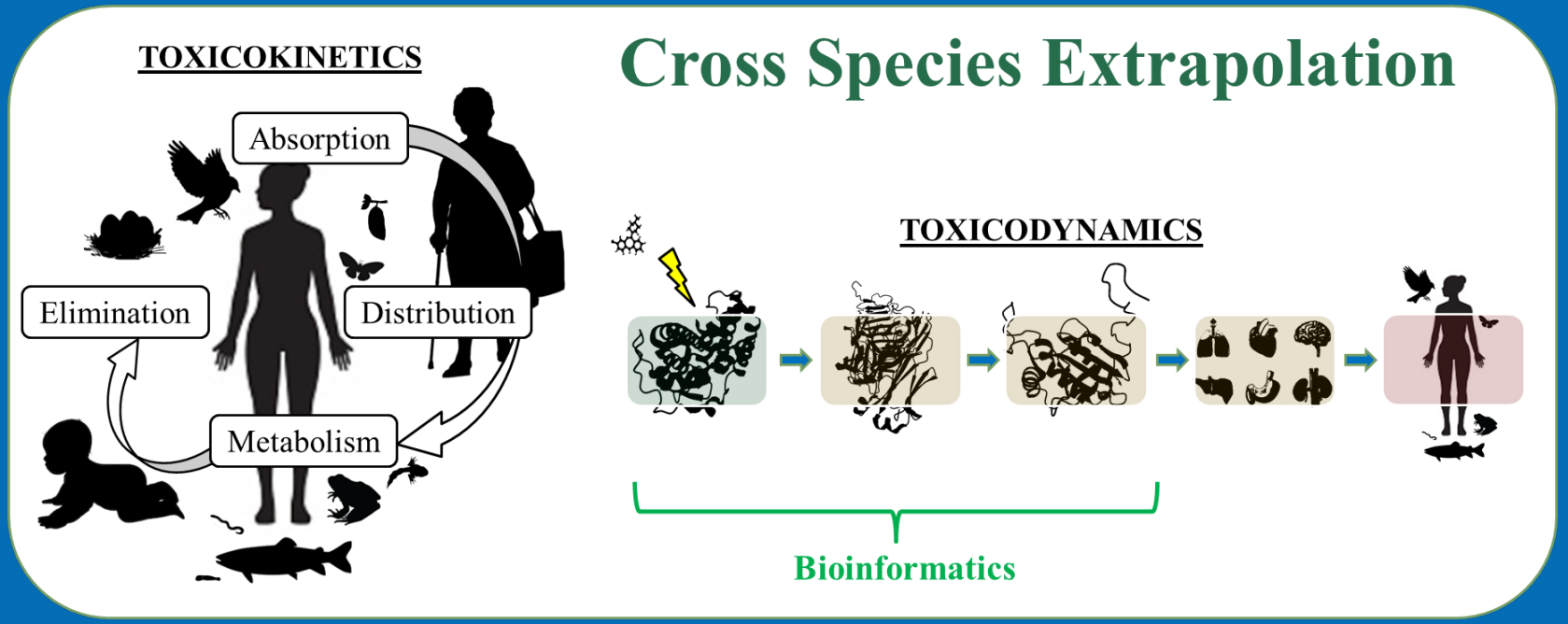
EPA
United States
Environmental Protection
Agency

Approaches and Models for Species Extrapolation

Carlie A. LaLone, Ph.D.
Research Bioinformaticist



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

February 2nd, 2021

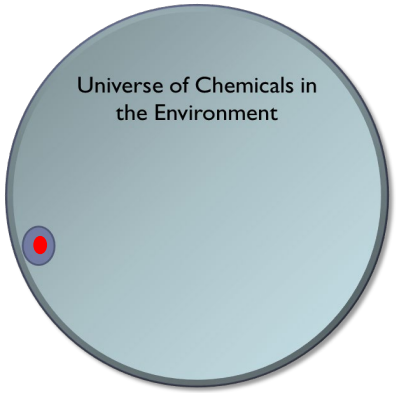

BuLi
Me
Me
C_6H_{13}

EPA United States Environmental Protection Agency

Chemical Safety Evaluation

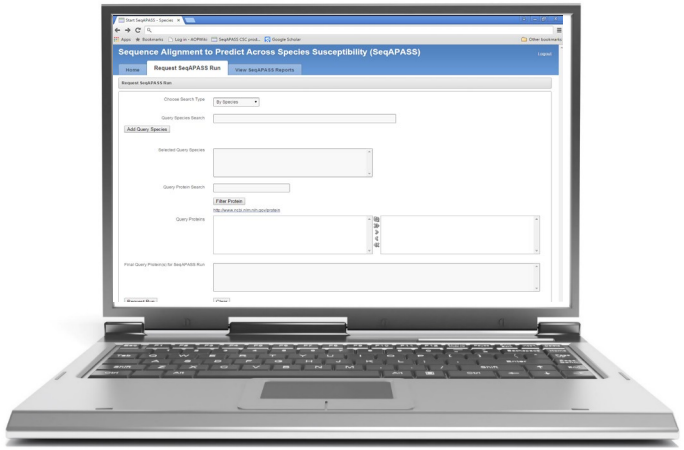
- Protect human health and the environment
  - Ensure that chemicals in the marketplace are reviewed for safety

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

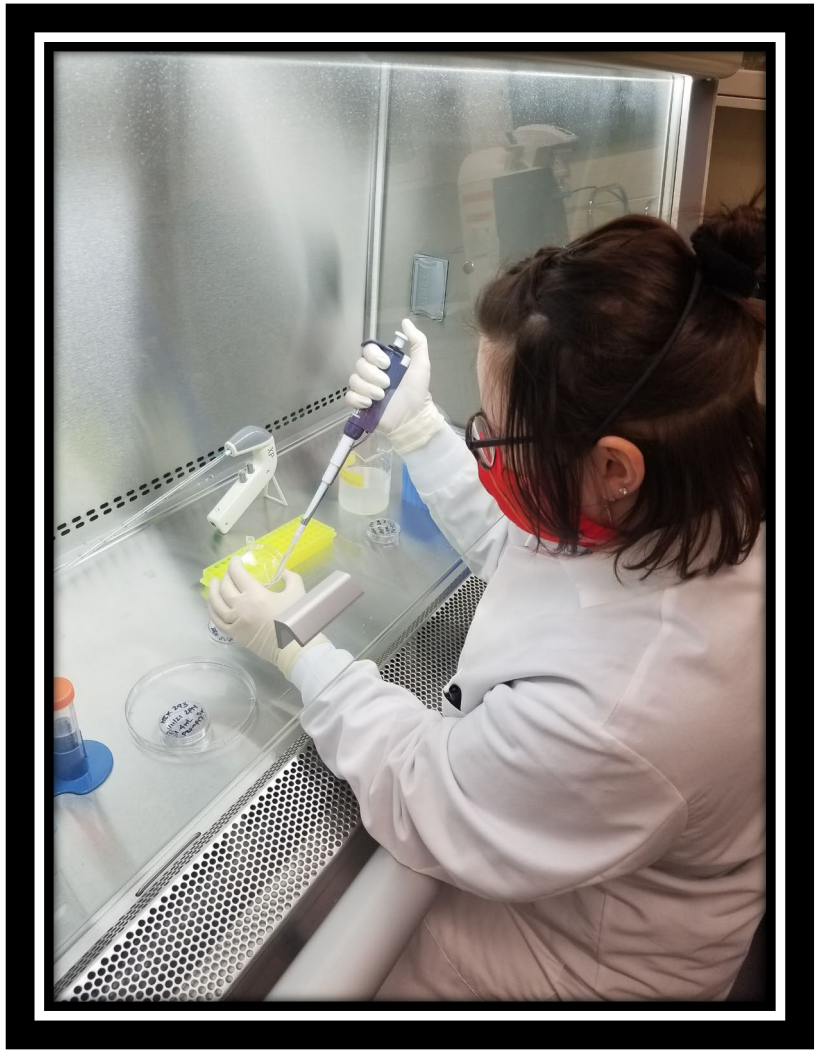


Surrogate species (model organism)

Strategic Approach to Species Extrapolation



Computational:
Bioinformatics (Session 2 Demo)
Systematic review

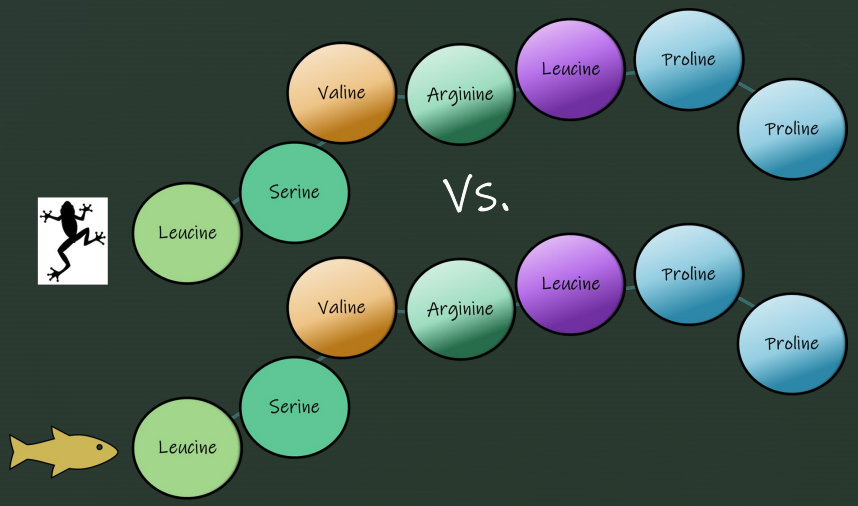


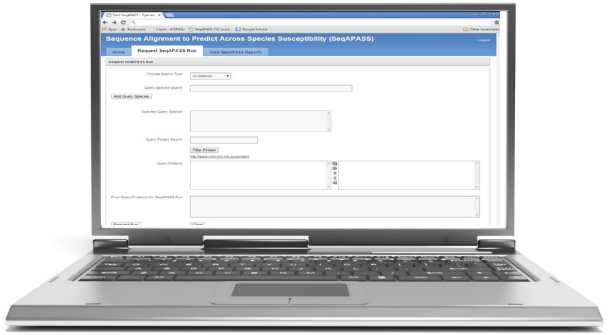
Experimental:
Site-directed mutagenesis
Attagene XS-2 Factorial assay (Dr. Blackwell)

Case Examples:
PFAS targets
Endocrine pathways
Pollinators

Sequence

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





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

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

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

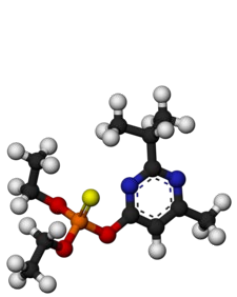
Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): A Web-Based Tool for Addressing the Challenges of Cross-Species Extrapolation of Chemical Toxicity

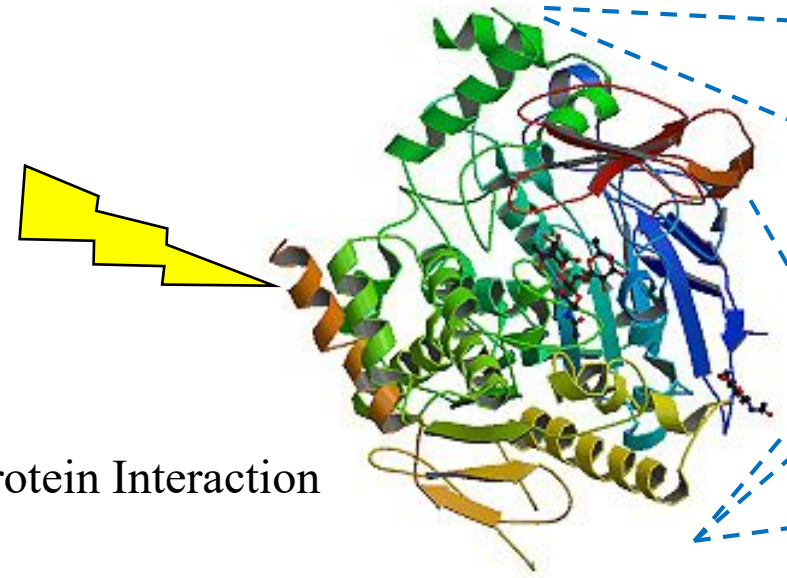
Carlie A. LaLone,*[,1] Daniel L. Villeneuve,* David Lyons,[†] Henry W. Helgen,[‡] Serina L. Robinson,[§,2] Joseph A. Swintek,[¶] Travis W. Saari,* and Gerald T. Ankley*

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

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)

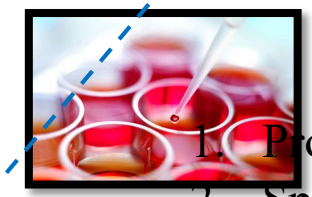
What information is required for a SeqAPASS query?





Chemical-Protein Interaction

Knowledge of a sensitive or targeted species

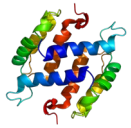


Knowledge of the model organism used in an *in vitro* assay

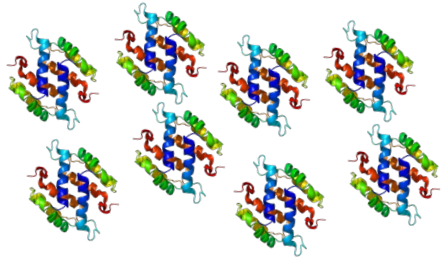
1. Protein
2. Species

Knowledge of the species for which the Key Event was developed

Chemical Molecular Target in Target Species



≅



Compare to Millions of Proteins
From Thousands of Species

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

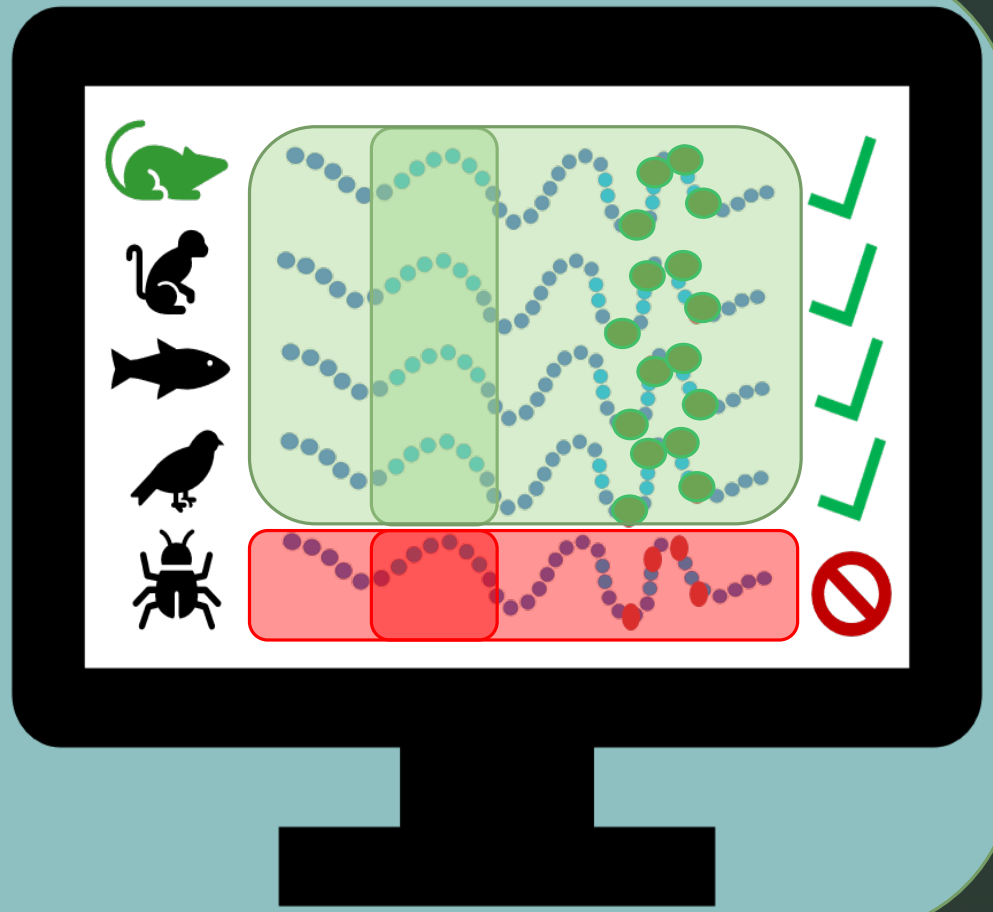
Flexible Analysis Based On Available Data

Level 1 Primary Amino Acid Sequence Alignments

Level 2 Conserved Functional Domain Alignments

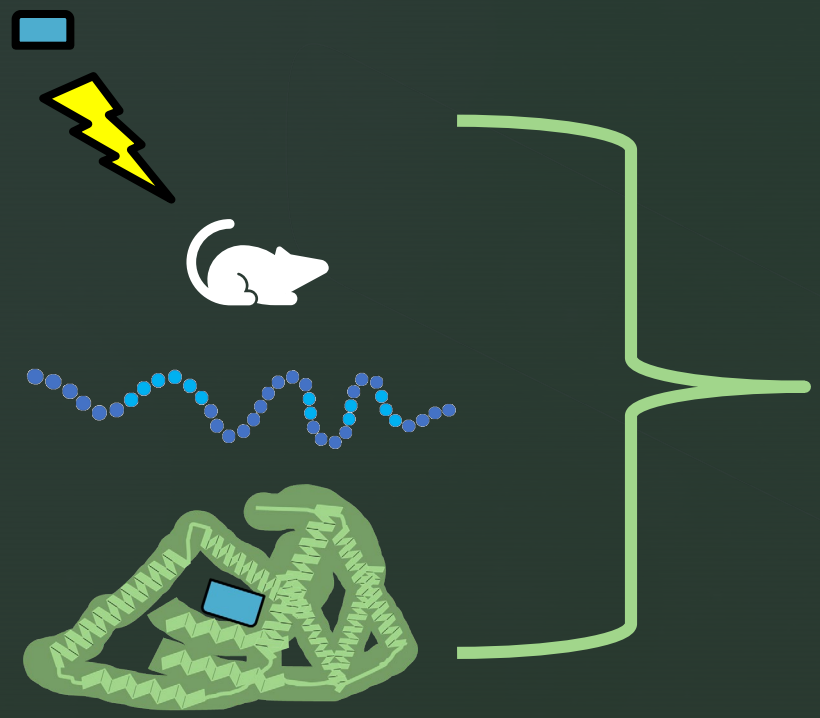
Level 3 Critical (Close Contact) Amino Acid Conservation

seqapass.epa.gov/seqapass/

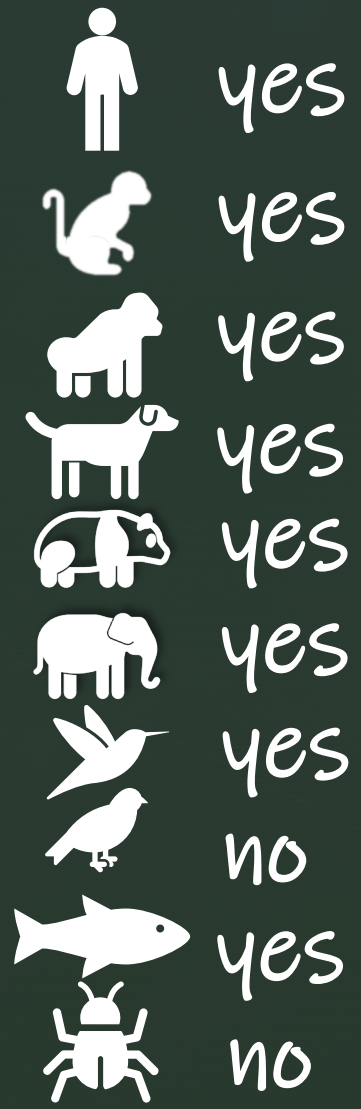


Gather Lines of Evidence Toward Protein Conservation

SeqAPASS Predicts Likelihood of Similar Susceptibility based on Sequence Conservation:



yes



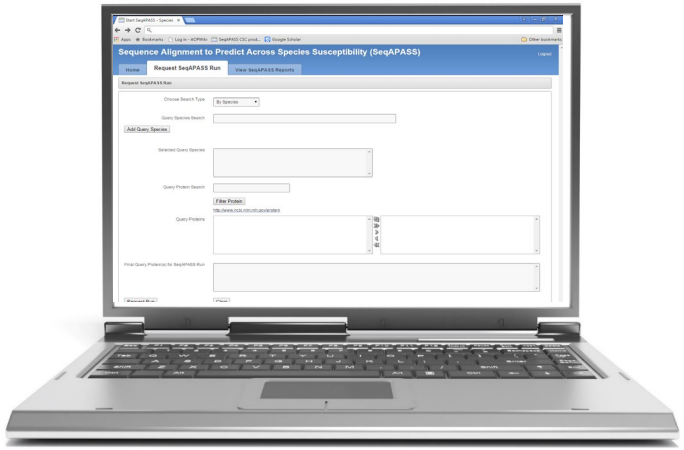
yes
yes
yes
yes
yes
yes
yes
no
yes
no

Line(s) of evidence indicate
- The protein is conserved
- The protein is NOT conserved

Evolution of the SeqAPASS tool

- V5.0 (Nov. 2020): Develop visualization (Level 3), Develop Decision Summary Report
- v4.0 (2019): Improve visualization, user guidance, summary tables, interoperability
- v3.0 (2018): Develop visualization (Level 1 & 2), automate Level 3 Susceptibility Predictions
- v2.0 (2017): develop Level 3 Susceptibility Predictions
- v1.0 (2016): Develop interface Level 1 & 2 and integrate essential functionality

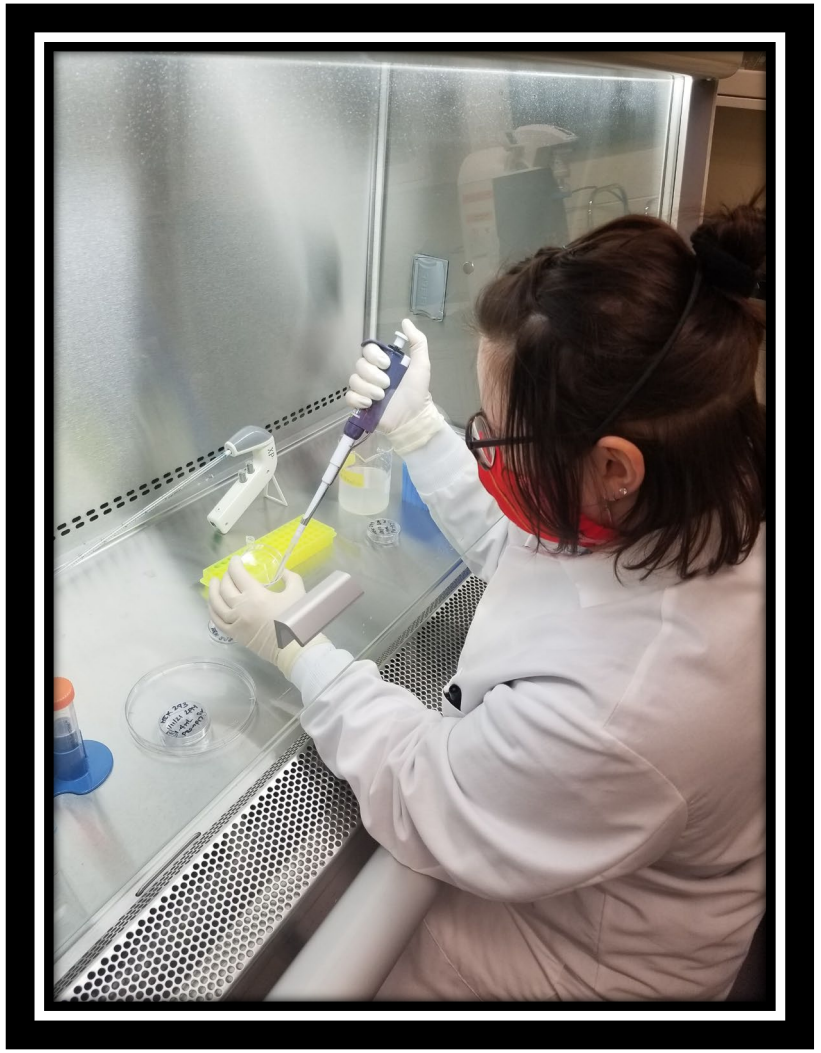
Strategic Approach to Species Extrapolation



Computational:

Bioinformatics (Session 2 Demo)

Systematic review



Experimental:

Site-directed mutagenesis

Attagene XS-2 Factorial assay (Dr. Blackwell)

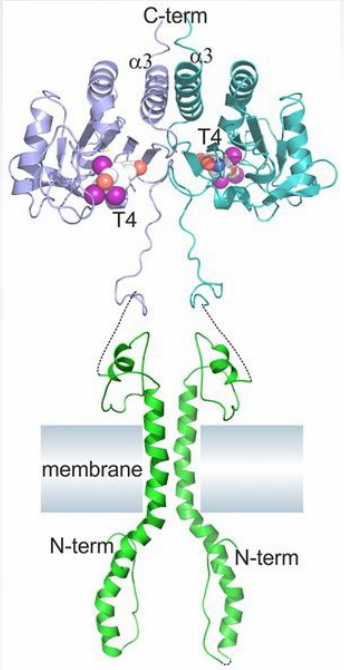
Case Examples:

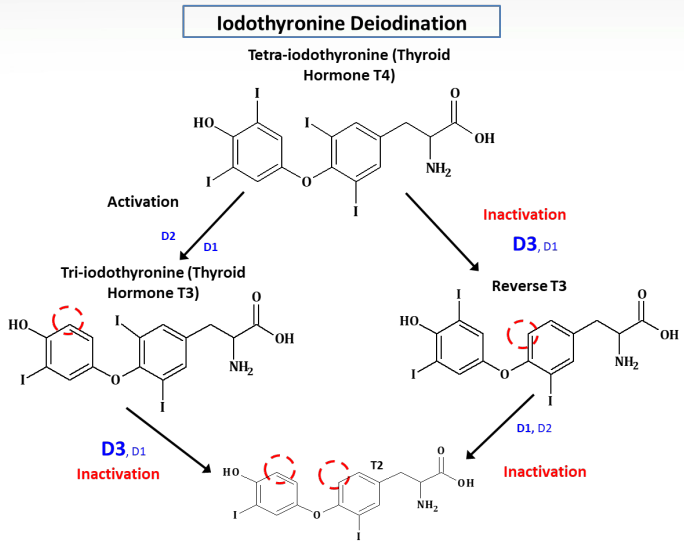
PFAS targets

Endocrine pathways

Pollinators

Deiodinase 3: Important enzyme in thyroid function



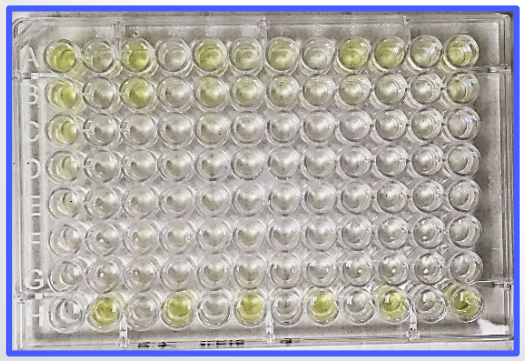


SeqAPASS Critical Amino Acid Comparison

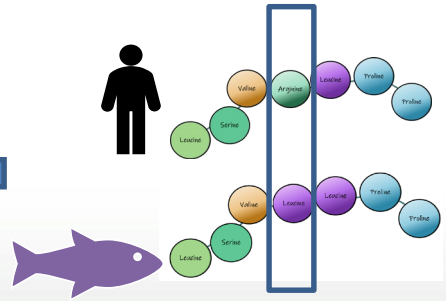
Total Match
Partial Match
Not a Match
Susceptible Yes
Susceptible No

Common Name	Similar Susceptibility	Amino Acid 1	Amino Acid 2	Amino Acid 3	Amino Acid 4	Amino Acid 5
Human	Y	168C	169T	239C	240A	257Y
Red drum	N	130C	131S	201S	202N	219Y
Blue tilapia	Y	123C	124T	194C	195L	212Y
Tongue sole	N	130C	131S	201G	202N	219Y
Zebra mbuna	N	130C	131S	201T	202N	219Y
Nile tilapia	N	130C	131S	201T	202N	219Y
Senegalese sole	N	130C	131S	201G	202N	219Y
Ballan wrasse	N	130C	131S	201S	202N	219Y
Gilthead seabream	N	130C	131S	201S	202N	219Y
Monterrey platyfish	N	124C	125T	195C	196R	213Y
Amazon molly	N	124C	125T	195C	196R	213Y
Shortfin molly	N	124C	125T	195C	196R	213Y
Sapphire devil	N	129C	130S	200G	201N	218Y
Southern platyfish	N	124C	125T	195C	196R	213Y
Goldlined spinefoot	N	130C	131S	201C	202E	219Y
Torafugu	N	130C	131S	201S	202N	219Y
Sailfin molly	N	130C	131S	201S	202N	219Y
Threespot wrasse	N	110C	111S	181S	182N	199Y
Princess parrotfish	N	97C	98S	168S	169N	186Y
Striped parrotfish	N	85C	86S	156S	157N	--
Frogs and toads	N	132C	133T	203C	204R	221Y
Two-lined caecilian	Y	129C	130T	200C	201P	218Y
Puerto Rican coqui	N	130C	131T	201C	202R	219Y
Caecilians	Y	129C	130T	200C	201P	218Y
Tropical clawed frog	N	130C	131T	201C	202R	219Y
African clawed frog	N	128C	129T	199C	200R	217Y
Gabon caecilian	Y	130C	131T	201C	202P	219Y
American bullfrog	N	131C	132T	202C	203R	220Y
Coelacanth	N	131C	132T	202C	203F	220Y
Sea lamprey	N	143G	144S	215C	216P	233A

Site Directed Mutagenesis to Probe SeqAPASS Level 3



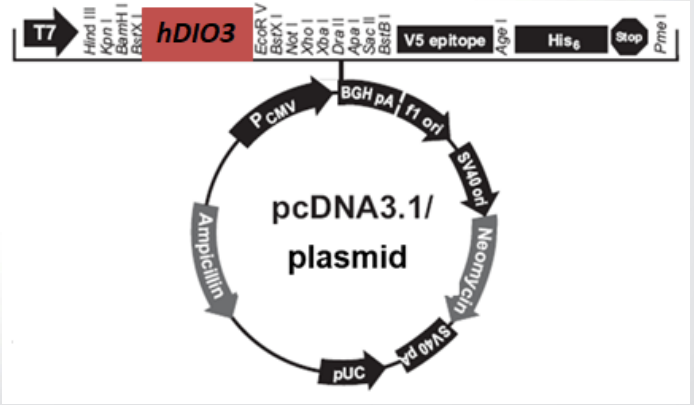
Mutate hDIO3



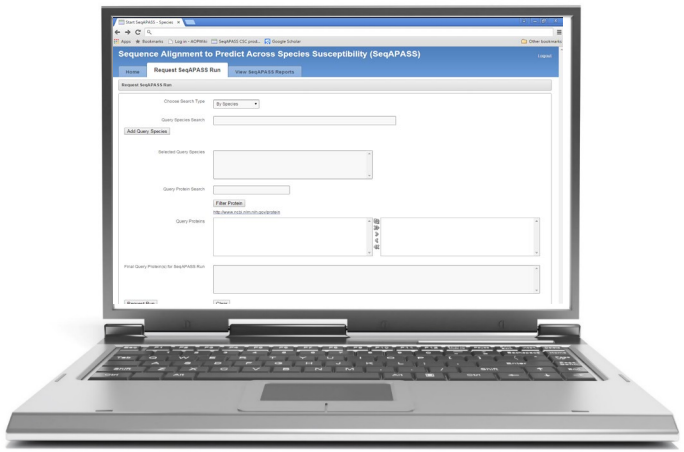
hDIO3 mutants:

Sea lamprey	C168G
Fish	T169S
Fish	C239S
Frog	A240R
Lungfish	Y257A
Sea lamprey	Y257F

Human Deiodinase 3 Enzyme



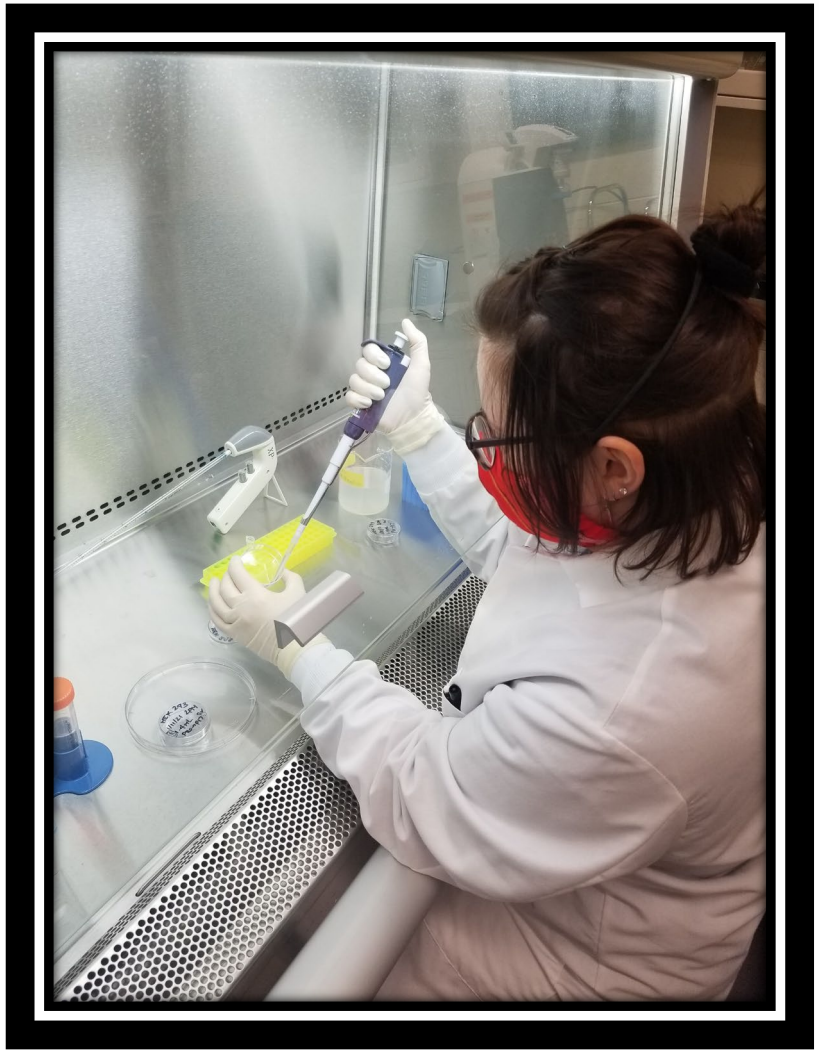
Strategic Approach to Species Extrapolation



Computational:

Bioinformatics (Session 2 Demo)

Systematic review



Experimental:

Site-directed mutagenesis

Attagene XS-2 Factorial assay (Dr. Blackwell)

Case Examples:

PFAS targets

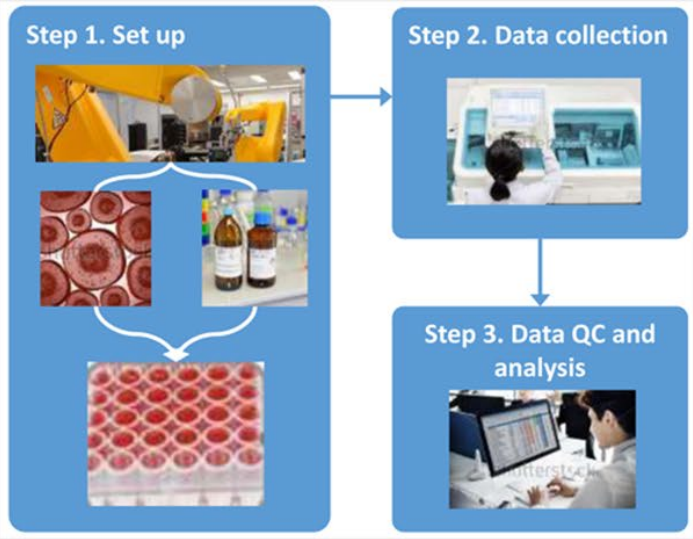
Endocrine pathways

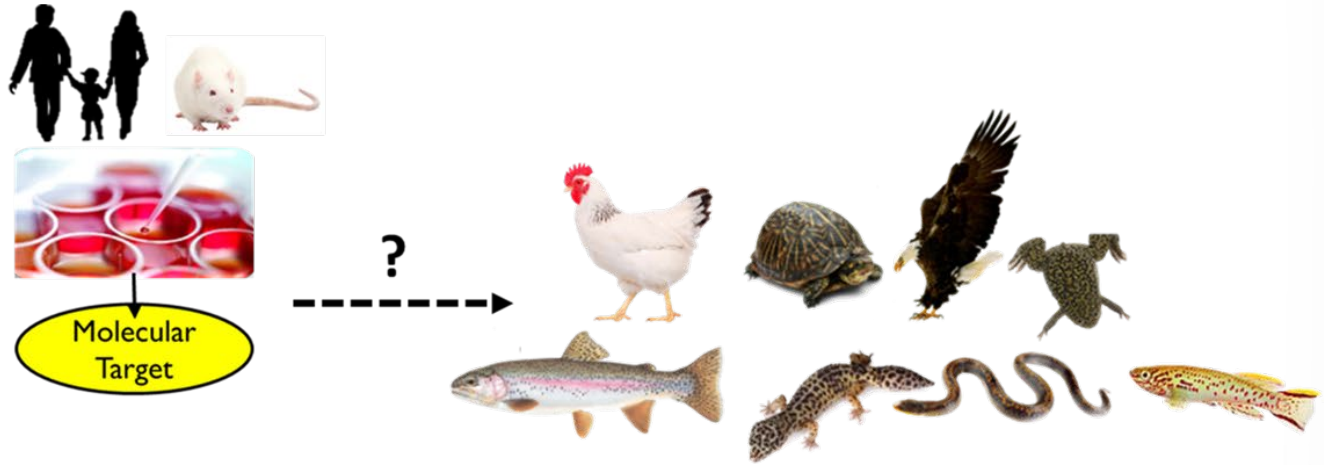
Pollinators

U.S. EPA Toxicity Forecaster (ToxCast)

U.S. EPA ToxCast Program:

US EPA ToxCast Program: Uses **mammalian cell-based assays** to rapidly screen chemicals, identify putative molecular targets, and identify those most likely to be endocrine disruptors





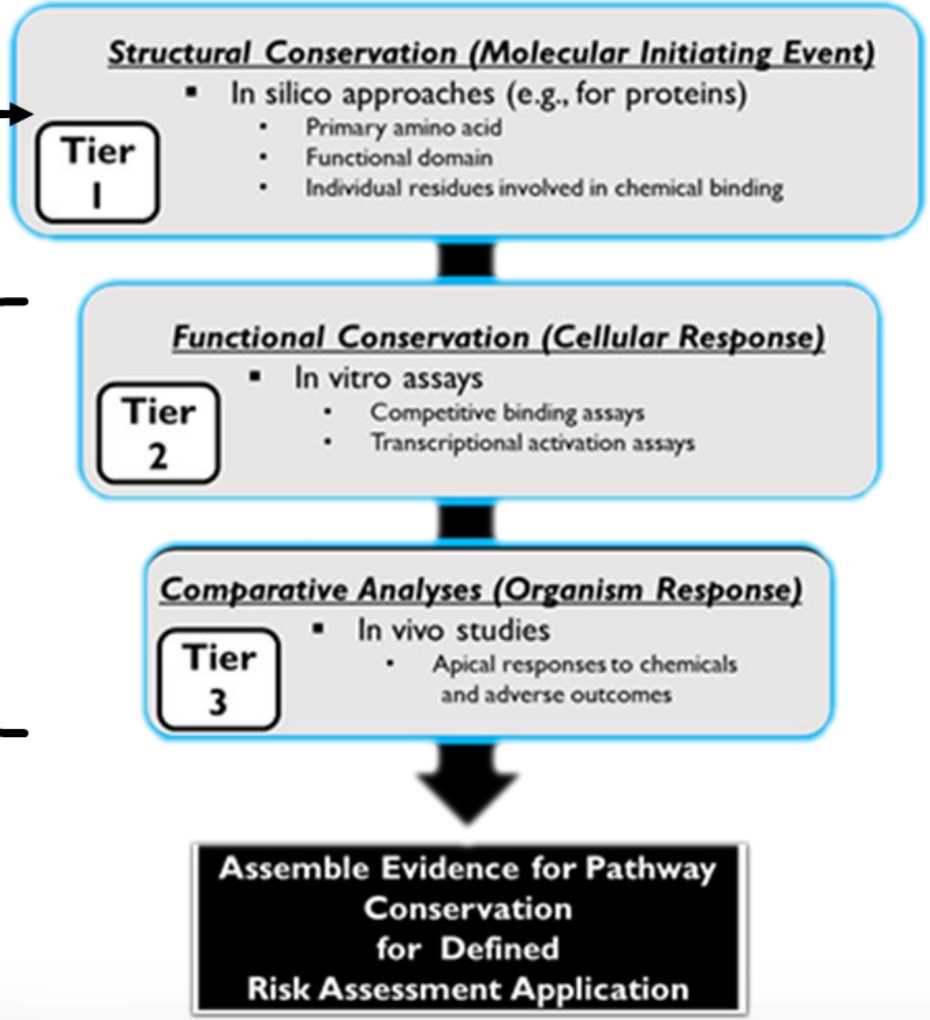
Key Questions for Consideration:

- How well does this mammalian-based prioritization approach reasonably reflect potential impacts on other vertebrates?
- Can we expect chemicals that interact with mammalian receptors to also interact with receptors of other species?

Hierarchal Framework for Evaluating Pathway Conservation Using Existing Evidence

Analysis of structure across species using in silico computational tools

Literature review and comparison of available in vitro and in vivo estrogen receptor data



> Environ Toxicol Chem. 2016 Nov;35(11):2806-2816. doi: 10.1002/etc.3456. Epub 2016 Jun 28.

Evaluation of the scientific underpinnings for identifying estrogenic chemicals in nonmammalian taxa using mammalian test systems

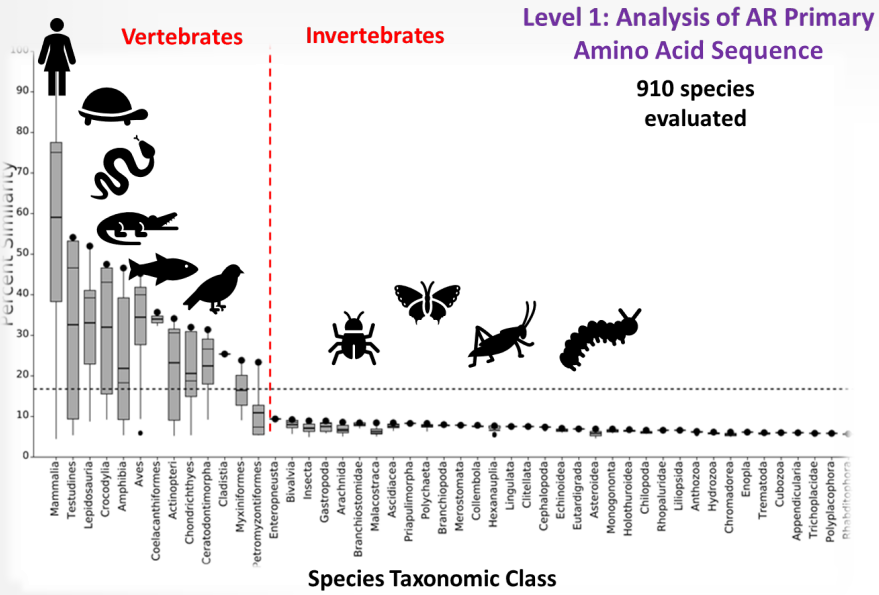
Gerald T Ankley [1], Carlie A LaLone [2], L Earl Gray [3], Daniel L Villeneuve [2], Michael W Hornung [2]

What other important endocrine targets have a large base of pre-existing structural, molecular target, and toxicity data?

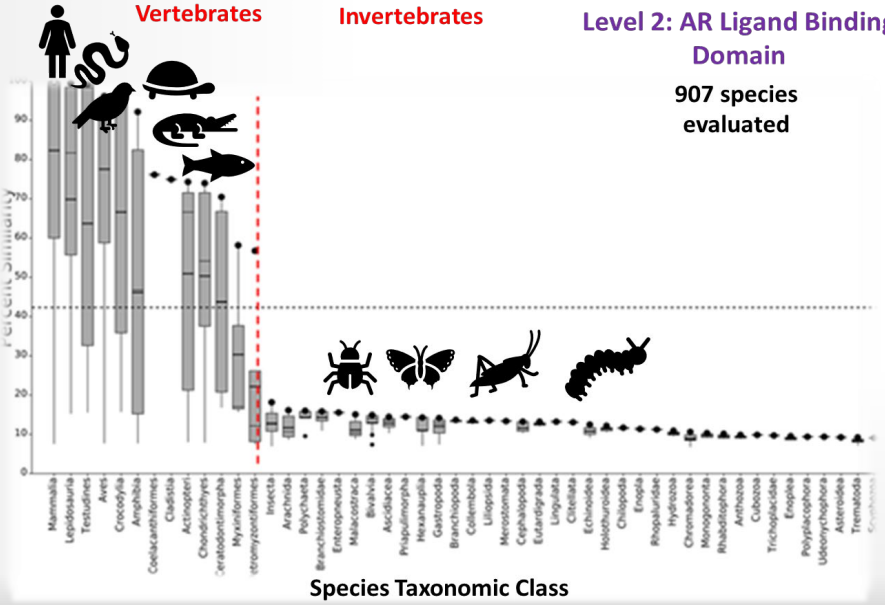
→ **Androgen Receptor (AR)**

Assessing AR Conservation Across Species Using the SeqAPASS Tool

1.



2.



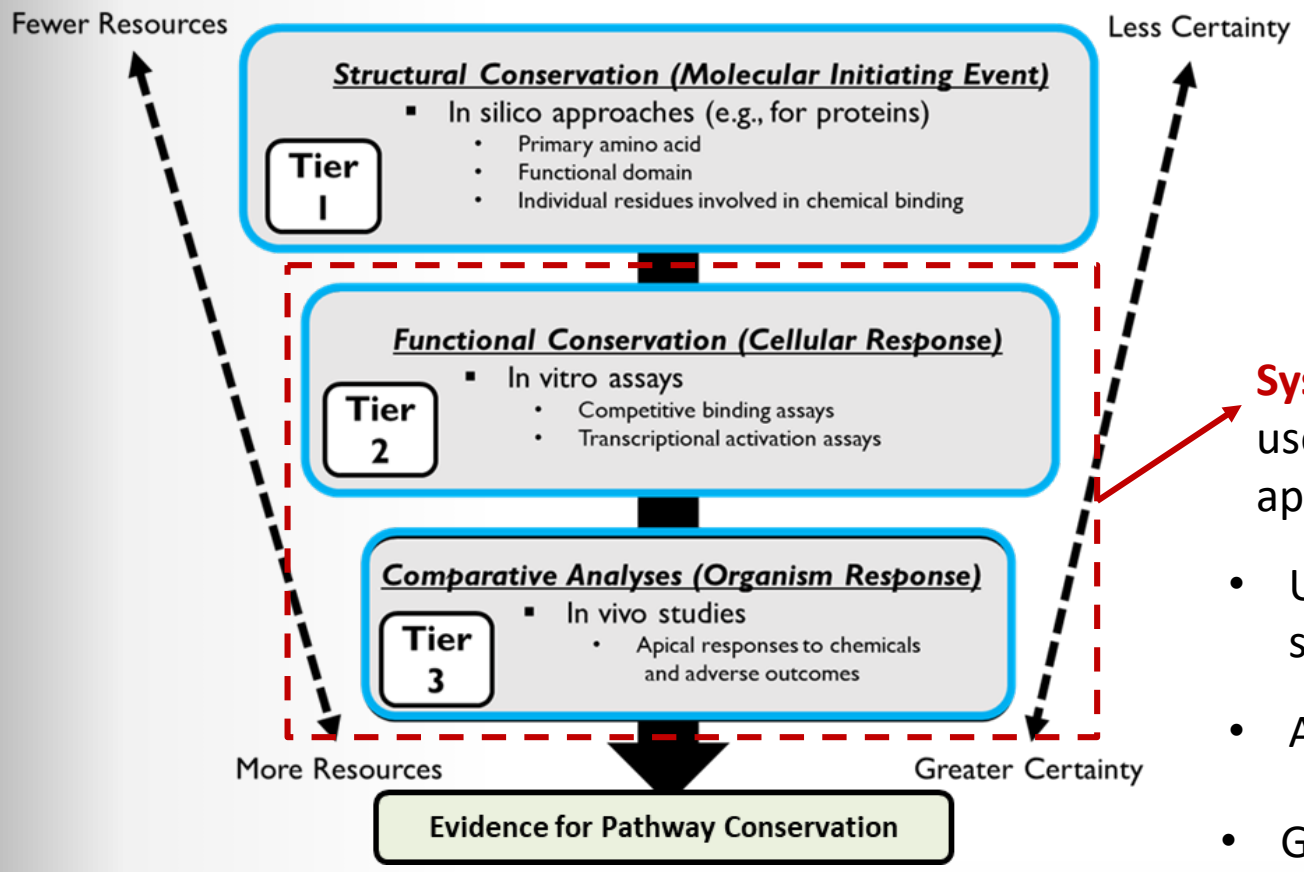
3. **Level 3: Analysis of Conservation of Individual Amino Acid Residues**

250 species evaluated

Taxonomic Group	# of Spp.	Shared Susceptibility
Mammals	117/1	Yes/No
Lizards, Snakes	11	Yes
Turtles	3	Yes
Birds	58	Yes
Crocodiles, Alligators	4	Yes
Amphibians	13	Yes
Coelacanths	2	Yes
Eel-shaped	1	Yes
Bony Fish	87/1	Yes/No
Sharks, Rays	4	Yes
Lungfish	2	Yes

- Across all three levels, SeqAPASS results suggest conservation of AR across vertebrate species
- Overall, these predictions suggest that chemicals that bind and activate AR in mammalian-based assays, are likely to interfere with AR in other vertebrate species
- Line of evidence for pathway conservation

Evaluating Existing Data to Extrapolate High-Throughput Androgen Receptor Screening Data Across Species



Environ Toxicol Chem. 2016 Nov;35(11):2806-2816. doi: 10.1002/etc.3456. Epub 2016 Jun 28.

Evaluation of the scientific underpinnings for identifying estrogenic chemicals in nonmammalian taxa using mammalian test systems

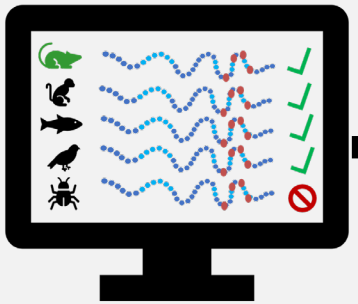
Gerald T Ankley [1], Carlie A LaLone [2], L Earl Gray [3], Daniel L Villeneuve [2], Michael W Hornung [2]

Systematic Literature Review: A type of literature review that uses systematic methods to collect secondary data, critically appraise research studies, and synthesize findings

- Using existing evidence (literature), we can evaluate the scientific basis of our cross-species predictions
- Advances in data science can improve this workflow
- Gathering in vivo and in vitro data from vertebrate species exposed to known androgenic compounds provides **additional lines of evidence** for the conservation of the biological pathway across species

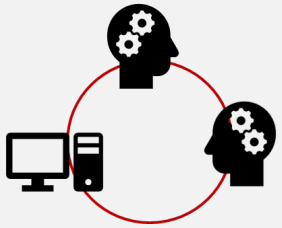
Evaluating Existing Data to Extrapolate High-Throughput Androgen Receptor Screening Data Across Species

SeqAPASS Evaluation of Structural Conservation Across Species

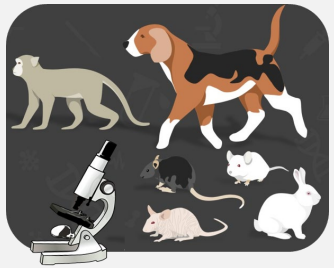


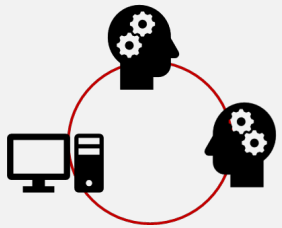
Conserved?
Yes
No

Systematic Evaluation of In Vitro Cross-Species Data



Systematic Evaluation of In Vivo Cross-Species Data

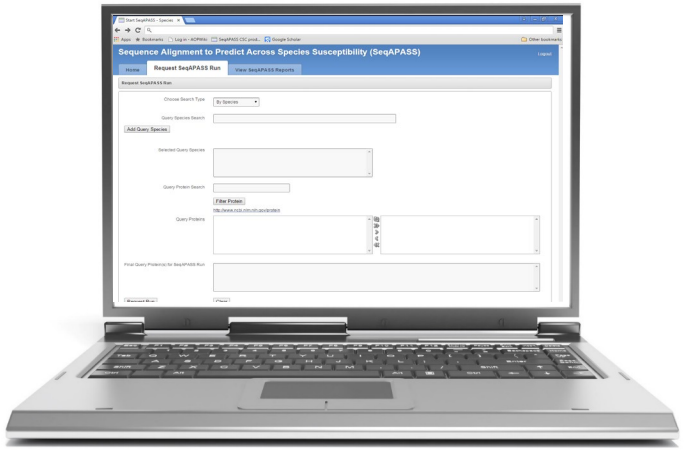




Weight of Evidence for Pathway Conservation Across Species for Defined Risk Assessment Applications

- Apply pathway to other targets of interest
- Repeat process to account for the emergence of new information

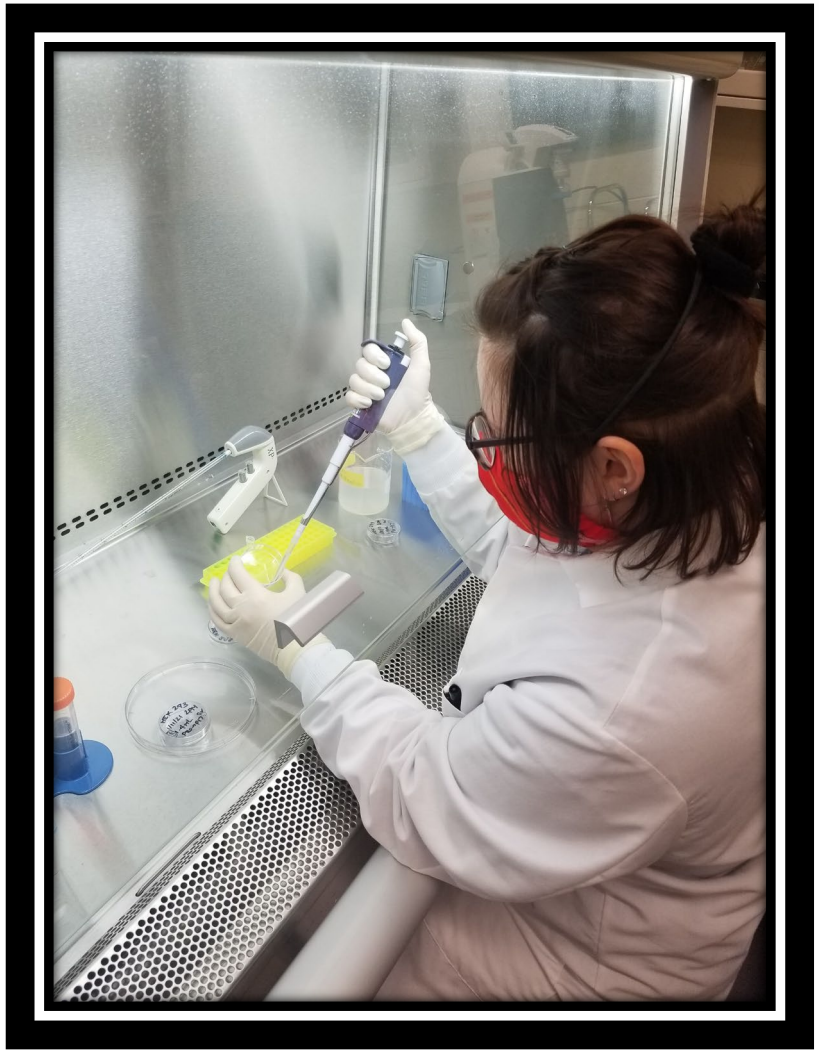
Strategic Approach to Species Extrapolation



Computational:

Bioinformatics (Session 2 Demo)

Systematic review



Experimental:

Site-directed mutagenesis

Attagene XS-2 Factorial assay (Dr. Blackwell)

Case Examples:

PFAS targets

Endocrine pathways

Pollinators

United States Environmental Protection Agency

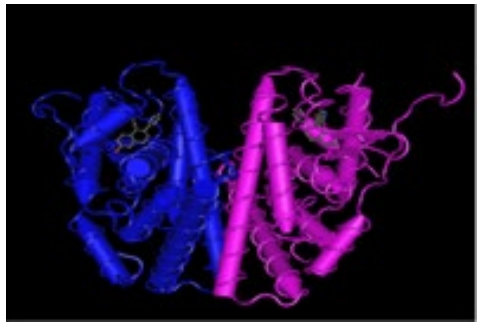
Final ID: **1.05.06** Donovan Blatz
Presentation Type: Poster
Final ID: **1.03.07** Sara Vliet
Presentation Type: Platform

Sequence

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

Yes or No
Susceptible or Not Susceptible

Structure

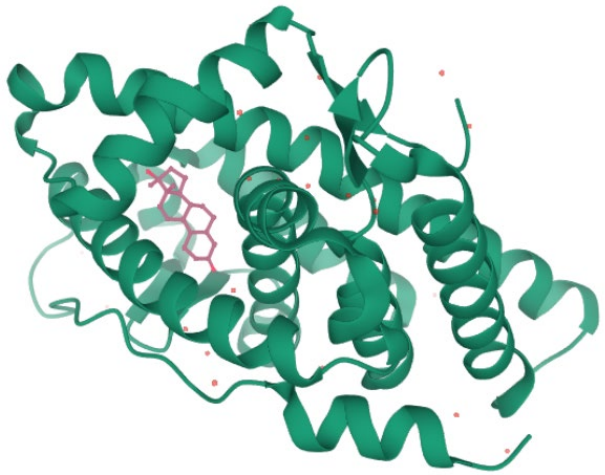


Structural-based
comparisons of similarity
Predicted binding affinity

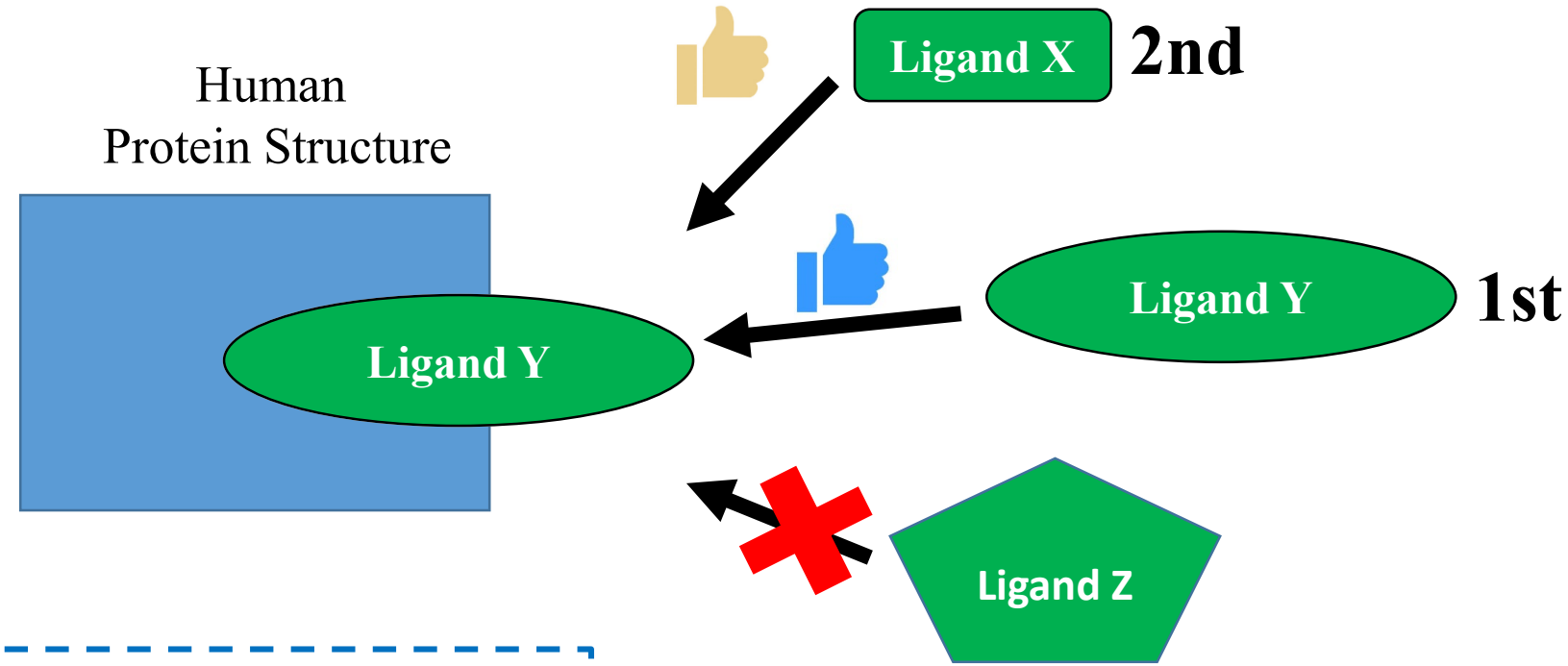
Function

Improvements in bioinformatics

Advances in Drug Discovery/Development



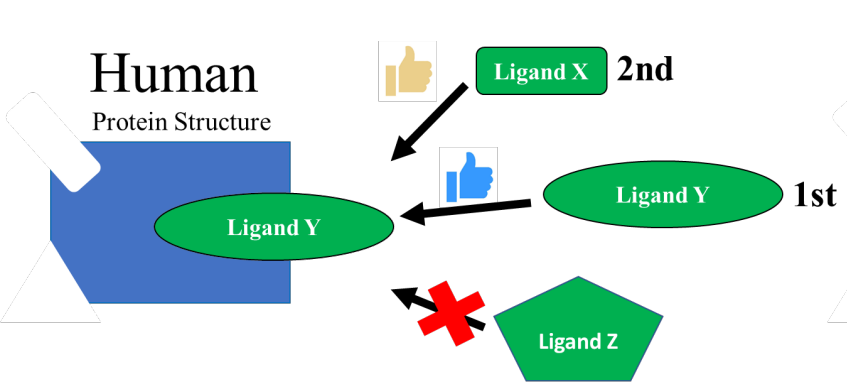
Structure derived from X-ray crystallography

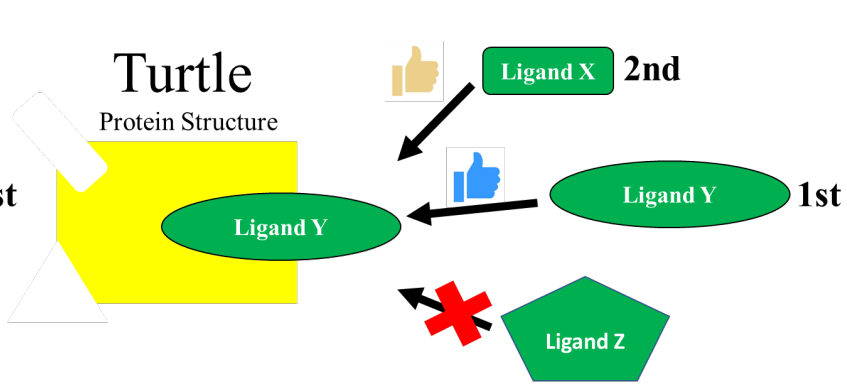


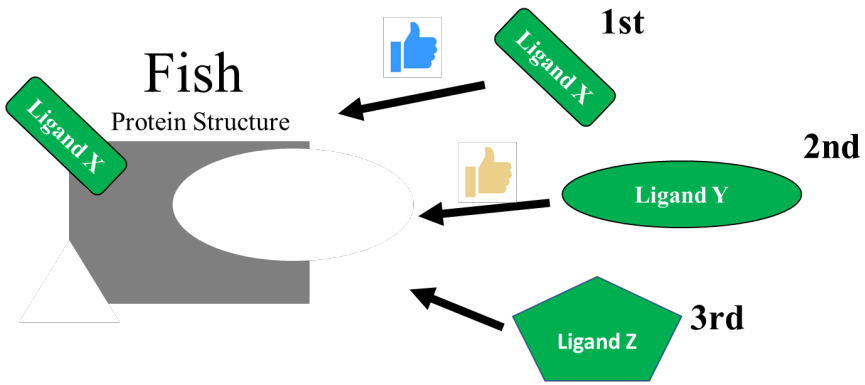
Bioinformatics Toolbox:

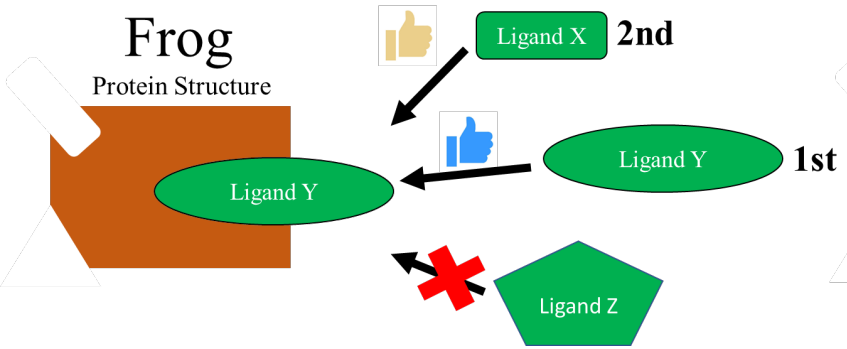
Molecular modeling
Molecular docking
Virtual screening
Molecular dynamic simulations

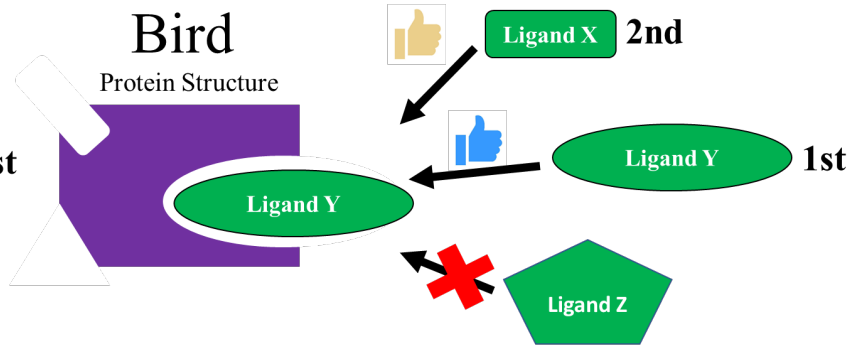
Application to Species Extrapolation

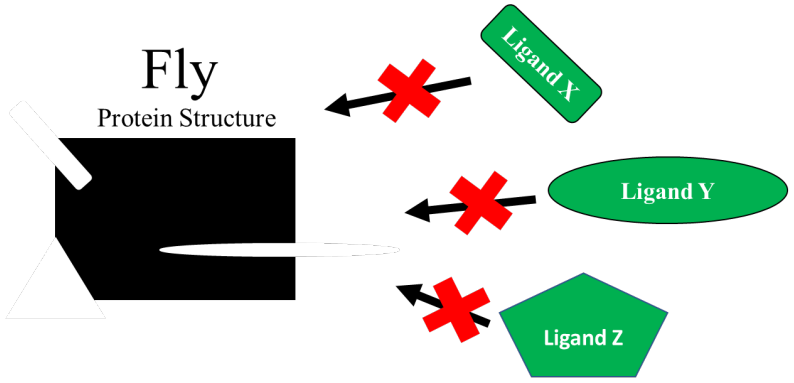












Bioinformatics Toolbox:

Molecular modeling

Molecular docking

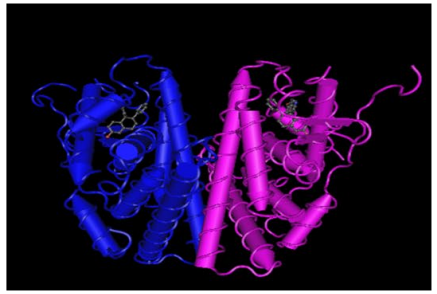
Virtual screening

Molecular dynamic simulations

Sequence

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

Structure



SeqAPASS Results from Level 1

Query Sequence FASTA + FASTA from 100s of Aligned Sequences Across Taxa

>NP_001434.1 Protein X [Homo sapiens]
MSFSGKYQLQSQENFEAFMKAIGLPEELIQKGKDI
KGVSEIVQNGKHFKFTITAGSKVIQNEFTVGEECE
LETMTGEKVKTVVQLEGDNKLVTTFKNIKSVTELN
GDIITNTMTLGDIVFKRISKRI

>NP_787011.1 Protein X [Bos taurus]
MNFSGKYQVQTQENYEAFMKAVGMPDDIIQKGK
DIKGVSEIVQNGKHFKFIITAGSKVIQNEFTLGEECE
MEFMTGEKIKAVVQQEGDNKLVTTFKGIKSVTEFN
GDTVTSTMTKGDVVFKRVSKRI

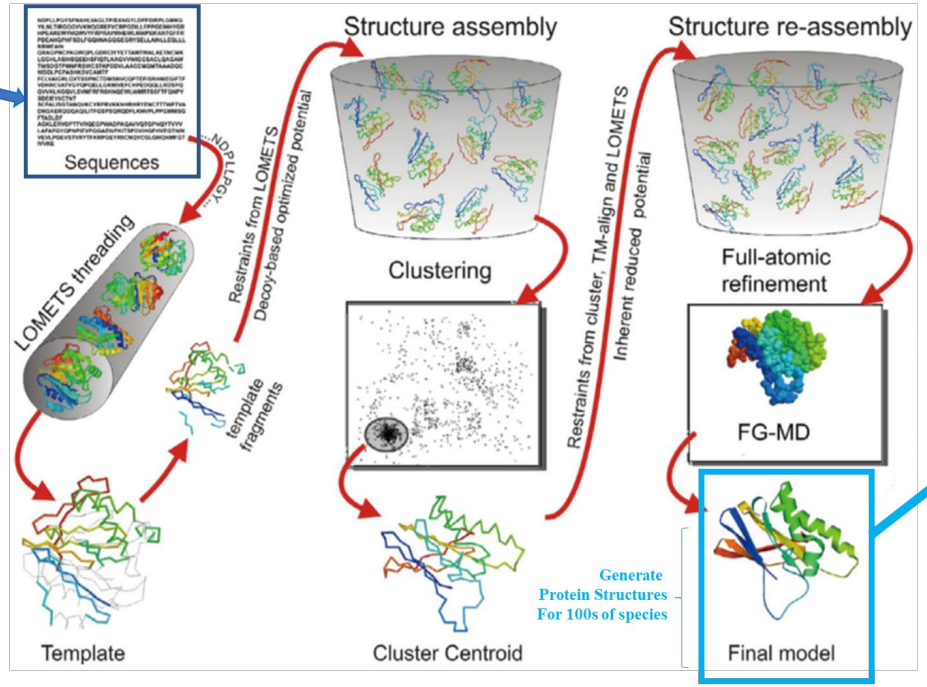
>KFQ76585.1 Protein X [Phoenicopterus ruber ruber]
MSFTGKYELQSQENFEPFMKALGLPDDQIQKGKD
IKSISEIVQDGKKFKVTVTTGSKVMQNEFTIGEECD
IEMLTGEKVKAVVQMEGNNRLVANLKGLKSVTEL
NGDIITHTMTMGDLTYKRISKRI

>NP_001116883.1 Protein X [Xenopus tropicalis]
MAFAGKYELVHQENFETFMKAIGLSDELIQKGKDV
KSVTEIQQNGKHFIVTVTTGSKVLRNEFTIGEEAE
LETPTGEKVKSVVKLEGDNKLVVQLKAITSTTELSG
DTITHVLTLNNLVFKRVSKRV

100s of FASTA

Iterative Threading ASSEmbly Refinement

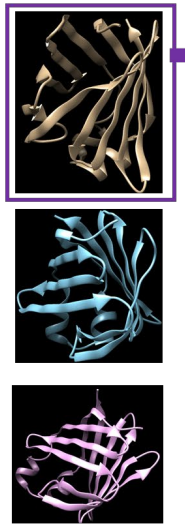
Develop Models for 100s of Species Based on Aligned Sequences
(I-TASSER; https://zhanglab.ccmb.med.umich.edu/I-TASSER/)



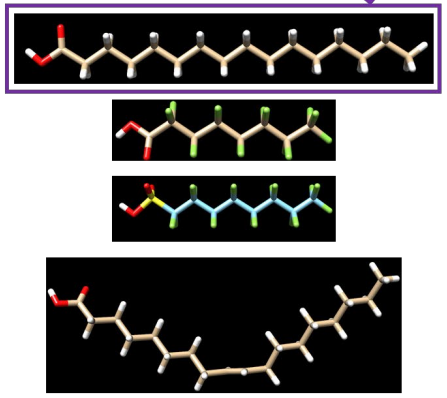
UCSF Chimera

DockPrep Structures and Minimize Ligands

Protein Structure Models From 100s of Species

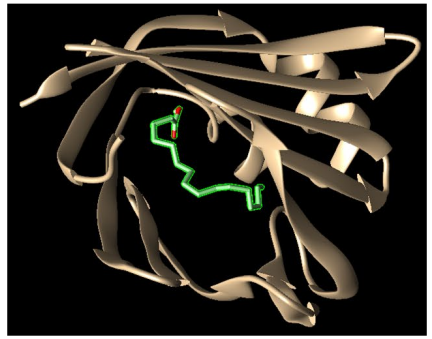


Ligands of Interest for Docking



AutoDock Vina

Dock Multiple Ligands to Protein Structures



Collect Predicted Binding Affinity

S	Score	RMSD lb.	RMSD 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  O2_18
```

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

Predicting Binding Affinity

Acknowledgements

U.S. EPA, ORD

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

GDIT

Thomas Transue
Cody Simmons
Audrey Wilkinson

SeqAPASS v5.0

LaLone.Carlie@epa.gov

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