Jun 18th, 9:00 AM - 10:20 AM

A Chemical Properties Simulator to Support IEM

Gene Whelan
U.S. Environmental Protection Agency, Whelan.Gene@epa.gov

Eric Weber
U.S. Environmental Protection Agency, weber.eric@epa.gov

Caroline Stevens
U.S. Environmental Protection Agency, stevens.caroline@epa.gov

Mitch Pelton
Pacific Northwest National Laboratory, mitch.pelton@pnnl.gov

Kurt Wolfe
U.S. Environmental Protection Agency, wolfe.kurt@epa.gov

See next page for additional authors

Presenter/Author Information
Gene Whelan, Eric Weber, Caroline Stevens, Mitch Pelton, Kurt Wolfe, Rajbir Parmar, Mike Galvin, Said Hilal, and Justin Babendreier

International Environmental Modelling and Software Society (iEMSs)
7th Intl. Congress on Env. Modelling and Software, San Diego, CA, USA,
Daniel P. Ames, Nigel W.T. Quinn and Andrea E. Rizzoli (Eds.)
http://www.iemss.org/society/index.php/iemss-2014-proceedings

A Chemical Properties Simulator to Support IEM

Gene Whelan[a], Eric Weber[a], Caroline Stevens[a], Mitch Pelton[b], Kurt Wolfe[a], Rajbir Parmar[a], Mike Galvin[a], Said Hilal[a], Justin Babendreier[a]

[a]U.S. Environmental Protection Agency, Athens, GA USA (Whelan.Gene@epa.gov, Weber.Eric@epa.gov,Stevens.Caroline@epa.gov, Wolfe.Kurt@epa.gov, Parmar.Rajbir@epa.gov, Galvin.Mike@epa.gov, Hilal.Said@epa.gov, Babendreier.Justin@epa.gov)
[b]Pacific Northwest National Laboratory, Richland, WA USA (Mitch.Pelton@pnnl.gov)

Abstract: Users of Integrated Environmental Modeling (IEM) systems are responsible for defining individual chemicals and their properties, a process that is time-consuming at best and overwhelming at worst, especially for new chemicals with new structures. A software tool is needed to allow users to define a chemical structure, predict transformation products within an environmental setting, and calculate relevant physicochemical properties. Independent software provides relevant chemical and environmental descriptors to parameterize IEM systems that support fate/transport of organics by integrating cheminformatic applications and software technologies. These 1) encode process science using SMART reaction strings, an extension of SMILES notation; 2) generate transformation products based on functional group analysis (nitroaromatics, azo aromatics, halogenated alkanes), environmental conditions (aerobic or anaerobic), and reaction processes (reduction, hydrolysis, photolysis, biodegradation); 3) generate molecular descriptors (partition coefficients, electron affinities) through calculators; 4) collect environmental descriptors (pH, Fe(II), dissolved organic carbon, soil organic carbon content) from a user or via web-based databases (National Water Quality Database); and 5) retrieve and analyze generated data (quantitative structure activity relationships) in structure-based databases. The results are a web-based tool where the user identifies the organic chemical by structure, common or IUPAC name, or CASID; selects reaction conditions and media; provides environmental descriptors from site-specific data; and chooses a specific transformation process from a reaction library to produce transformation pathways and products, with their physicochemical properties, for IEM consumption.

***Keywords**:* Chemical properties; Cheminformatics; Transformation products; Functional groups

1. INTRODUCTION

Physicochemical properties are necessary to model fate, transport, exposure, and impacts of organic chemicals from source to receptor. The U.S. Environmental Protection Agency (EPA) has invested in a number of tools to support impacts of exposure to chemicals (e.g., DSSTox, ACToR, ExpoCast™, ExpoCastDB, ToxCast™, ToxCastDB, ToxRefDB, and ToxPi; EPA, 2014). While tools like ACToR, an online warehouse of publicly available data, are focused on toxicity-related data, there is a need for chemical properties that are more useful in fate and transport exposure modeling. Current support databases and delivery products do not provide site-specific parent-progeny relationships for new or existing organics and do not allow assessors to develop the necessary physicochemical properties automatically for fate and transport modeling from different sources including chemical databases, quantitative structure activity relationships (QSARs), and monitoring. There is a need, therefore, for a tool, preferably web accessible, that quickly helps to automate the process that

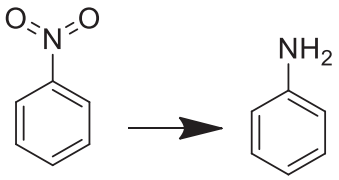
- allows users to define an organic chemical (existing or new) by structure, scientific name, or Simplified Molecular-Input-Line-Entry System (SMILES) string (Wikipedia_SMILES, 2014);
- defines environmental conditions at the user-specified location (e.g., anywhere in the United States) where the chemical has been released;

- adjusts chemical properties to existing environmental conditions in various environmental media that impact its form (e.g., ionization), mobility, transformation, and toxicity;
- defines the potential transformation products, based on the latest chemical reaction rules, chemical characteristics, and environmental conditions;
- provides physicochemical properties from a number of sources, including chemical databases, QSARs, and monitoring for the parent chemical and predicted transformation products; and
- delivers chemical-specific data through interactive or batch means that can be readily consumed by fate and transport exposure modeling models and assessors.

The objective is to use chemical structure, knowledge of process-based chemical reaction science, and site-specific environmental conditions to predict transformation products and physicochemical properties (e.g., mobility, reactions rates) used in fate and transport modeling, where concentrations and risk calculations can be performed for parent chemicals and predicted transformation products.

2. METHODS AND MATERIALS

The conceptual design describes the structure of organic chemicals in a universal manner and predicts transformation products and physicochemical properties, based on functional groups (e.g., nitro, methyl, azo, amine), process science (reaction rules for reduction, hydrolysis, oxidation, etc.), and environmental conditions [e.g., pH, temperature, soluble Fe(II)]. For example, Figure 1 presents six different reactions with functional groups that are susceptible to reduction. By classifying chemicals and their functional groups, their transformation products and properties (e.g., transformation rate coefficients, sorption properties, etc.) can be predicted -- if the chemical reaction rules can be codified -- based on molecular descriptors (e.g., partition coefficients, electron affinities), functional groups, and environmental conditions. An example of a molecular descriptor that can be calculated is ionization constants (pKa values) which are important because ionization will greatly impact distribution and, thus, the reactivity of organic chemicals in aquatic ecosystems. A simple example of a reaction rule is the ability to recognize reduction of the nitroaromatic functional group to the aromatic amine, as described by a SMART Reaction string: O=N(=O)C1=CC=CC=C1>>NC1=CC=CC=C1. The challenge is to determine if the predicted transformation will occur under the existing environmental conditions. A process to predict transformation products and their physicochemical properties is shown in Figure 2; an explanation of Figure 2 is presented as follows:



1. Develop and capture rules for transformation products of relevant organic compounds, based on molecular descriptors and functional groups.
2. Provide an option for chemical entry. The structure and subsequent name of the parent organic chemical is defined, resulting in a SMILES string descriptor.
3. Consider environmental properties from data-mining for environmental descriptors which are based on oxidizing versus reducing environments, environmental medium (e.g., groundwater, surface water), and environmental chemical conditions [e.g., pH, temperature, soluble Fe(II)], and which are also a function of geographic location.
4. Generate potential transformation products based on environmental conditions and transformation processes such as reduction and hydrolysis.
5. Provide parameterization and execution of QSARs and algorithms to approximate additional environmental properties (e.g., rate constants).
6. Summarize physicochemical properties for parent/transformation products.
7. Populate a database with calculated and measured physicochemical properties of parent and potential transformation products for future access and retrieval.

3. RESULTS AND DISCUSSION

Based on chemical structure and knowledge of the environmental system of interest (e.g., surface water versus anaerobic sediment), it is possible to use process science to estimate physicochemical properties of parent organic chemicals and their transformation products, and then transfer these

properties to models that need them. This requires knowledge of functional groups susceptible to reactions such as abiotic reduction (e.g., nitroaromatics, azo aromatics, halogenated alkanes), which is supplied by reaction libraries designed by experts; molecular descriptors for predicting mobility and reaction rates, which require software technology to access physicochemical calculators; environmental descriptors and conditions necessary for predicting reaction rates, which require software technology to access online databases; and parameterization (e.g., chemical property input requirements, including metadata) of environmental fate and transport models, which requires software technology to provide seamless parameterization and transfer between production of physicochemical properties and model consumption.

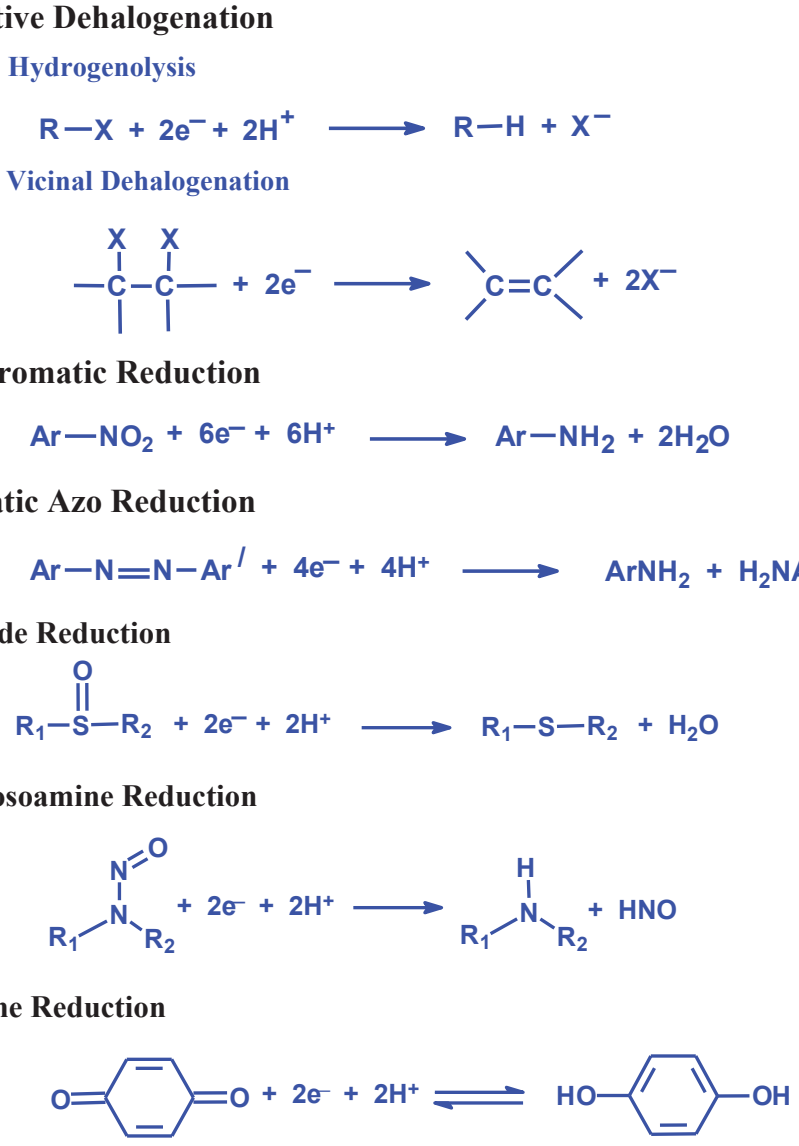


Figure 1. Examples of functional groups that are susceptible to reduction where Ar and R represent aromatic and non-aromatic groups, respectively.

The foundation of a chemical fate simulator is its ability to codify reaction rules, based on laboratory experimentation and expert knowledge that underlie the process science. These rules are captured in

software reaction libraries of one-step reactions based on functional group transformation for reduction, hydrolysis, photolysis, oxidation by water disinfectants, and aerobic and anaerobic biodegradation. Reaction libraries describe the functional groups associated with the reaction and the order in which they react. Examples of process science include -Br, -Cl, -NH_2, -OH, and OCH_3 groups being ortho-directing, and alkyl groups being para-directing for reduction of 2,4-dinitroaromatics where this process science is encoded in a SMART Reaction String (i.e., [F,Cl,Br,I:11][c:1]1[cH:2][cH:3][c:4]([cH:5][c:6]1[N:7](=[O:12])=[O:13])[N:8](=[O:9])=[O:10]>>[NH2:7][c:6]1[cH:5][c:4]([cH:3][cH:2][c:1]1[F,Cl,Br,I:11])[N:8](=[O:9])=[O:10], as noted earlier. The encoded process science for an expert-developed reaction library for reductive transformation would, for example, include hydrogenolysis, vicinal dehalogenation, aromatic azo reduction, sulfoxide reduction, N-nitrosomine reduction, quinone reduction, reductive dealkylation, and isoxazole cleavage.

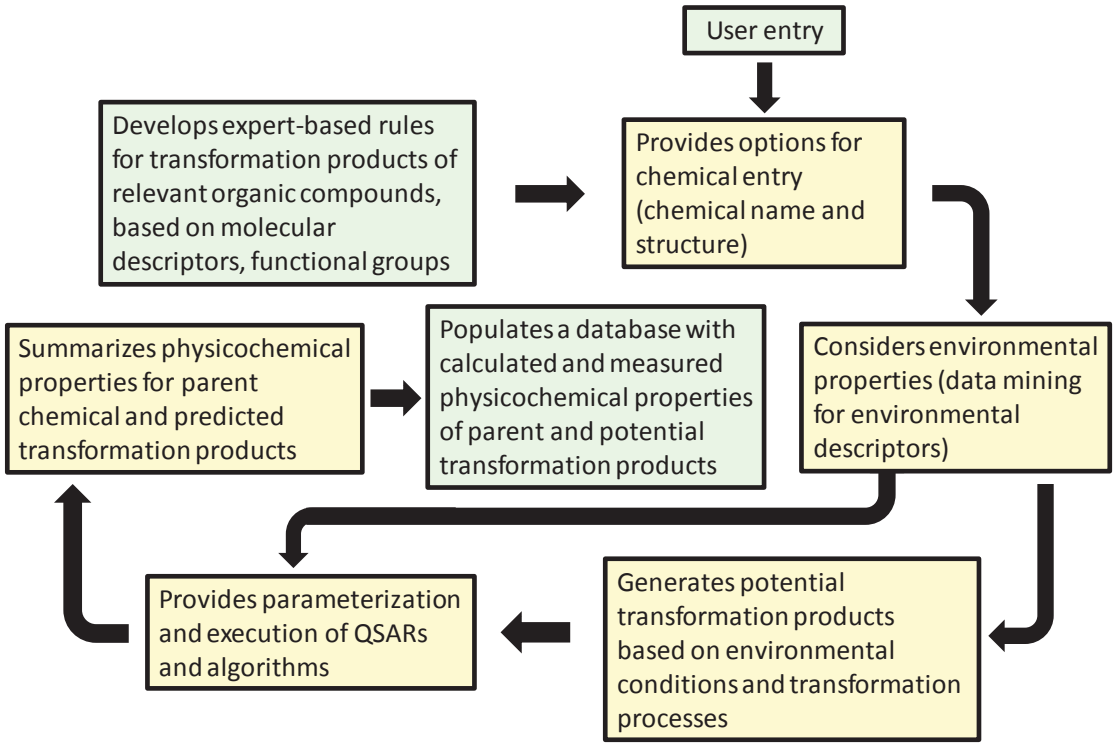


Figure 2. Components of a chemical fate simulator

The foundation of a chemical fate simulator is its ability to codify reaction rules, based on laboratory experimentation and expert knowledge that underlie the process science. These rules are captured in software reaction libraries of one-step reactions based on functional group transformation for reduction, hydrolysis, photolysis, oxidation by water disinfectants, and aerobic and anaerobic biodegradation. Reaction libraries describe the functional groups associated with the reaction and the order in which they react. Examples of process science include -Br, -Cl, -NH_2, -OH, and OCH_3 groups being ortho-directing, and alkyl groups being para-directing for reduction of 2,4-dinitroaromatics where this process science is encoded in a SMART Reaction String (i.e., [F,Cl,Br,I:11][c:1]1[cH:2][cH:3][c:4]([cH:5][c:6]1[N:7](=[O:12])=[O:13])[N:8](=[O:9])=[O:10]>>[NH2:7][c:6]1[cH:5][c:4]([cH:3][cH:2][c:1]1[F,Cl,Br,I:11])[N:8](=[O:9])=[O:10], as noted earlier. The encoded process science for an expert-developed reaction library for reductive transformation would, for example, include hydrogenolysis, vicinal dehalogenation, aromatic azo reduction, sulfoxide reduction, N-nitrosomine reduction, quinone reduction, reductive dealkylation, and isoxazole cleavage.

Once reaction rules are codified, the system uses them (as illustrated by a series of screen captures in Figure 3) to automate the process of predicting transformation products and physicochemical properties of parents and progeny. The Chemical Editor (Figure 3a) provides options for chemical

entry (chemical identifier). The user has the option of constructing a chemical structure which is useful for new and emerging chemicals, or supplying a CASID or IUPAC name or SMILES string. Based on user-supplied information, other identifiers are determined, when available, as well as the chemical formula and molecular weight.

To determine transformation products (Figure 3b), a Define Transformation Pathways screen requires the user to supply certain environmental descriptors such as whether the reaction is under aerobic versus anaerobic conditions; the medium of interest (i.e., surface water, benthic sediment, surface soil, vadose zone/groundwater, and solid waste); and the transformation process of interest (e.g., reduction, hydrolysis, photolysis, and aerobic and anaerobic biodegradation). The system provides consistency checks between respiration and media categories: for example, aerobic and anaerobic environments are usually associated with surface waters and benthic sediments, respectively.

Using expert-derived reaction rules contained in the software reaction libraries, the Reaction Pathway Simulator (Figure 3c) generates potential transformation products based on information supplied by the user on the transformation pathways screen. This screen presents parent chemical and transformation products by level, with SMILES strings and IUPAC names. The user can limit the number of levels associated with progeny since an increase in the number of levels produces an increase in potential branching associated with potential routes for transformation. The expression describes major and minor pathways and physicochemical properties associated with chemical transformation pathways and conditions under which a given pathway will be expressed.

By combining results from the Reaction Pathway Simulator, which presents all possibilities for transformation products based on the selected transformation process, with additional environmental conditions/descriptors (Figure 3d), the most relevant physicochemical properties for that location can be determined. Key environmental parameters include water temperature, pH, soluble Fe(II), fraction organic content, and dissolved organic content. Data-mining techniques allow access to necessary databases (e.g., USGS National Water Quality Database; USGS, 2014) for collection of the required environmental descriptors. The user identifies the latitude and longitude of the site from mapping software or by inserting the coordinates and the software accesses, retrieves, and stores the data for future use or user modification.

The Physicochemical Properties Calculator (Figure 3e), coupled with a Reaction Rate Calculator that uses QSARs and other algorithms to parameterized rate coefficients and other properties (e.g., sorption), accesses chemical property calculators [e.g., ChemAxon (ChemAxon, 2014); SPARC (EPA, 2013a); EPISuite (EPA, 2013b); FRAMES (PNNL, 2008)] to list and compare values of common calculated and measured physicochemical properties. The user can store properties of parent and potential transformation products in a structure-based database, providing the ability to roll-up chemical-specific, physiochemical data from multiple sources; prioritize data; weight data sources and combine values; express data as a distribution, variance, and expected value from all sources; and pre-populate input files of fate and transport models.

4. SUMMARY

Our effort provides physicochemical properties of parent and progeny for eventual seamless consumption by modeling tool sets to assess environmental exposure and subsequent human/ecological receptor health risks associated with loading and fate/transport of organic chemicals and their transformation products. A chemical fate simulator develops and captures expert-developed rules for transformation products of relevant organic compounds based on molecular descriptors, functional groups, and environmental conditions to help determine physicochemical properties such as transformation rate coefficients, sorption properties, etc. These rules are expressed to resonate easily with users, such as chemical structures, reaction products, unique SMILES strings, etc. Although the long-term goal is to develop the ability to simulate all major environmental transformation processes (hydrolysis, reductive transformation, biodegradation, photolysis, irreversible sorption through covalent binding, and transformation by drinking water disinfectants), the initial focus is simulation of the transformation pathways associated with reductive transformation, hydrolysis, and irreversible sorption through covalent binding. The design allows multiple researchers to register their experiments and allows models to have real-time access to data

to utilize physicochemical properties based on changing environmental conditions. This capability supports evaluation of existing and new organic compounds, including military relevant (energetic) compounds, and to quantitatively estimate environmental concentrations and their relationships to compliance limits

ACKNOWLEDGMENTS

This paper has been reviewed in accordance with the United States Environmental Protection Agency's peer and administrative review policies and approved for presentation and publication.

REFERENCES

ChemAxon, 2014. ChemAxon. http://www.chemaxon.com/ (last accessed 23.02.14.).

EPA (U.S. Environmental Protection Agency), 2013a. SPARC performs automated reasoning in chemistry. http://www.epa.gov/athens/research/sparc.html (last accessed 23.02.14.).

EPA (U.S. Environmental Protection Agency), 2013b. Exposure assessment tools and models, Estimation Program Interface (EPI) Suite. http://www.epa.gov/oppt/exposure/pubs/episuite.htm (last accessed 23.02.14.).

EPA (U.S. Environmental Protection Agency), 2014. Computational toxicology research. http://epa.gov/ncct/index.html (last accessed 23.02.14.).

PNNL (Pacific Northwest National Laboratory), 2008. EARTH: Environmental assessment & risk tools homepage – Software & development tools - MMEDE http://mepas.pnnl.gov/FramesV1/mmede.stm (last accessed 23.02.14.).

USGS (U.S. Geological Survey), National Water-Quality Assessment (NAWQA) Program. 2014; https://water.usgs.gov/nawqa/ (last accessed 23.02.14.).

Wikipedia_SMILES, 2014. Simplified molecular-input line-entry system http://en.wikipedia.org/wiki/Simplified_molecular-input_line-entry_system (last accessed 23.02.14.).

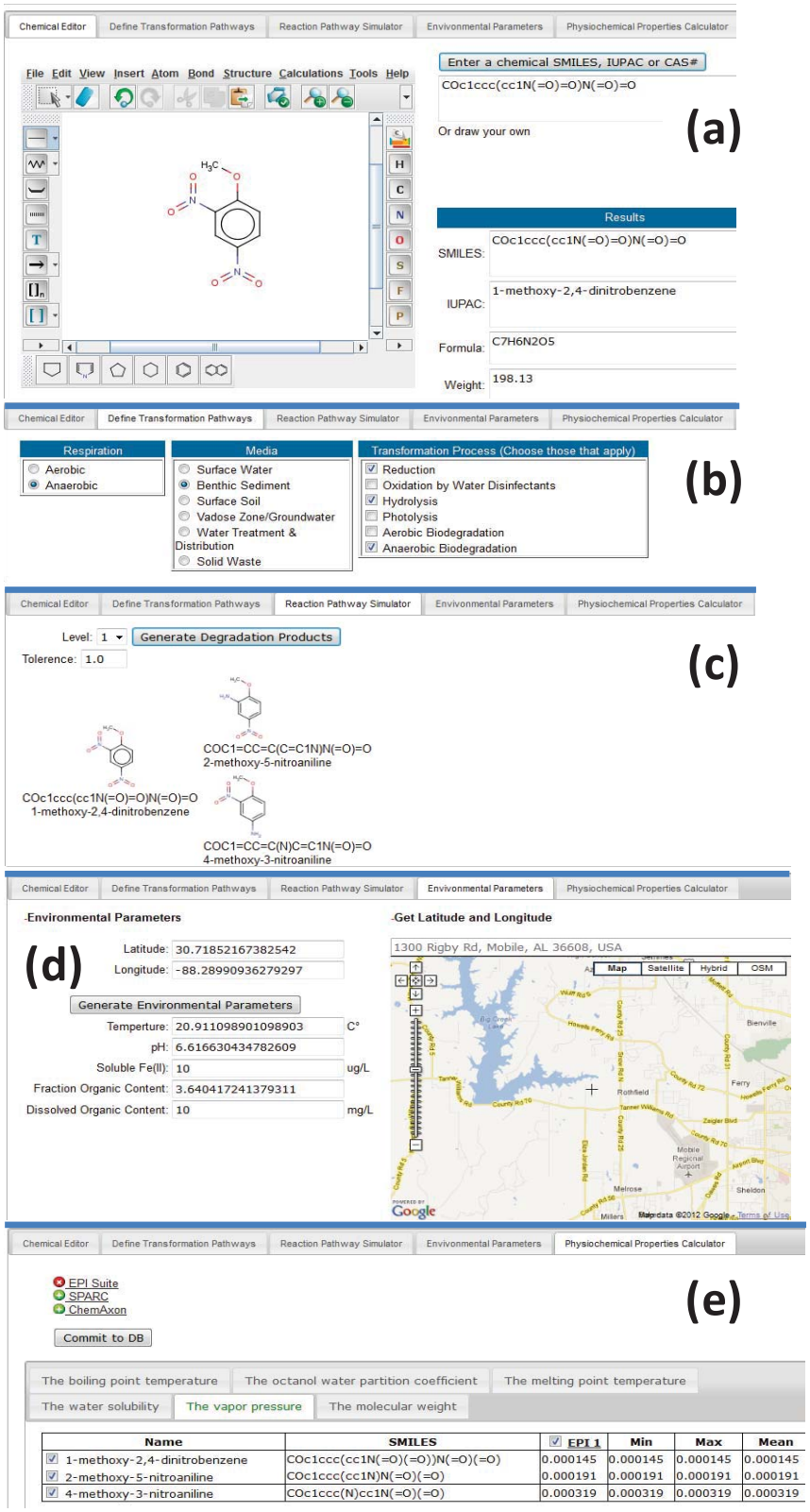


Figure 3. Design of a chemical fate simulator: (a) Chemical editor, (b) Define transformation pathways, (c) Reaction pathway simulator, (d) Environmental properties, and (e) Physiochemical properties calculator.