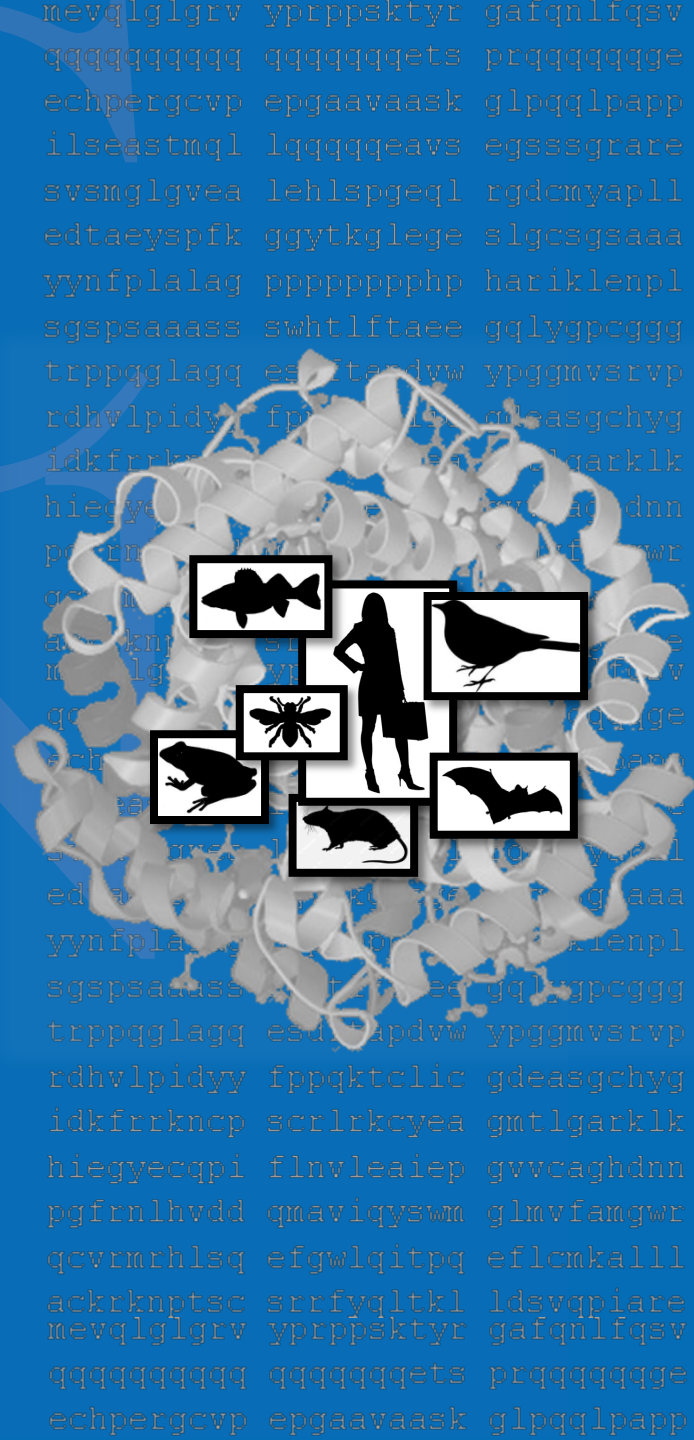
United States Environmental Protection Agency

New Approach Methods: Application of the Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) tool



Carlie A. LaLone, Ph.D.

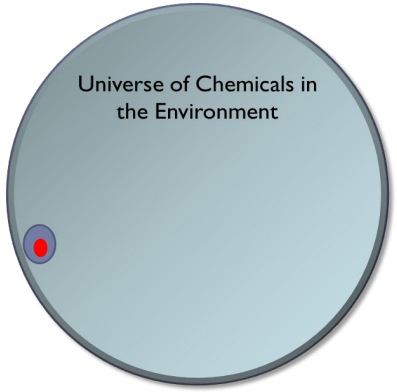
Research Bioinformaticist

Office of Research and Development
National Health and Environmental Effects Research Laboratory, Mid-continent Ecology Division

Chemical Safety Evaluation

- Core Mission at EPA: Protect human health and the environment
  - Ensure that chemicals in the marketplace are reviewed for safety

- Challenging mission:
  - Only a fraction of the compounds around us have been tested for safety
    - Chemicals used by U.S. consumers and industry: +50,000
    - Tested: ~300
    - *Chemicals tested across species: Even more sparse*



Regulatory Decision-Making and Guideline Documents

- Capitalize on the toxicity information that is available and, where we can, request additional toxicity tests

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 21st 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

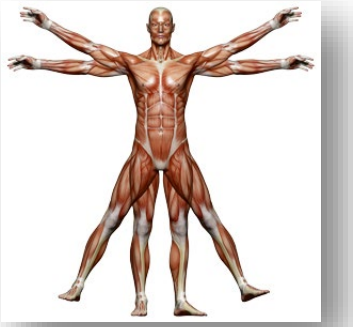
Confidence in our decisions, with limited data available, limited resources for testing, strong backing to reduce animal use

Transformation of Toxicity Testing

Historically:

Whole animal test

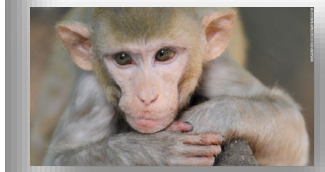
- **Observe Toxic Outcome**



Cannot Test

≅

- Assumed that sensitivity of species to a chemical is a function of their relatedness
- Human Health Risk Assessment

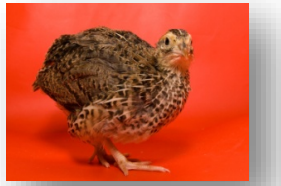


Use of Surrogates

- Ecological Risk Assessment

Cannot Test

≅



Representative species across a diversity of organism classes

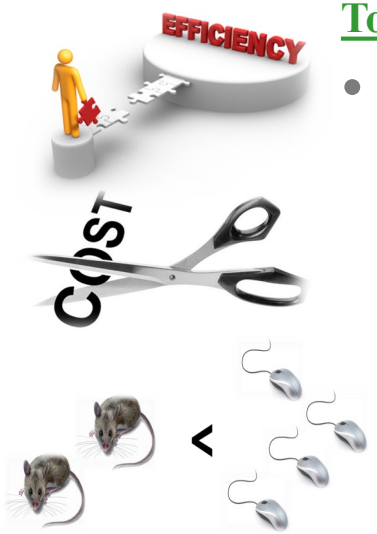
Transformation of Toxicity Testing

Historically:

Whole animal test

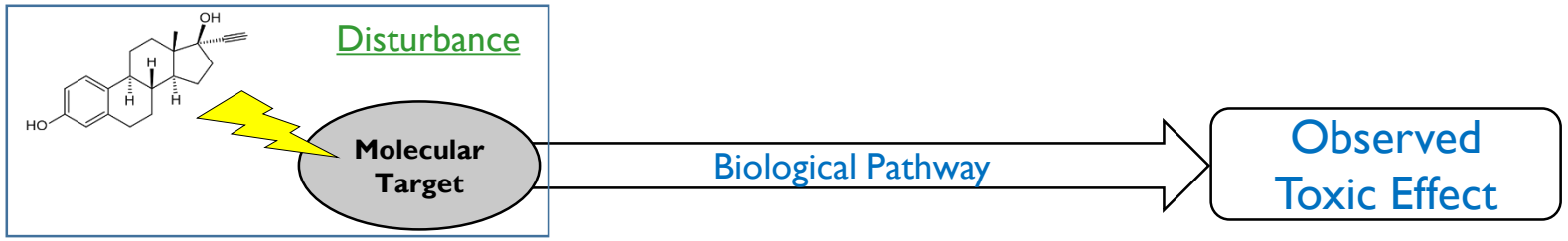
- **Observe Toxic Outcome**
  - **Examples**
    - **tumor development**
    - **mortality**

  Resource intensive



Toxicity Testing in the 21st Century:

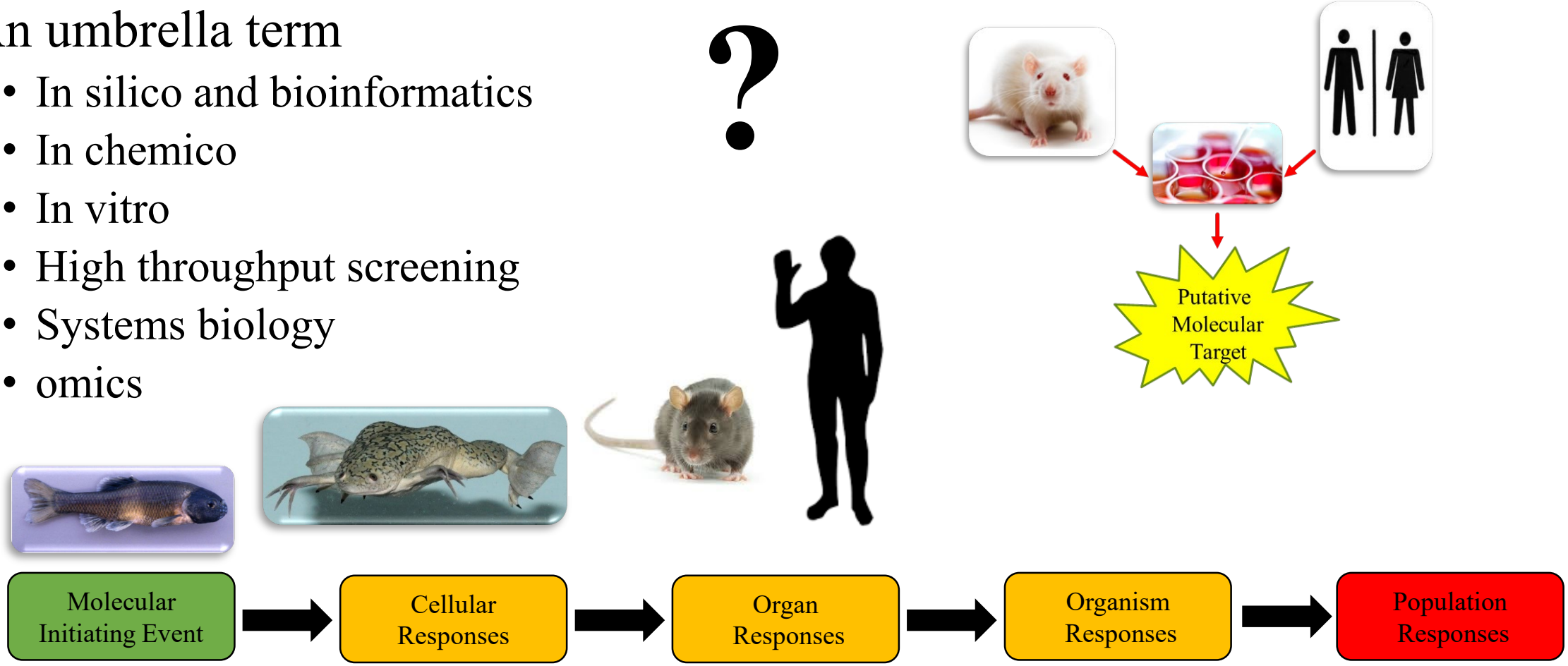
- ***In vitro* and *in silico* methods**
  - **Pathway-based approaches**
  - **Focus on disturbance of the biological pathway**
    - **Predictive of the observable toxic effects**



Enabled by evolution of the science and technology

New Approach Methods for Regulatory Decision-Making

- An umbrella term
  - In silico and bioinformatics
  - In chemico
  - In vitro
  - High throughput screening
  - Systems biology
  - omics



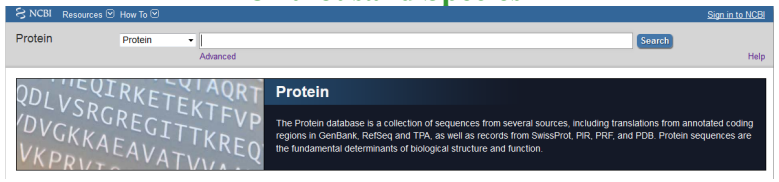
Need for species extrapolation

New Approach Methods: Species Extrapolation

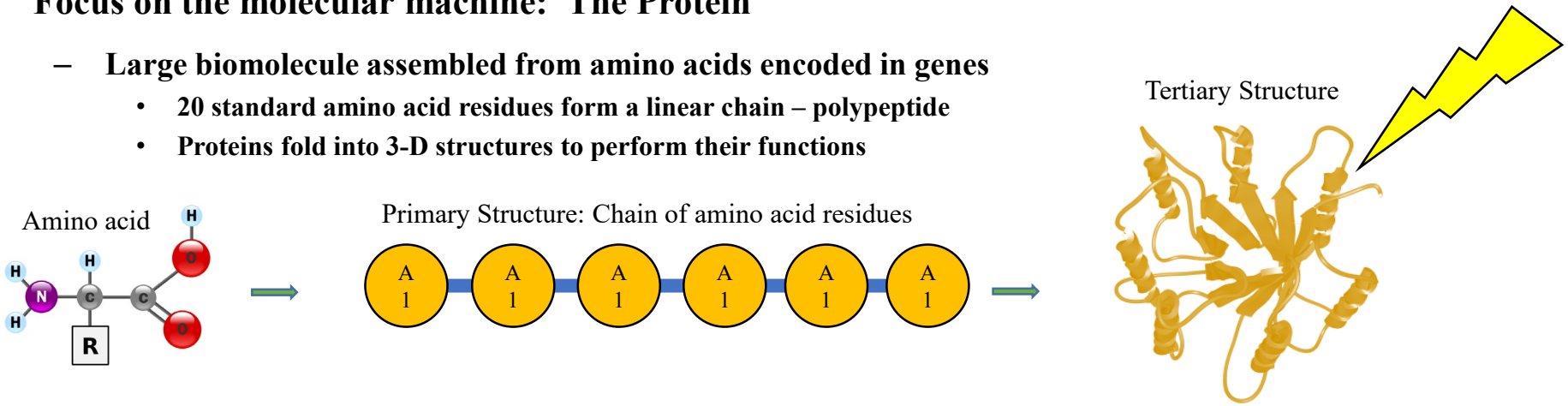
New tools and technologies have emerged

- Improved sequencing technologies
- Large databases of sequence data

As of this week
~121 million Proteins
~84 thousand Species



- **Focus on the molecular machine: The Protein**
  - **Large biomolecule assembled from amino acids encoded in genes**
    - **20 standard amino acid residues form a linear chain – polypeptide**
    - **Proteins fold into 3-D structures to perform their functions**



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

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

EPA
United States
Environmental Protection
Agency

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

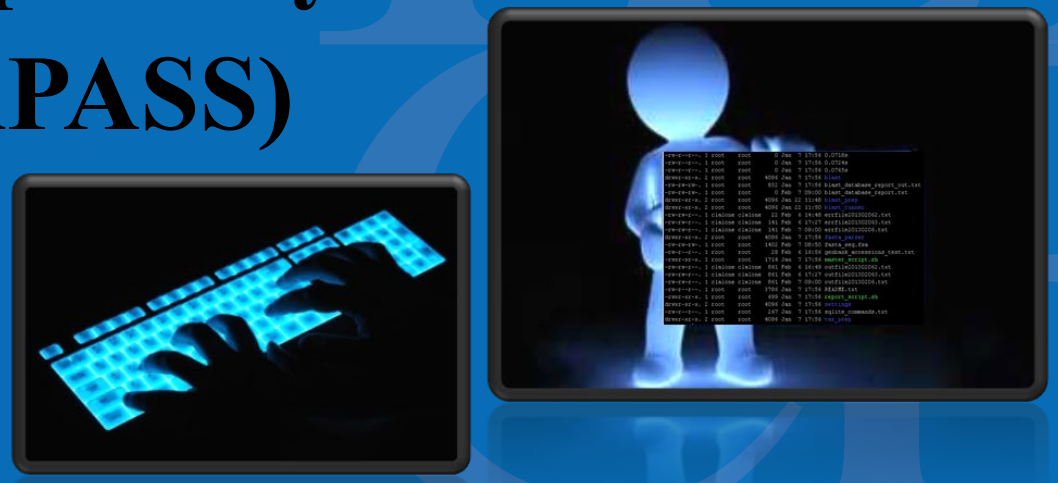
TOXICOLOGICAL SCIENCES, 153(2), 2016, 228–245

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

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

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

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)



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

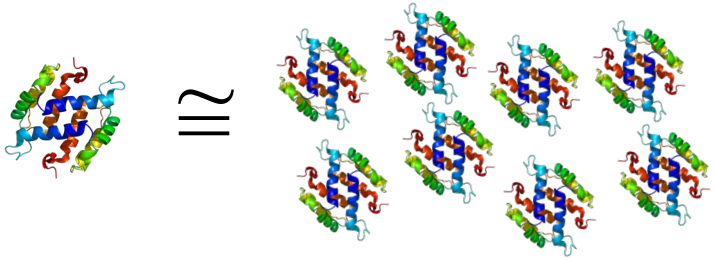
Understanding Protein Conservation

Sequence Alignment to Predict Across Species Susceptibility

- Computational Assessment of Protein Similarity: Quantitative Metrics
  - A comparison of species at the molecular level

1. Must know the molecular target (e.g., pharmaceuticals, pesticides)
2. Must identify target species or have knowledge of sensitive species

Chemical Molecular Target in Target Species

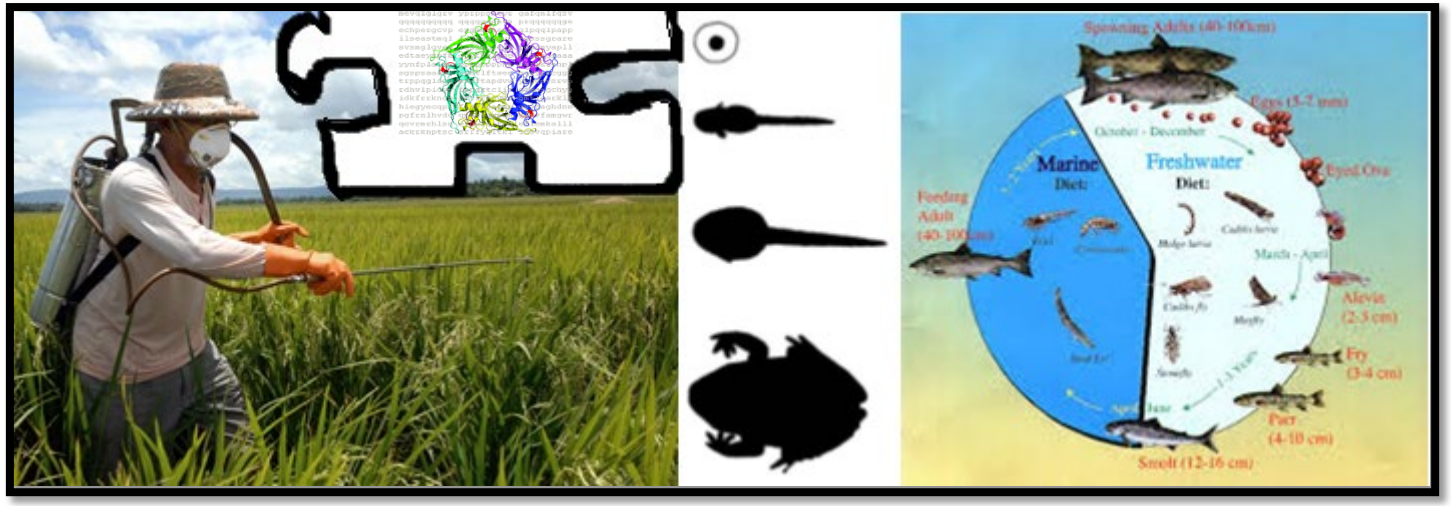


Compare to Millions of Proteins From Thousands of Species

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

Predict Relative Intrinsic Susceptibility

- Intrinsic susceptibility can be defined as the vulnerability (or lack thereof) of an organism to chemical insult due to its inherent biological composition
  - Receptor/enzyme (protein) available for the chemical to act upon
- Relative:
  - Molecular target conservation is but a component of multiple determinants of species susceptibility

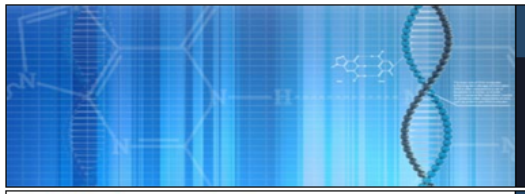


Developing SeqAPASS:

Available Databases and Tools

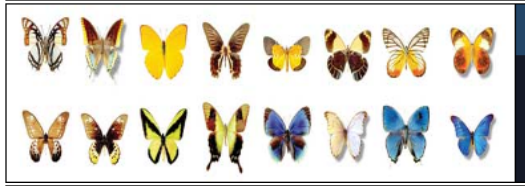
- National Center for Biotechnology Information
- Established in 1988: a division of National Library of Medicine at NIH

Existing Curated Databases



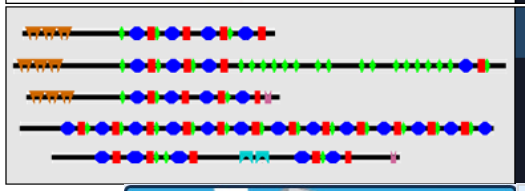
RefSeq: NCBI Reference Sequence Database

A comprehensive, integrated, non-redundant, well-annotated set of reference sequences including genomic, transcript, and protein.



Taxonomy

The Taxonomy Database is a curated classification and nomenclature for all of the organisms in the public sequence databases. This currently represents about 10% of the described species of life on the planet.



CDD

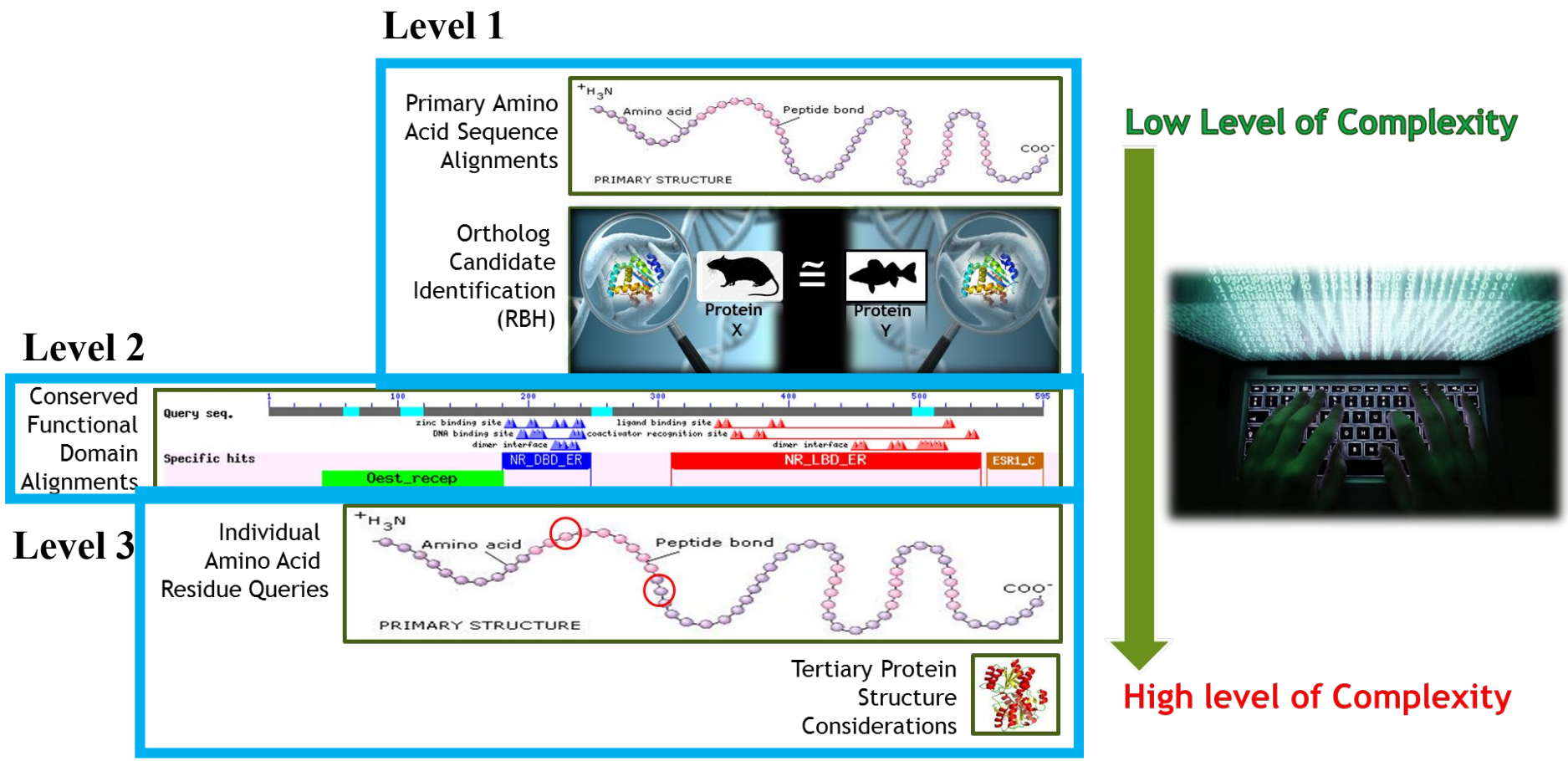
The Conserved Domain Database is a resource for the annotation of functional units in proteins. Its collection of domain models includes a set curated by NCBI, which utilizes 3D structure to provide insights into sequence/structure /function relationships.

Existing Sequence Alignment Tools

The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

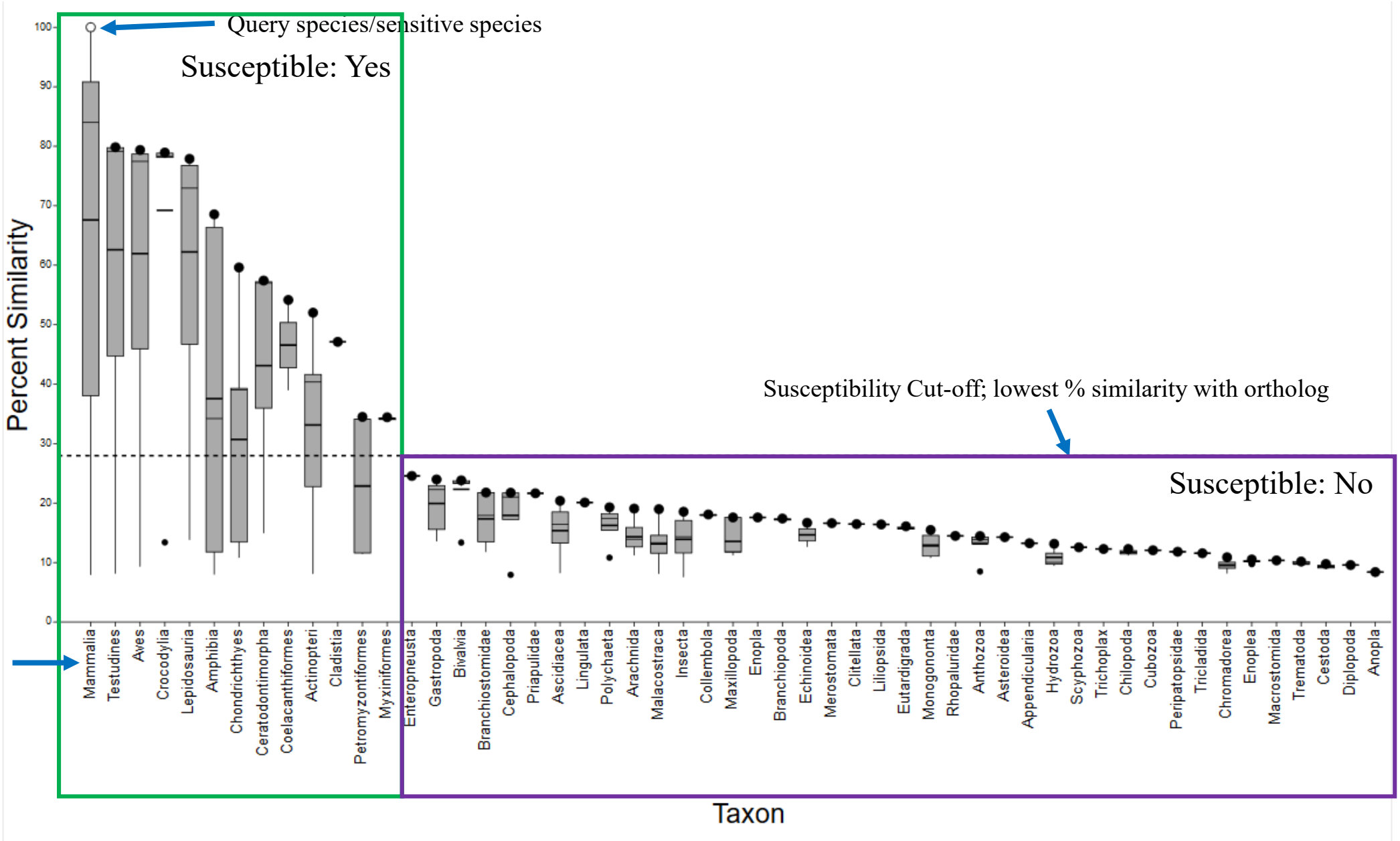
COBALT — **Constraint-based Multiple Alignment Tool**

Strategic Automated Approach for Assessing Protein Similarity



Developed with both researchers and risk assessors in mind

Integrate Interactive Data Visualization



Taxonomic group: Class

Focus on Structure: Level 3 Susceptibility Predictions

OXFORD
SOT Society of Toxicology
www.toxsci.oxfordjournals.org
ToxSci 20 Years

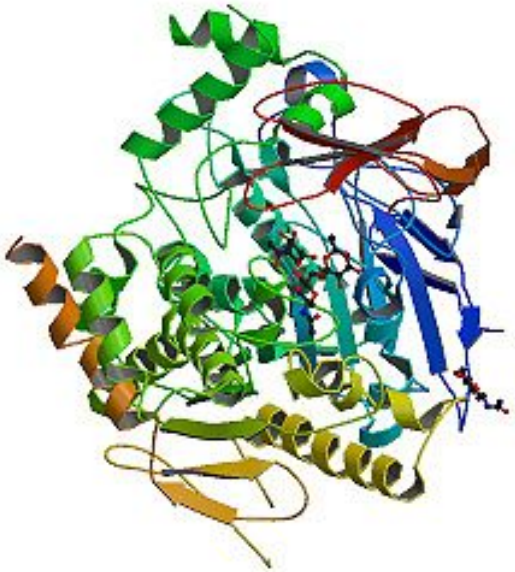
TOXICOLOGICAL SCIENCES, 2018, 1–15

doi: 10.1093/toxsci/kfy186
Dryad Digital Repository DOI: https://doi.org/10.5061/dryad.2tg69
Advance Access Publication Date: July 27, 2018
Research Article

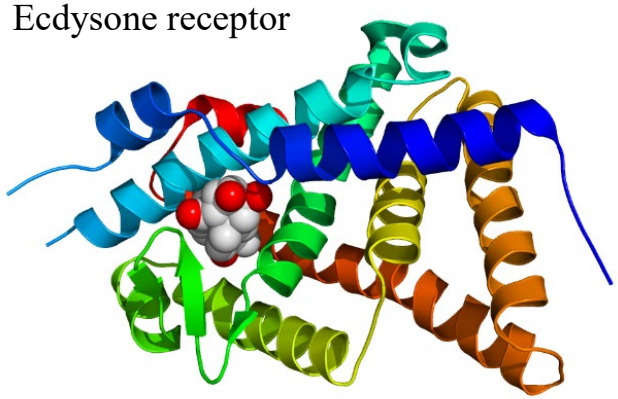
In Silico Site-Directed Mutagenesis Informs Species-Specific Predictions of Chemical Susceptibility Derived From the Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) Tool

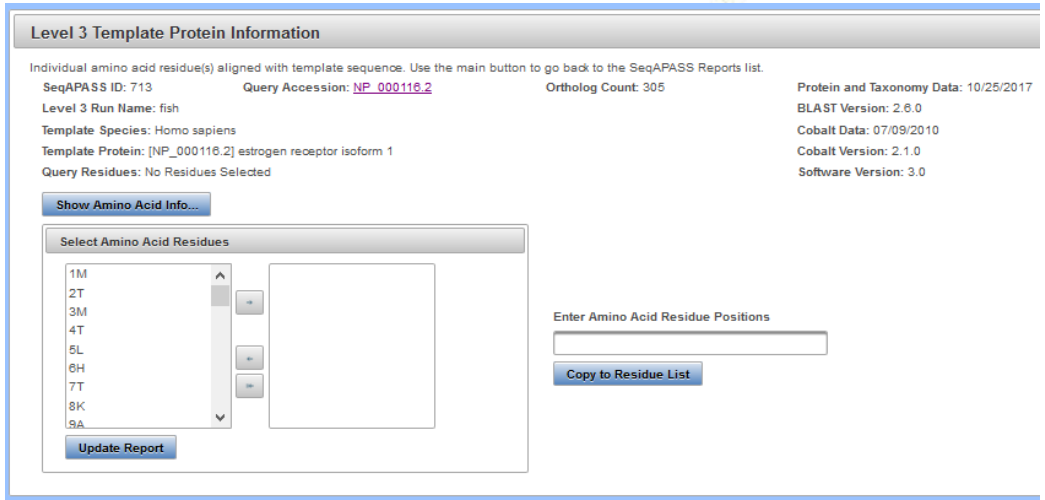
Jon A. Doering,[*,†] Sehan Lee,[‡,§] Kurt Kristiansen,[¶] Linn Evenseth,[¶] Mace G. Barron,[‡] Ingebrigt Sylte,[¶] and Carlie A. LaLone[*,1]

Acetylcholinesterase



Ecdysone receptor





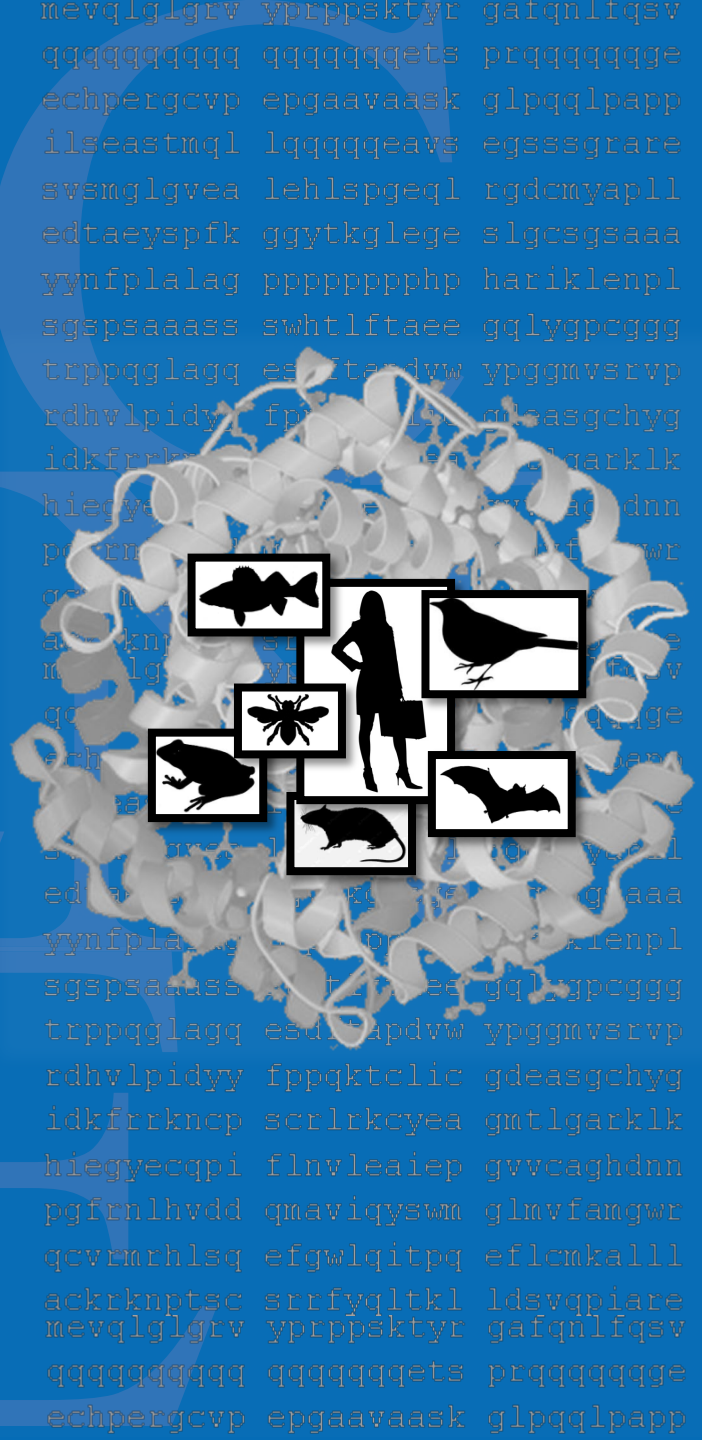
Application of SeqAPASS

SeqAPASS Applications: Case Studies

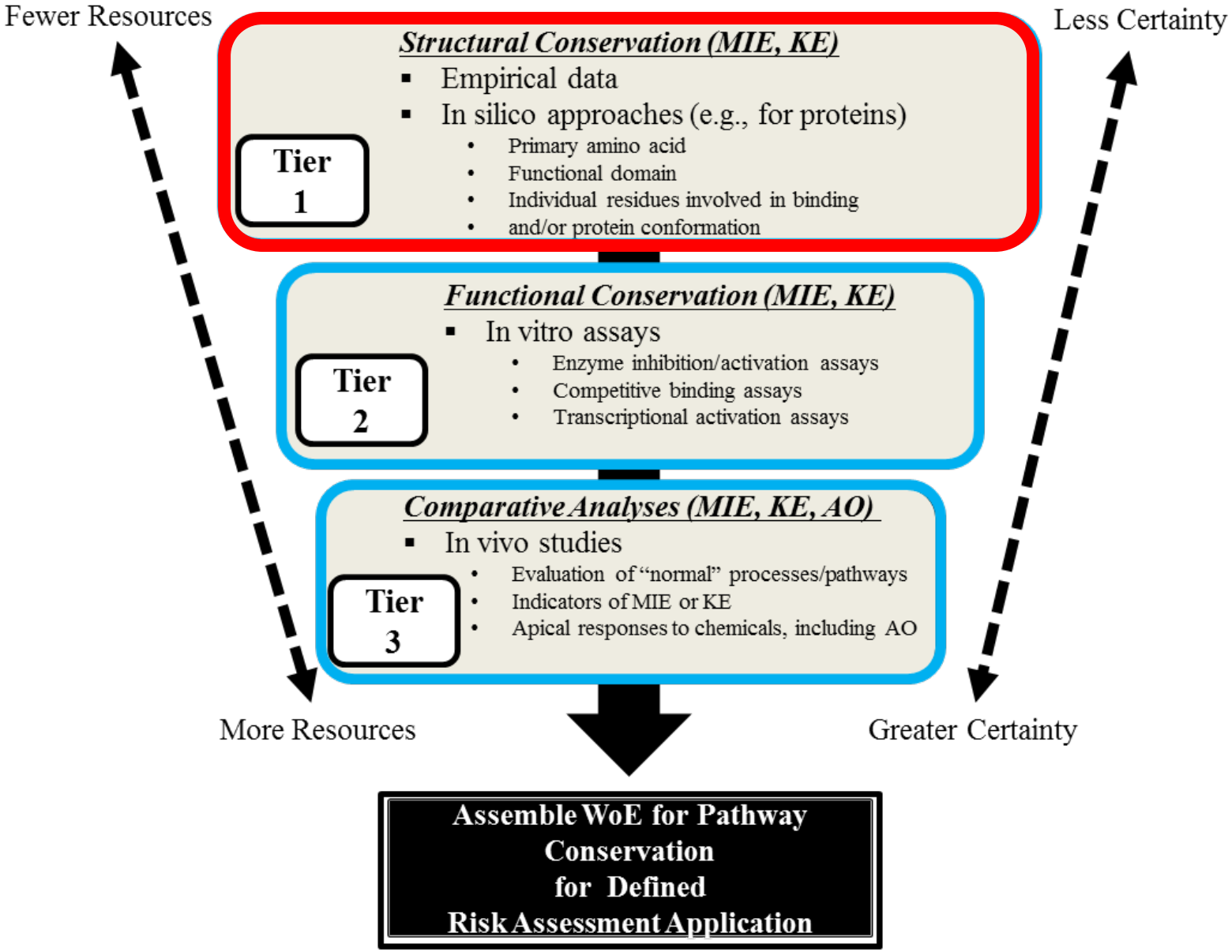
- **Extrapolate adverse outcome pathway knowledge across species**
  - Organophosphates and Carbamates
  - Conazole fungicides
  - Invertebrate endocrine disruptors
- **Extrapolate high throughput screening data**
  - Chemicals that target human estrogen receptor alpha, androgen receptor, steroidogenic enzymes, thyroid axis proteins, and all ToxCast Assay targets
- **Predict relative intrinsic susceptibility**
  - Pesticides: Neonicotinoids, permethrin, methoxyfenozide, and tebufenozide
  - Pharmaceuticals: ethinyl estradiol, 17β-trenbolone, spironolactone
- **Predict chemical bioaccumulation across species**
  - PFAS
- **Generate research hypotheses** Strobilurin fungicides
- **Prioritization strategies** Pharmaceuticals

United States Environmental Protection Agency

Application of SeqAPASS to evaluate HTS assays



Hierarchal Framework for Evaluating Pathway Conservation

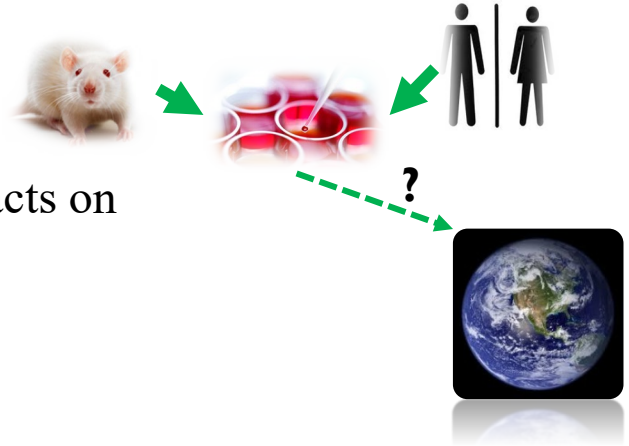


Ankley et al., 2016. EVALUATION OF THE SCIENTIFIC UNDERPINNINGS FOR IDENTIFYING ESTROGENIC CHEMICALS IN NONMAMMALIAN TAXA USING MAMMALIAN TEST SYSTEMS. ET&C

Active endocrine disruptor screening and testing program (EDSP)

- **Legislated mandate to US EPA: evaluate ~10,000 chemicals for possible endocrine activity**
- **Need to <u>prioritize</u> chemicals for resource-intensive Tier 1 screening and Tier 2 testing (EDSP21)**
  - U.S. EPA ToxCast
    - Screen chemicals – Identify chemicals most likely to be endocrine disruptors (Estrogen receptor, Androgen receptor, Steroidogenesis, Thyroid)

Key Question for Consideration:



Does this mammalian-based prioritization approach reasonably reflect potential impacts on other vertebrates?

- Estrogen receptor-α ✓

Homo sapiens Estrogen Receptor alpha

SeqAPASS Results

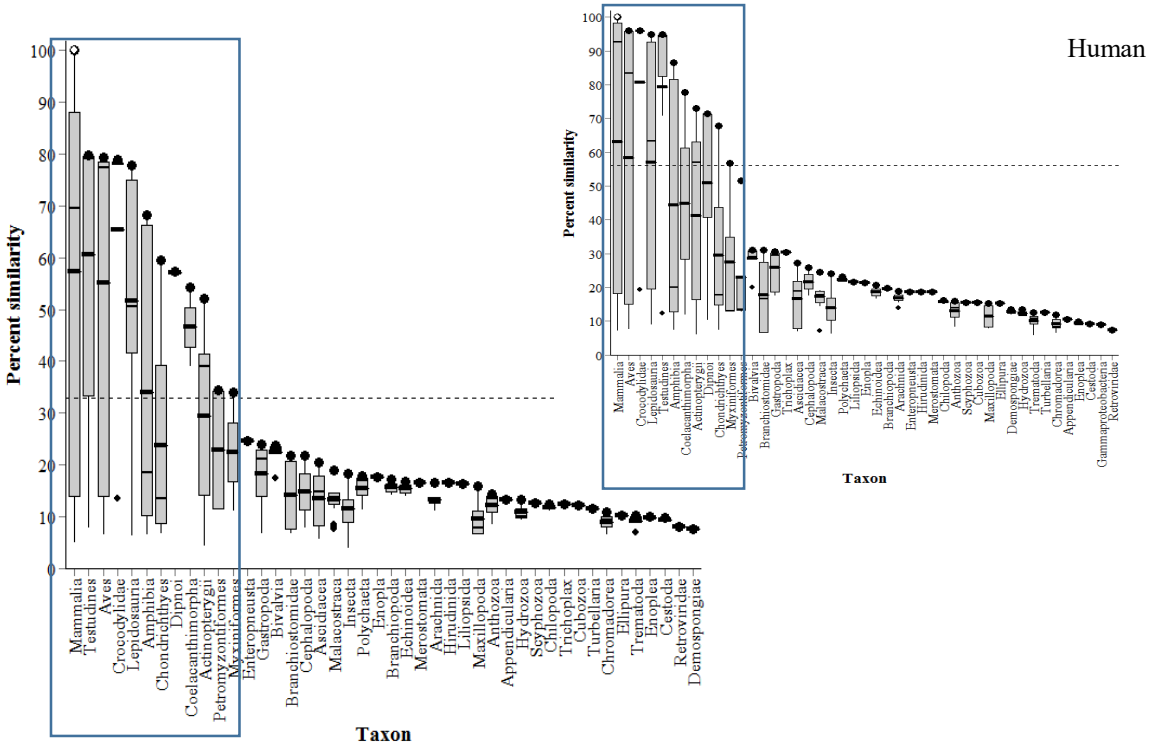
Tier 1

Structural Conservation (MIE, KE)

- Empirical data
- In silico approaches (e.g., for proteins)
  - Primary amino acid
  - Functional domain
  - Individual residues involved in binding
  - and/or protein conformation

Level 1: Primary Amino Acid Sequence

Level 2: Ligand Binding Domain



Human

Level 3: Individual Amino Acid Residues

Species Name	Class	Protein Name	D351	E353	K362	V364	R394	H524
Homo sapiens	Mammalia	estrogen receptor isoform 1	D	E	K	V	R	H
Pan paniscus	Mammalia	estrogen receptor alpha	D	E	K	V	R	H
Papio anubis	Mammalia	estrogen receptor	D	E	K	V	R	H
Trachypithecus obscurus	Mammalia	estrogen receptor alpha	D	E	K	V	R	H
Chlorocebus aethiops	Mammalia	estrogen receptor alpha	D	E	K	V	R	H
Gyps africanus	Aves	estrogen receptor alpha	D	E	K	V	R	H
Torgos tracheliotus	Aves	estrogen receptor alpha	D	E	K	V	R	H
Gallus gallus	Aves	estrogen receptor alpha	D	E	K	V	R	H
Coturnix japonica	Aves	estrogen receptor alpha	D	E	K	V	R	H
Zonotrichia albicollis	Aves	estrogen receptor 1	D	E	K	V	R	H
Taeniopygia guttata	Aves	estrogen receptor	D	E	K	V	R	H
Crocodylus niloticus	Crocodylidae	estrogen receptor alpha	D	E	K	V	R	H
Alligator mississippiensis	Crocodylidae	estrogen receptor a	D	E	K	V	R	H
Caiman crocodilus	Crocodylidae	estrogen receptor	D	E	K	V	R	H
Elaphe quadrivirgata	Lepidosauria	estrogen receptor alpha	D	E	K	V	R	H
Protobothrops flavoviridis	Lepidosauria	estrogen receptor alpha	D	E	K	V	R	H
Aspidoscelis uniparens	Lepidosauria	estrogen receptor	D	E	K	V	R	H
Eublepharis macularius	Lepidosauria	estrogen receptor alpha	D	E	K	V	R	H
Gekko swinhonis	Lepidosauria	estrogen receptor alpha	D	E	K	V	R	H
Pseudemys nelsoni	Testudines	estrogen receptor alpha	D	E	K	V	R	H
Chrysemys picta	Testudines	estrogen receptor	D	E	K	V	R	H
Lepidochelys olivacea	Testudines	estrogen receptor alpha	D	E	K	V	R	H
Andrias japonicus	Amphibia	estrogen receptor alpha	D	E	K	V	R	H
Xenopus laevis	Amphibia	estrogen receptor 1	D	E	K	V	R	H
Xenopus (Silurana) tropicalis	Amphibia	estrogen receptor alpha	D	E	K	V	R	H
Amietophrynus rangeri	Amphibia	estrogen receptor alpha	D	E	K	V	R	H
Glandirana rugosa	Amphibia	estrogen receptor 1	D	E	K	V	R	H
Pimephales promelas	Actinopterygii	estrogen receptor alpha	D	E	K	V	R	H
Salmo salar	Actinopterygii	estrogen receptor	D	E	K	V	R	H
Oryzias latipes	Actinopterygii	estrogen receptor	D	E	K	L	R	H
Oncorhynchus mykiss	Actinopterygii	estrogen receptor	D	E	K	V	R	H
Danio rerio	Actinopterygii	estrogen receptor 1	D	E	K	V	R	H
Protopterus annectens	Dipnoi	estrogen receptor alpha	D	E	K	V	R	H
Protopterus dolloi	Dipnoi	estrogen receptor alpha	D	E	K	V	R	H

Conservation across vertebrates consistent with published literature

Ankley et al., 2016. EVALUATION OF THE SCIENTIFIC UNDERPINNINGS FOR IDENTIFYING ESTROGENIC CHEMICALS IN NONMAMMALIAN TAXA USING MAMMALIAN TEST SYSTEMS. ET&C

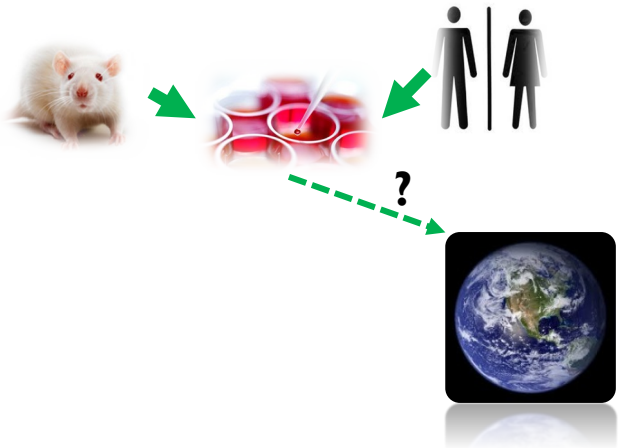
Active endocrine disruptor screening and testing program (EDSP)

- **Legislated mandate to US EPA: evaluate ~10,000 chemicals for possible endocrine activity**
- **Need to <u>prioritize</u> chemicals for resource-intensive Tier 1 screening and Tier 2 testing (EDSP21)**
  - U.S. EPA ToxCast
    - Screen chemicals – Identify chemicals most likely to be endocrine disruptors (Estrogen receptor, Androgen receptor, Steroidogenesis, Thyroid)

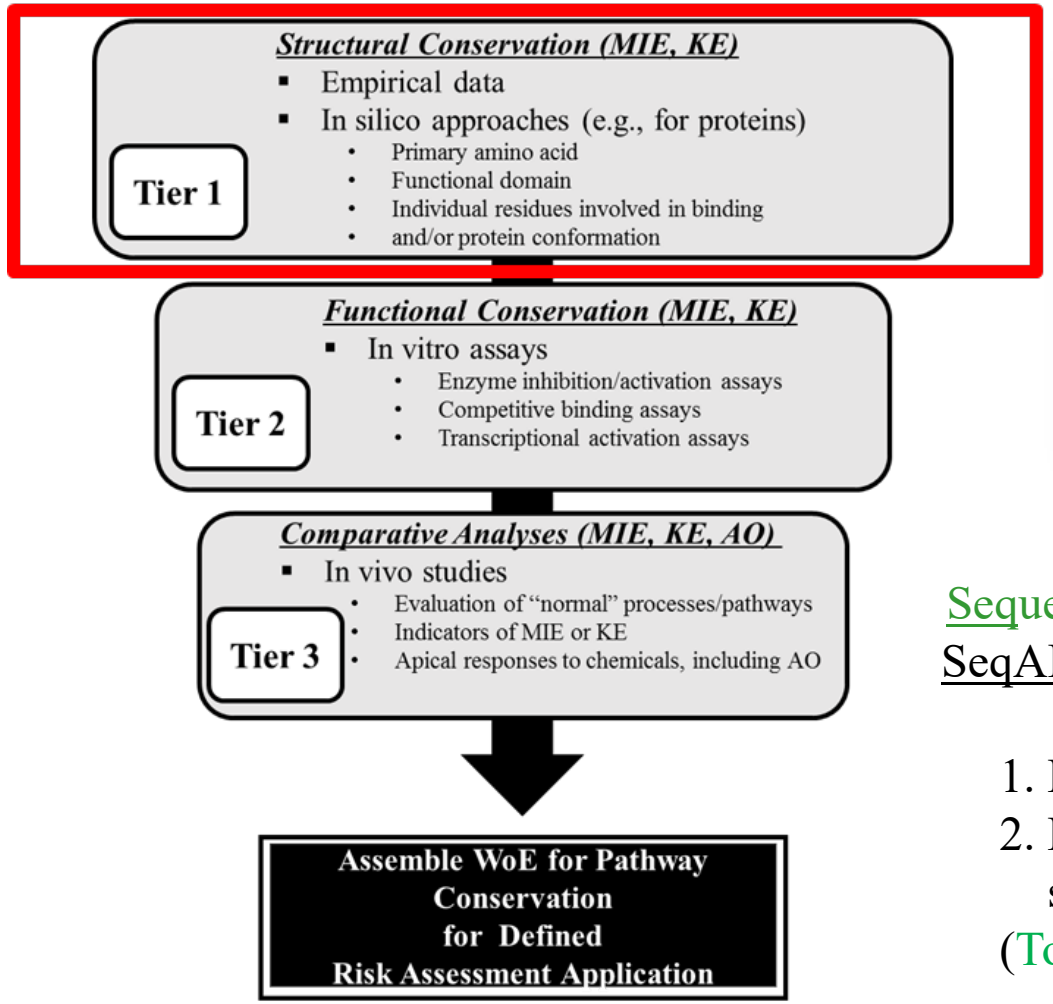
Key Question for Consideration:

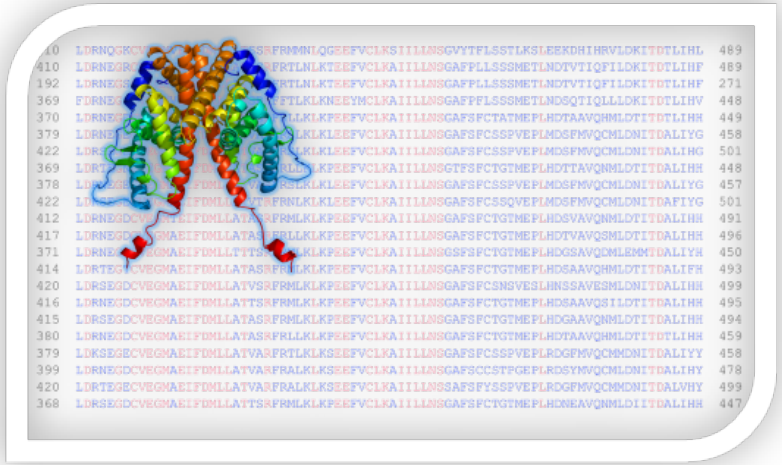
Does this mammalian-based prioritization approach reasonably reflect potential impacts on other vertebrates?

- Estrogen receptor-α
- Steroidogenic CYPs
- Thyroid axis
- Androgen Receptor



Tier 1: Evaluation of Structural Conservation of the MIE





Sequence Alignment to Predict Across Species Susceptibility

SeqAPASS: Computational Assessment of Protein Similarity: Quantitative Metrics

1. Must know the molecular target (ToxCast Assay Target)
2. Must identify target species or have knowledge of sensitive species

(ToxCast Assay Model Organism)

United States Environmental Protection Agency

HTS Assays – SeqAPASS Level 1 & 2

Tier 1

Structural Conservation (MIE, KE)

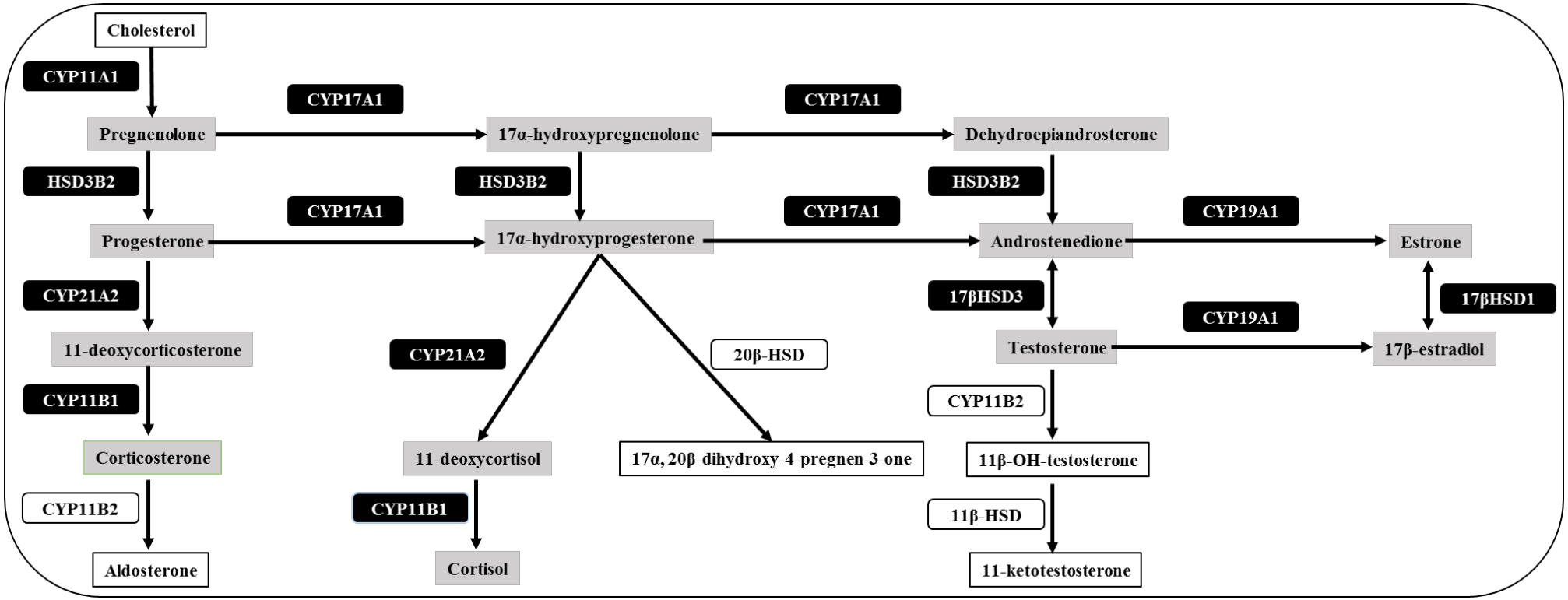
- Empirical data
- In silico approaches (e.g., for proteins)
  - Primary amino acid
  - Functional domain
  - Individual residues involved in binding
  - and/or protein conformation

Conserved across vertebrates

Assay Name	Assay Target	Model organism	SeqAPASS Query Accession	Domain evaluated (NCBI Accession)
ATG_TRANS	Androgen receptor, AR	Human (*Homo sapiens*)	P10275	Ligand binding domain (cd07073)
NVS_NR_hAR	Androgen receptor, AR	Human (*Homo sapiens*)	P10275	Ligand binding domain (cd07073)
OT_AR_ARELUC_AG_1440	Androgen receptor, AR	Human (*Homo sapiens*)	P10275	Ligand binding domain (cd07073)
OT_AR_ARSRC1_0480	Androgen receptor, AR	Human (*Homo sapiens*)	P10275	Ligand binding domain (cd07073)
OT_AR_ARSRC1_0960	Androgen receptor, AR	Human (*Homo sapiens*)	P10275	Ligand binding domain (cd07073)
TOX21_AR_BLA_Agonist	Androgen receptor, AR	Human (*Homo sapiens*)	P10275	Ligand binding domain (cd07073)
TOX21_AR_BLA_Antagonist	Androgen receptor, AR	Human (*Homo sapiens*)	P10275	Ligand binding domain (cd07073)
TOX21_AR_LUC_MDAKB2_Agonist	Androgen receptor, AR	Human (*Homo sapiens*)	P10275	Ligand binding domain (cd07073)
TOX21_AR_LUC_MDAKB2_Antagonist	Androgen receptor, AR	Human (*Homo sapiens*)	P10275	Ligand binding domain (cd07073)
NVS_NR_cAR	Androgen receptor, AR	Chimpanzee (*Pan troglodytes*)	O97775	Ligand binding domain (cd07073)
NVS_NR_rAR	Androgen receptor, AR	Norway rat (*Rattus norvegicus*)	P15207	Ligand binding domain (cd07073)
HT-H295R	Cholesterol side-chain cleavage enzyme; CYP11A1	Human (*Homo sapiens*)	P05108	NA
HT-H295R	Steroid 17α-hydroxylase/17,20 Lyase; CYP17A1	Human (*Homo sapiens*)	P05093	NA
HT-H295R	3β-hydroxysteroid dehydrogenase; HSD3B2	Human (*Homo sapiens*)	P26439	NA
HT-H295R	Aromatase; CYP19A1	Human (*Homo sapiens*)	P11511	NA
NVS_ADME_hCYP19A1	Aromatase; CYP19A1	Human (*Homo sapiens*)	P11511	NA
TOX21_Aromatase_Inhibition	Aromatase; CYP19A1	Human (*Homo sapiens*)	P11511	NA
HT-H295R	17β-hydroxysteroid dehydrogenase type1; 17βHSD1	Human (*Homo sapiens*)	P14061	NA
HT-H295R	17β-hydroxysteroid dehydrogenase type3; 17βHSD3	Human (*Homo sapiens*)	P37058	NA
HT-H295R	Steroid 21-hydroxylase; CYP21A2	Human (*Homo sapiens*)	AFK10138	NA
HT-H295R	Steroid 11-beta-hydroxylase; CYP11B1	Human (*Homo sapiens*)	AAA35741	NA
ATG_THRa1_TRANS	Thyroid hormone receptor alpha; THRA	Human (*Homo sapiens*)	P10827	Ligand binding domain (cd06935)
NVS_NR_hTRa_Antagonist	Thyroid hormone receptor alpha; THRA	Human (*Homo sapiens*)	P10827	Ligand binding domain (cd06935)
TOX21_TR_LUC_GH3_Agonist	Thyroid hormone receptor alpha; THRA	Human (*Homo sapiens*)	P10827	Ligand binding domain (cd06935)
ATG_THRb_TRANS2	Thyroid hormone receptor beta; THRB	Human (*Homo sapiens*)	P10828	Ligand binding domain (cd06935)
NCCT_TPO_AUR	Thyroid peroxidase, TPO	Norway rat (*Rattus norvegicus*)	P14650	Thyroid _peroxidase (cd09825)
NCCT_TPO_GUA	Thyroid peroxidase, TPO	Pig (Sus scrofa)	P09933	Thyroid _peroxidase (cd09825)
hNIS-HEK293T-EPA	Sodium/iodide cotransporter, NIS	Human (*Homo sapiens*)	Q92911	NA
HTS assay in development	Type I iodothyronine deiodinase, D1	Human (*Homo sapiens*)	P49895	NA
HTS assay in development	Type II iodothyronine deiodinase, D2	Human (*Homo sapiens*)	Q92813	NA
HTS assay in development	Type III iodothyronine deiodinase, D3	Human (*Homo sapiens*)	P55073	NA
HTS assay in development	Thyroid stimulating hormone receptor, TSHR	Human (*Homo sapiens*)	P16473	N-terminal domain (cd15964)
HTS assay in development	Iodotyrosine deiodinase, IYD	Human (*Homo sapiens*)	Q6PHW0	Iodotyrosine_dehalogenase domain (cd02144)

Steroidogenesis – H295R

Human H295R cell line to quantify steroid hormone production: Profiles of steroidogenic disruption



If chemicals are identified through the H295R assay that disrupt steroidogenesis, could those results be extrapolated to other vertebrates?

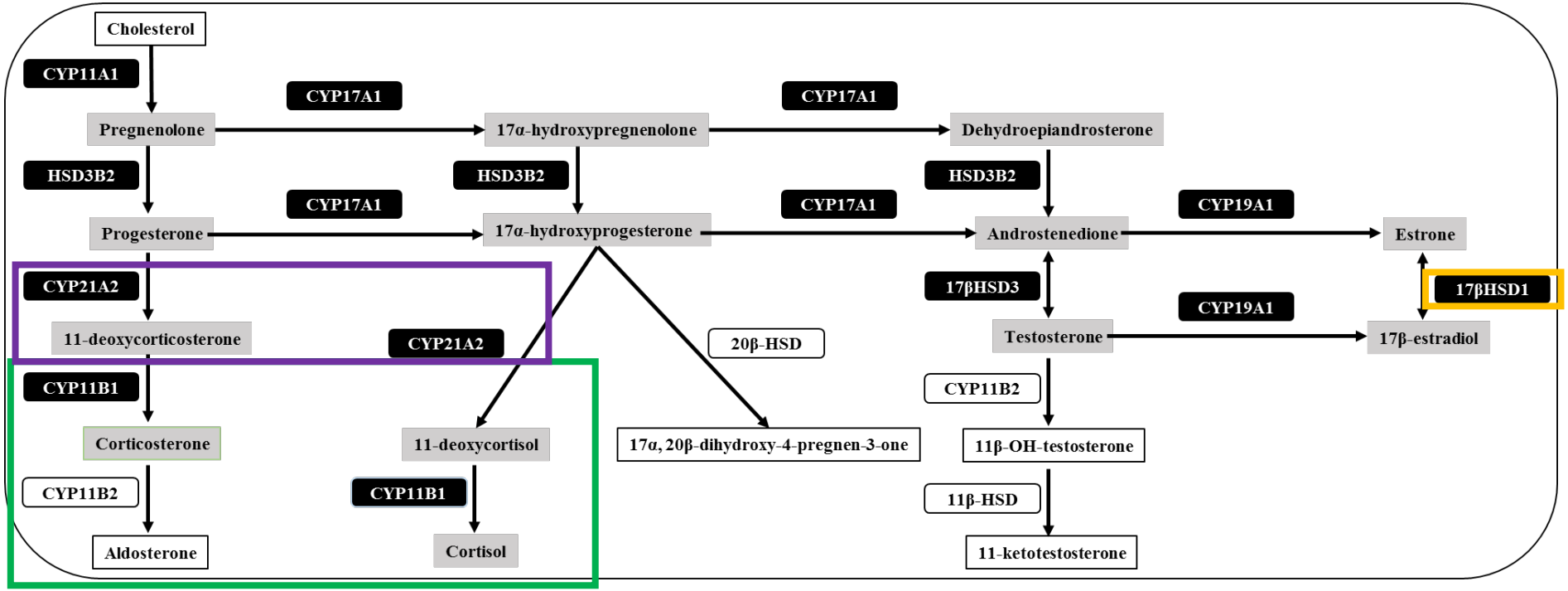
Steroidogenesis – SeqAPASS L1

	Query Protein [b]SeqAPASS Cut-off	CYP11A1 32.92	CYP17A1 30.95	HSD3B2 16.16	CYP19A1 46.79	17βHSD1 38.91	17βHSD3 38.40	CYP21A2 33.59	CYP11B1 37.60
Vertebrates	Mammalia	Yes *(115 of 127)*	Yes *(113 of 153)*	Yes *(117 of 118)*	Yes *(111 of 121)*	Yes *(110 of 116)*	Yes *(112 of 113)*	Yes *(96 of 142)*	Yes *(108 of 127)*
	Actinopteri	Yes *(58 of 74)*	Yes *(63 of 118)*	Yes *(59 of 59)*	Yes *(137 of 147)*	Yes *(19 of 49)*	Yes *(42 of 49)*	Yes *(20 of 118)*	Yes *(38 of 76)*
	Amphibia	Yes *(8 of 8)*	Yes *(5 of 6)*	Yes *(4 of 4)*	Yes *(8 of 10)*	Yes *(3 of 3)*	Yes *(3 of 4)*	Yes *(3 of 6)*	Yes *(6 of 8)*
	Aves	Yes *(45 of 65)*	Yes *(61 of 71)*	Yes *(73 of 73)*	Yes *(65 of 67)*	Yes *(16 of 66)*	Yes *(45 of 66)*	Yes *(4 of 71)*	No *(0 of 70)*
	Chondrichthyes	Yes *(4 of 5)*	Yes *(3 of 4)*	Yes *(4 of 4)*	Yes *(4 of 5)*	No *(0 of 2)*	Yes *(2 of 2)*	Yes *(1 of 4)*	No *(0 of 5)*
	Coelacanthiformes	Yes *(1 of 1)*	Yes *(1 of 1)*	Yes *(1 of 1)*	-	Yes *(1 of 1)*	Yes *(1 of 1)*	Yes *(1 of 1)*	Yes *(1 of 1)*
	Crocodylia	Yes *(2 of 4)*	Yes *(4 of 4)*	Yes *(4 of 4)*	Yes *(4 of 4)*	Yes *(3 of 3)*	Yes *(4 of 4)*	No *(0 of 4)*	Yes *(1 of 3)*
	Lepidosauria	Yes *(4 of 8)*	Yes *(7 of 8)*	Yes *(7 of 7)*	Yes *(6 of 6)*	No *(0 of 7)*	Yes *(7 of 7)*	Yes *(4 of 8)*	Yes *(5 of 8)*
	Myxiniformes	Yes *(1 of 1)*	-	-	-	-	-	-	No *(0 of 1)*
	Testudines	Yes *(3 of 3)*	Yes *(4 of 6)*	Yes *(3 of 3)*	Yes *(7 of 7)*	Yes *(2 of 3)*	Yes *(3 of 3)*	No *(0 of 6)*	Yes *(1 of 4)*
[a]Invertebrates	Anthozoa	No *(0 of 5)*	Yes *(3 of 5)*	Yes *(3 of 4)*	No *(0 of 5)*	No *(0 of 5)*	No *(0 of 5)*	No *(0 of 5)*	No *(0 of 5)*
	Arachnidia	No *(0 of 15)*	No *(0 of 11)*	Yes *(8 of 9)*	No *(0 of 15)*	No *(0 of 2)*	No *(0 of 14)*	No *(0 of 10)*	No *(0 of 15)*
	Bivalvia	No *(0 of 8)*	No *(0 of 7)*	Yes *(4 of 4)*	No *(0 of 10)*	No *(0 of 1)*	No *(0 of 4)*	No *(0 of 7)*	No *(0 of 5)*
	Branchiopoda	No *(0 of 2)*	No *(0 of 2)*	Yes *(1 of 2)*	No *(0 of 2)*	-	No *(0 of 2)*	No *(0 of 2)*	No *(0 of 2)*
	Branchiostomidae	No *(0 of 2)*	Yes *(2 of 2)*	Yes *(2 of 2)*	No *(0 of 2)*	-	No *(0 of 2)*	No *(0 of 2)*	No *(0 of 2)*
	Gastropoda	No *(0 of 5)*	No *(0 of 3)*	Yes *(2 of 3)*	No *(0 of 6)*	No *(0 of 1)*	No *(0 of 5)*	No *(0 of 3)*	No *(0 of 8)*
	Insecta	No *(0 of 176)*	No *(0 of 137)*	Yes *(8 of 56)*	No *(0 of 172)*	No *(0 of 4)*	No *(0 of 117)*	No *(0 of 135)*	No *(0 of 189)*

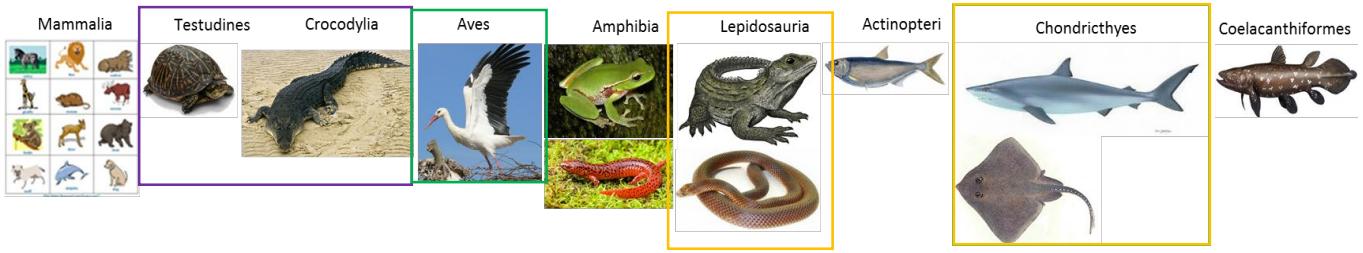
Steroidogenesis – H295R

Structural Conservation (MIE, KE)
- Empirical data
- In silico approaches (e.g., for proteins)
  - Primary amino acid
  - Functional domain
  - Individual residues involved in binding
  - and/or protein conformation

Tier 1



Steroidogenic enzymes conserved across vertebrates – Line of evidence to extrapolate H295R results



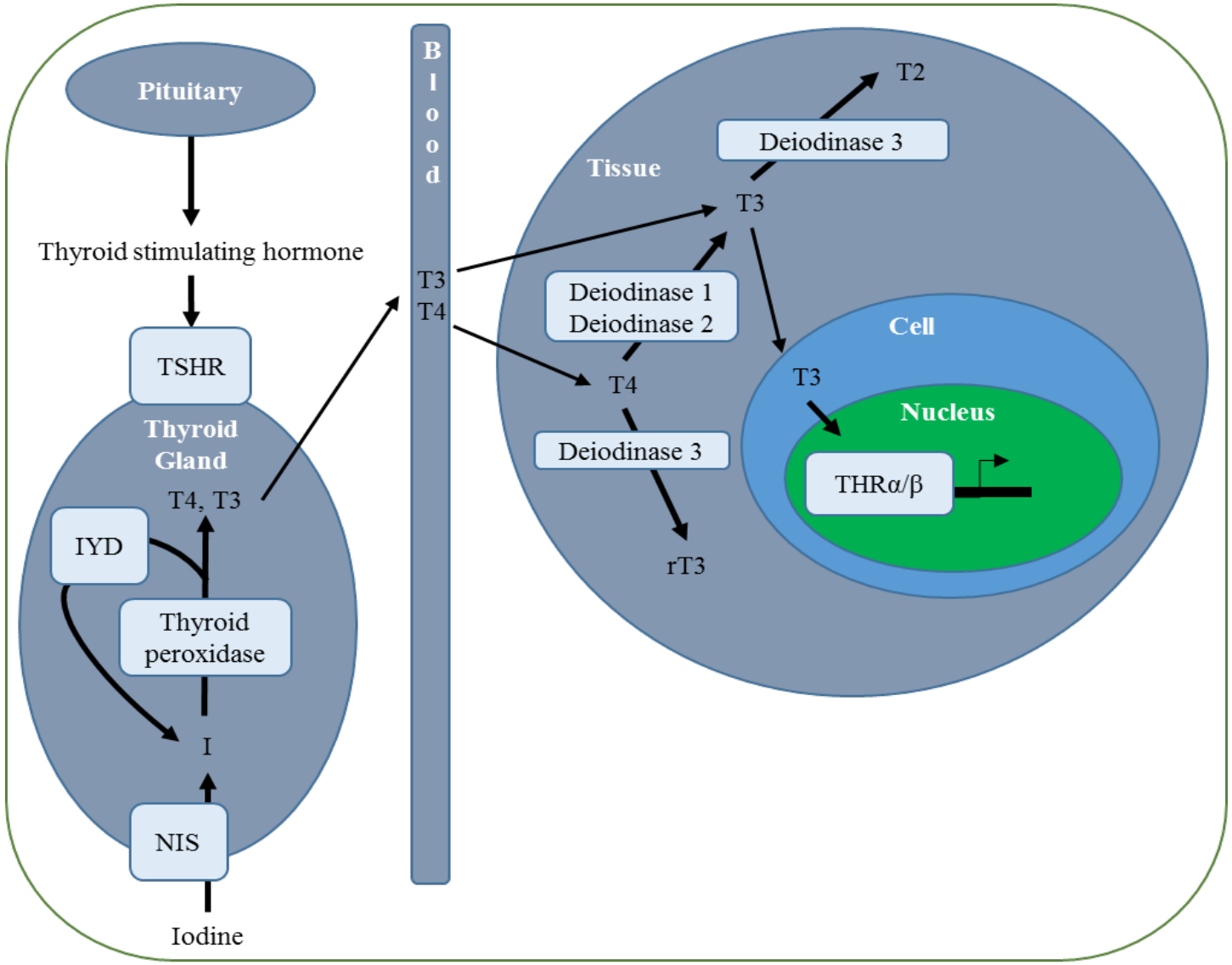
HTS Assays - SeqAPASS

Structural Conservation (MIE, KE)
- Empirical data
- In silico approaches (e.g., for proteins)
  - Primary amino acid
  - Functional domain
  - Individual residues involved in binding
  - and/or protein conformation

Tier 1

Assay Name	Assay Target	Model organism	SeqAPASS Query Accession	Domain evaluated (NCBI Accession)
ATG_TRANS	Androgen receptor, AR	Human (*Homo sapiens*)	P10275	Ligand binding domain (cd07073)
NVS_NR_hAR	Androgen receptor, AR	Human (*Homo sapiens*)	P10275	Ligand binding domain (cd07073)
OT_AR_ARELUC_AG_1440	Androgen receptor, AR	Human (*Homo sapiens*)	P10275	Ligand binding domain (cd07073)
OT_AR_ARSRC1_0480	Androgen receptor, AR	Human (*Homo sapiens*)	P10275	Ligand binding domain (cd07073)
OT_AR_ARSRC1_0960	Androgen receptor, AR	Human (*Homo sapiens*)	P10275	Ligand binding domain (cd07073)
TOX21_AR_BLA_Agonist	Androgen receptor, AR	Human (*Homo sapiens*)	P10275	Ligand binding domain (cd07073)
TOX21_AR_BLA_Antagonist	Androgen receptor, AR	Human (*Homo sapiens*)	P10275	Ligand binding domain (cd07073)
TOX21_AR_LUC_MDAKB2_Agonist	Androgen receptor, AR	Human (*Homo sapiens*)	P10275	Ligand binding domain (cd07073)
TOX21_AR_LUC_MDAKB2_Antagonist	Androgen receptor, AR	Human (*Homo sapiens*)	P10275	Ligand binding domain (cd07073)
NVS_NR_cAR	Androgen receptor, AR	Chimpanzee (*Pan troglodytes*)	O97775	Ligand binding domain (cd07073)
NVS_NR_rAR	Androgen receptor, AR	Norway rat (*Rattus norvegicus*)	P15207	Ligand binding domain (cd07073)
HT-H295R	Cholesterol side-chain cleavage enzyme; CYP11A1	Human (*Homo sapiens*)	P05108	NA
HT-H295R	Steroid 17α-hydroxylase/17,20 Lyase; CYP17A1	Human (*Homo sapiens*)	P05093	NA
HT-H295R	3β-hydroxysteroid dehydrogenase; HSD3B2	Human (*Homo sapiens*)	P26439	NA
HT-H295R	Aromatase; CYP19A1	Human (*Homo sapiens*)	P11511	NA
NVS_ADME_hCYP19A1	Aromatase; CYP19A1	Human (*Homo sapiens*)	P11511	NA
TOX21_Aromatase_Inhibition	Aromatase; CYP19A1	Human (*Homo sapiens*)	P11511	NA
HT-H295R	17β-hydroxysteroid dehydrogenase type1; 17βHSD1	Human (*Homo sapiens*)	P14061	NA
HT-H295R	17β-hydroxysteroid dehydrogenase type3; 17βHSD3	Human (*Homo sapiens*)	P37058	NA
HT-H295R	Steroid 21-hydroxylase; CYP21A2	Human (*Homo sapiens*)	AFK10138	NA
HT-H295R	Steroid 11-beta-hydroxylase; CYP11B1	Human (*Homo sapiens*)	AAA35741	NA
ATG_THRa1_TRANS	Thyroid hormone receptor alpha; THRA	Human (*Homo sapiens*)	P10827	Ligand binding domain (cd06935)
NVS_NR_hTRa_Antagonist	Thyroid hormone receptor alpha; THRA	Human (*Homo sapiens*)	P10827	Ligand binding domain (cd06935)
TOX21_TR_LUC_GH3_Agonist	Thyroid hormone receptor alpha; THRA	Human (*Homo sapiens*)	P10827	Ligand binding domain (cd06935)
ATG_THRb_TRANS2	Thyroid hormone receptor beta; THRB	Human (*Homo sapiens*)	P10828	Ligand binding domain (cd06935)
NCCT_TPO_AUR	Thyroid peroxidase, TPO	Norway rat (*Rattus norvegicus*)	P14650	Thyroid _peroxidase (cd09825)
NCCT_TPO_GUA	Thyroid peroxidase, TPO	Pig (Sus scrofa)	P09933	Thyroid _peroxidase (cd09825)
hNIS-HEK293T-EPA	Sodium/iodide cotransporter, NIS	Human (*Homo sapiens*)	Q92911	NA
HTS assay in development	Type I iodothyronine deiodinase, D1	Human (*Homo sapiens*)	P49895	NA
HTS assay in development	Type II iodothyronine deiodinase, D2	Human (*Homo sapiens*)	Q92813	NA
HTS assay in development	Type III iodothyronine deiodinase, D3	Human (*Homo sapiens*)	P55073	NA
HTS assay in development	Thyroid stimulating hormone receptor, TSHR	Human (*Homo sapiens*)	P16473	N-terminal domain (cd15964)
HTS assay in development	Iodotyrosine deiodinase, IYD	Human (*Homo sapiens*)	Q6PHW0	Iodotyrosine_dehalogenase domain (cd02144)

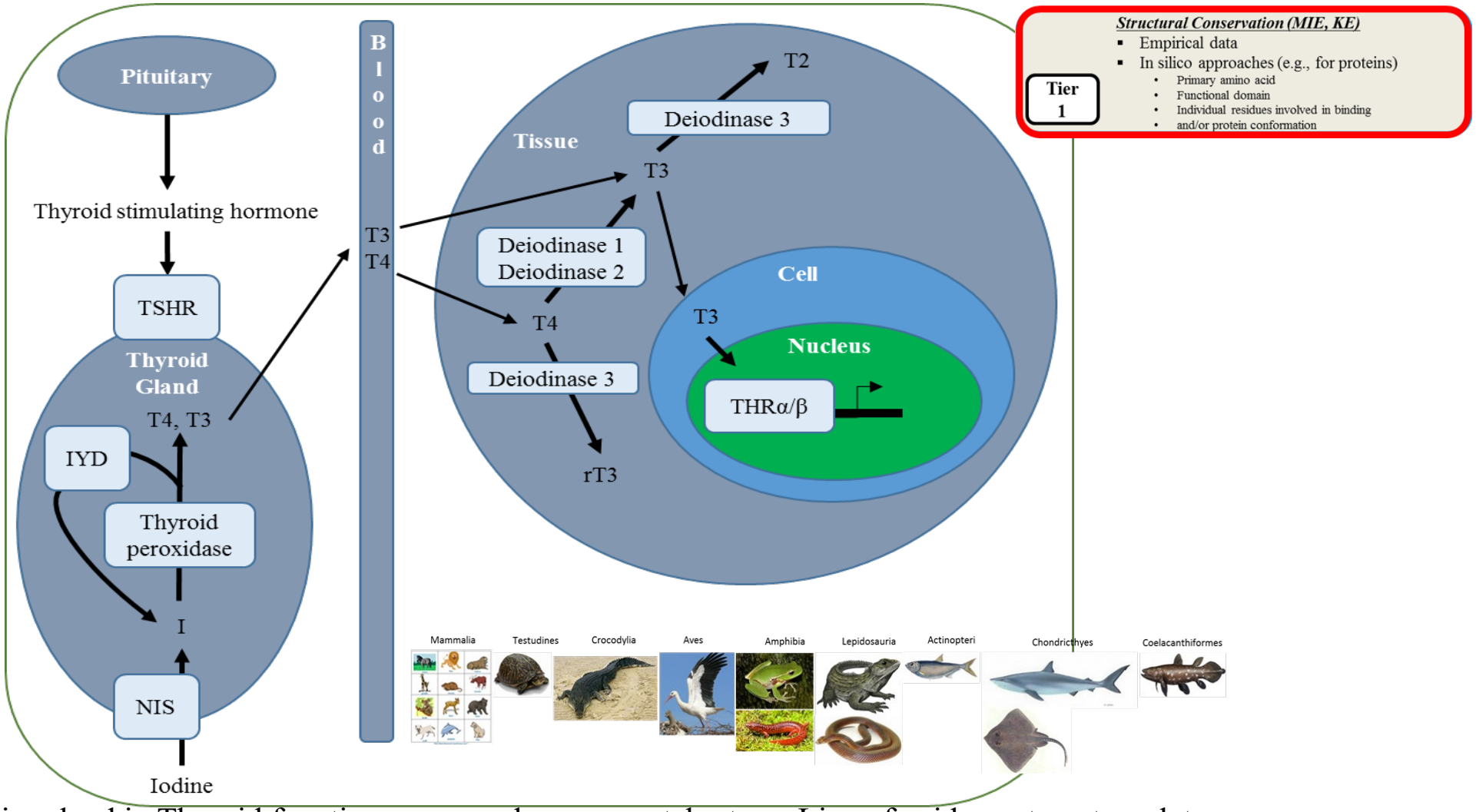
Thyroid Axis – HTS Assays developed or in development



Thyroid Axis – SeqAPASS Level 1&2

	Query Protein	THRα		THRβ		TPO		NIS	DIO1	DIO2	DIO3	TSHR		IYD	
	SeqAPASS Level	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**1**	**1**	**1**	**1**	**2**	**1**	**2**
	[b]SeqAPASS Cutoff	34.53	70.03	33.15	28.04	38.75	54.65	44.54	25.83	40.00	16.77	48.42	82.57	24.53	33.62
Vertebrates	Mammalia	Yes *(113 of 121)*	Yes *(115 of 162)*	Yes *(115 of 167)*	Yes *(113 of 152)*	Yes *(94 of 117)*	Yes *(84 of 141)*	Yes *(100 of 110)*	Yes *(126 of 127)*	Yes *(123 of 127)*	Yes *(127 of 127)*	Yes *(114 of 157)*	Yes *(113 of 181)*	Yes *(106 of 106)*	Yes *(111 of 111)*
Vertebrates	Actinopteri	Yes *(68 of 121)*	Yes *(71 of 127)*	Yes *(71 of 120)*	Yes *(78 of 117)*	Yes *(17 of 53)*	Yes *(19 of 63)*	Yes *(48 of 63)*	Yes *(62 of 70)*	Yes *(52 of 70)*	Yes *(69 of 70)*	Yes *(52 of 1409)*	Yes *(1 of 2169)*	Yes *(45 of 45)*	Yes *(45 of 45)*
Vertebrates	Amphibia	Yes *(12 of 24)*	Yes *(13 of 29)*	Yes *(15 of 23)*	Yes *(28 of 28)*	Yes *(3 of 3)*	Yes *(3 of 3)*	Yes *(3 of 3)*	Yes *(4 of 10)*	Yes *(4 of 10)*	Yes *(7 of 10)*	Yes *(2 of 16)*	Yes *(3 of 16)*	Yes *(3 of 3)*	Yes *(3 of 3)*
Vertebrates	Aves	Yes *(68 of 80)*	Yes *(68 of 75)*	Yes *(69 of 80)*	Yes *(75 of 80)*	Yes *(58 of 68)*	Yes *(58 of 68)*	Yes *(21 of 67)*	Yes *(70 of 70)*	Yes *(63 of 70)*	Yes *(70 of 70)*	Yes *(67 of 108)*	Yes *(68 of 137)*	Yes *(66 of 137)*	Yes *(66 of 68)*
Vertebrates	Ceratodontimorpha	No *(0 of 2)*	No *(0 of 2)*	No *(0 of 2)*	No *(0 of 2)*	-	-	-	Yes *(1 of 1)*	Yes *(1 of 1)*	Yes *(1 of 1)*	No *(0 of 1)*	No *(0 of 4)*	-	-
Vertebrates	Chondrichthyes	Yes *(3 of 6)*	Yes *(3 of 6)*	Yes *(3 of 7)*	Yes *(5 of 7)*	Yes *(2 of 2)*	Yes *(2 of 3)*	Yes *(2 of 4)*	Yes *(3 of 3)*	Yes *(3 of 3)*	Yes *(3 of 3)*	Yes *(1 of 9)*	Yes *(1 of 11)*	Yes *(2 of 2)*	Yes *(2 of 2)*
Vertebrates	Coelacanthiformes	Yes *(1 of 2)*	Yes *(1 of 2)*	Yes *(1 of 2)*	Yes *(1 of 1)*	No *(0 of 1)*	No *(0 of 1)*	Yes *(1 of 1)*	Yes *(1 of 1)*	Yes *(1 of 1)*	Yes *(1 of 1)*	Yes *(1 of 1)*	Yes *(1 of 1)*	Yes *(1 of 1)*	Yes *(1 of 1)*
Vertebrates	Crocodylia	Yes *(4 of 6)*	Yes *(4 of 5)*	Yes *(4 of 6)*	Yes *(4 of 5)*	Yes *(4 of 4)*	Yes *(4 of 4)*	Yes *(3 of 4)*	Yes *(4 of 4)*	Yes *(4 of 4)*	Yes *(4 of 4)*	Yes *(4 of 4)*	Yes *(4 of 4)*	Yes *(3 of 3)*	Yes *(4 of 4)*
Vertebrates	Lepidosauria	Yes *(8 of 15)*	Yes *(8 of 13)*	Yes *(8 of 15)*	Yes *(9 of 11)*	Yes *(6 of 6)*	Yes *(6 of 7)*	Yes *(4 of 6)*	Yes *(9 of 9)*	Yes *(7 of 9)*	Yes *(8 of 9)*	Yes *(6 of 341)*	Yes *(7 of 422)*	Yes *(6 of 6)*	Yes *(6 of 6)*
Vertebrates	Myxiniformes	[c]Yes *(0 of 1)*	[c]Yes *(0 of 2)*	No *(0 of 1)*	Yes *(1 of 2)*	No *(0 of 1)*	No *(0 of 1)*	-	-	-	-	No *(0 of 1)*	No *(0 of 1)*	-	-
Vertebrates	Petromyzontiformes	Yes *(1 of 4)*	Yes *(1 of 4)*	Yes *(1 of 4)*	Yes *(4 of 4)*	-	No *(0 of 1)*	-	Yes *(1 of 1)*	Yes *(1 of 1)*	Yes *(1 of 1)*	No *(0 of 5)*	No *(0 of 5)*	-	-
Vertebrates	Testudines	Yes *(3 of 9)*	Yes *(3 of 5)*	Yes *(3 of 9)*	Yes *(3 of 4)*	Yes *(3 of 3)*	Yes *(3 of 3)*	Yes *(2 of 3)*	Yes *(3 of 3)*	Yes *(3 of 3)*	Yes *(3 of 3)*	Yes *(3 of 4)*	Yes *(3 of 26)*	Yes *(3 of 3)*	Yes *(3 of 3)*
[a]Invertebrates	Ascidiacea	No *(0 of 5)*	No *(0 of 5)*	Yes *(1 of 5)*	Yes *(2 of 5)*	No *(0 of 2)*	Yes *(1 of 2)*	No *(0 of 1)*	Yes *(2 of 2)*	No *(0 of 2)*	Yes *(2 of 2)*	No *(0 of 1)*	No *(0 of 1)*	-	-
[a]Invertebrates	Asteroidea	No *(0 of 1)*	No *(0 of 1)*	Yes *(1 of 1)*	Yes *(1 of 1)*	No *(0 of 1)*	No *(0 of 1)*	No *(0 of 1)*	No *(0 of 1)*	No *(0 of 1)*	No *(0 of 1)*	No *(0 of 2)*	No *(0 of 3)*	Yes *(1 of 1)*	Yes *(1 of 1)*
[a]Invertebrates	Bivalvia	No *(0 of 10)*	No *(0 of 10)*	Yes *(2 of 10)*	Yes *(3 of 10)*	No *(0 of 6)*	No *(0 of 8)*	No *(0 of 4)*	Yes *(4 of 5)*	No *(0 of 5)*	Yes *(4 of 5)*	No *(0 of 4)*	No *(0 of 5)*	-	-
[a]Invertebrates	Branchiostomidae	No *(0 of 4)*	No *(0 of 4)*	Yes *(3 of 4)*	Yes *(4 of 4)*	No *(0 of 3)*	No *(0 of 4)*	No *(0 of 2)*	Yes *(2 of 2)*	No *(0 of 2)*	Yes *(2 of 2)*	No *(0 of 4)*	No *(0 of 4)*	Yes *(2 of 2)*	Yes *(2 of 2)*
[a]Invertebrates	Dictyosteliida	-	-	-	-	-	-	-	No *(0 of 6)*	No *(0 of 6)*	Yes *(1 of 5)*	-	-	-	-
[a]Invertebrates	Enteropneusta	No *(0 of 1)*	No *(0 of 1)*	Yes *(1 of 1)*	Yes *(1 of 1)*	No *(0 of 1)*	No *(0 of 1)*	No *(0 of 1)*	No *(0 of 1)*	No *(0 of 1)*	No *(0 of 1)*	No *(0 of 1)*	No *(0 of 1)*	-	Yes *(1 of 1)*
[a]Invertebrates	Gastropoda	No *(0 of 9)*	No *(0 of 9)*	Yes *(2 of 2)*	Yes *(8 of 9)*	No *(0 of 3)*	No *(0 of 4)*	No *(0 of 3)*	No *(0 of 2)*	No *(0 of 2)*	No *(0 of 2)*	No *(0 of 5)*	No *(0 of 5)*	-	-
[a]Invertebrates	Lingulata	No *(0 of 1)*	No *(0 of 1)*	Yes *(1 of 1)*	Yes *(1 of 1)*	No *(0 of 1)*	No *(0 of 1)*	No *(0 of 1)*	Yes *(1 of 1)*	No *(0 of 1)*	Yes *(1 of 1)*	[c]Yes *(0 of 1)*	[c]Yes *(0 of 1)*	Yes *(1 of 1)*	Yes *(1 of 1)*
[a]Invertebrates	Polychaeta	No *(0 of 2)*	No *(0 of 1)*	Yes *(1 of 2)*	Yes *(1 of 1)*	No *(0 of 5)*	No *(0 of 10)*	No *(0 of 1)*	No *(0 of 1)*	No *(0 of 1)*	No *(0 of 1)*	No *(0 of 2)*	No *(0 of 2)*	Yes *(1 of 1)*	Yes *(1 of 1)*

Thyroid Axis – HTS Assays developed or in development



Proteins involved in Thyroid function conserved across vertebrates – Line of evidence to extrapolate

Additional research to understand role of proteins in invertebrates

Tier 1: Evaluation of Structural Conservation of the MIE

Tier 1

Structural Conservation (MIE, KE)

- Empirical data
- In silico approaches (e.g., for proteins)
  - Primary amino acid
  - Functional domain
  - Individual residues involved in binding
  - and/or protein conformation

Tier 2

Functional Conservation (MIE, KE)

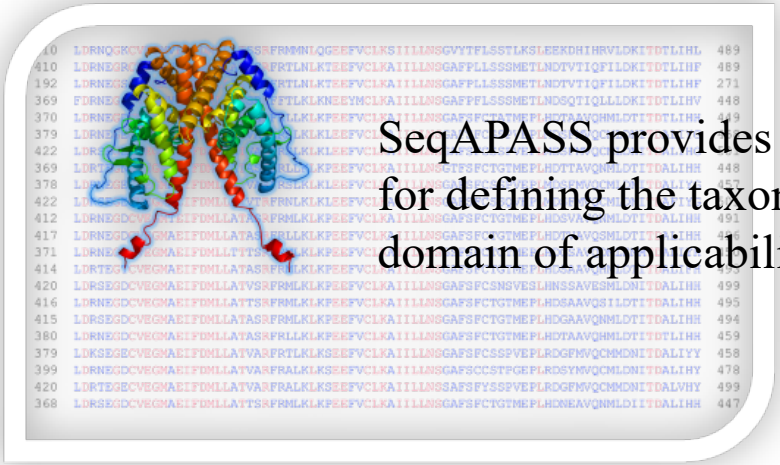
- In vitro assays
  - Enzyme inhibition/activation assays
  - Competitive binding assays
  - Transcriptional activation assays

Tier 3

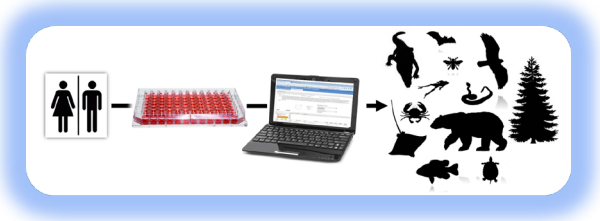
Comparative Analyses (MIE, KE, AO)

- In vivo studies
  - Evaluation of "normal" processes/pathways
  - Indicators of MIE or KE
  - Apical responses to chemicals, including AO

Assemble WoE for Pathway Conservation for Defined Risk Assessment Application



SeqAPASS provides an initial line of evidence for defining the taxonomic domain of applicability for HTS results



Environmental Science & Technology

Environ. Sci. Technol. | Environ. Sci. Technol. Lett.

Browse the Journal | Articles ASAP | Current Issue | Submission & Review | Open Access | About the Journal

Ecotoxicology and Human Environmental Health

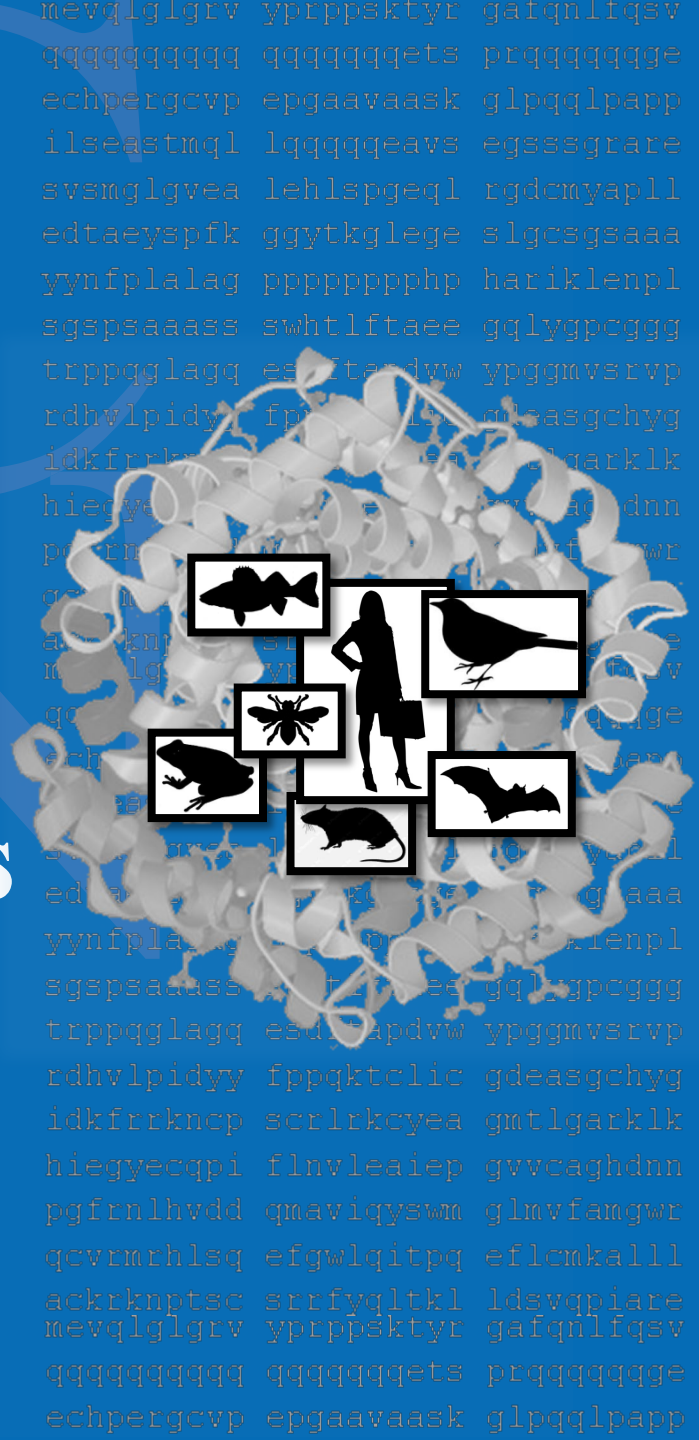
Evidence for Cross Species Extrapolation of Mammalian-Based High-Throughput Screening Assay Results

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

Environ. Sci. Technol., Just Accepted Manuscript
DOI: 10.1021/acs.est.8b04587
Publication Date (Web): October 10, 2018

United States Environmental Protection Agency

Application of SeqAPASS to Extrapolating Toxicity and Bioaccumulation of PFAS Across Species



SeqAPASS for understanding toxicity and bioaccumulation

- 75 PFAS chemicals are being evaluated in ToxCast to identify molecular targets
  - e.g. PPARα, PPARγ, CYP2C9

- Accumulation in blood and liver are largely due to protein binding
  - Serum albumin
  - Liver fatty acid binding protein (LFABP)

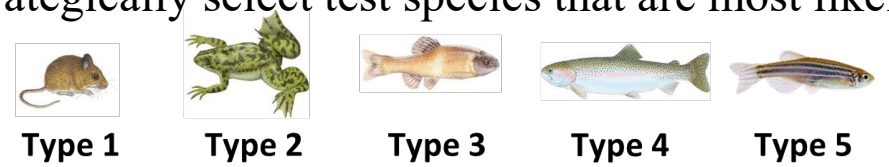
- Differences in $T_{1/2}$ due to species and gender differences in activity of renal transporters
  - Organic anion transporters (OATs)

SeqAPASS evaluation of PFAS targets

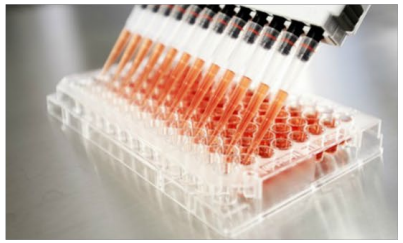
- Initial line of evidence for extrapolation of ToxCast results
  - 75 PFAS chemicals (NCCT)
  - Identify putative molecular targets
  - Use **existing SeqAPASS results** to understand conservation across species
- **PFAS specific:**
  - Expand to Level 3 individual amino acid residue comparisons for greater taxonomic resolution

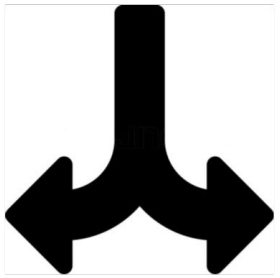
Example: Jon Doering (NRC Postdoc) using SeqAPASS to inform future testing based on PPARγ similarity

- Used SeqAPASS level 3 to identify genotypes to strategically select test species that are most likely to span the sensitivity distribution of PPARγ agonists



Cross-species *In Vitro* Screening Assays

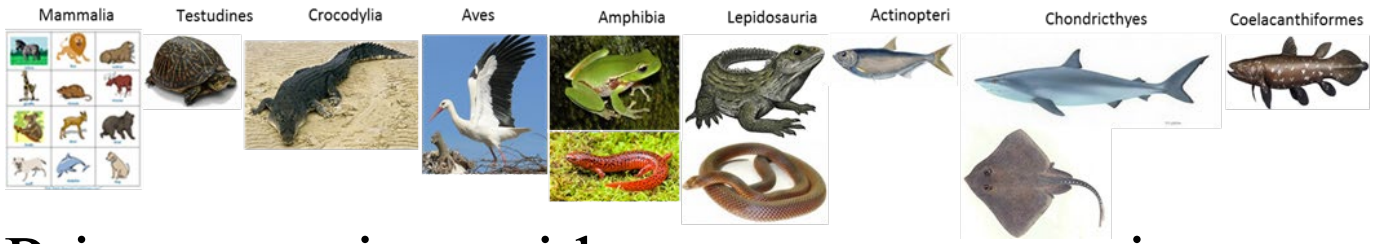




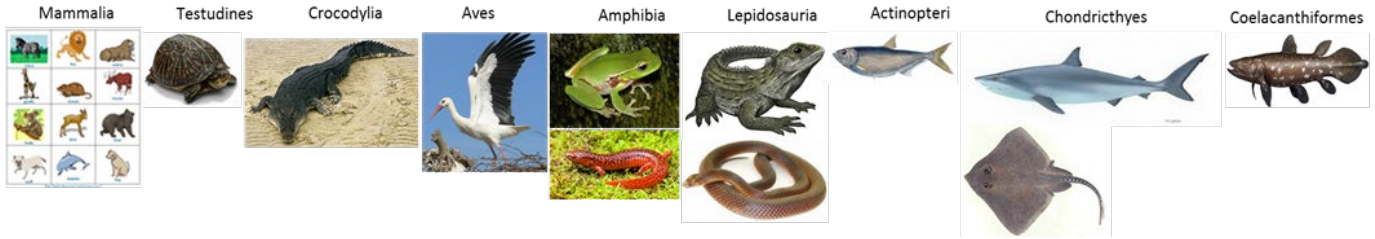
Cross-species *In Vivo* Bioassays

SeqAPASS for Bioaccumulation Across Species

- Level 1: Primary amino acid sequence comparison
  - **Human OAT1** – renal transporter, activity important for T1/2
    - Initial line of evidence: **Conserved across vertebrates**

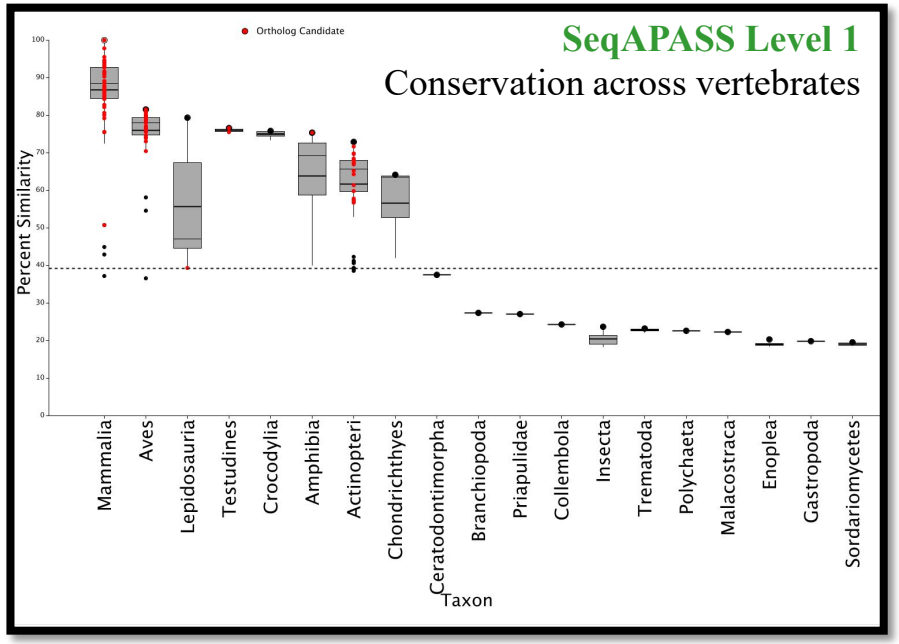


- Level 1: Primary amino acid sequence comparison
  - **Human Serum Albumin** – PFAS binding important in blood
    - Initial line of evidence: **Conserved across vertebrates**



Initial line of evidence: PFAS less likely to interact with invertebrate proteins

Human Liver Fatty Acid Binding Protein (LFABP)



ligands	hLFABP H-bond interaction	hLFABP largest energy contribution	rLFABP H-bond interaction	rLFABP largest energy contribution
PFBA	ARG 122, SER 124	ARG 122, SER 39, ILE 52	–	SER 57, LYS 58, LYS 32
PFPA	ARG 122, SER 124	ARG 122, VAL 83, PHE 50	–	ARG 122, TYR 55, ILE 53
PFHxA	ARG 122, SER 124	ARG 122, SER 39, SER 124	TYR 120	ARG 122, ILE 53, LYS 58
PFHpA	ARG 122, SER 124	ARG 122, SER 39, ILE 52	–	ARG 122, ILE 60, MET 74
PFOA	ARG 122, SER 124	ARG 122, SER 39, ILE 52	–	ARG 122, TYR 55, ILE 60
PFNA	ARG 122, SER 124	ARG 122, SER 39, ILE 52	–	ARG 122, ILE 60, ILE 53
PFBS	ARG 122, SER 39	ARG 122, SER 124, LEU 9	ARG 122, SER 39	ARG 122, SER 100, LEU 71
PFHxS	ARG 122	ARG 122, SER 124, SER 39	–	ARG 122, ASN 111, LEU 51
PFOS	ARG 122, SER 124	ARG 122, SER 124, ILE 52	TYR 120	ARG 122, ILE 60, ILE 53
EEA	ARG 122, SER 39	ARG 122, SER 39, ASN 111	–	ARG 122, MET 74, ILE 60
GenX	THR 102	ARG 122, ASN 111, THR 73	–	ARG 122, MET 74, ILE 53
ADONA	ARG 122, SER 124	ARG 122, SER 39, SER 124	–	ARG 122, MET 74, TYR 55
2m-PFOA	SER 100	ARG 122, SER 100, ASN 111	–	ARG 122, TYR 120, ILE 60
F-53	ARG 122, SER 124	ARG 122, PHE 50, SER 39	TYR 120	ARG 122, SER 124, ILE 53
F-53B	ARG 122, SER 124	ARG 122, SER 124, SER 39	TYR 120	ARG 122, TYR 55, ILE 60

Environmental Science & Technology
Article
Cite This: Environ. Sci. Technol. 2018, 52, 7972–7980
pubs.acs.org/est

Predicting Relative Protein Affinity of Novel Per- and Polyfluoroalkyl Substances (PFASs) by An Efficient Molecular Dynamics Approach

Weixiao Cheng and Carla A. Ng*

Identification of
Individual amino acid residues
Critical for binding PFAS

SeqAPASS Level 3

- Species specific predictions

Differences in amino acid residues across taxa

Class	# Species	Susceptible	50	93
Primates/Ungulates/whales/dolphins	32	Yes	F	T
Rodents/other mammals/fish/amphibians/testudines	132	No	I/L/V	T
Zebrafish, Aves, Lepidosauria, Chondrichthyes	72	No	V/I	A
Crocodylia	4	No	F	A

In silico site-directed mutagenesis and docking support predictions:

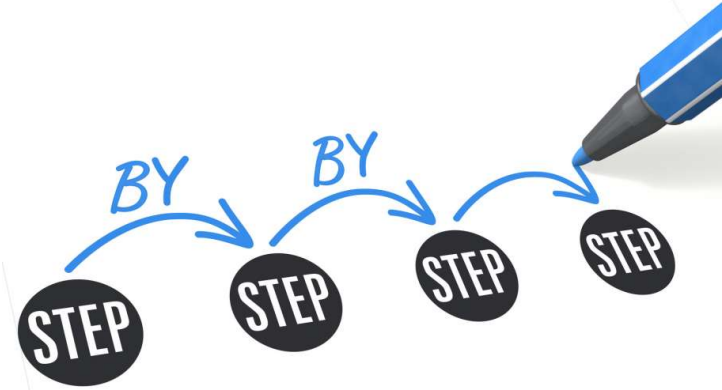
PFAS binding to zf different than other fish
Explore the prediction experimentally

≠

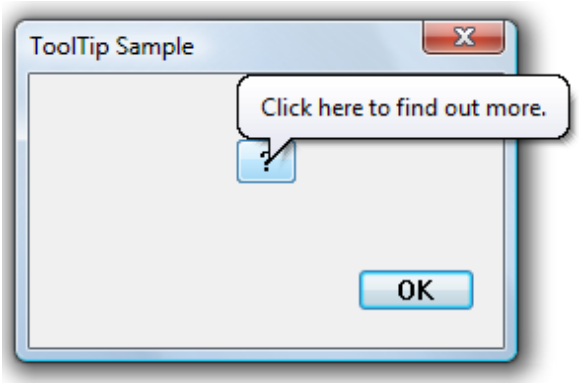
Emphasis on prediction "Validation"

1. Build the SeqAPASS tool
2. Develop case studies to demonstrate application
   - Compare predictions to available empirical data
3. Implement Training
   - Gather user information
   - Improve tool based on user feedback
4. "Validation" of SeqAPASS predictions in the laboratory: Research Efforts
   - Cross species
     - In vitro
     - In vivo
     - Chemical proteomics
5. *In silico* structural evaluation
   - Create an automated pipeline, Level 4?
     - Homology modelling
     - Docking

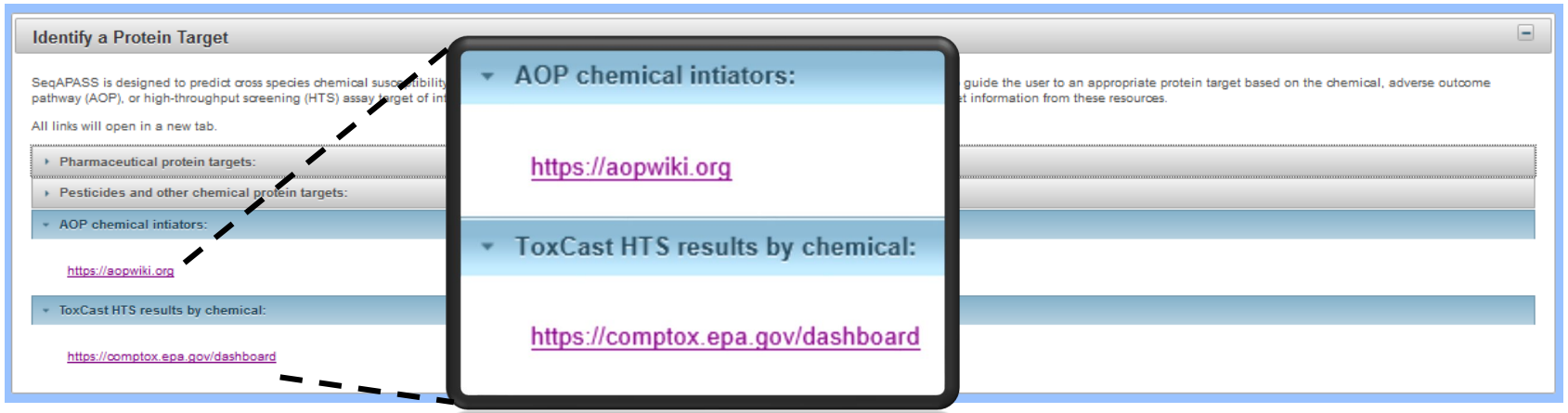
Looking for a talented postdoc



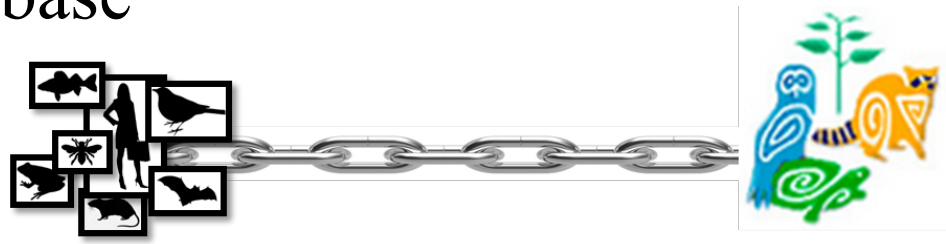
Technical Advances: FY19



- Integrated help menus, tool tips, and user guide
- Guidance for identification of protein targets and individual amino acid residues



- Risk Assessor Report and data summary tables
- Interoperability with ECOTOX Knowledgebase



Acknowledgements

U.S. EPA, ORD

Gerald Ankley
Daniel Villeneuve
Sigmund Degitz
Jon Doering
Colin Finnegan
Brett Blackwell
Michael Hornung
L. Earl Gray
Kendra Dean
Shane Poole

GDIT

Thomas Transue
Cody Simmons

Badger Technical Services

Joe Swintek

University of Pittsburg

Carla Ng
Weixiao Cheng

SeqAPASS v3.1

LaLone.Carlie@epa.gov

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