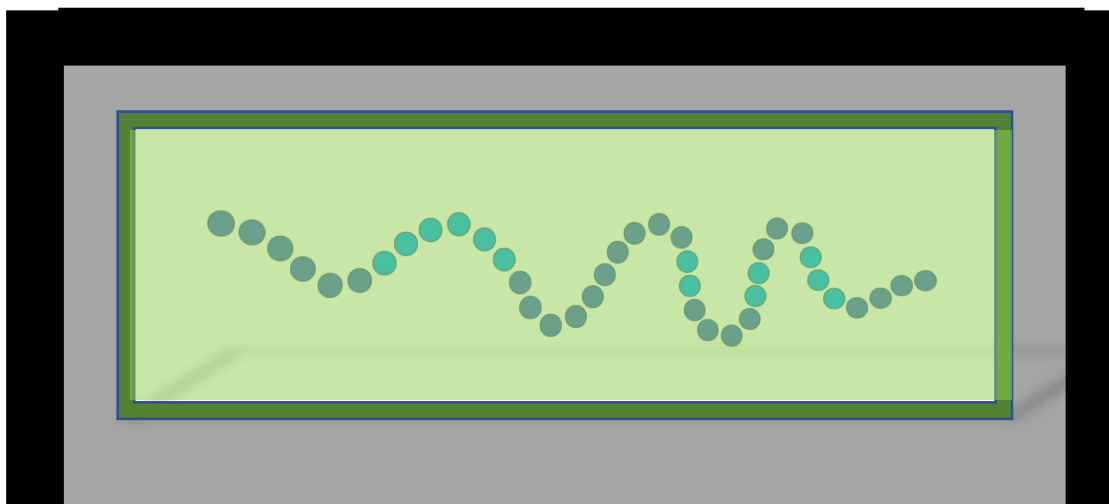
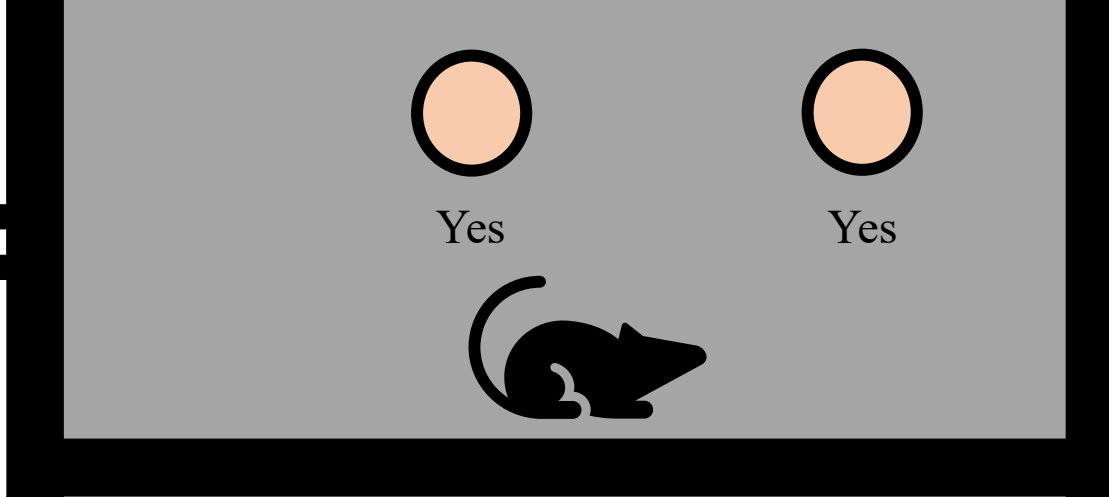


Human Functional Domain(s)





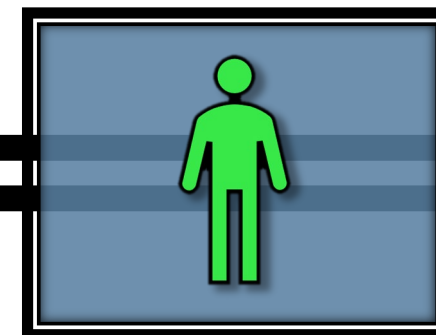
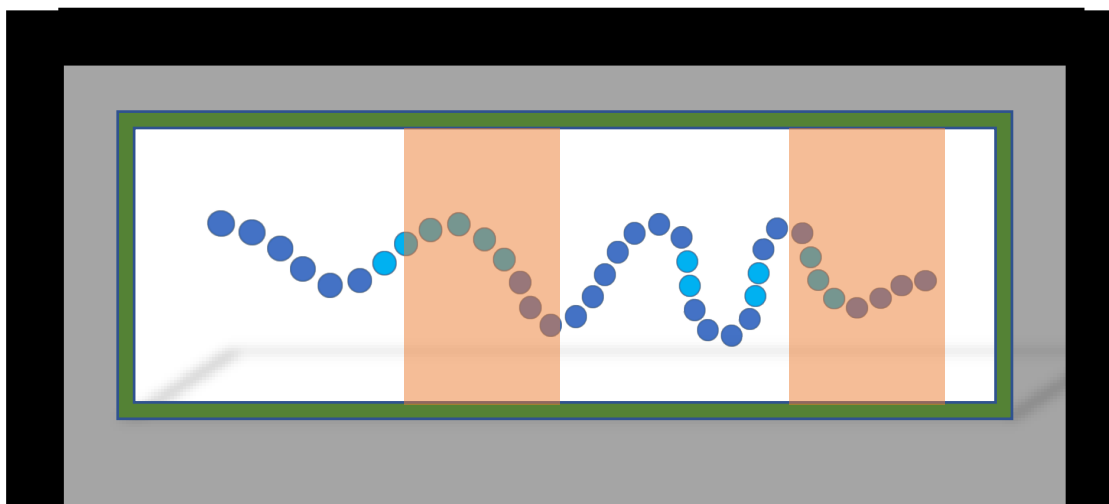
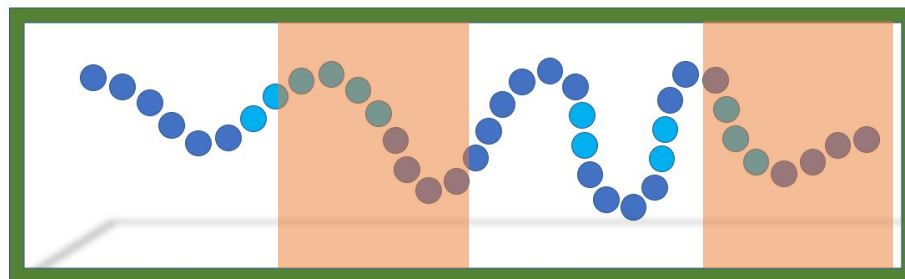
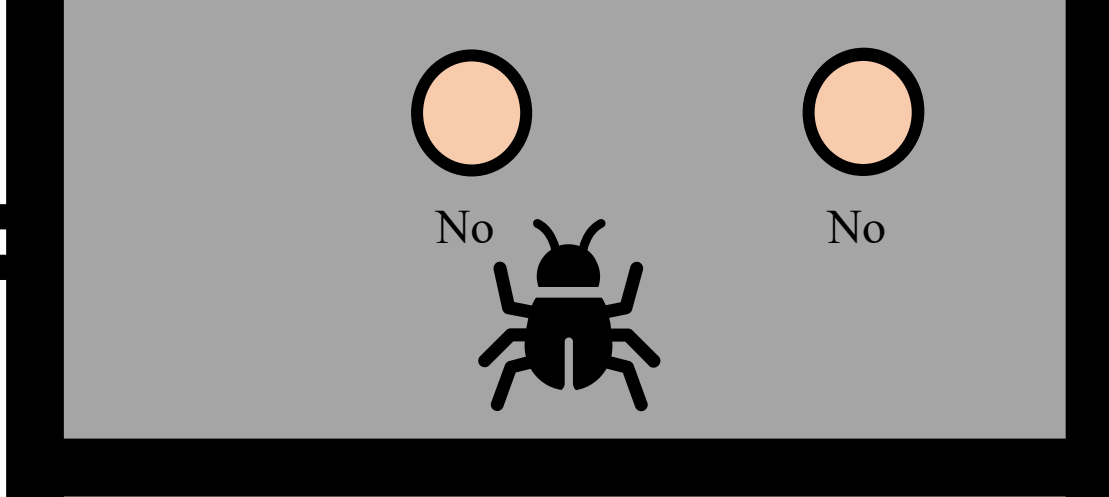
Line of Evidence:  
Domain  
Conserved



Human Functional Domain(s)

# Level 2

Line of Evidence:  
Domain  
Conserved



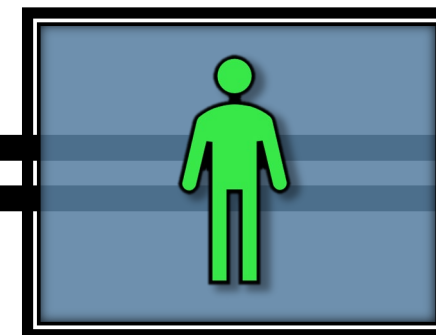
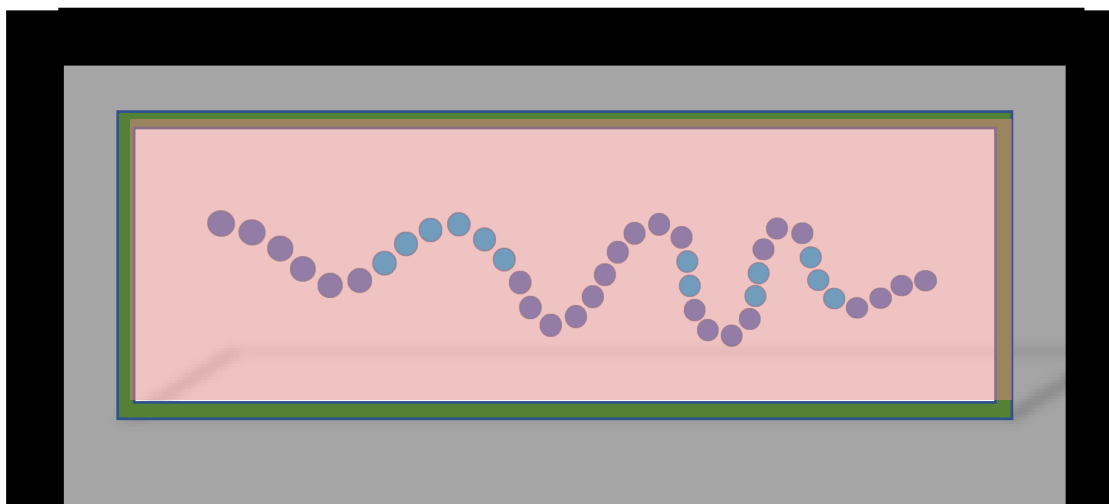
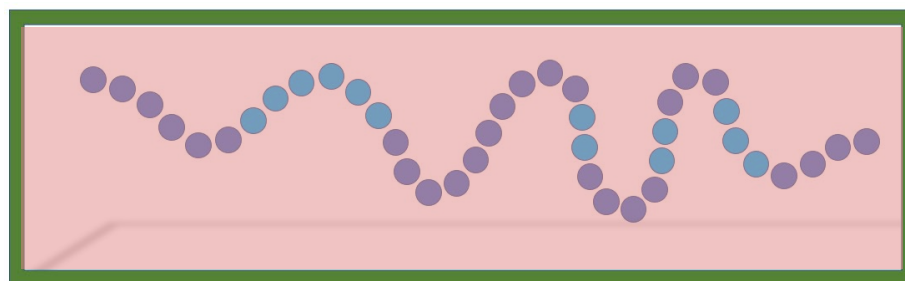
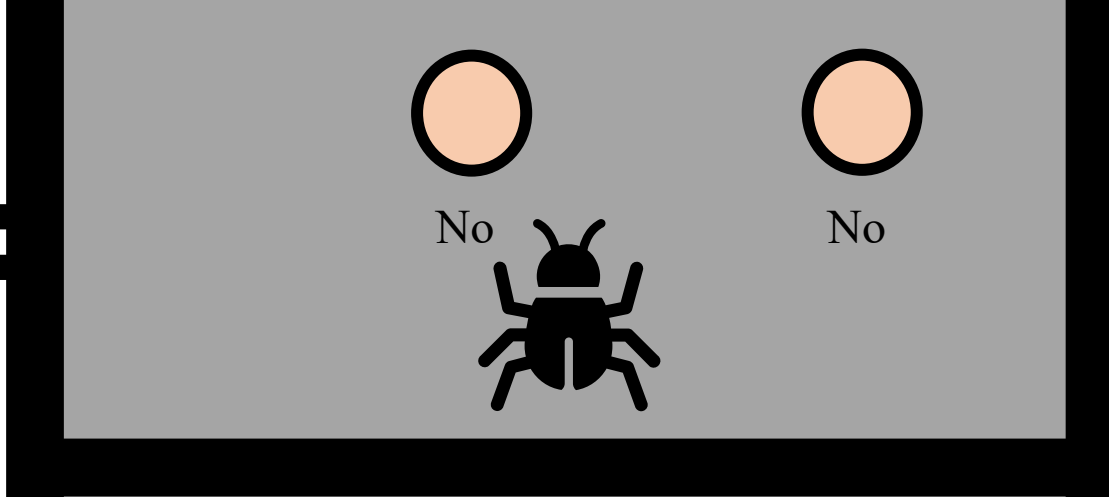
Human Functional Domain(s)

Level  
2

Line of Evidence:

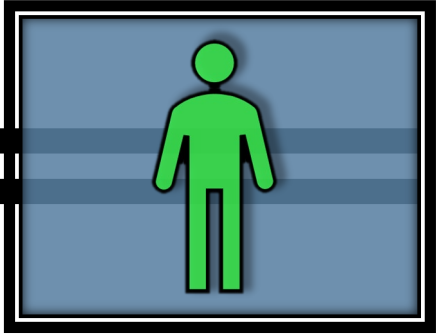
Domain

Not Conserved

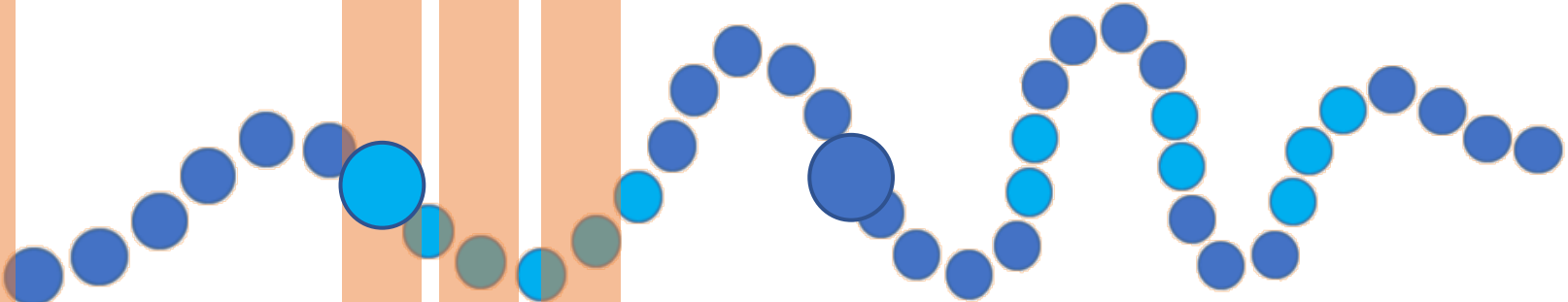
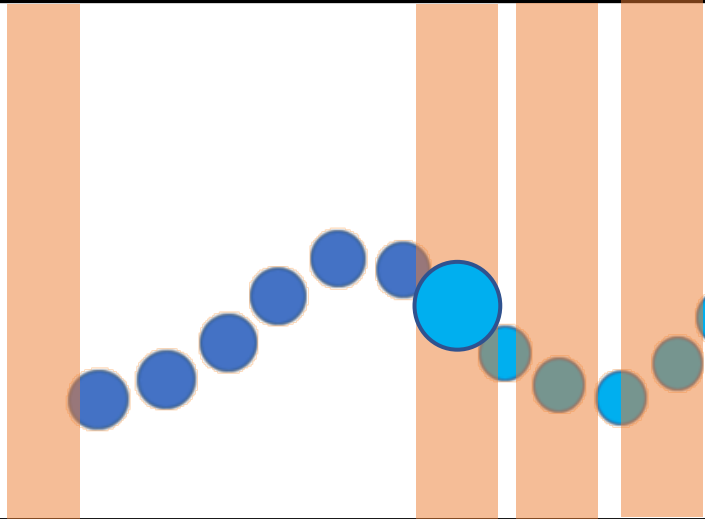
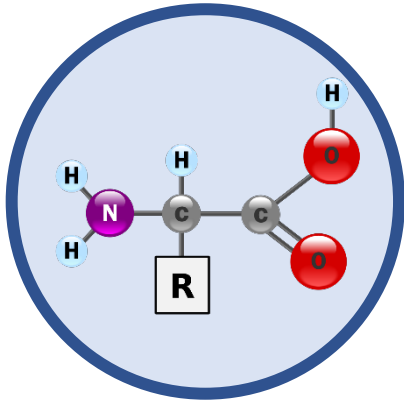
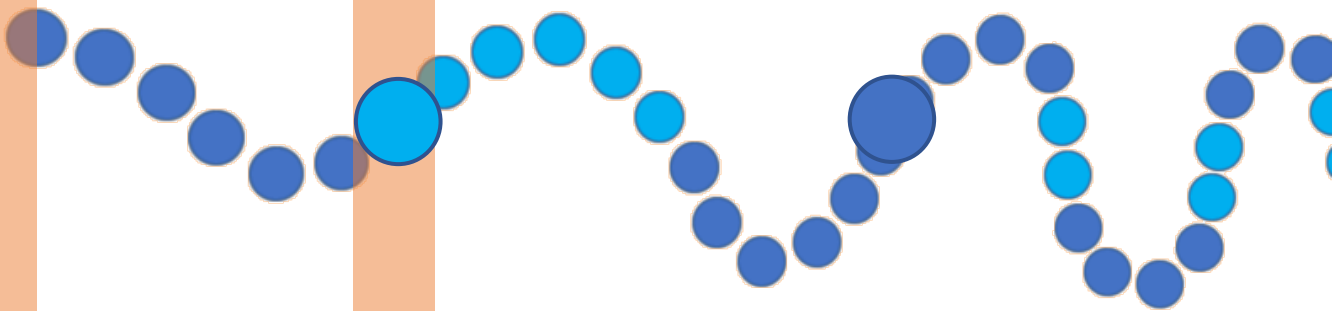
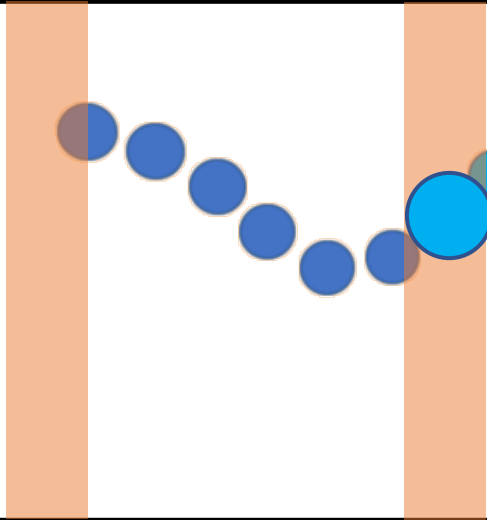
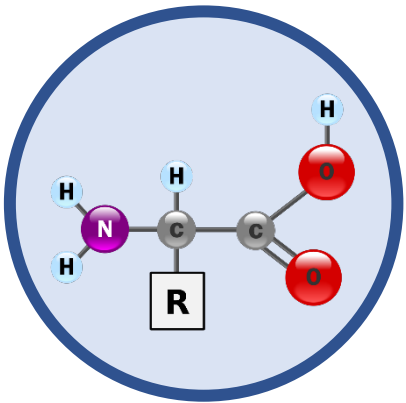


Human Functional Domain(s)



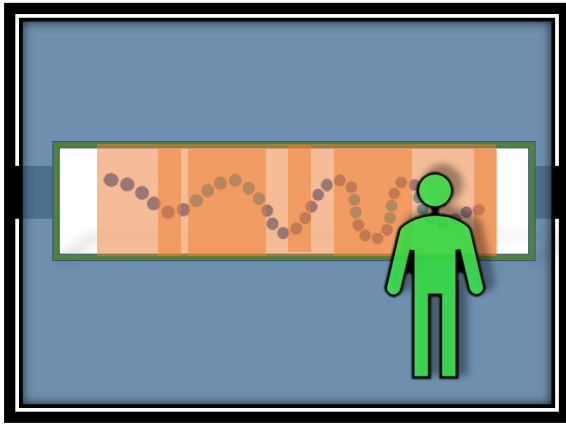


Human Critical Amino Acids



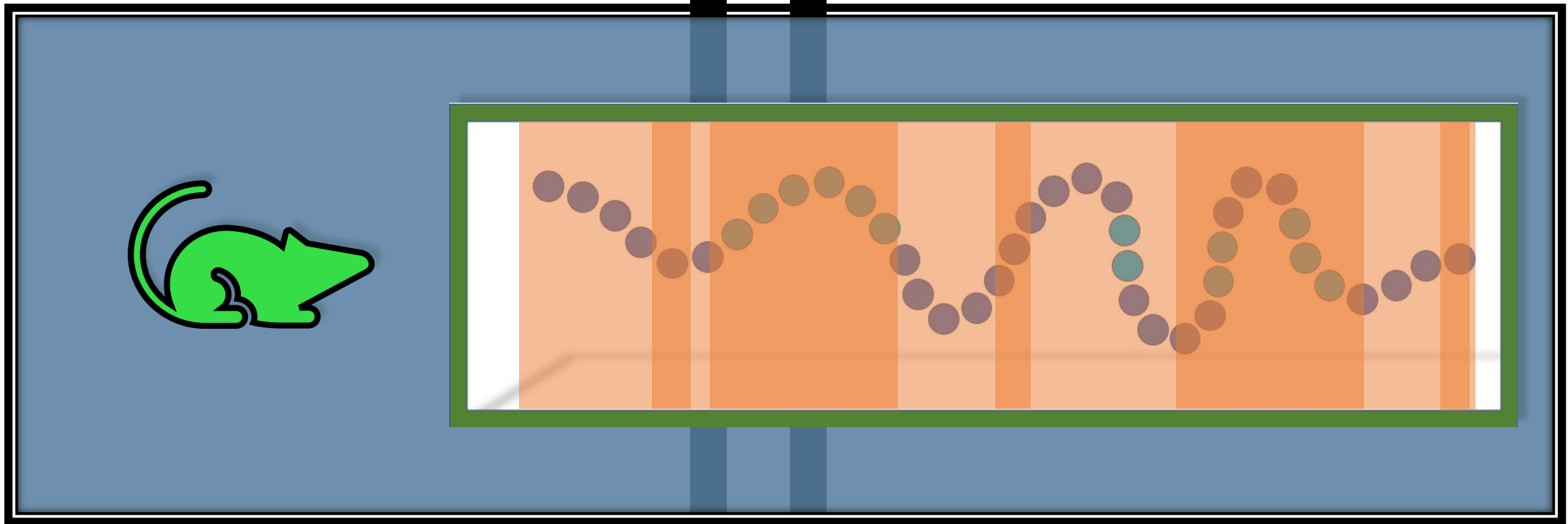
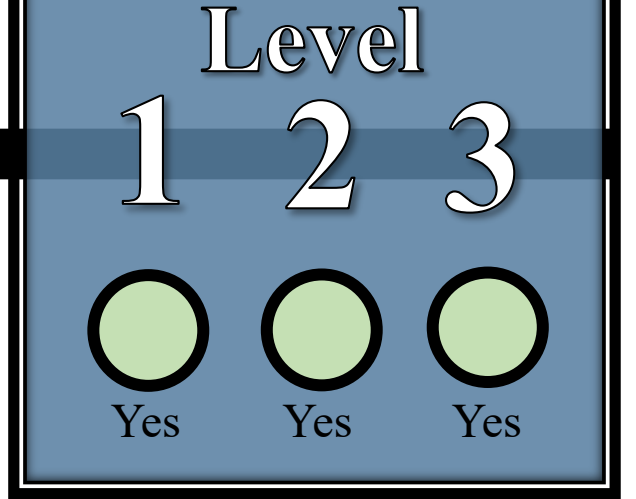
Line of Evidence: Conserved





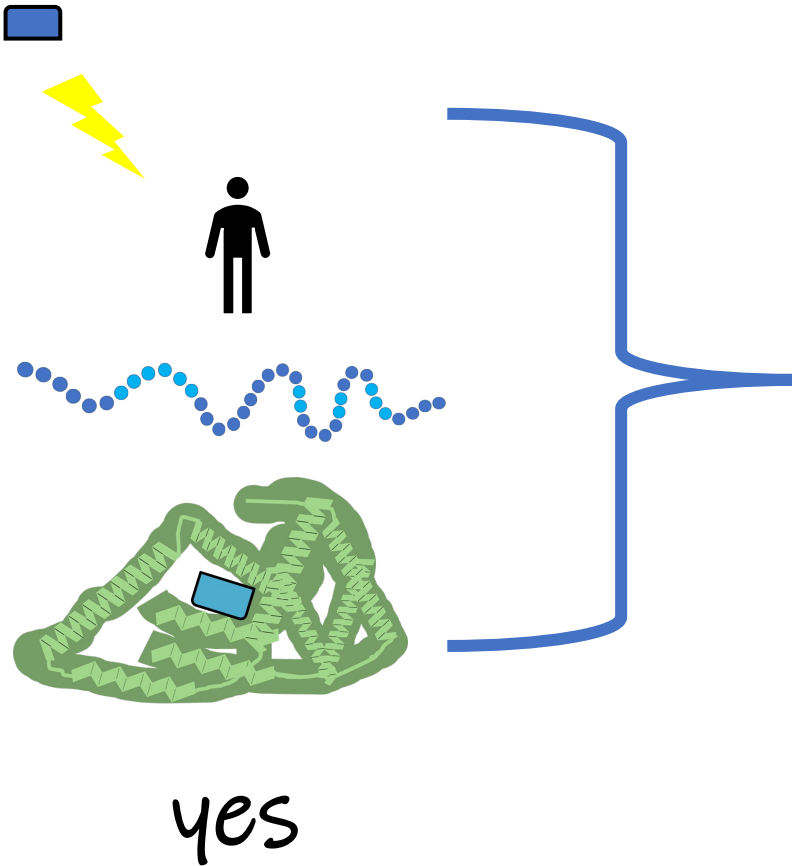
# SeqAPASS











## 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:



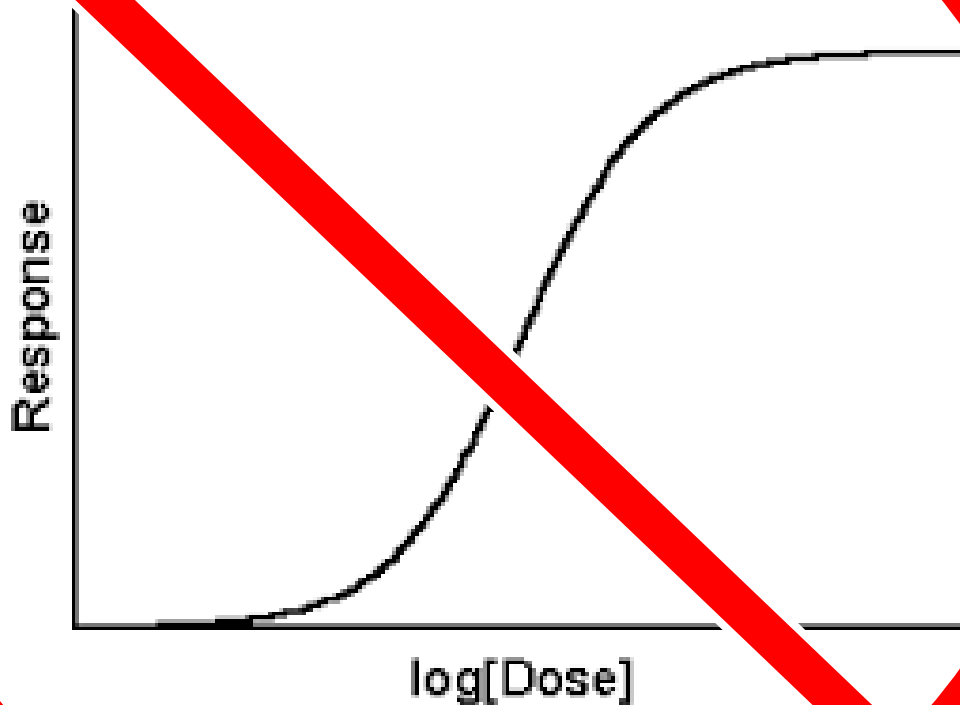
	yes
	yes
	yes
	yes
	yes
	yes
	yes
	no
	yes
	no

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.

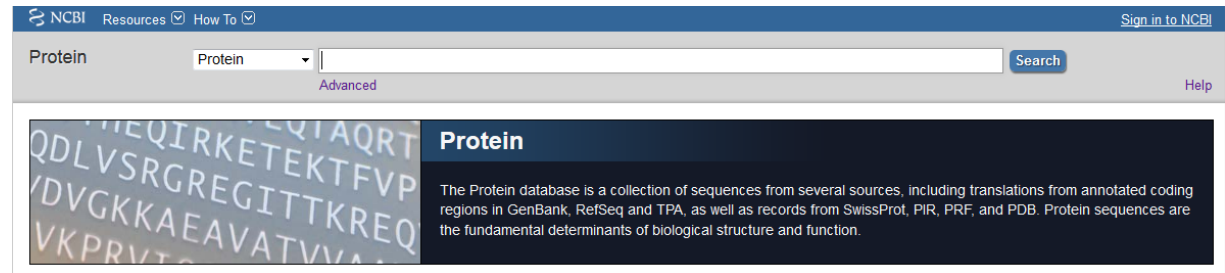


# Strengths of SeqAPASS

## New tools and technologies

- Improved sequencing technologies
- Large databases of sequence data

**NCBI: 197,232,209 Proteins representing 108,257 Organisms**



- 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

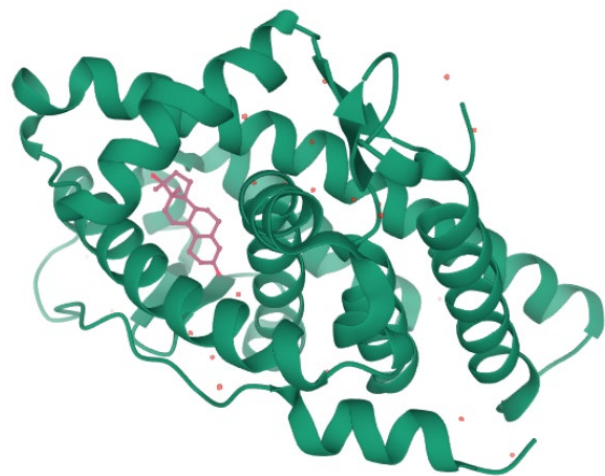
Advances in Bioinformatics –  
Future of SeqAPASS

# Always Look Several Steps Ahead



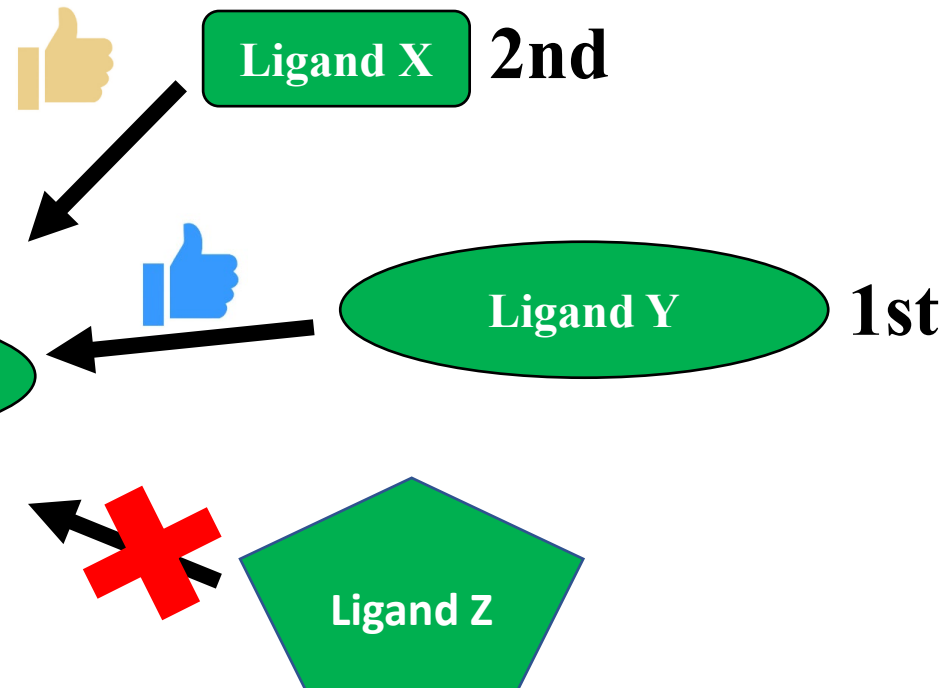
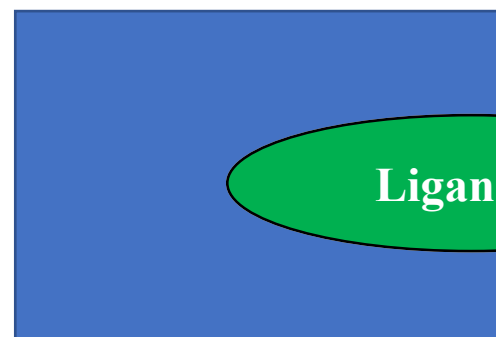
# Advances in Drug Discovery/Development

(COVID-19 has led to advances)



Structure derived  
from X-ray  
crystallography

Human  
Protein Structure



## Bioinformatics Toolbox:

Molecular modeling

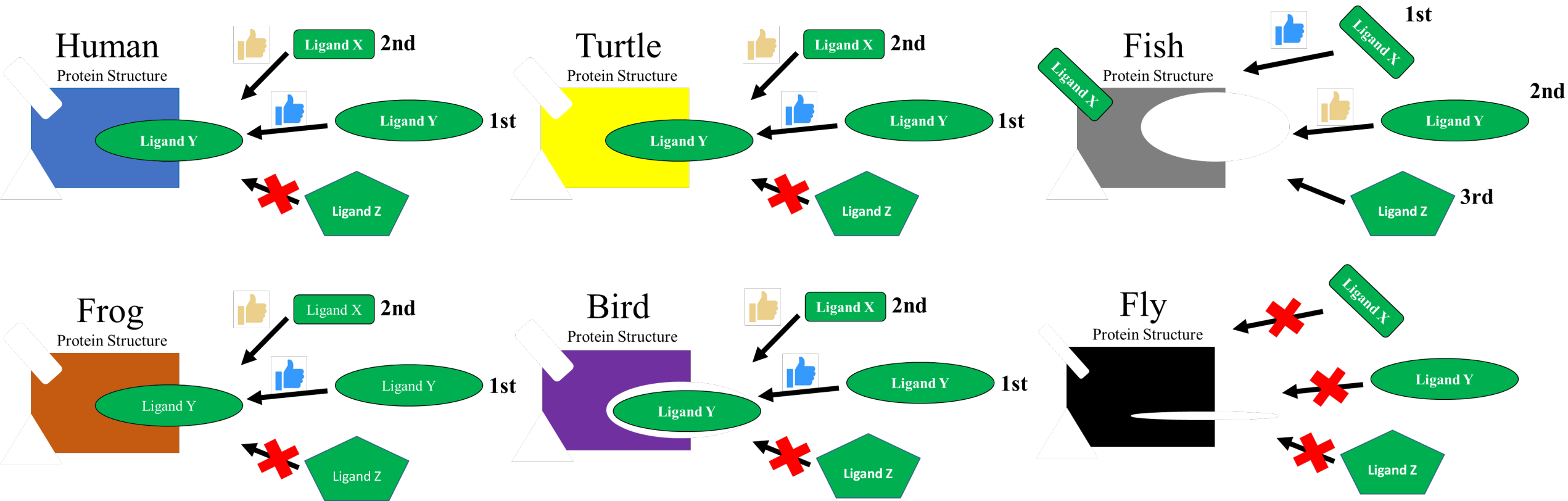
Molecular docking

Virtual screening

Molecular dynamic simulations



# Application to Species Extrapolation



## Bioinformatics Toolbox:

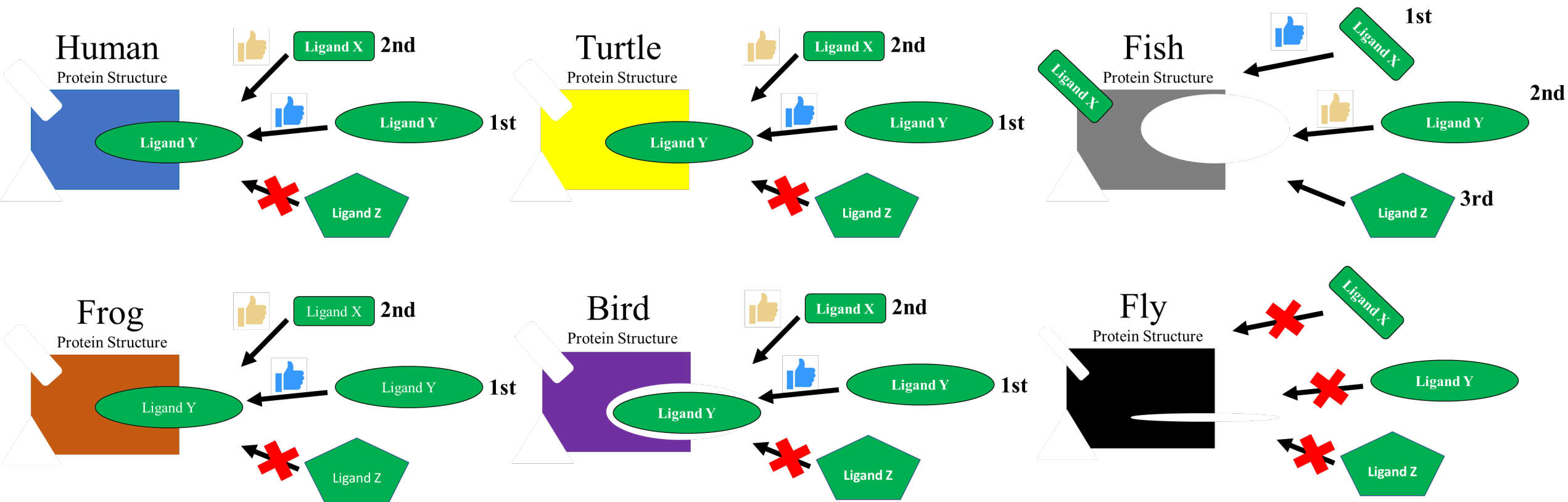
Molecular modeling

Molecular docking

Virtual screening

Molecular dynamic simulations

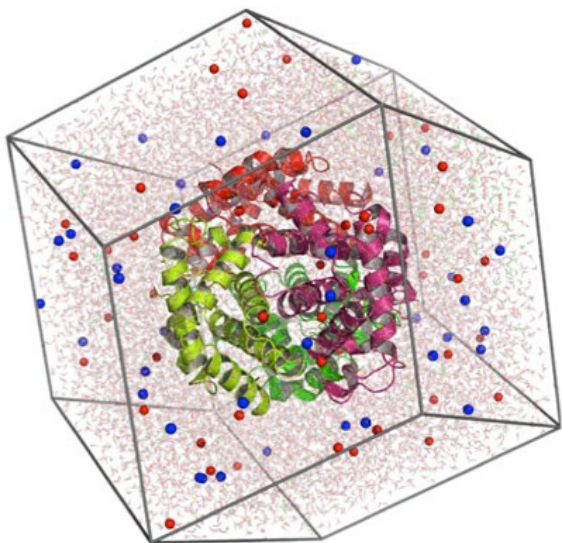
# Application to Species Extrapolation



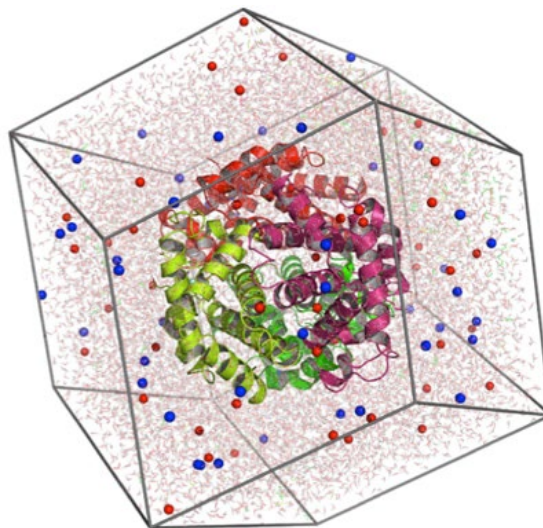
**Bioinformatics Toolbox:**  
Molecular modeling  
Molecular docking  
Virtual screening  
Molecular dynamic simulations

*Thousands/Millions/Billions  
of  
Chemicals*

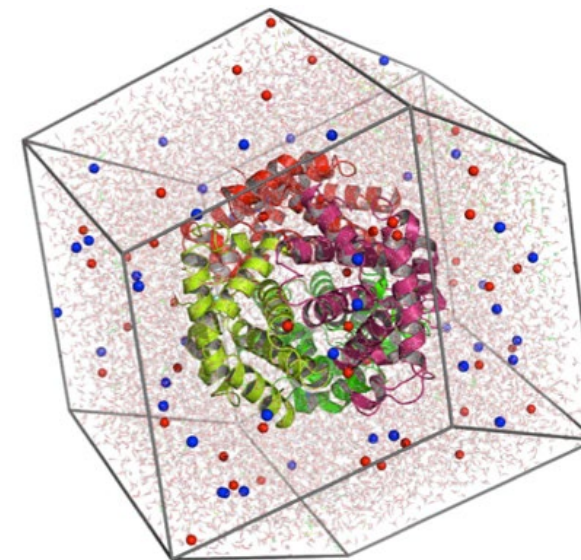
# Application to Species Extrapolation



Human = free energy ( $\Delta G$ )



Fish = free energy ( $\Delta G$ )



Chicken = free energy ( $\Delta G$ )

- Positions of all the atoms in a biomolecular system (e.g., a protein surrounded by water and perhaps a lipid bilayer)
- Calculate the force exerted on each atom by all the other atoms
- Newton's laws of motion to predict the spatial position of each atom as a function of time
- Three-dimensional movie that describes the atomic-level configuration of the system at every point during the simulated time interval
- Provide substantially more accurate estimates of ligand binding affinities (free energies) than other computational approaches such as docking

## Bioinformatics Toolbox:

Molecular modeling

Molecular docking

Virtual screening

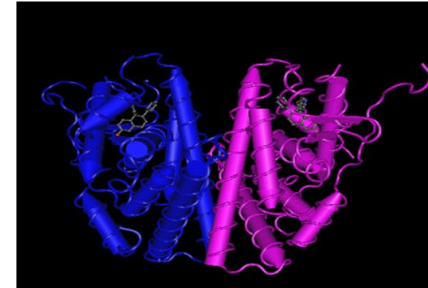
Molecular dynamic simulations

# Sequence

```
MTMTLHTKASGMALLHQIQGNELEPLNRPQLKIPLERPLGE
VYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYGTGLPYG
PGSEAAAFSGNSLGGFPPLNSVSPSPLMLLHPPQLSPFLQ
PHGQQVPYYLENEPSGYTVREAGPPAFYRPNSDNRRQGGR
ERLASTNDKSGMAMESAKETRYCAVCNDYASGYHYGVWSC
EGCKAFFKRSIQGHNDYMCPTNQCTIDKNRRKSCQACRLR
KCYEVGMMKGGIRKDRRGGRLMKHRQRDDGEGRGEVG
SAGDMRAANLWPSPLMIKRSKKNLSLSTADQMVSALLA
EPPILYSEYDTPRPFSEASMMGLLTNLADRELHMINWAKV
PGFVDLTLDQVHLLCAWLEILMIGLVWRSMEHPGKLLFA
PNLLDRNGQKCGVEGMVEIFDMLLATSSRFMMNLQGEFF
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  
LETMTGEKVTVVQLEGDNKLVTFKNIKSVTELN  
GDIITNTMTLGDIVFKRISKRI

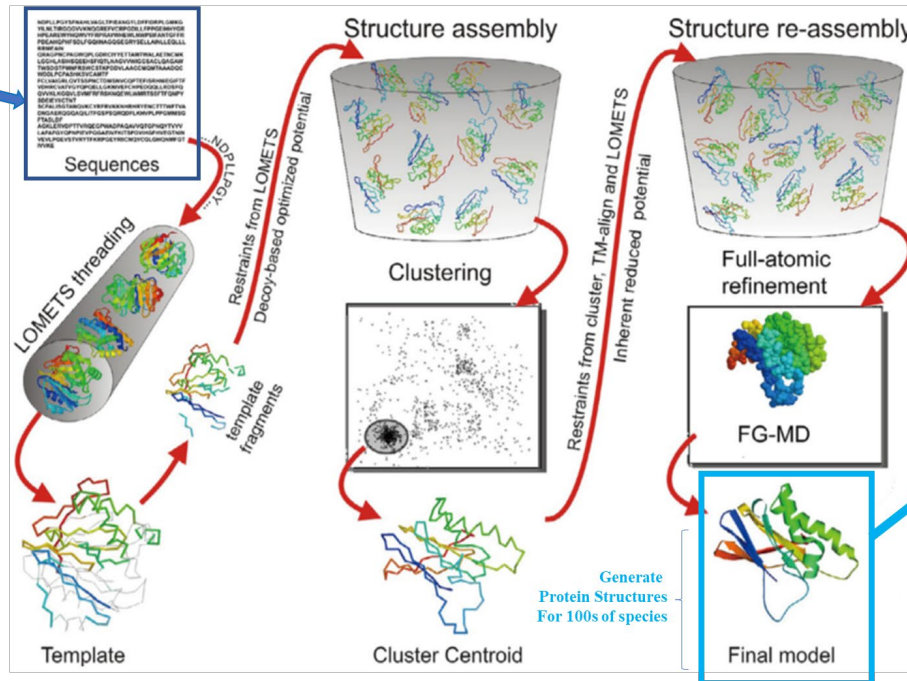
>NP\_787011.1 Protein X [Bos taurus]  
MNFSGKYQLQSQENFEAFMKAIGLPELIQKGD  
KGVSEIVQNGKHFKFTITAGSKVIQNEFTVGEECE  
MEFMTGEKIKAVVQLEGDNKLVTFKNIKSVTEFN  
GDTVSTMTKGDVVKRISKRI

>KFQ76585.1 Protein X [Phoenixopterus ruber  
ruber]  
MSFTGKYLQSQENFEAFMKAIGLPELIQKGD  
KGVSEIVQNGKHFKFTITAGSKVIQNEFTVGEECE  
MEFMTGEKIKAVVQLEGDNKLVTFKNIKSVTEFN  
GDTVSTMTKGDVVKRISKRI

>NP\_001116883.1 Protein X [Xenopus  
tropicalis]  
MAFAGKYELVHQENFEAFMKAIGLPELIQKGD  
KGVSEIVQNGKHFKFTITAGSKVIQNEFTVGEECE  
LETPTGKVKSVVQLEGDNKLVQKAITSTELSG  
DTITHVLTNNLVFKRISKRI

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/>)

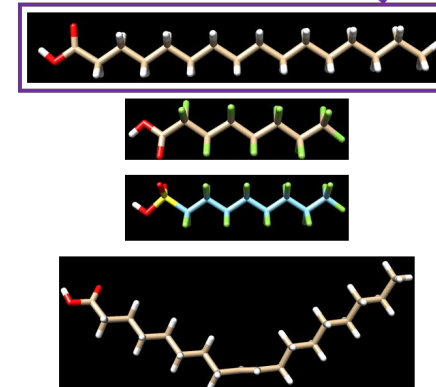


## UCSF Chimera DockPrep Structures and Minimize Ligands

### 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	RMSD Lb	RMSD ub	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





# Application of SeqAPASS

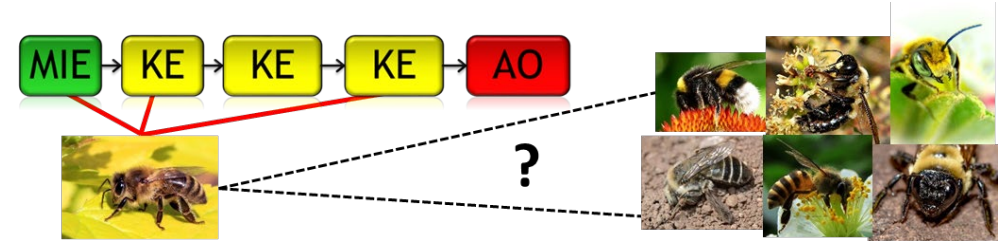




# Applications of Bioinformatics: Case Studies

- **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

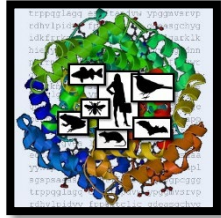
- **Prioritization strategies** Pharmaceuticals



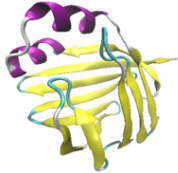
# Predict chemical bioaccumulation across species

## LFABP structures across 7 different species

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



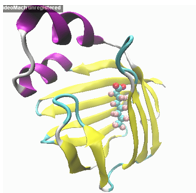
Sequence comparisons  
SeqAPASS



Homology modeling  
(Phyre2) or PDB



Molecular docking  
(Autodock vina)



Molecular dynamics  
(AMBER 14)

## 9 PFAS structures

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

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

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

# Using Bioinformatics For Species Extrapolation Can be A Big Leap From Current Practices



New things can  
be SCARY!



But  
Transformative!



And eventually  
comfortable!



# Pillars for Taking on a New Approach

TIME/CHAMPIONS



**CONFIDENCE**



**TRAINING**

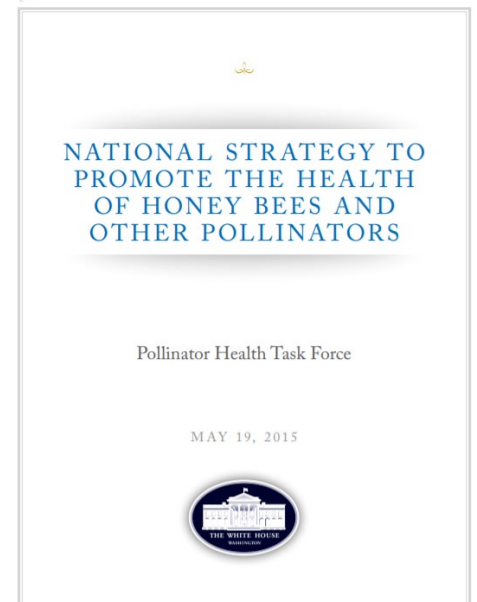
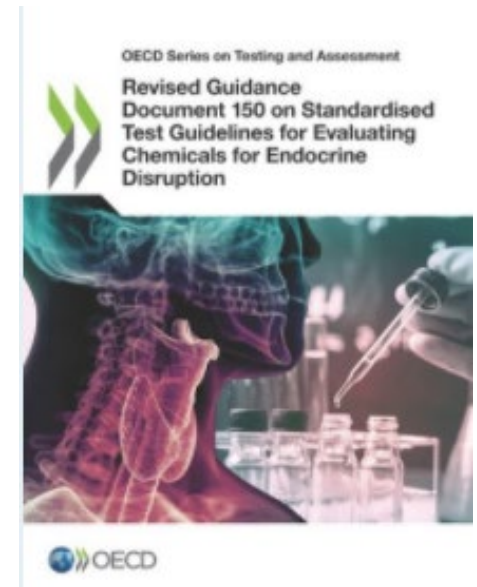


**SUPPORT**



# Bioinformatics in Species Extrapolation for Chemical Safety Evaluation

- The Organisation for Economic Co-operation and Development (OECD) published the Revised Guidance Document 150 on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption
- Appendices of the White House National Strategy to Promote the Health of Honey Bees and Other Pollinators



Government

Industry

# Consortium to Advance Cross Species Extrapolation in Regulation

## Steering Committee:

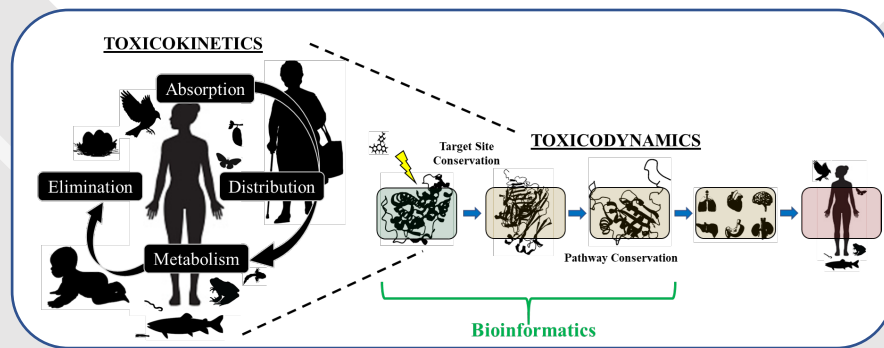
Carlie LaLone (US EPA)  
Geoff Hodges (Unilever)  
Nil Basu (McGill U)  
Steve Edwards (RTI)  
Fiona Sewell (NC3Rs)  
Michelle Embry (HESI)  
Patience Browne (OECD)

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)

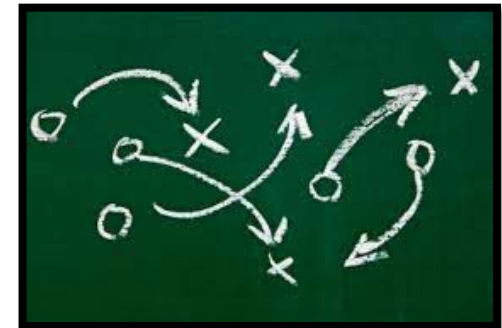
Academia

NGO



# Where do we go from here?

- Connect bioinformatics with general informatics including systematic methods
  - ECOTOX Knowledgebase
    - Olker et al.
  - Systematic literature review for WOE
    - Vliet et al.
- Advance structural evaluations (computing power and storage)
  - Move from docking to virtual screening
    - Blatz et al.
    - Mayasich et al.
- Specific laboratory studies to support bioinformatics
  - Site-directed mutagenesis
    - Mayasich et al.
    - Vliet and Cavallin et al.



# Acknowledgements

## U.S. EPA, ORD

Donovan Blatz (ORISE)

Sara Vliet (ORISE)

Sally Mayasich (ORISE)

Marissa Jensen (Univ. Minnesota Duluth)

## GDIT

Thomas Transue

Cody Simmons

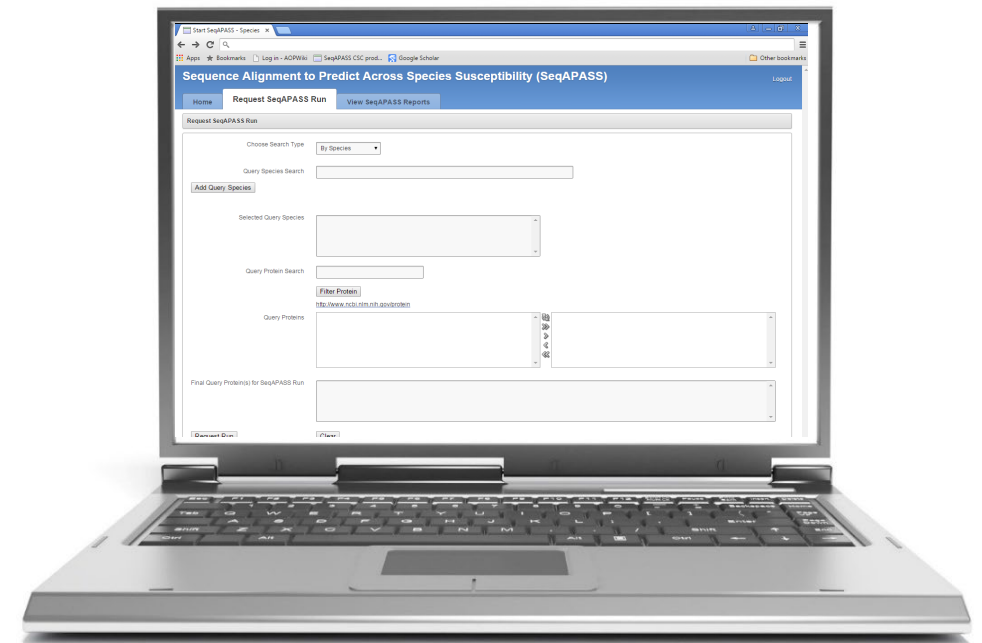
Audrey Wilkinson

## University of Pittsburgh

Carla Ng

Weixiao Cheng

SeqAPASS v6.0



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

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