A Computational Tool for Defining Conservation of Molecular Initiating Events Across Species

> U.S. EPA Carlie LaLone 8/22/13

FFKRSIQGHNDY+CPATNQCTID+NRRKSCQACRLRKCYEVGMMKGGIRKDR GGR ++

Chemical Safety for Sustainability

Problem Statement

Although chemicals are essential to modern life, we lack innovative, systematic, effective, and efficient approaches and tools to inform decisions that reduce negative environmental and societal impacts of chemicals while increasing economic value.



Toxicity Testing in the 21st Century



- "Transform toxicity testing from a system based on whole-animal testing to one founded primarily on in vitro [or in silico] methods that evaluate changes in biologic processes using cells, cell lines, or cellular components, preferably of human origin"
- "The vision emphasizes the development of suites of predictive, high-throughput assays" (p. 7)
- "The mix of tests in the vision include tests that assess critical mechanistic endpoints involved in the induction of overt toxic effects rather than the effects themselves." (p. 121)

Increase the efficiency and speed of chemical evaluations



Identify putative molecular initiating events: chemical - biomolecule



Adverse outcome pathway:

Links molecular initiating event to adverse outcome relevant to risk assessment
Hazard prediction from inherency or toxicity-pathway data

Application of 21st C Paradigm to Ecotoxicology – challenge of species extrapolation



Taxonomic applicability domain

- Inherency tools
- Molecular screening data
- AOPs (to some extent)



Chemicals with well defined MIEs

Pharmaceuticals and Pesticide

- Designed to act on specific molecular targets to provide therapeutic benefits or exterminate pests
 - Molecular targets

 Genomic information and translated protein sequence information
 How do we translate this knowledge to potential effects on ecological species?

Molecular Target Similarity– Species Extrapolation

Molecular Target Similarity: Species Extrapolation

• Ability to identify organism classes with differing sensitivity based on molecular target homology (Kostich and Lazorchak, Science of the Total Environment. 2008, 389, 329-339)

• Distribution of orthologs

(Gunnarsson et al. Environ Sci Technol. 2008; 42(15):5807-13)



• Which proteins are conserved and to what degree?

Example Supporting Sequence-based Intrinsic Susceptibility Predictions

Estrogen Receptor (OW/ORD Emerging Contaminants Workgroup, 2008)

Animal Kingdom	Genus	Common name	Chronic value (ng/L)
	Danio	Zebrafish	<1.1
Vortobrotoo oguatia	Pimephales	Fathead minnow	1.5
vertebrates, aqualic	Oryzias	Medaka	3.2
	Oncorhynchus	Rainbow trout	<16
	Potamopyrgus	Snail	50
	Gammarus	Freshwater shrimp	>7,600
Invertebrates,	Daphnia	Water flea	45,000
aquatic	Tisbe	Copepod	>100,000
	Chironomus	Midge	320,000
	Brachionus	Rotifer	800,000

- Moving from qualitative understanding of molecular target conservation to quantitative measures
 - New tools and technologies have emerged

Developing an Automated Computational Methodology



Protein Sequence Similarity Tool



Predict relative intrinsic susceptibility

- Relative in that it is dependent on which species is selected as the target (query species)
 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





Computational Assessment of Protein Similarity: Quantitative Metrics

 Assume that presence of molecular target in non-target species is one critical route via which a chemical could cause adverse effects

• Target species vs. Non-target species (NCBI)

- Align amino acid sequences and conserved domains
- Assume greater similarity = greater likelihood
 - interact with molecular target in non-target species

VS.









Toxin, Toxin-Target Database

- Pesticide Properties DataBase, T3DB, Veterinary Substances DataBase, & DrugBank : identifies molecular targets for pesticide/pharmaceutical
- Link to NCBI GenBank: Protein accession



• <u>Molecular Target Similarity Tool</u>

- Automated BLASTp and Conserved domain database
- Query target species protein accession against all organism classes for vertebrates, invertebrates, and plants

Output from Sequence Analysis

	Number of Proteins in NCBI Database	Organism Class	Species		NCBI Protein Accession	Erotein Name		Most	simila	r	val	ue % Simil	lanity	Number of Conserved Domains	
	35 859	Incerta	Aedes neovoti		A (7837024-1	whise-sated para-bi	ke sodium chann	to tar	get sp	ecie	S –	100.0		4	
	115	Incerta	Culey niniens n	nilaus	RAT770181	nam-sodim channel				3064	0	88.8		4	
	20 202	hoecta	Cular animanal	ha cintur	DA177017 1	pur sodim chinici				2062	0	00.0		4	
Spe	ie	NCI	I Protein	Brotein N	ame	para-sociali chariter			Bit	E-val	hue	9/6		7	
		A							666070			Similari	it.	3	
									1				- Y	4	
														3	
Aed	es aegypti	ACE	370241	voltage-g	ated para-like s	odium channel			4464	0		100.0		2	
Hor	no saniens	NP	066287.2	sodim ch	annel protein ty	pe 2 subunit abha i	isoform 1		1639	0		36.7		3	
	212240	msecta		estoni	EL/W62207.1	1422241				2411	U	/0.9		3	
	30,028	Insecta	Drosophila viri	ās	EDW65213.1	paralytic				3410	0	76.4		3	
Sma		NCT	T Protein	Dratain N					Di4	E wa		46		Number of	
spe	1100	NCE A	a Frotein	Froteinity	ame				DIL	E-val	ue			Compared of	
		Acce	ssion						scor			Similari	īÿ	Conserved	
														Domains	
Aed	es aegypti	ACE	37024.1	voltage-g	ated para-like s	odium channel			4464	0		100.0		4	
Aed Dro	es aegypti sophila erecta	ACE	37024.1 46862.1	voltage-ga GG19319	ated para-like s	odium channel			446-	0 0		100.0 76.9	ſ	4 0	
Aed Dro	es aegypti sophila erecta	ACE	37024.1 46862.1	voltage-ga GG19319	ated para-like s	odium channel			446 3431	0		100.0 76.9	[4 0	
Aed Dro	es aegypti sophila erecta 324	ACE EDV	37024.1 46862.1 Perinlaneta am	voltage-g GG19319	ated para-like s	odium channel	channel Nam I almha		446 3431	0 0 3218	Ŭ	100.0 76.9	Ç	4 0	
Aed Dro	es aegypti sophila erecta 324 16.324	ACE EDV Insecta	37024.1 46862.1 Periplaneta am Danaus y lexiny	voltage-g: GG19319 ericana	ated para-like s ACX44801.1 EHT74501.1	odium channel	charmel Nav I alphy charmel alpha suba	 1. subunit 1. nit	446- 3431	0 0 3218 3212	0	100.0 76.9 72.1 72.0	ς	4 0 4	
Aed Dro	es aegypti sophila erecta 324 16,324 15,129	ACE EDV Insecta Insecta	37024.1 46862.1 Periplaneta am Danaus plenipp Harnemathos s	voltage-g; GG19319 ericana w abatar	ated para-like s ACX44801.1 EHJ74501.1 EFJ78501.1	odium channel	charanel Nav I alpha charanel Nav I alpha charanel alpha subu in para	na A subunit Init	446- 3431	0 0 3218 3212 3212	0 0 0	100.0 76.9 72.1 72.0 720	C	4 0 4 4	
Aed Dro	es aegypti sophila erecta 324 16,324 15,129 31,591	ACE EDV nsecta Insecta Insecta Btrachioneda	37024.1 46862.1 Periplaneta am Danaus plexipp Harpegnathos s Danksio sulex	voltage-g GG19319 ericana us saltator	ated para-like s ACX44801.1 EHJ74501.1 EFJ86793.1 EFFX81393.1	odium channel	charmel Navi a alpha charmel Navi a alpha charmel alpha subu in para DAPPIID RAFT 50	na Asubunit Init	446- 3431	0 0 3218 3212 3212 2652	0 0 0	100.0 76.9 72.1 72.0 72.0 72.0 59.4	C	4 0 4 4 3	
Aed Dro	es aegypti sophila erecta 324 16,324 15,129 31,591 20	ACE EDV Insecta Insecta Insecta Branchiopoda Malacostraca	37024.1 46862.1 Periplaneta am Danaus plexipp Harpegnathos s Daphnia pulex Concer bareabi	voltage-g GG19319 ericana us adator	ACK44801.1 EHU74501.1 EFU86793.1 EFX81393.1 ARL 10360.2	odium channel	channel Nav I alpha channel Nav I alpha channel alpha subu in para DAPPUDRAFT_50 channel	n subunit. nit 150	446- 3431	0 0 3218 3212 3212 2652 2523	0 0 0 0	100.0 76.9 72.1 72.0 72.0 59.4 56.5	C	4 0 4 4 3 3	
Aed Dro	es aegypti sophila erecta 324 16,324 15,129 31,591 20 41,714	ACE EDV Insecta Insecta Insecta Branchiopoda Malacostraca Arachmida	37024.1 46862.1 Periplaneta am Danaus plenipp Harpegnathos s Daphnia pulen Cancer doreali: Kodes scanilor	voltage-g GG19319 ericana us saltator	ACX44801.1 EHU74501.1 EFU74501.1 EFN86793.1 EFX81393.1 ABL10360.2 EFE03677.1	odium channel	channel Nav I alpha channel alpha subu in para DAPPUDRAFT_50 channel p channel alpha sub	n subumit. mit. 150	446- 3431	0 0 3218 3212 3212 2652 2523 2416	0 0 0 0 0	100.0 76.9 72.1 72.0 72.0 72.0 59.4 56.5 54.1	C	4 4 4 3 3 3	
Aed Dro	es aegypti sophila erecta 324 16,324 15,129 31,591 20 41,714 212	ACE EDV Insecta Insecta Insecta Branchiopoda Malacostraca Arachnida	37024.1 46862.1 Periplaneta am Danaus plenipp Harpegnathos s Daphnia pulen Cancer doreali: kodes scapular	voltage-g; GG19319 ericana us adator s tis hunidi	ACX44801.1 EHD74501.1 EFD74501.1 EFFN86793.1 EFFX81393.1 ABL10360.2 EEEC03677.1 ABH12275.1	odium channel voltage-gated sodium o voltage-gated sodium Sodium channel protei hypothetical protein I voltage-gated sodium o skeletal muscle sodium untatine no base-sated	channel Nav I alphy channel alpha subu in para DAPPUDRAFT_50 channel a channel alpha sub sodium channel	n subunit. mit 150 umit, putstive	446- 3431	0 0 3218 3212 3212 2652 2523 2416 2273	0 0 0 0 0 0	100.0 76.9 72.1 72.0 72.0 59.4 56.5 54.1 50.9	C	4 4 4 3 3 3 2	
Aed Dro	es aegypti sophila erecta 324 16,324 15,129 31,591 20 41,714 212 25,335	ACE EDV Insecta Insecta Insecta Branchiopoda Malacostraca Arachnida Mammalia	37024.1 46862.1 Periplaneta am Danaus plexipp Harpegnathos s Daphnia pulex Cancer doreali: kodes scapular Haplopetna sch Chilikiiy incek	voltage-g; GG19319 ericana us altator s tis twadti	ACX44801.1 EHD74501.1 EHD74501.1 EFN86793.1 EFX81393.1 ABL10360.2 EEC03677.1 ABH12275.1 VP. 002749429.1	odium channel voltage-gated sodium o voltage-gated sodium o sodium channel protei hypothetical protein I voltage-gated sodium o skeletal muscle sodium putative voltage-gated DREDICTED: sodium	channel Nav I alpha channel alpha subu in para DAPPUDRAFT_50 channel a channel alpha sub I sodium channel channel autorio tw	na nationait nait 150 arait, patative ne 2 subanait alaba	446- 3431	0 0 3218 3212 3212 2652 2523 2416 2273 1640	0 0 0 0 0 0 0 0	100.0 76.9 72.1 72.0 72.0 72.0 72.0 59.4 56.5 54.1 50.9 36.7	C	4 4 4 3 3 3 2 2	
Aed Dro	es aegypti sophila erecta 324 16,324 15,129 31,591 20 41,714 212 25,335	ACE EDV Insecta Insecta Insecta Branchiopoda Malacostraca Arachnida Arachnida Mammalia	37024.1 46862.1 Periplaneta am Danaus plenipp Harpegnathos s Daphnia pulen Cancer doreali: hodes scapular Haplopetna sch Callithrix jacch	voltage-g; GG19319 ericana us altator s tis tris triadti uus	ACX44801.1 EHJ74501.1 EFJN86793.1 EFJN86793.1 ABL10360.2 EEC03677.1 ABH12275.1 XP_002749429.1	odium channel voltage-gated sodium o voltage-gated sodium o Sodium channel protei hypothetical protein I voltage-gated sodium o skeletal muscle sodium putative voltage-gated PREDICTED : sodium	channel Nav I alpha channel alpha subu in para DAPPUDRAFT_50 channel n channel alpha sub I sodium channel channel protein typ	na subunit nit 150 unit, putstive 2 subunit alphy	446- 3431	0 0 3218 3212 3212 2652 2523 2416 2273 1640	0 0 0 0 0 0 0 0 0	100.0 76.9 72.1 72.0 72.0 72.0 59.4 56.5 54.1 50.9 36.7	C	4 4 4 3 3 3 2 2 2	
Aed Dro	es aegypti sophila erecta 324 16,324 15,129 31,591 20 41,714 212 25,335 25,661	ACE EDV Insecta Insecta Insecta Branchiopoda Malacostraca Arachnida Arachnida Mammalia	37024.1 46862.1 Periplaneta am Danaus y lenipy Haryegnathos s Day knia puler Cancer doreali: Kodes scapular Hayloyedma sch Callithrin jacch Macaca mulatto	voltage-gi GG19319 enicana us sadator s sis kwidti sus	ACX44801.1 EHU74501.1 EFU74501.1 EFX86793.1 EFX81393.1 ABL10360.2 EEC03677.1 ABH12275.1 XP_002749429.1 XP_001100011.1	odium channel voltage-gated sodium o voltage-gated sodium o voltage-gated sodium o sodium charmel protein I voltage-gated sodium o skeletal muscle sodium o skeletal muscle sodium putative voltage-gated PREDICTED : sodium PREDICTED : sodium	channel Nav I alpha channel alpha subu in para DAPPUDRAFT_50 channel n channel alpha sub I sodium channel channel protein typ channel protein typ	a subunit. nit. 150 unit, putative pe 2 subunit alphy pe 2 subunit alphy	446. 3431 1 isoform 2 1. like isofo	0 0 3218 3212 3212 2652 2523 2416 2273 1640 1639	0 0 0 0 0 0 0 0	100.0 76.9 72.1 72.0 72.0 72.0 59.4 56.5 54.1 50.9 36.7 36.7	C	4 4 4 3 3 3 2 2 2	
Aed Dro	es aegypti sophila erecta 324 16,324 15,129 31,591 20 41,714 212 25,335 25,661 596,962	ACE EDV Insecta Insecta Insecta Branchiopoda Malacostraca Arachnida Arachnida Mammalia Mammalia	37024.1 46862.1 Peripianeta am Danaus pieripp Harpegnathos s Daphnia puler Cancer doreali: Inodes scapular Haplopetna sol Callithrix jacch Macaca mulatio Homo sapiens	voltage-g; GG19319 ericana us altator s is is knidti us	ACX44801.1 EH74501.1 EFN86793.1 EFX81393.1 ABL10360.2 EEC03677.1 ABH12275.1 XP_002749429.1 XP_001100011.1 NP 066287.2	odium channel voltage-gated sodium o voltage-gated sodium o sodium channel protei hypothetical protein I voltage-gated sodium o skeletal muscle sodium putative voltage-gated PREDICTED : sodium 6 sodium channel protei	channel Nav I alpha channel alpha subu in para DAPPUDRAFT_50 channel n channel alpha sub I sodium channel channel protein ty channel protein ty channel protein ty	a subunit nit 150 omit, putative se 2 subunit alpho se 2 subunit alpho se 2 subunit alpho	446- 3431	0 0 3218 3212 3212 2652 2523 2416 2273 1640 1639 1639	0 0 0 0 0 0 0 0 0	100.0 76.9 72.1 72.0 72.0 72.0 72.0 72.0 72.0 72.0 72.0	C	4 4 4 3 3 3 2 2 2 2 2	
Aed Dro	es aegypti sophila erecta 324 16,324 15,129 31,591 20 41,714 212 25,335 25,661 596,962 40,939	ACE EDV Insecta Insecta Insecta Branchiopoda Malacostraca Arachnida Arachnida Mammalia Mammalia	37024.1 46862.1 Periplaneta am Danaus plenipp Harpegnathos s Daphnia pulen Cancer doreali: Inodes scapular Haplopebna sch Callithrin jacch Macaca mulatto Homo sapiens Sus scrofa	voltage-g; GG19319 ericana us saltator s s s s s s s s s s s s s s s s s s s	ACX44801.1 EHU74501.1 EFU86793.1 EFX81393.1 ABL10360.2 EEC03677.1 ABH12275.1 XP_002749429.1 XP_001100011.1 NP_066287.2 XP_003133492.1	odium channel	channel Nav I alpha channel Nav I alpha channel alpha subu in para DAPPUDRAFT_50 channel n channel alpha sub I sodium channel channel protein typ channel protein typ n type 2 subunit alp	n subunit nit 150 unit, putative 9e 2 subunit alphy 9e 2 subunit alphy 9a isoform 1	446. 3431	0 0 3218 3212 3212 2652 2523 2416 2273 1640 1639 1639		100.0 76.9 72.1 72.0 72.0 59.4 56.5 54.1 50.9 36.7 36.7 36.7 36.7	C	4 4 4 3 3 3 2 2 2 2 2 2 2	

Setting the Susceptibility Cut-off

Ortholog identification

Ortholog = A sequence diverged after a speciation event
Reciprocal best hit BLAST



Orthologs Identified

• Set susceptibility cut-off

• Protein with lowest percent similarity that was identified as an ortholog

Taxonomic Lineage	Species	Reciprocal Best Hit	NCBI Protein Accession	Protein Name	Bit score	E-value	% similarity
Insecta	Aedes aegypti	Target Species	ACB37022	voltage-gated para-like sodium channel	4449	0	100
Insecta	Harpegnathos saltator	Ortholog	EFN86793	Sodium channel protein para	3223	0	72.4
Arachnida	Ixodes scapularis	-	XP_002407119	skeletal muscle sodium channel alpha subunit, putative	2417	0	54.3
Insecta	Solenopsis invicta	Ortholog	EFZ17857	hypothetical protein SINV_07049	1551	0	34.9
Ascidiacea	Ciona intestinalis		XP_002123673	similar to sodium channel, voltage-gated, type III, alpha	1403	0	31.5
Gastropoda	Aplysia californica	-	NP_001191637	sodium channel alpha-subunit SCAP1	890	0	20.0
Chondrichthyes	Heterodontus francisci	-	ADV73289	voltage-dependent sodium channel 2	843	0	18.9

Analysis Overview



Case Examples: Illustration of Concept

Ethinylestradiol



17β -trenbolone



Permethrin



LaLone et al. Molecular Target Sequence Similarity as a Basis for Species Extrapolation to Assess the Ecological Risk of Chemicals with Known Modes of Action. 2013. Submitted.



Permethrin-Insecticide

Molecular Target: Voltage-gated sodium channel



VGSC Slide material: Ke Dong, Michigan State University

Acute neurotoxicity: Muscle spasm, paralysis, death
Target Species: Lice, ticks, fleas, mites, scabies (www.drugbank.ca),

mosquitoes (U.S. EPA, Registration Eligibility Decision, RED)



Tool: Example Query



Conservation Across Species

Target Species: *Aedes aegypti* Molecular Target: Voltage-gated para-like sodium channel

Define taxonomic applicability domain



Comparing Relative Intrinsic Species Susceptibility Predictions to Empirical Toxicity Data





Mining the ECOTOX database Summarizes Toxicity Data for Individual Chemicals

	<u>Chemical</u>	CAS
	Permethrin	52645-53-1
	(+-)-cis-Permethrin	61949-76-6
	(+-)-trans-Permethrin	61949-77-7
Chemical (Analytical measurements, Chemical purity)	(+)-cis-Permethrin	54774-45-7
	(-)-cis Permethrin	54774-46-8
	(+)-trans Permethrin	51877-74-8
	(-)-trans Permethrin	54774-47-9
	• EC50/LC50 • Mortality / Survivabil	ity
Endpoint (Desired Acute Response in Target Species = Death)	 Acute durations (48 hr Water Exposure (No feedback) 	invertebrate/96 hr vertebrate) eding or diet)

Test of concept



Taking a Closer Look

Delving into sequence comparisons

SYDASDCSFLSEDISMSLSDGDVVGFDMEWPPLYNRGKLGKVALIQLCVSESKCYL HVSSMSVFPQGLKMLLENKAVKKAGVGIEGDQWKLLRDFDIKLKNFVELTDVAN (KLKCTETWSLNSLVKHLLGKQLLKDKSIRCSNWSKFPLTEDQKLLYAATDAYAGFI YRNLEILDDTVQRFAINKEEEILLSDMNKQLTSISEEVMDLAKHLPHAFSKLENPRR /SILLKDISENLYSLRRMIIOSTNIETELRPSNNINI ISFEDISTIGGVQQKQIREHEVLI

IVEDETWDPTLDHI DITEHELOILEOOS **SPNDNENDTSYVIES** EDDENEANEGEED RDNVAVMATGYG **JSAQSENVLTDIKLC** HDFRDSFRKLGSL .YLEVRRKTGNILOI *JTYHAGMSFSTRKD* **YQEIGRAGRDGLQS** ISSRCRRQIILSHFED **JPQAFKLLSAVDILC** AFSROLITEGFLVEV LLPSSKTVSSGTKE SPEKAYSSSOPVIS/ **AKMRPTTVENVKRI**



KLKENMERACLMS ESDEDLEMEMLKHL **KMERNLGLPTKEE (PVQWKVIHSVLEE** VLOLKMSNIPACFL TLIAVDEAHCISEW RNPQITCTGFDRPN QVTGELRKLNLSC OVIHYGAPKDMESY KLKMMAKMEKYL YSMDDSEDTSWDF FGTGKDQTESWWK OSLILOANEELCPKK KISSGSNISKKSIMV PPAILATNKILVDM TDLFSSTKPQEEQK

SLVAKNKICTLSQSMAITYSLFQEKKMPLKSIAESRILPLMTIGMHLSQAVKAGCPL DLERAGLTPEVQKIIADVIRNPPVNSDMSKISLIRMLVPENIDTYLIHMAIEILKHGPD GLOPSCDVNKRRCFPGSEEICSSSKRSKEEVGINTETSSAERKRRLPVWFAKGSDTS

Logic behind development



Conserved Domains: The Next Level

Original Query Target Species: *Aedes aegypti* Molecular Target: Voltage-gated para-like sodium channel NCBI Accession: <u>ACB37024.1</u> Bit Score: 4464 Total Conserved Domains: 4 Organism Class: Insecta

Organism Class	Non- Target Species	Protein Name	Calculated % similarity	Common Domains	Conserved domain IDs	Domain type	Sequence coverage of conserved domains	Conserved Domain % Similarity
Insecta	<u>Apis mellifera</u>	paralytic	73	3	<u>pfam11933</u> <u>pfam08016</u> <u>pfam00520</u>	Unknown function Polycystin cation channel Ion transport protein	1-222, same 217-423, same 1-194, same	61 92 93



Knowledge of individual residues



Resistance and sensitivity: Single amino acid residue substitutions

- Targeted mutagenesis
- Voltage clamp techniques

Targeted sequence comparisons

		Constraint-based Multiple Alignment Tool My NCBI	2
Home	Recent R	esults Help [Sign In] [[Register]
123404.	vs.	vs. Species A, B, C	
✓ <u>ACB37024</u>	1651	NMLTMTLDHYKQTDTFSAVLDYLNMIFICIFSSECLMKIFALRYHYFIEPWNLFDFVVVILSILGLVLSDLIEKYF	VSPT 1730
ACV87000	1619	NMLTMTLDHYQQTQTFSDVLDYLNMIFIYIFTSECLMKIFALRYHYFKEPWNLFDFVVVILSILGLVLSDIIEKYF	VSPT 1698
✓ ACB37024	1731	LLRVVRVAKVGRVLRLVKGAKGIRTLLFALAMSLPALFNICLLLFLVMFIFAIFGMSFFMHVKYKSGLDDVYNFKT	FGQS 1810
✓ <u>ACV87000</u>	1699	LLRVVRVAKVGRVLRLVKGAKGIRTLLFALAMSLPALFNICLLLFLVMFIFAIFGMSFFMHVKDKSGLDDVYNFKT	FGQS 1778
🗹 <u>ACB37024</u>	1811	MILLFQMSTSAGWDGVLDGIINEDECLPPDNDKGYPGNCGSATIGITYLLAYLVISFLIVINMYIAVILENYSQAT	EDVQ 1890
✓ <u>ACV87000</u>	1779	MILLFQMSTSAGWDGVLDGIINEEDCQEPNNEIGYPGNCGSSTIGIAYLLSYLVISFLIVINMYIAVILENYSQAT	EDVQ 1858

Automate the process of querying a/multiple specific residue position/s

- Use knowledge for species susceptibility predictions
- Hypotheses generation

Knowledge of molecular target conservation complements the AOP construct



Established AOP



- Established mechanistic linkage with quantitative or semi-quantitative data
- ----- Plausible linkage with limited data



NDC 0603-5763-21 SPIRONOLACTONE TABLETS, USP 25 mg Rx only 100 TABLETS

Qualitest

Androgen Receptor

Conserved Molecular Target



Environmental Toxicology and Chemistry, Volume 30, Issue 6, pages 1376-1382, 2011

Observed Adverse Effect in Aquatic Species

Molecular Target Conservation: Species Susceptibility Prediction



• Hypothesis:

- Based on homology to the human androgen receptor
 - Small fish likely to be susceptible
 - Invertebrates unlikely to be susceptible

AR Activation in Small Fish



Small fish species sensitive



Secondary Sex0.5 µg/LReproduction5 µg/L



5 μg/L 50 μg/L





Key Findings

• Evidence that MIE conservation and knowledge of sequence similarity across species can be useful for defining taxonomic domain of applicability for AOP • Consistency of AOP across small fish species • Predictable adverse effects • Further examination of functional domains, individual residue queries, and protein structure between species may enhance predictive utility



• Manuscripts:

- Molecular Target Sequence Similarity as a Basis for Species Extrapolation to Assess the Ecological Risk of Chemicals with Known Modes of Action. Submitted
- Cross species sensitivity to a novel androgen agonist of environmental concern, spironolactone. ET&C 2013. Published online

• Sequence Similarity Tool:

- Transferrable tool
 - Automated with capabilities described throughout talk
 - Primary protein sequence, conserved domains, and individual residue query capabilities



Current Analyses: Sequence Similarity Tool

- Honey bee sensitivity- focus on pesticide MIE nAChR
- Acetylcholine esterase
- Avian AOP for Conazoles CYP51
- 27 Pharmaceutical case study
- Estrogen receptor-sequence similarity vs. binding of in vitro recombinant ER across species



Future Directions: Sequence Similarity Tool

• **Test** the predictive utility:

- Establishing quantitative relationships between target similarity and initiation of responses using comparative in vitro systems
- Confirmation that in silico predictions correspond with in vivo responses
- Use tool to provide cross-species insights as to ADME to support PBPK modeling

 Improve the tool: Develop automated computational methods for assessing tertiary structure across species

Acknowledgments

- USEPA (NHEERL)– Duluth, MN
 - C. LaLone, G. Ankley, D. Villeneuve, C. Russom, J. Berninger, J. Tietge, J. Cavallin, M. Severson
- USEPA (NCEA) RTP, NC
 - Lyle Burgoon
- Computer Science Corporation
 - H. Helgen, D. Lane, S. Watala
- USEPA-RTP, NC
 - V. Wilson, E. Gray, P. Hartig
- USEPA (NERL)– Athens, GA
 - T. Collette, D. Ekman
- USEPA-Chicago, IL (GLNPO)
 - T. Smith