The development and use of the Chemical Transformation Simulator (CTS) for new and existing chemicals

Marcy Card
Office of Pollution Prevention and Toxics
US EPA Office of Chemical Safety and Pollution Prevention

Caroline Tebes Stevens
Center for Environmental Measurement and Modeling
US EPA Office of Research and Development

Toxic Substances Control Act

Amended by the 2016 Frank R. Lautenberg Chemical Safety for the 21st Century Act

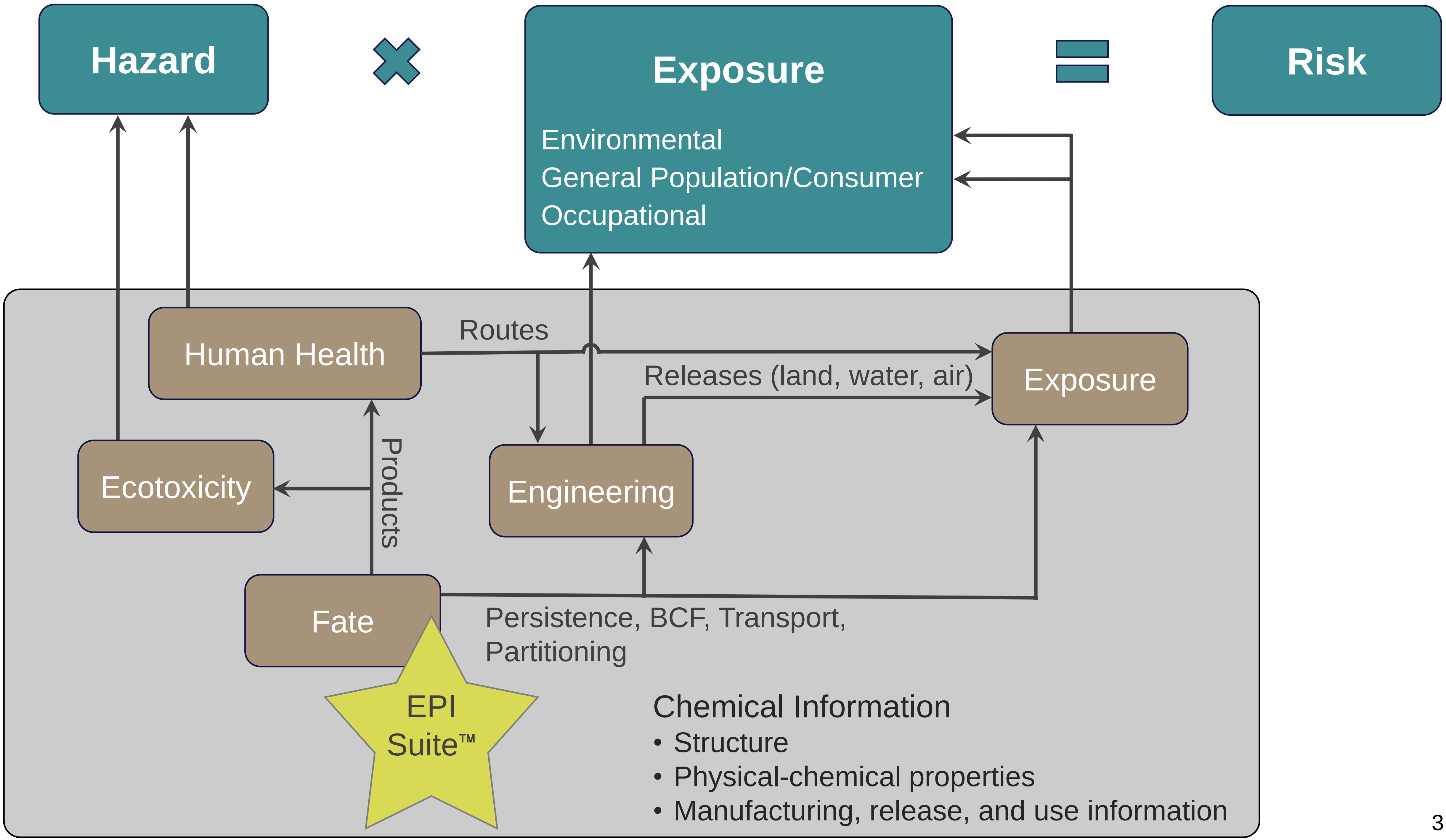
→ Does a chemical present unreasonable risks to human health or the environment?

New Chemicals

- Limited data available. Applications require:
  - Chemical name and structure
  - Production volume
  - Use(s)
  - Any test data owned or reasonably ascertainable by the submitter
- Short timelines (90 days)
- Screening-level assessments
- Data gap filling by read-across from analogues and/or models

Existing Chemicals

- Many chemicals are data-rich. Evaluations include:
  - Systematic review of publicly-available information
  - Data submitted by industry and other stakeholders
- Longer timelines (3 years)
- Deeper assessments to characterize risks, including higher-tier modeling
- Data gap filling by test orders, read-across from analogues, and/or models


Hazard
Exposure
Environmental
General Population/Consumer
Occupational
Risk
Human Health
Routes
Releases (land, water, air)
Exposure
Ecotoxicity
Products
Engineering
Fate
Persistence, BCF, Transport, Partitioning
EPI Suite™
Chemical Information
• Structure
• Physical-chemical properties
• Manufacturing, release, and use information

P-Chem Data Needs

- Air-water partitioning (Henry's Law)
- Autoflammability, flash point
- Density
- Dissociation constants (pK_a, pK_b)
  - P-chem properties for ionized species
- Melting point and boiling point
- Morphology, particle size distribution
- Octanol-air partition coefficient (K_{OA})
- Octanol-water partition coefficient (K_{OW})
- Organic carbon partition coefficient (K_{OC})
- Vapor density
- Vapor pressure
- Water solubility
- P-chem properties for compounds outside of the EPI Suite™ chemical domain
  - Perfluorinated substances, nanomaterials, metals, polymers

Highlighting indicates endpoints not estimated by EPI Suite™

Fate Data Needs

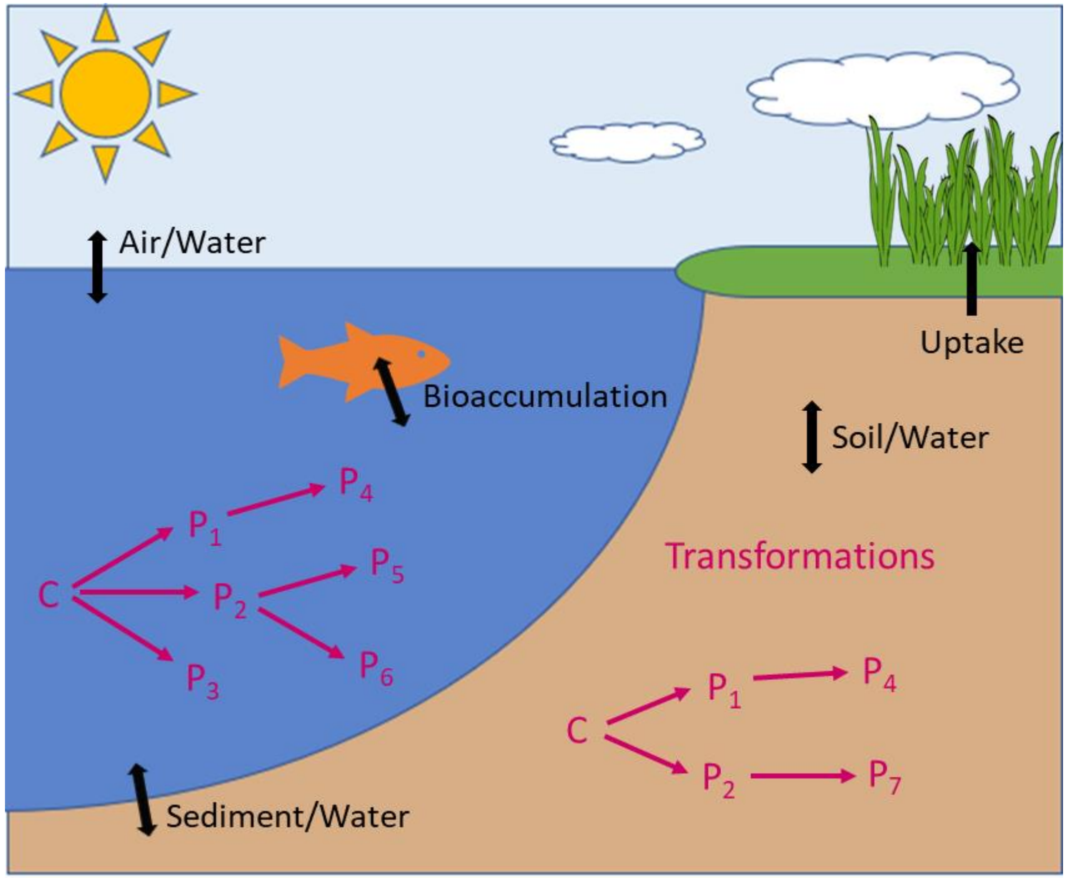
- Transformation processes
  - Aerobic, anaerobic biodegradation rates, products
  - Metabolism rates, products
  - Hydrolysis rates, products
  - Indirect and direct photolysis rates, products
  - Oxidation/reduction rates, products
- Overall persistence
- Bioconcentration, bioaccumulation in fish
  - BCF, BAF for organisms other than fish
  - Trophic magnification, BSAF
- Transport and partitioning among media (e.g., fugacity, mobility in groundwater)
  - Other transport and partitioning issues, e.g. long-range atmospheric transport
- Speciation (e.g., ionization, spontaneous polymerization)
- Removal in wastewater treatment and incineration
- Fate information for compounds outside of the EPI Suite™ chemical domain
  - Perfluorinated substances, nanomaterials, metals, polymers

Highlighting indicates endpoints not estimated by EPI Suite™

Chemical Transformation Simulator (CTS)

How does CTS address OPPT data needs?

- Identifies **likely transformation products** of organic chemicals in environmental and biological systems
  - Toxicity and properties of products may differ from those of the parent
- Provides estimated and measured **physicochemical property values** for both the parent and products
  - Properties control movement within and between environmental media compartments
  - Properties control partitioning into biological tissues



CTS Application

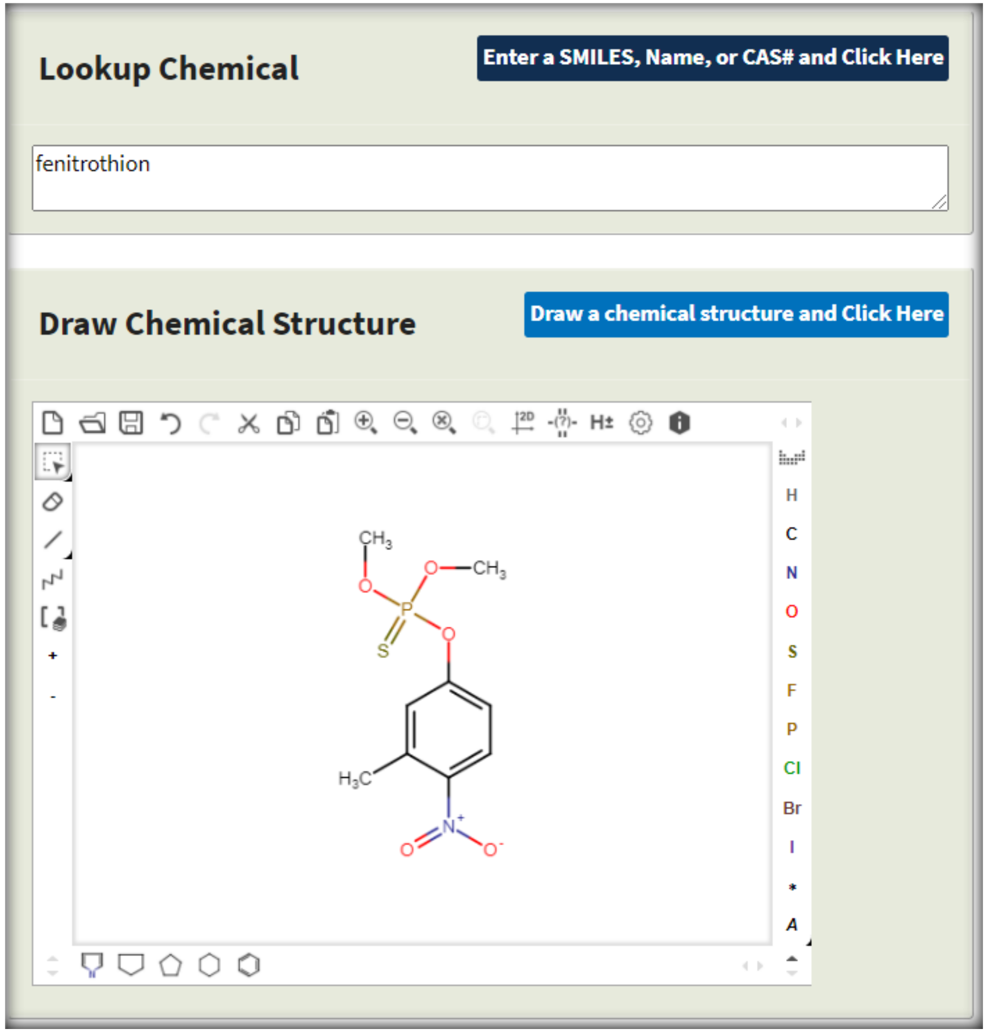
CTS is a web-based tool run on a browser

https://qed.epa.gov/cts/

Three workflows have been developed for user convenience:

- Calculate Chemical Speciation → Provides ionization constants (pKa values), species distribution as a function of pH, stereoisomer structures, tautomer distribution at selected pH
- Calculate Physicochemical Properties → Provides calculated physicochemical properties and measured property values, if available
- Generate Transformation Products → Provides likely environmental transformation products and metabolites of organic chemicals

First Step for All Workflows: Input Chemical of Interest



Chemical Editor: Identify the chemical of interest by name, CAS ID #, SMILES string, or by drawing the molecular structure.

CTS retrieves chemical identifying information from the EPA CompTox Chemicals Dashboard database:

- Name (preferred & IUPAC)
- SMILES
- Formula
- CAS # (preferred & associated)
- DTXSID
- Mass (Average & Monoisotopic)

SMILES standardization

Input for “Calculate Chemical Speciation” Workflow

Calculate Ionization Constants (p*Ka*) Parameters	
Number of decimals for pKa:	2
pH Lower Limit:	0
pH Upper Limit:	14
pH Step Size:	0.2
Generate Major Microspecies at pH:	7.0
Isoelectric Point (pI) pH Step Size for Charge Distribution:	0.5

pKa calculations
Speciation as a function of pH

Calculate Dominant Tautomer Distribution	
Maximum Number of Structures:	100
at pH:	7.0

Dominant tautomer distribution

Calculate Stereoisomers	
Maximum Number of Structures:	100

Stereoisomers

Output for "Calculate Chemical Speciation" Workflow

☑ Calculate Ionization Constants (p*K*a) Parameters	
Number of decimals for pKa:	2
pH Lower Limit:	0
pH Upper Limit:	14
pH Step Size:	0.2
Generate Major Microspecies at pH:	7.0
Isoelectric Point (pI) pH Step Size for Charge Distribution:	0.5

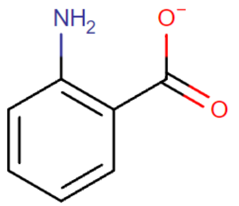
☐ Calculate Dominant Tautomer Distribution	
Maximum Number of Structures:	100
at pH:	7.0

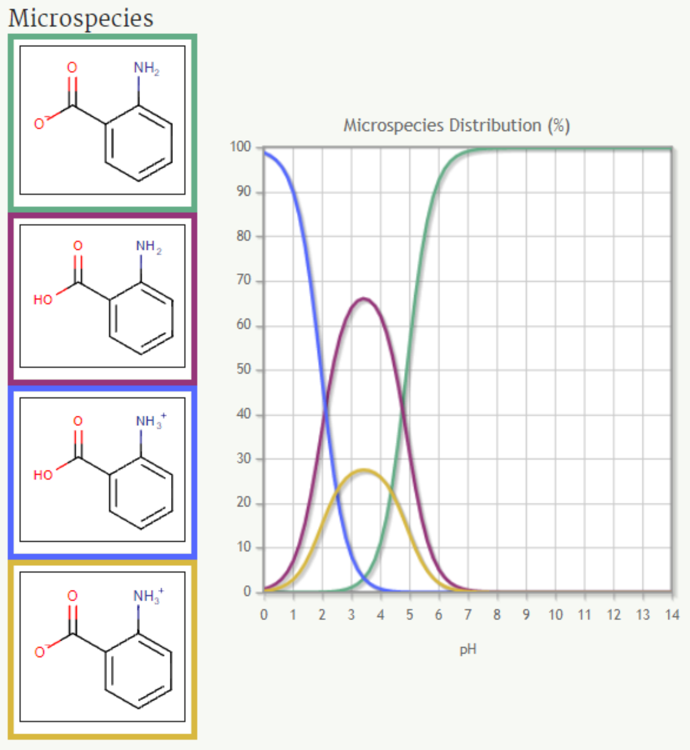
☐ Calculate Stereoisomers	
Maximum Number of Structures:	100

Example:
2-aminobenzoic acid

pKa = 4.89
pKb = 1.95

Major Microspecies at pH: 7.0





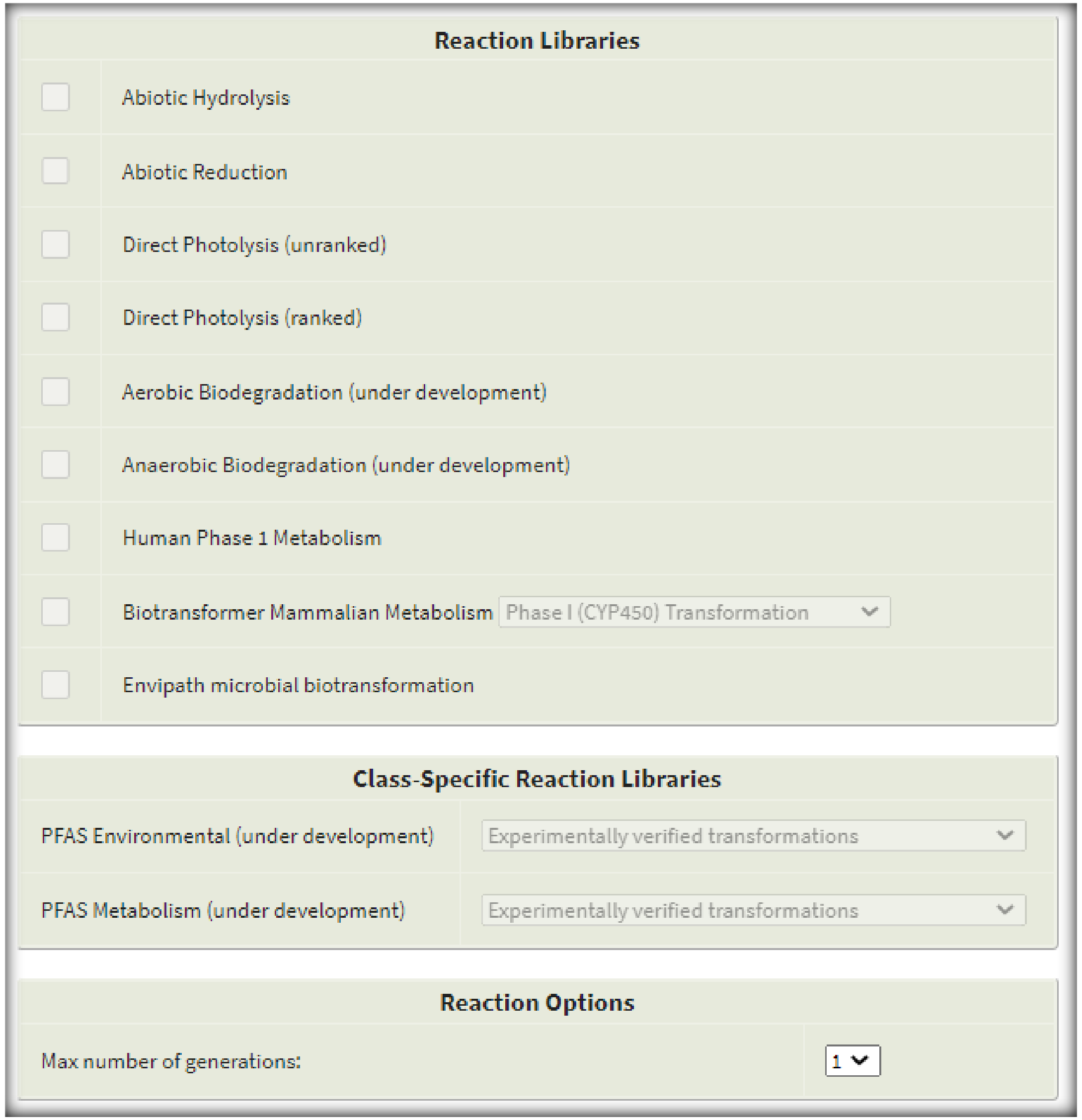
“Calculate Physicochemical Properties” Workflow

Physicochemical property calculator(s):

ChemAxon = Marvin Plug-In Calculators
EPI Suite™ = Estimation Programs Interface Suite
T.E.S.T. = Toxicity Estimation Software Tool
OPERA = OPEn structure–activity/property Relationship App

- Measured physicochemical property values from PHYSPROP database
- Geometric mean calculated value provided as consensus estimated property value
- pH-adjusted values provided for K_{ow} and water solubility

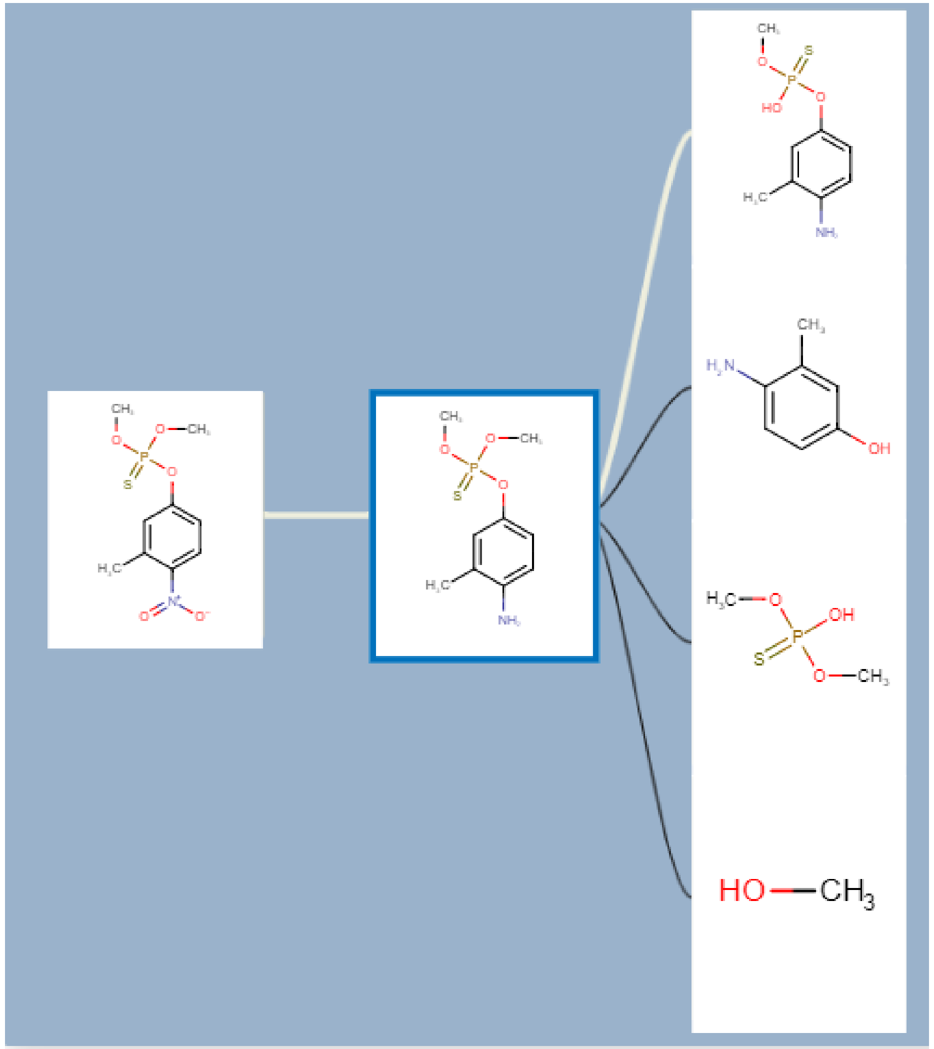
Input for “Generate Transformation Products” Workflow



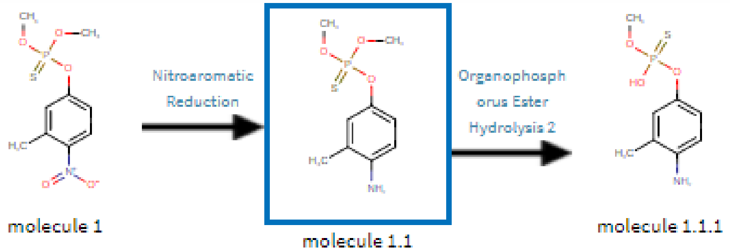
Reaction Library Selection:

- CTS reaction libraries for environmental transformations
  - Expert system to guide selection of one or more libraries based on scenario of interest
- Class-specific libraries for environmental and metabolic transformations of PFAS
- Human Phase 1 Metabolism (ChemAxon)
- Externally developed prediction tools
  - **BioTransformer** metabolism prediction modules
  - **EnviPath** for aerobic microbial biotransformations

Output for "Generate Transformation Products" Workflow

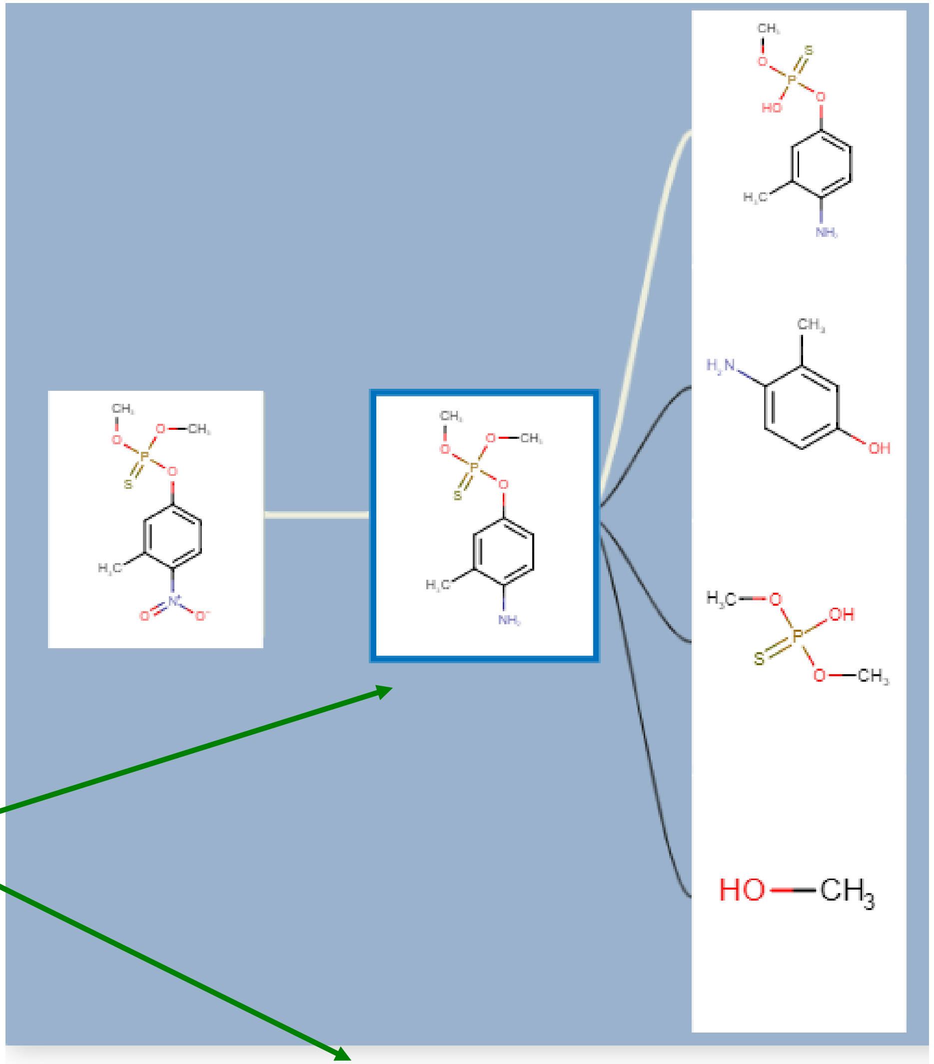


Tree structure diagram showing up to four generations of products



Sequence of transformation reactions

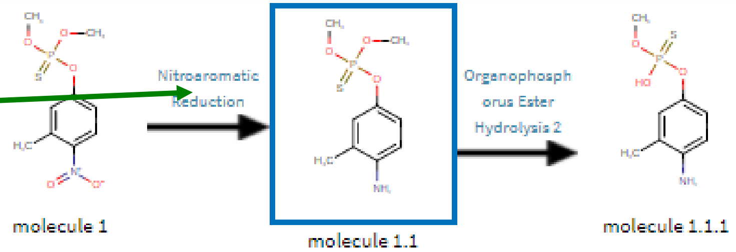
Output for “Generate Transformation Products” Workflow



Tree Structure diagram showing up to four generations of products

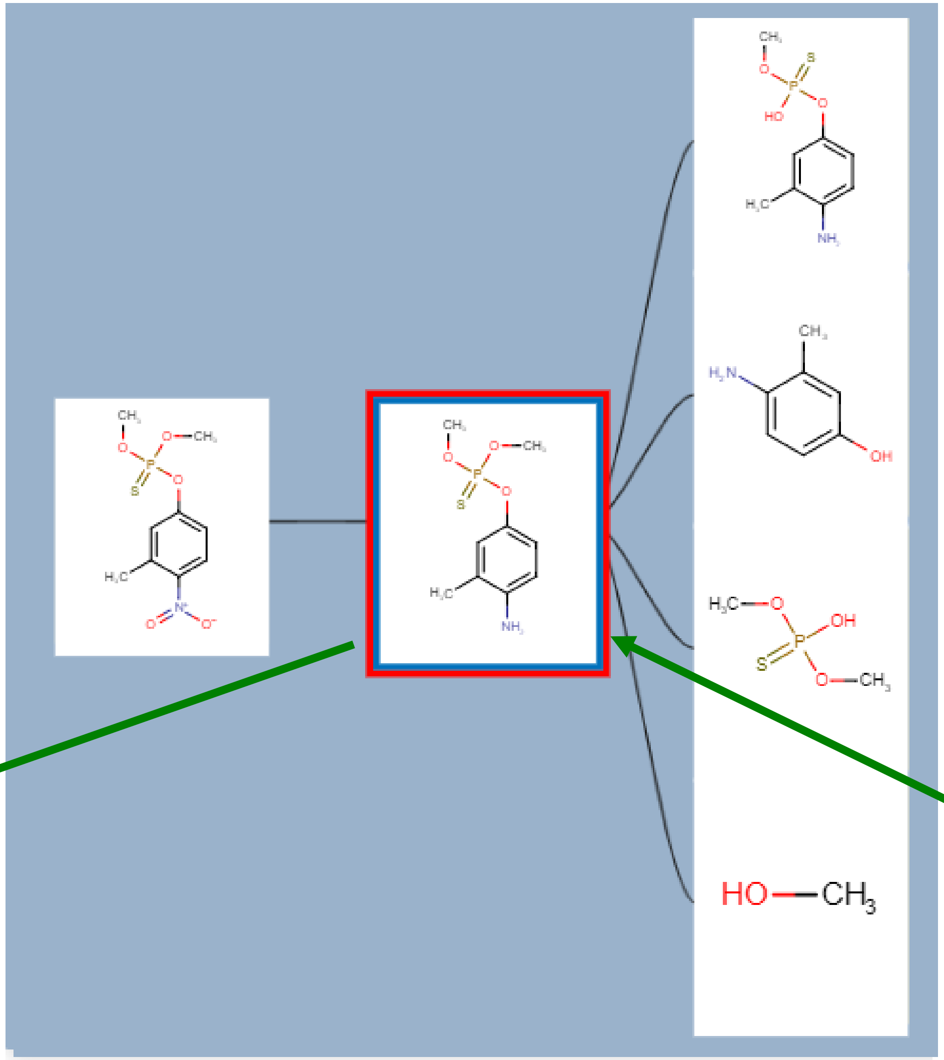
Likely products outlined in blue (predicted accumulation > 10%)

Hyperlink to reaction scheme documentation (reaction scheme, examples with citations)



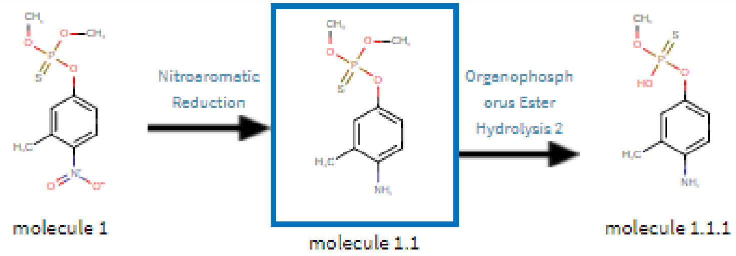
Sequence of transformation reactions

Output for "Generate Transformation Products" Workflow



	ChemAxon	EPI Suite	TEST	OPERA	Geometric Mean	Measured
All						
Neutral Species Properties:						
Melting Point (°C)						
Boiling Point (°C)						
Water Solubility (mg/L)						
Vapor Pressure (mmHg)						
Ionization Constant						
Henry's Law Constant (atm-m³/mol)						
Octanol/Water Partition Coefficient (log)						
Organic Carbon Partition Coefficient (log(L/kg))						
Bioconcentration Factor (log(L/kg))						
Bioaccumulation Factor (log(L/kg))						

Right-click a molecule of interest for chemical identification and option to calculate physicochemical property values



CTS Reaction Library Development

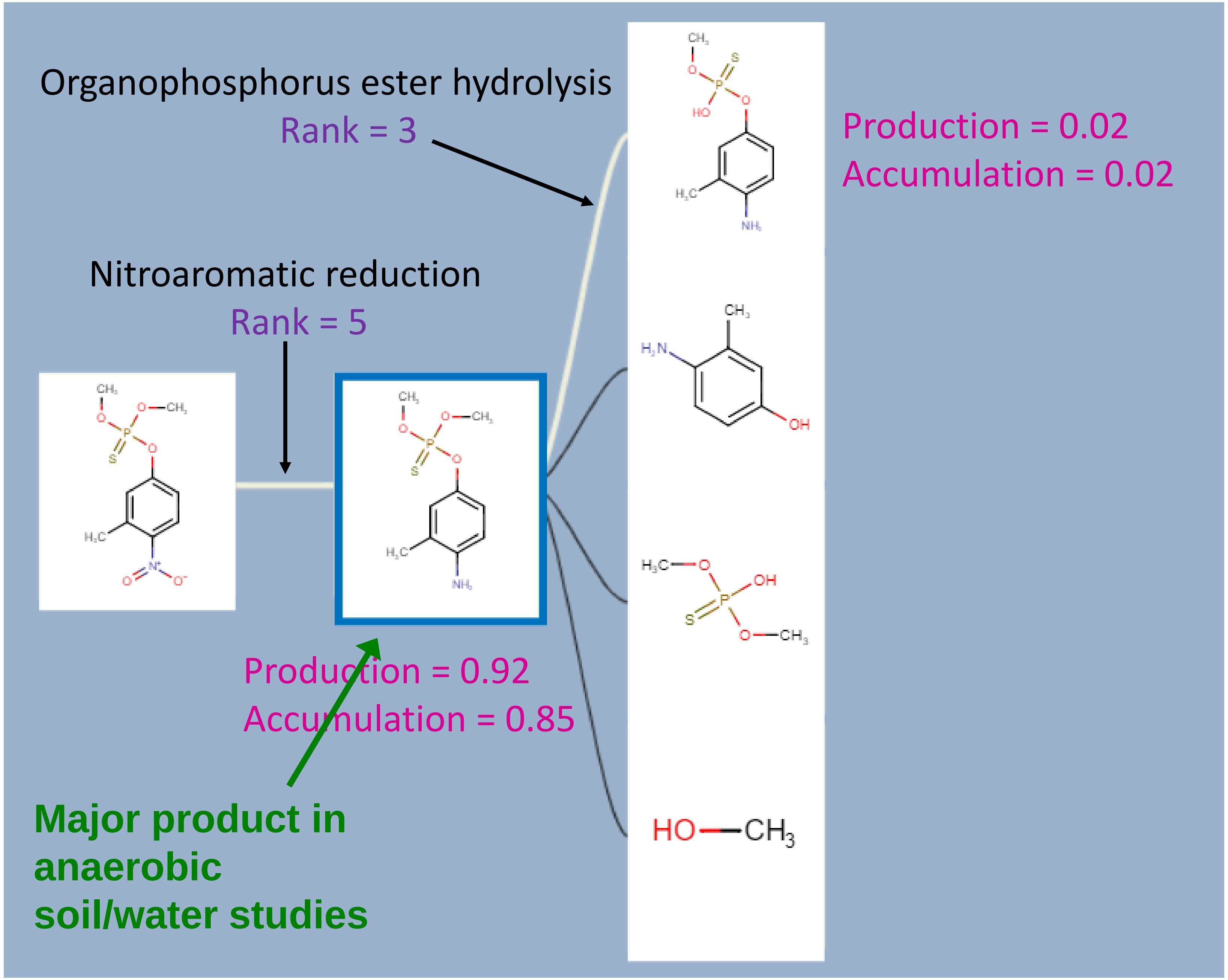
Reaction Library: a collection of schemes showing how structural groups on a molecule are modified by a particular transformation process

- Compile observed transformations (parent/product pairs) and rates from journal papers and regulatory reports
- Develop reaction schemes from journal-reported transformations
- Assign scheme ranks based on journal-reported transformation rates
- Assess performance against reported transformations in reports
- Refine schemes and ranks as needed

Rank	Half-life Range
7	< 30 min
6	30 min - 200 min
5	200 min - 24 hours
4	24 hours - 7 days
3	7 days - 60 days
2	60 days to 1 year
1	> 1 year

Predicted products of fenitrothion from CTS reaction libraries for abiotic hydrolysis and abiotic reduction

Without ranking, there would be 5 products in the first generation and 6 in the second.



Assessing Reaction Library Performance

Predictions from CTS Reaction Libraries are assessed against observed products reported in the scientific literature or regulatory reports.

Recall (a.k.a. Sensitivity)

- What proportion of the reported transformation products are predicted?

$$Recall = \frac{\#\ of\ correctly\ predicted\ products}{Total\ \#\ of\ reported\ products}$$

Precision (a.k.a. Specificity)

- What proportion of the predicted products have been reported?

$$Precision = \frac{\#\ of\ correctly\ predicted\ products}{Total\ \#\ of\ predicted\ products}$$

How does CTS address OPPT data needs?

- CTS identifies **likely transformation products** of organic chemicals in environmental and biological systems
  - CTS reaction libraries have been developed for various environmental transformation processes, with additional libraries under development
  - Strategies to identify the most likely transformation products
    - Ranking of reaction schemes based on reported transformation rates
    - Reaction rules to constrain reactivity and/or identify the most likely transformation site
  - CTS links to externally developed tools for prediction of mammalian metabolism and aerobic biodegradation
  - Exploratory work on the use of QSAR-predicted transformation rate constants to more accurately predict fractional formation of products

How does CTS address OPPT data needs? (cont)

- CTS provides estimated and measured **physicochemical property values** for both the parent and products
  - Selected properties from four physicochemical property calculators: EPI Suite™, T.E.S.T., OPERA, and ChemAxon
  - Measured values from PHYSPROP
  - Future Work: Integrate additional open-source property calculators? Provide measured physicochemical property values from other sources?

CTS User Community

- **Chemical exposure and risk assessors** can use CTS to address data gaps associated with chemical registration and assessment and to support alternative assessment activities for manufactured chemicals.
- **Environmental exposure assessment modelers** can use CTS as a parameterization tool for models that simulate/predict environmental fate, transport, bioaccumulation and dose estimation.
- **Laboratory scientists** can use CTS predictions to help interpret the results of field and laboratory studies.

CTS Development Team

Current Team Members

Caroline Tebes Stevens

Eric Weber

Kurt Wolfe

Rachel Gladstone (ORISE)

Jovian Lazare (ORISE postdoc)

Nick Pope (Contractor)

Former Team Members

Marcy Card (ORISE postdoc)

Jack Jones (retired EPA)

Michaela Koopmans (SSC)

Meredith Martin (SSC)

John Olmstead (SSC)

Jay Patel (ORISE)

Alok Shah (SSC)

Chenyi Yuan (ORISE postdoc)

Questions?

stevens.caroline@epa.gov

Office: 706-355-8218

Mobile: 706-540-3936

Try it out: https://qed.epa.gov/cts/

We welcome your feedback and suggestions for improvement!

Extra Slides

CTS Reaction Libraries

- Abiotic Hydrolysis (Tebes-Stevens et al, 2017)
  - 24 schemes; rank assignments at pH 5, 7, & 9
- Abiotic Reduction
- Direct Photolysis (Yuan et al, 2020 & 2021)
  - 154 schemes; ranked and unranked versions
- Anaerobic Biodegradation
  - currently testing and refining library of 80+ schemes
- PFAS (Weber et al, paper in preparation)
  - Focused on a chemical class, rather than a transformation process
  - Environmental transformations of PFAS (42 schemes)
  - Metabolic transformations of PFAS (48 schemes)

24 in common

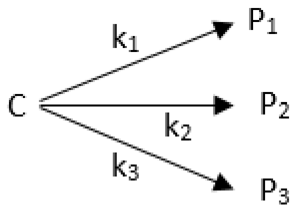
CTS Linkage to External Predictive Tools

- ChemAxon Metabolizer library for human metabolism predictions
  - 159 ranked reaction schemes, limited to phase I metabolism
- BioTransformer for metabolism prediction (http://biotransformer.ca/)
  - Four modules accessed through webservices:
    - Phase I (CYP450) Transformation
    - Enzyme Commission (EC)-Based Transformation
    - Phase II Transformation
    - Human Gut Microbial Transformation
- EnviPath for prediction of microbial biotransformation products under aerobic conditions (https://envipath.org/)
  - Rule-based, with relative reasoning to identify most likely product
  - Predictions retrieved from EnviPath servers through webservices

Use of Ranks to Predict Likelihood

CTS implements ChemAxon's Metabolizer algorithm for estimating likelihood of formation, which is based on first-order transformation kinetics.

Example with three transformation products:



$$\frac{d[C]}{dt} = -(k_1 + k_2 + k_3)[C]$$

$$\frac{d[P_i]}{dt} = k_i[C]$$

Production: $P_i = \frac{[P_i]}{[C]_0 - [C]} = \frac{k_i}{k_1 + k_2 + k_3}$

Accumulation: $Accum_i = \frac{(k_i - k_{deg,i})}{k_1 + k_2 + k_3}$

For each scheme, the characteristic rate constant (k_i) is calculated as an exponential function of the rank.

ChemAxon's Metabolizer Algorithm

For each scheme, a characteristic rate constant (a.k.a. "formation") is calculated as an exponential function of the rank:

$$k_i \stackrel{\text{def}}{=} 7^{Rank_i}$$

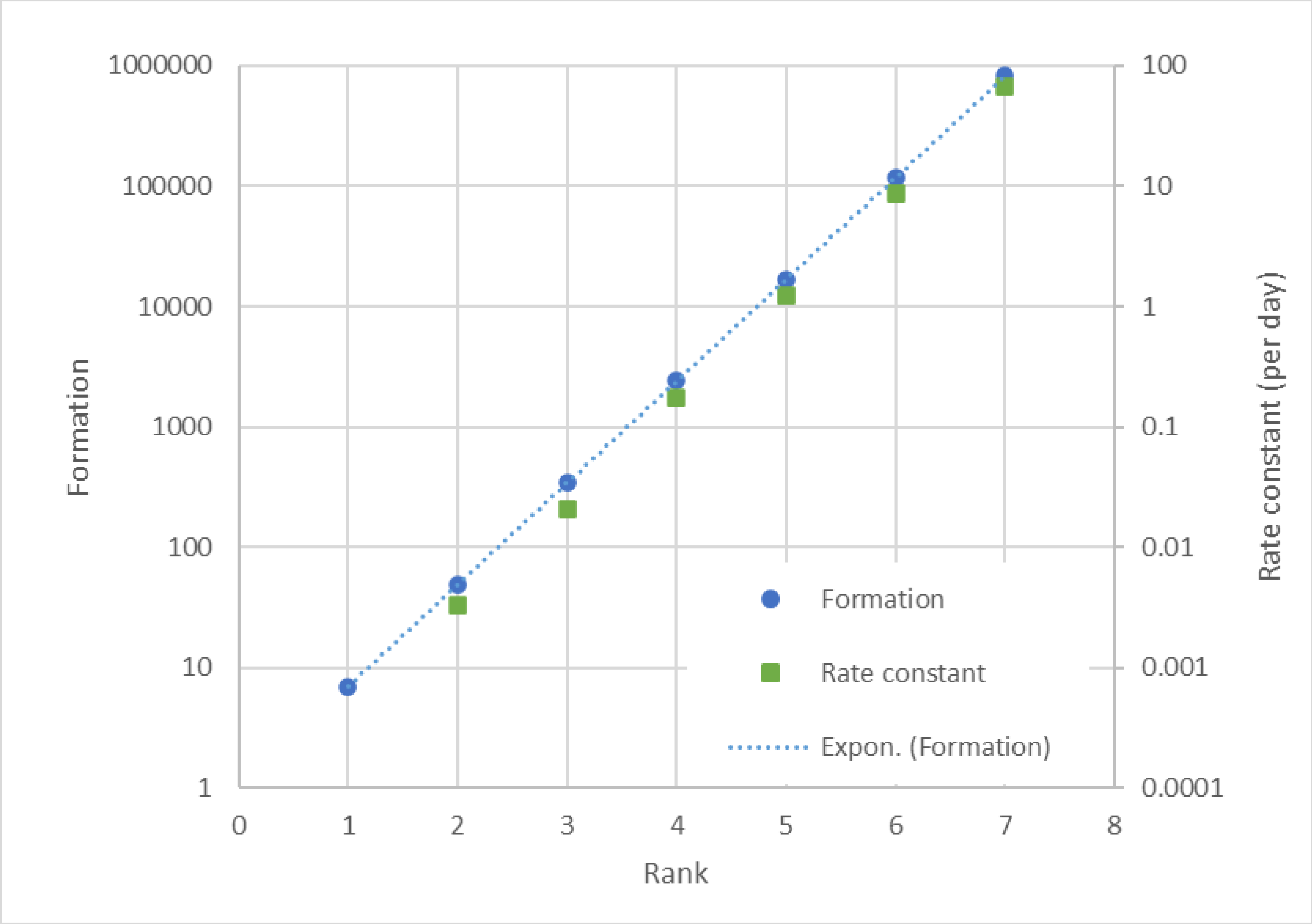
These dimensionless characteristic rate constants are used to calculate production and accumulation:

Production: $P_i = \frac{k_i}{\sum_{j=1}^{N} k_j}$

Accumulation: $Accum_i = \frac{(k_i - k_{deg,i})}{\sum_{j=1}^{N} k_j}$

To be consistent with the Metabolizer algorithm, the half-life ranges for rank assignments in CTS Reaction Libraries were selected so that the rate constants associated with the midpoints of each half-life range would lie along a line that was an exponential function of the rank.

ChemAxon's Metabolizer Algorithm (cont)



Formation:

$$k_i \stackrel{\text{def}}{=} 7^{Rank_i}$$

Rate constant:

$$k_i = \frac{\ln(2)}{\left(t_{1/2,mid}\right)_i}$$

References

Tebes-Stevens, et al (2017) Prediction of Hydrolysis Products of Organic Chemicals under Environmental pH Conditions. *Environ Sci Technol*, 51(9): 5008-5016.

Tebes-Stevens, et al (2018) Demonstration of a consensus approach for the calculation of physicochemical properties required for environmental fate assessments. *Chemosphere*, 194: 94-106.

Yuan, et al (2020) Reaction Library to Predict Direct Photochemical Transformation Products of Environmental Organic Contaminants in Sunlit Aquatic Systems. *Environ Sci Technol*, 54: 7271-7279.

Yuan, et al (2021) Prioritizing Direct Photolysis Products Predicted by the Chemical Transformation Simulator: Relative Reasoning and Absolute Ranking. *Environ Sci Technol*, 55: 5950–5958.