SeqAPASS: Sequence alignment to predict across-species susceptibility

Carlie A. LaLone[1], Daniel L. Villeneuve[2], Henry W. Helgen[3], and Gerald T. Ankley[2]

[1]University of Minnesota Water Resources Center, 6201 Congdon Blvd. Duluth, MN, 55804
[2]U.S. Environmental Protection Agency, 6201 Congdon Blvd. Duluth, MN, 55804
[3]Computer Sciences Corporation, 6201 Congdon Blvd. Duluth, MN, 55804
E-mail contact: LaLone.Carlie@epa.gov

1. Introduction

Efforts to shift the toxicity testing paradigm from whole organism studies to those focused on the initiation of toxicity and relevant pathways have led to increased utilization of in vitro and in silico methods. Hence the emergence of high through-put screening (HTS) programs, such as U.S. EPA ToxCast, and application of the adverse outcome pathway (AOP) framework for identifying and defining biological key events triggered upon perturbation of molecular initiating events and leading to adverse outcomes occuring at a level of organization relevant for risk assessment [1]. With these recent initiatives to harness the power of "the pathway" in describing and evaluating toxicity comes the need to extrapolate data beyond the model species.

Sequence alignment to predict across-species susceptibilty (SeqAPASS) is a web-based tool that allows the user to begin to understand how broadly HTS data or AOP constructs may plausibly be extrapolated across species, while describing the relative intrinsic susceptibiltiy of different taxa to chemicals with known modes of action (e.g., pharmaceuticals and pesticides). The tool rapidly and strategically assesses available molecular target information to describe protein sequence similarity at the primary amino acid sequence, conserved domain, and individual amino acid residue levels. This in silico approach to species extrapolation was designed to automate and streamline the relatively complex and time-consuming process of comparing protein sequences in a consistent, logical, and criteria driven manner intended for predicting across species susceptibility to a chemical perturbation.

To define the domain of applicability and enhance the utility of the SeqAPASS tool, multiple case studies have been explored, including the derivation of predictions for across species susceptibilty to chemicals that target the human estrogen receptor, bovine androgen receptor, mosquito voltage-gated sodium channel, fungus cytochrome P450 51, and honey bee nicotinic acetylcholine receptor. These examples highlight the utility of the SeqAPASS tool for researchers and regulators alike.

2. Materials and methods

The susceptibility of an organism to chemical perturbation is determined by a number of key factors, including the availability of a protein target for the chemical to act upon. Chemicals such as pharmaceuticals and pesticides have relatively well defined molecular targets, with a majority of the sequence information curated in the National Center for Biotechnology Information (NCBI) protein database. Upon identification of a protein target and target species for a given chemical (e.g., human or veterinary animal protein sequence in the case of pharmaceuticals; Figure 1a) SeqAPASS compares that protein to all other proteins available in the NCBI database and identifies the best matching sequences. Evaluation of the similarity between the primary amino acid sequences (Figure 1b) (including ortholog candidate identification; Figure 1d), conserved domains (Figure 1c), and individual amino acid residues (Figure 1e) lead to the prediciton of susceptibility across species (Figure 1f).

3. Results and discussion

3.1. Class level susceptibility predictions: primary amino acid sequence alignments

SeqAPASS was designed to implement a strategic approach for predictions of susceptibility using protein sequence-based comparisons at various levels of sequence detail. Our initial case studies examining the human estrogen receptor, mosquito voltage-gated para-like sodium channel, and bovine androgen receptor to predict relative intrinsic susceptibility to ethinyl estradial, permethrin, and trenbolone acetate across species focused on the least complex measure of protein similarity at the primary amino acid sequence level [2]. The results of these analyses allowed for organism class-level predictions with the ability to describe

differences in potential susceptibility between vertebrate and invertebrate species. Due to the abundance of empirical toxicity data for exposures to EE2 and permethrin, predicted susceptibility versus measured sensitivity were evaluated, concluding that predictions fit well with laboratory derived toxcity values.

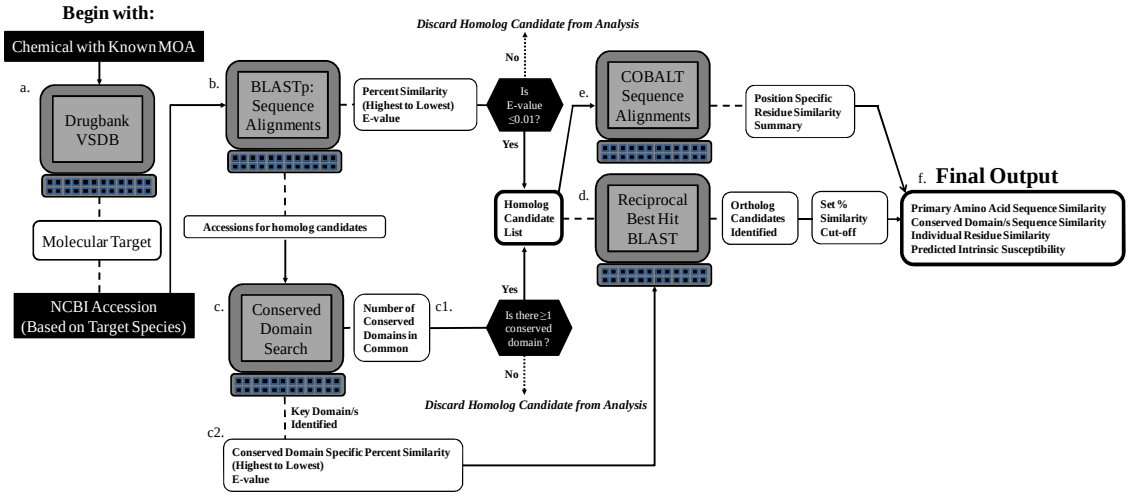

Figure 1: SeqAPASS workflow used to derive relative intrinsic susceptibility predictions, which initiates with the identification of a chemical of interest with a known molecular target. Black boxes indicate necessary input for the computer algorithms and white ovals represent output derived from computer programs. Black hexagons indicate automated data filters designed to assess the quality of alignment criteria. Workflow begins in the upper left and moves to the final output on the middle right. Bold-outlined rectangles represent output used for full sequence similarity analysis (homolog candidate list) and ortholog identification leading to predictions of susceptibility (final output). Acronym definitions: MOA, mode of action; VSDB, Veterinary Substance Database; BLASTp, Basic Local Alignment Search Tool for proteins; and NCBI, National Center for Biotechnology Information. Adapted figure from LaLone et al. [2].

3.2. Species level susceptibility predictions: Conserved domains and residue analyses

As a means to enhance susceptibility predictions, SeqAPASS allows for conserved domain analysis, as well as the ability to query individual residue positions across species when knowledge exists related to secondary or tertiary protein structure and function. A prediction of honey bee susceptibility to neonicotinoids based on evaluation of the multisubunit nicotinic acetylcholine receptor compared to targeted insect species was obtained utilizing SeqAPASS. This analysis illustrated the value of ligand binding domain comparisons and the identification of site specific residue similarities across species as a means to provide further evidence for the likelihood of honey bee susceptibility.

4. Conclusions

The conservation of a molecular target in a non-target species is a key determinant in predicting chemical susceptibility. The SeqAPASS tool has been designed to rapidly evaluate protein similarity at multiple levels of complexity to predict across species susceptibility at the organism class level down to the species level when sufficient information about the molecular target structure exists. Through a variety of case studies assessing relatively simple to highly complex multi-subunit proteins we demonstrate the various applications of the predictions as well as define the challenges associated with such evaluations.

5. References

[1] Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, Mount DR, Nichols JW, Russom CL, Schmieder PK, Serrrano JA, Tietge JE, Villeneuve DL. 2010. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environ Toxicol Chem 29:730-741.

[2] LaLone CA, Villeneuve DL, Burgoon LD, Russom CL, Helgen HW, Berninger JP, Tietge JE, Severson MN, Cavallin JE, Ankley GT. 2013. Molecular Target Sequence Similarity as a Basis for Species Extrapolation to Assess the Ecological Risk of Chemicals with Known Modes of Action. Aquatic Toxicology 144-145: 141-154

Acknowledgement - The authors thank D. Lane and S. Walata for aiding in the computer programming in the development of the SeqAPASS tool.