ID:1.10P.3

Protein structural similarity for extrapolation of toxicity knowledge across species

C.A. LaLone†, D. Blatz#, S.M.F. Vliet#, C. Simmons, and T. Transue

†U.S. Environmental Protection Agency, Office of Research and Development, Great Lakes Toxicology and Ecology Division, Duluth, MN 55804; #Oak Ridge Institute for Science and Education, Duluth, MN 55804;

Carlie A. LaLone | lalone.carlie@epa.gov | 218-529-5038

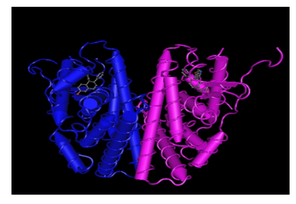
Introduction

- Bioinformatic approaches for understanding protein structural similarity are advancing rapidly. In fact, computational "virtual" screening for bioactive molecules using advanced molecular modeling approaches is common practice in the pharmaceutical industry, mitigating required time, costs, and risks for late-stage drug-candidate failure. Typically, these methods are used to understand which structures, from hundreds, can optimally bind to a given protein target to select candidates to move forward in drug development.
- In principle, these same virtual screening pipelines could be applied in support of chemical safety assessments and used to evaluate a chemical's interaction with hundreds to thousands of proteins representing both the diversity of toxicologically-relevant protein targets in the body as well as the variations in those proteins among species (e.g., vertebrates, invertebrates, plants, fungi, etc.). Here we describe a pipeline that can be used to 1) fill knowledge gaps in instances where chemical molecular targets are unknown by "virtually" identifying if and how chemical(s) bind to proteins across species and 2) provide quantitative information for chemical-protein interactions in species where limited or no chemical toxicity data exist.
- Harnessing the power of computing is the future for cross-species chemical safety screening which is why the US Environmental Protection Agency Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) tool was created and continues to evolve. SeqAPASS predicts chemical susceptibility across species based on protein-sequence similarity. The output from this tool provides an initial line of evidence for cross species extrapolation, however currently does not provide a numerical value for how well the chemical is expected to bind to the protein target in species predicted susceptible.
- Computational advances in drug-discovery are available to take this evaluation further to specifically examine protein structural interactions with chemicals using available x-ray crystallography and advanced bioinformatic methods to predict binding affinity. Such virtual methods are intended to significantly reduce the cost associated with current high-throughput in vitro screening methods and address the limitations in using model organisms which are primarily used as surrogates due to convenience or convention, rather than appropriateness. This presentation describes a computational pipeline to evaluate protein sequence and structural similarity in the context of biological pathways to gather lines of evidence for conservation of chemical targets for cross species extrapolation.

Sequence

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

Structure

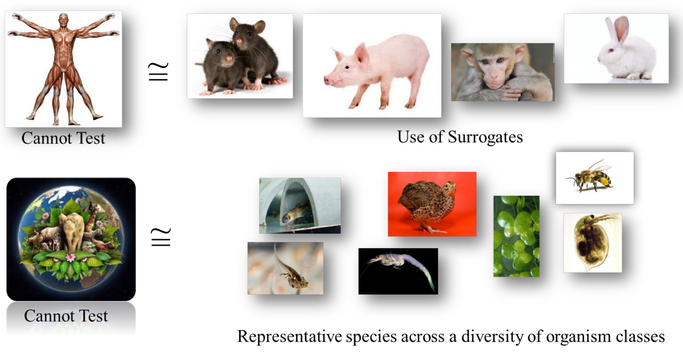


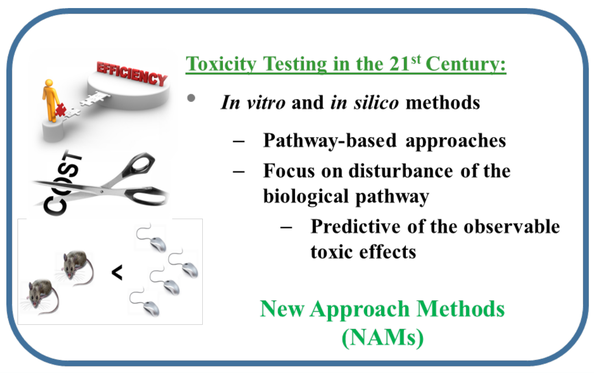
Need for Alternatives to Animal Testing

US Environmental Protection Agency Administrator Andrew Wheeler signed directive (September 10th, 2019) to Reduce Animal Testing

Calls for the Agency to:

1. Reduce its request for, and funding of, mammal studies by 30% by 2025
   - That is ~5 years from today
2. Eliminate all mammal study requests and funding by 2035
   - That is ~15 years from today

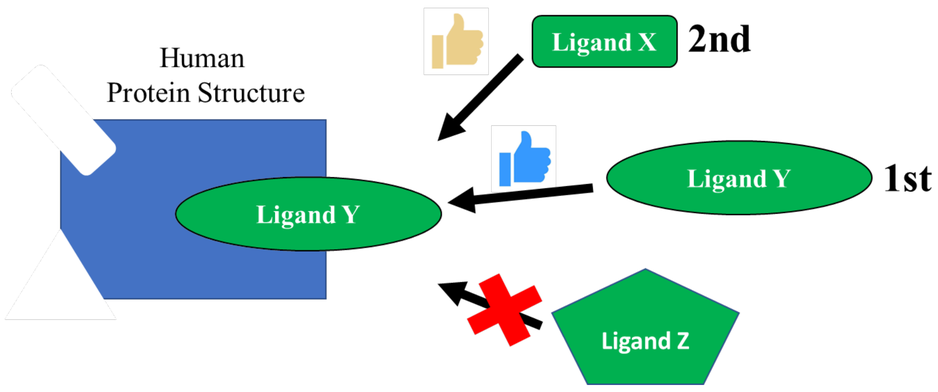




Moving away from animal testing to NEW APPROACH METHODS -Still need to understand how broadly data can be extrapolated across species

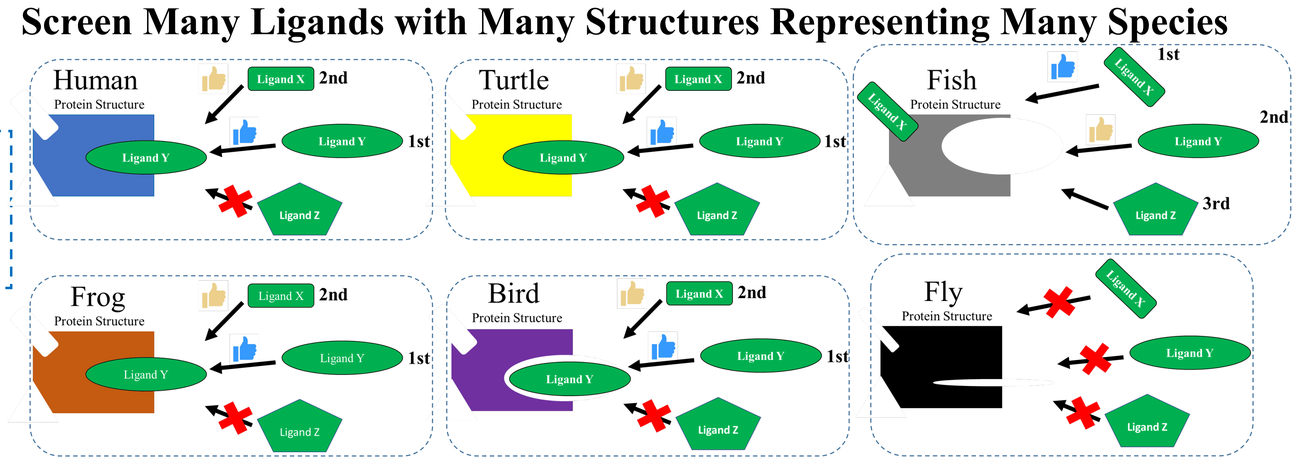
Structure-Based Virtual Screening Could Inform Cross Species Extrapolation

Screen Many Ligands with One Structure



Build Upon Advances In Bioinformatics

Screen Many Ligands with Many Structures Representing Many Species



Tools for Expansion of Sequence Based Predictions of Susceptibility to Incorporate Structure

Sequence Alignment to Predict Across Species Susceptibility [1]

SeqAPASS (https://seqapass.epa.gov/seqapass/): VERSION 4.0

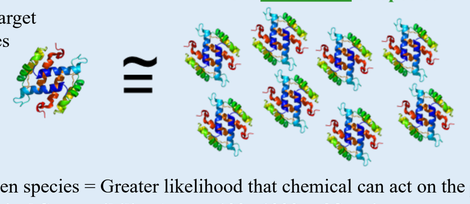
Necessary information for submitting a SeqAPASS query:

1. Must know the chemical molecular target
2. Must identify target species or have knowledge of sensitive species

} Protein Accession

Compare to Millions of Proteins From Thousands of Species

Chemical Molecular Target in Sensitive Species



Greater similarity between species = Greater likelihood that chemical can act on the protein

Predict Potential Chemical Susceptibility Across 100s-1000s of Species

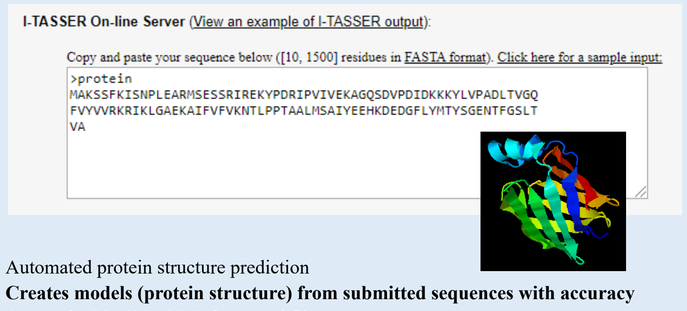
Output for Pipeline: FASTA files for each species

Iterative Threading ASSembly Refinement [2]

I-TASSER (https://zhanglab.ccmb.med.umich.edu/I-TASSER/)

Necessary information for Running I-TASSER:

1. FASTA formatted protein sequence



Automated protein structure prediction

Creates models (protein structure) from submitted sequences with accuracy

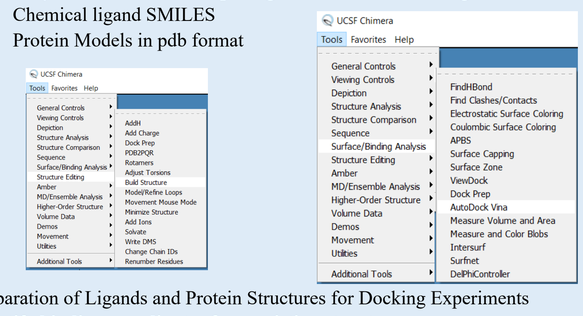
Output for Pipeline: PDB formatted file

University of California, San Francisco Chimera [3]

UCSF Chimera (https://www.cgl.ucsf.edu/chimera/)

Necessary information for Preparing for Molecular Docking:

1. Chemical ligand SMILES
2. Protein Models in pdb format

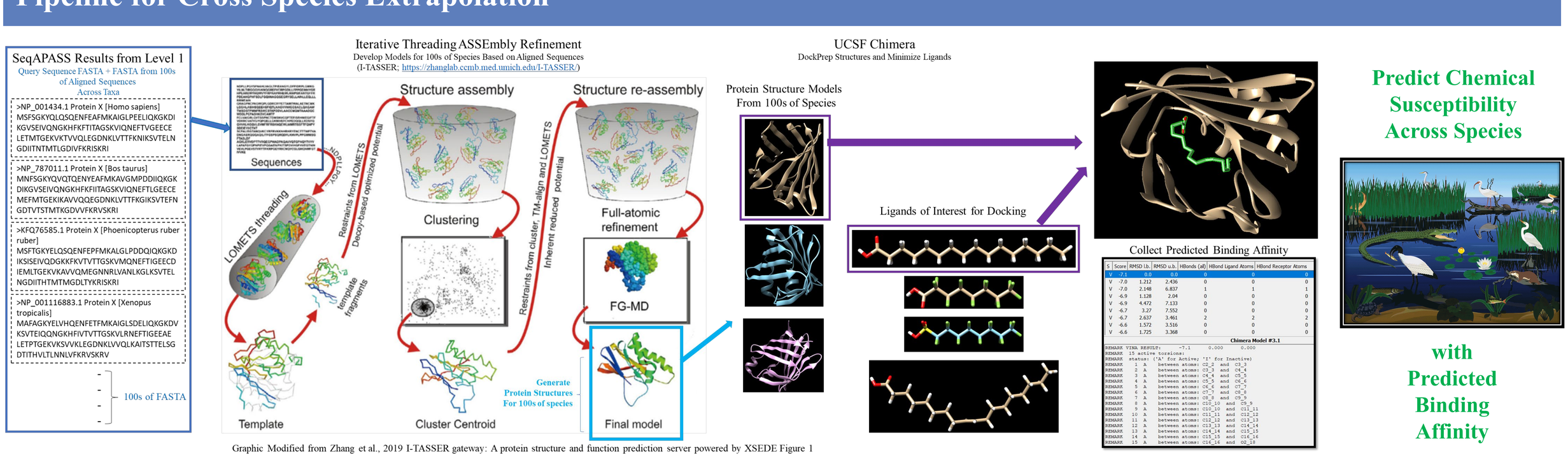


Preparation of Ligands and Protein Structures for Docking Experiments

Identify binding coordinates from existing structures

Output for Pipeline: Predicted binding affinity (ΔG (kcal/mol))

Pipeline for Cross Species Extrapolation



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

Conclusions

- The field of bioinformatics continues to advance rapidly
- Approaches used for structure-based virtual screening show promise for applications in cross species extrapolation
- Publicly accessible tools can be linked together to evaluate structural similarity and perform docking studies
- Research demonstrating the utility and limitations of the pipeline are ongoing

References / Acknowledgements

(1) LaLone, C.A. et al. (2016). Editor's Highlight: Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): A Web-Based Tool for Addressing the Challenges of Cross-Species Extrapolation of Chemical Toxicity. Toxicol Sci 153 (2), 228-245.

(2) Zhang, W. et al., (2019). I-TASSER gateway: A protein structure and function prediction server powered by XSEDE. Future Generation Computer Systems 99, 73-85.

(3) Pettersen E.F. et al. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 25 (13), 1605-1612.