



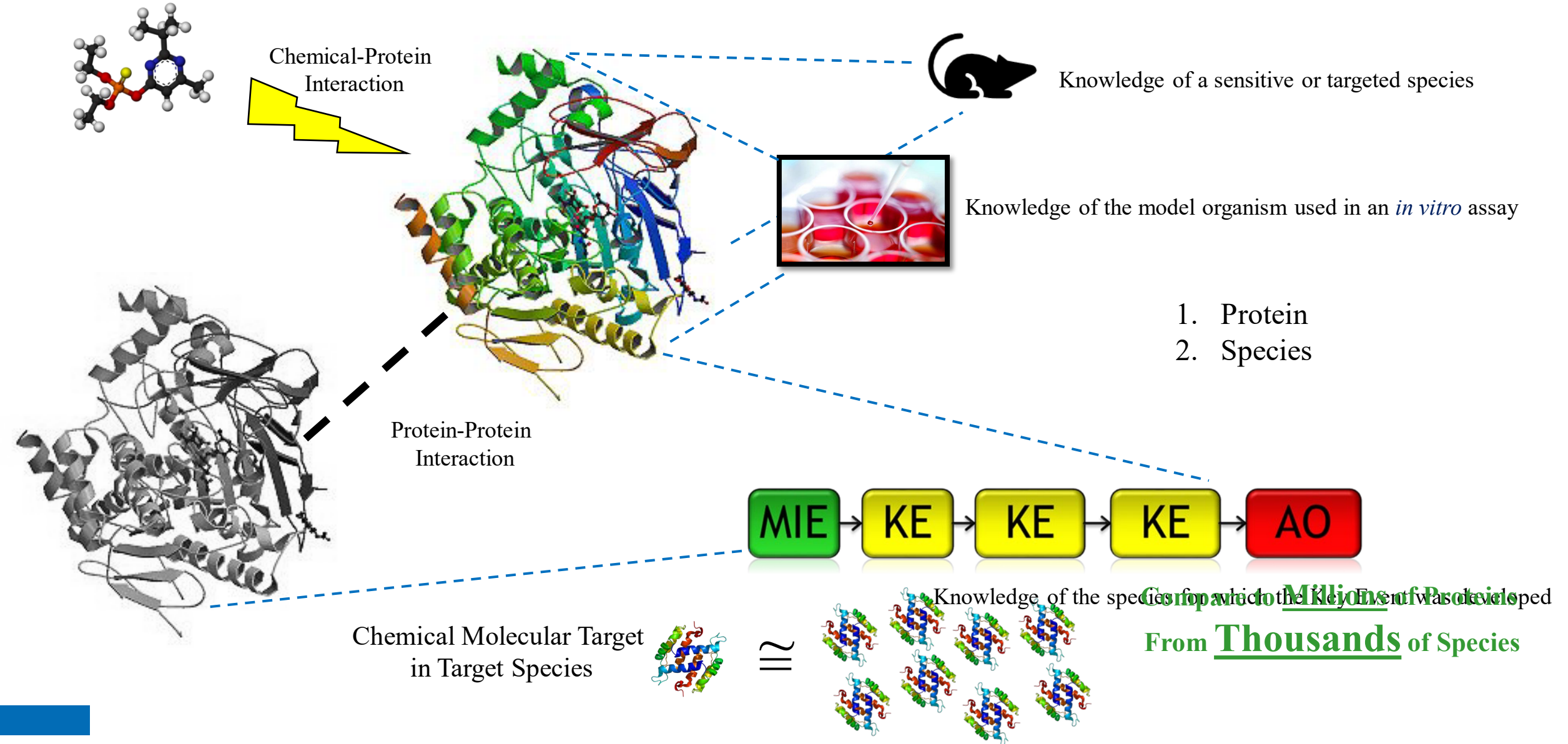
United States
Environmental Protection
Agency

SeqAPASS implements strategic connections of knowledge streams to inform species extrapolation for chemical safety

Carlie A. LaLone, Ph.D.
Research Bioinformaticist

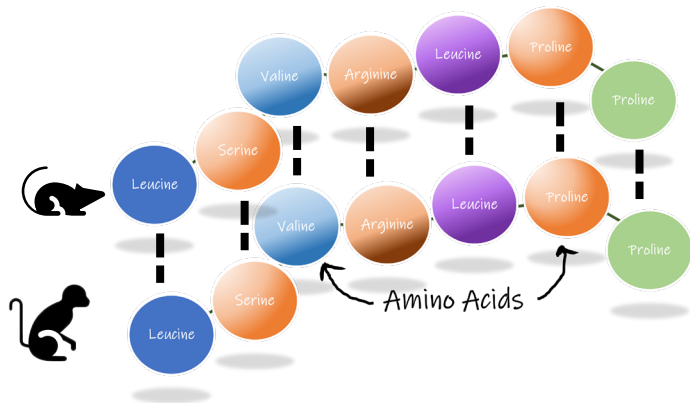


What information is required for a SeqAPASS query?



Flexible Analysis Based On Available Data

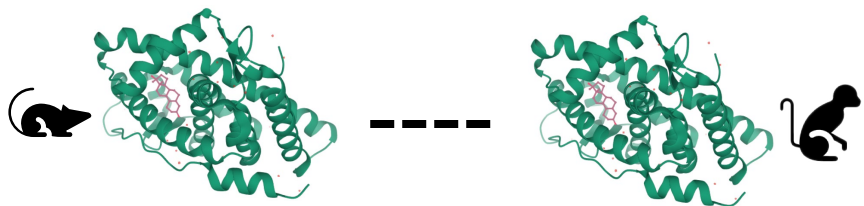
- Level 1** Primary Amino Acid Sequence Alignments
- Level 2** Conserved Functional Domain Alignments
- Level 3** Critical (Close Contact) Amino Acid Conservation



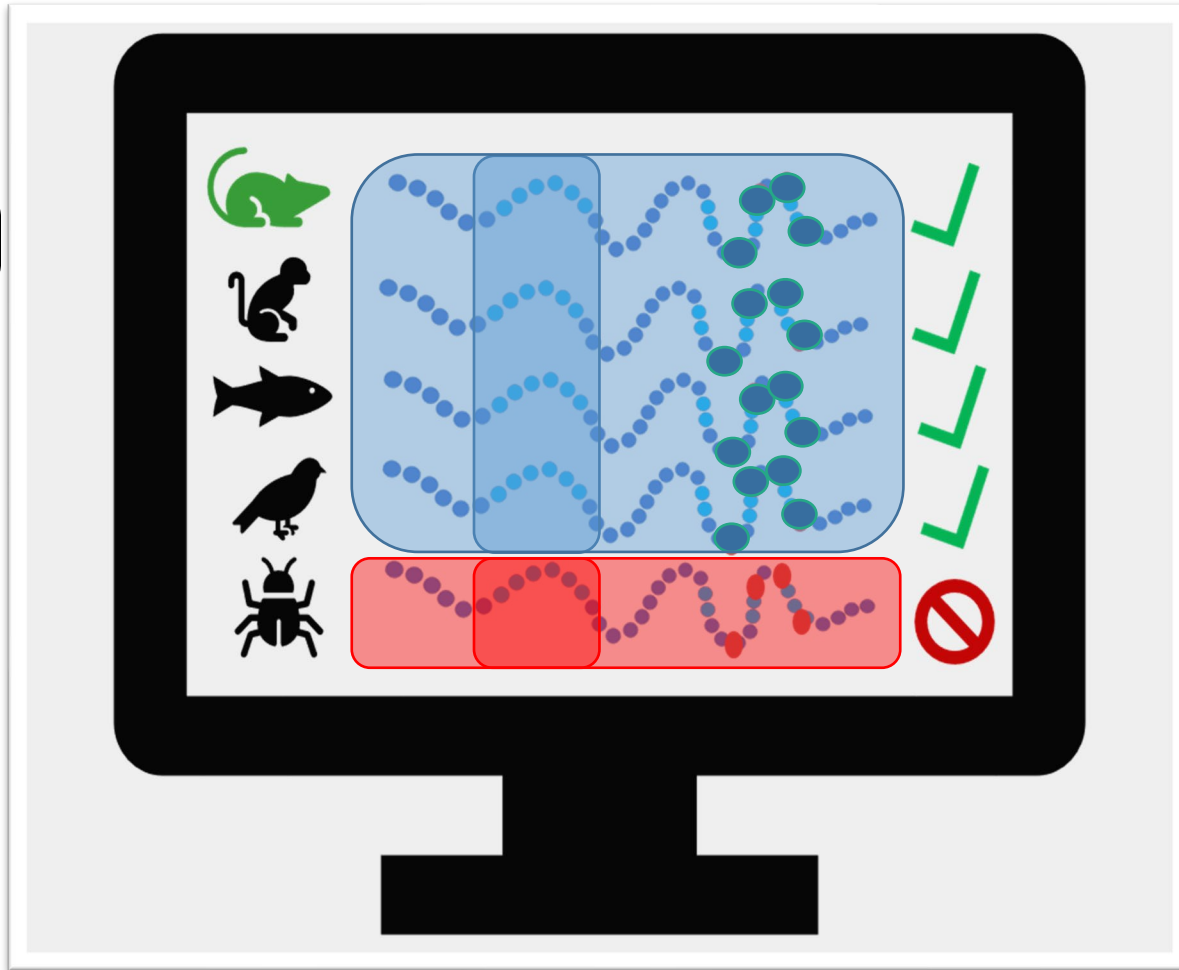
SeqAPASS v7.0
(Coming 2023)

Level 4 (expert users only)

Structure Alignments



Gather Lines of Evidence Toward Protein Conservation



Strategic Interoperability

Guide the User to Sources to Identify Protein Targets:

- DrugBank - <https://www.drugbank.ca>
- VSDB: Veterinary Substances DataBase - <http://sitem.herts.ac.uk/aeru/vsdb/index.htm>
- Therapeutic Target Database - <http://db.idrblab.net/ttd/>
- The Toxin and Toxin-Target Database - <http://www.t3db.ca>
- AOP-Wiki - <https://aopwiki.org>
- CompTox Chemicals Database - <https://comptox.epa.gov/dashboard>

Provide Transparency for Source Data and Executables:

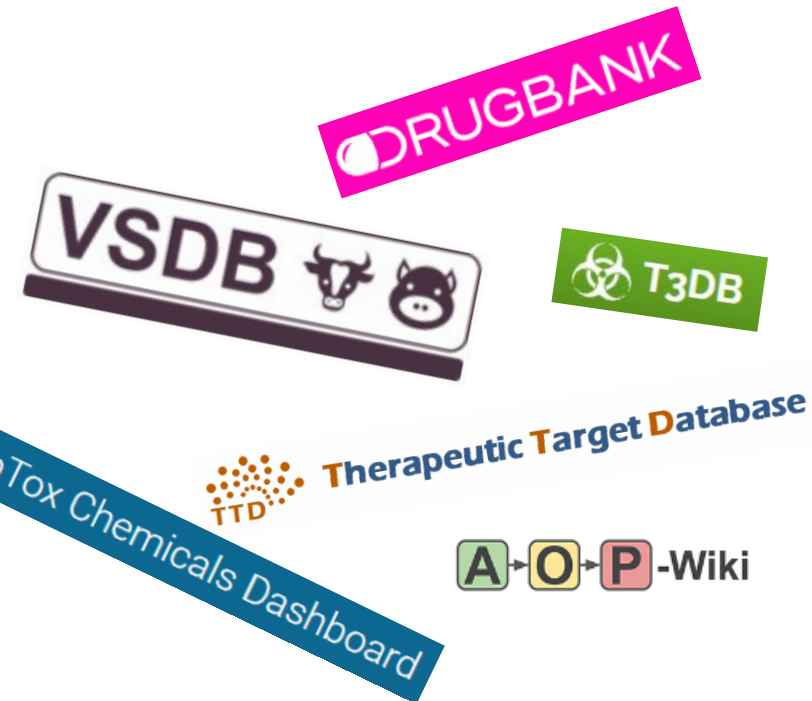
- NCBI Taxonomy Database - <http://www.ncbi.nlm.nih.gov/taxonomy>
- NCBI Protein Database - <http://www.ncbi.nlm.nih.gov/protein>
- NCBI Conserved Domain Database - <http://www.ncbi.nlm.nih.gov/Structure/cdd/>
- NCBI COBALT - <http://www.st-va.ncbi.nlm.nih.gov/tools/cobalt/>

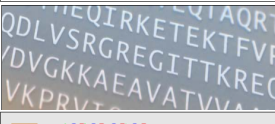
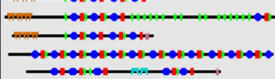
Guide the User to Appropriate Resources for Individual Amino Acid Comparisons:

- Google Scholar - <https://scholar.google.com/>

Aids for Data Synthesis and Connection:

- ECOTOX Knowledgebase - <https://cfpub.epa.gov/>
- U.S. Fish & Wildlife Environmental Conservation Online System - <https://ecos.fws.gov/ecp/>



	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.
	Protein The Protein database is a collection of sequences from several sources, including translations from annotated coding regions in GenBank, RefSeq and TPA, as well as records from SwissProt, PIR, PRF, and PDB. Protein sequences are the fundamental determinants of biological structure and function.
	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.



Acknowledgements

U.S. EPA, ORD

Marissa Jensen (University of Minnesota Duluth)

Sally Mayasich (University of Wisconsin)

Monique Hazemi (ORISE)

Sara Vliet (past ORISE 2021)

Jon Doering (past NRC 2018)

Colin Finnegan (past ORISE 2018)

Donovan Blatz (past ORISE 2021)

GDIT

Cody Simmons

Wilson Menendez

Audrey Wilkinson (past GDIT 2022)

Thomas Transue (past GDIT 2022)

SeqAPASS v7.0 (Coming 2023)



LaLone.Carlie@epa.gov

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