An integrative method for identification and prioritization of constituents of concern in produced water from onshore oil and gas extraction

Cloelle Danforth[a,*], **Weihsueh A. Chiu**[b], **Ivan Rusyn**[b], **Kim Schultz**[c], **Ashley Bolden**[c], **Carol Kwiatkowski**[c], **Elena Craft**[d,*]

[a]Environmental Defense Fund, 2060 Broadway, Suite 300, Boulder, CO 80302, USA

[b]Department of Veterinary Integrative Biosciences, Texas A&M University, 4458 TAMU, College Station, TX 77843, USA

[c]The Endocrine Disruption Exchange, PO Box 54, Eckert, CO 81418, USA

[d]Environmental Defense Fund, 301 Congress Ave #1300, Austin, TX 78701, USA

Abstract

In the United States, onshore oil and gas extraction operations generate an estimated 900 billion gallons of produced water annually, making it the largest waste stream associated with upstream development of petroleum hydrocarbons. Management and disposal practices of produced water vary from deep well injection to reuse of produced water in agricultural settings. However, there is relatively little information with regard to the chemical or toxicological characteristics of produced water. A comprehensive literature review was performed, screening nearly 16,000 published articles, and identifying 129 papers that included data on chemicals detected in produced water. Searches for information on the potential ecotoxicological or mammalian toxicity of these chemicals revealed that the majority (56%) of these compounds have not been a subject of safety evaluation or mechanistic toxicology studies and 86% lack data to be used to complete a risk assessment, which underscores the lack of toxicological information for the majority of chemical constituents in produced water. The objective of this study was to develop a framework to identify potential constituents of concern in produced water, based on available and predicted toxicological hazard data, to prioritize these chemicals for monitoring, treatment, and research. In order to integrate available evidence to address gaps in toxicological hazard on the chemicals in produced water, we have catalogued available information from ecological toxicity studies, toxicity screening databases, and predicted toxicity values. A Toxicological Priority Index (ToxPi) approach was applied to integrate these various data sources. This research will inform stakeholders and decision-makers on the potential hazards in produced water. In addition, this

*Corresponding authors at: Environmental Defense Fund, 301 Congress Ave, Suite 1300, Austin, TX 78701, USA. cdanforth@edf.org (C. Danforth), ecraft@edf.org (E. Craft).

work presents a method to prioritize compounds that, based on hazard and potential exposure, may be considered during various produced water reuse strategies to reduce possible human health risks and environmental impacts.

Keywords

Oil-field wastewater; Unconventional; Hydraulic fracturing; ToxPi; Ecotoxicology; Human health

1. Introduction

Produced water is the largest volumetric waste stream in oil and gas production; it is a combination of water that occurs from within or proximate to the hydrocarbon-bearing zone as well as any injected fluid used to stimulate or maintain production activities (Fakhru'l-Razi et al., 2009). This wastewater has complex chemistry (Stringfellow and Camarillo, 2019) and is generated during all types of petroleum hydrocarbon extraction activities, including conventional and unconventional production. As oil and gas production expands, so does the challenge to manage the estimated hundreds of billion gallons of wastewater generated per year from these activities (Veil, 2015). While approximately 90% of produced water is reinjected deep into the subsurface for enhanced recovery or disposal, there are increased efforts to recycle produced water in the oilfield (Liden et al., 2018). In addition, in arid-prone areas there are incentives to consider alternative management strategies of produced water, including for agricultural purposes. However, there remain outstanding questions as to the chemical composition and potential human and ecological hazards of produced water.

Produced water characteristics from both conventional and unconventional oil and gas development vary by formation and geology, time since well-emplacement, and types of chemicals used in operation. Generally, produced water contains dissolved and dispersed oil compounds, salts, metals, naturally occurring radioactive materials, stimulation and production chemicals, as well as the degradants and transformation compounds from these additives (Hoelzer et al., 2016). Currently, the majority of oil and gas producing states require disclosure of hydraulic fracturing additive chemicals, and most of these states allow for, or require, the use of FracFocus—a chemical registry for hydraulic fracturing in the United States—to meet these reporting requirements. The concentrations of produced water constituents can differ by orders of magnitude (Estrada and Bhamidimarri, 2016; Fakhru'l-Razi et al., 2009). Individual chemicals and their concentrations can be difficult to quantify or even identify. Many of the compounds associated with either hydraulic fracturing fluids or those identified in produced water lack established analytical methods (Oetjen et al., 2017; U.S. EPA, 2018). Advanced research-based analytical methods have been used to detect and quantify organic compounds in produced water, including homologous series of surfactants (Nell and Helbling, 2018; Thurman et al., 2017, 2014), biocides (Nell and Helbling, 2018), and compounds that are suspected additives but which are unreported to FracFocus (Hoelzer et al., 2016; Sitterley et al., 2018). Overall, there is not a comprehensive understanding of produced water composition (Ferrer and Thurman, 2015; Nell and Helbling, 2018; Oetjen et al., 2017).

Current Federal Effluent Limitation Guidelines (ELG), under Title 40 of the Code of Federal Regulation (CFR), limit the direct discharge of produced waters to surface water, with an exception for western states, where the only defined ELG is for oil and grease at 35 mg/L (40 CFR §435.53, 1995). Produced water discharge may include a wide array of treatment options, ranging from robust, multi-step treatment systems that use ultra-filtration and reverse osmosis units (CDPHE, 2011) to those that discharge water that has been minimally treated using oil/water separators (CDPHE, 2012). The U.S. EPA study of offsite Centralized Waste Treatment (CWT) facilities, which can be authorized to treat produced water for discharge throughout the United States under 40 CFR §437, concluded that CWTs without multi-step treatment are often deficient and may not remove many of the pollutants found in produced water (U.S. EPA, 2018). There are considerable ambiguities in how produced water should comply with existing federal regulations under the Clean Water Act (CWA) and many practices of disposal of produced water that do not involve surface water discharge, such as road de-icing, dust suppression, and irrigation, are not subject to the CWA (U.S. EPA, 2019a).

Practices that intentionally discharge produced water to the environment have the potential to increase risks to human health and the environment by introducing a new exposure pathway to an understudied wastewater (Faber et al., 2019). To date, the majority of studies on chemical hazards associated with onshore produced water have focused on hydraulic fracturing chemicals used in the stimulation process rather than the produced water generated by it and other development practices (Colborn et al., 2011; Elsner and Hoelzer, 2016; Gordalla et al., 2013; Rogers et al., 2015; Yost et al., 2016a, 2017).

Research on toxicity of produced water itself has been challenging due to lack of access to oil and gas wastewater samples (Santos et al., 2019). One research team at the University of Alberta has conducted integrated studies on the characterization and toxicity of flowback and produced water from a hydraulically fractured gas well in the Upper Devonian-aged Duvernay Formation (Blewett et al., 2017a, 2017b; He et al., 2018, 2017). Their work has indicated that there are potential impacts on a variety of aquatic organisms from additives and their transformation products, as well as from naturally occurring organic and inorganic constituents. Additionally, the findings indicate that some toxicity effects are partially associated with particulate matter, likely due to organic sorption to the particulates, and not from salinity alone.

Prior to their finalized 2016 report examining potential impacts from the hydraulic fracturing water cycle on drinking water (U.S. EPA, 2016), U.S. EPA released draft reports on the study, which included approximately 130 chemicals in flowback or produced water identified in literature reviews (U.S. EPA, 2015). These data were used by some research groups to assess the potential hazard of produced water by reviewing associated chemical toxicity data. Elliot et al. (2017) searched the REPROTOX database, which is comprised of reproductive and developmental toxicity data and reported that some toxicity information was available for only 55% of these compounds. Yost et al. (2016) conducted a review of chronic oral toxicity values and found that only 62% of the chemicals have some toxicity values available.

Our aim was to develop an integrative method to identify and prioritize chemical constituents in produced water, thereby addressing existing gaps in our knowledge on the potentially hazardous chemicals in produced water. First, we performed a comprehensive literature search to aggregate available information on chemical constituents detected in onshore produced water in North America. Next, data on the potential ecotoxicological and human health hazards were identified for these chemicals and used to create a prioritized list of the produced water constituents of concern. Finally, we determined whether analytical methods for the chemicals of concern are available to support future exposure assessment efforts. Collectively, our approach represents a comprehensive framework to identify potential constituents of concern in produced water in order to prioritize these chemicals for monitoring, treatment, and research based on available and predicted toxicological hazard data.

2.   Methods

2.1.   A literature search for chemicals identified in produced water

A comprehensive literature search was designed to identify studies that performed a chemical analysis of flowback or produced water generated onshore during oil and gas development. Searches were conducted electronically using Web of Science and PubMed for all years through March 8, 2018. Search keywords and logic are reported in SI Table 1. Additional articles were identified using cited references from review articles. A list of review articles can be found in SI Table 2. Two independent reviewers screened titles and abstracts for relevance using DistillerSR. To be considered for inclusion, the studies had to present primary results for chemical analysis of flowback or produced water from onshore oil and gas production sites in Canada, Mexico, or the United States and be in the English language. Two independent reviewers resolved discrepancies regarding inclusion through discussion; if needed, a third reviewer was also consulted. The articles that met inclusion criteria underwent full text review and data extraction for chemical data, including Chemical Abstract Service (CAS) number and concentration, if reported in the study. Often, CAS Registry Numbers were not included in the literature; if provided by the author, CAS numbers were extracted as reported and their authenticity confirmed using the U.S. EPA CompTox Chemistry Dashboard ("Chemistry Dashboard," 2019). Otherwise, the Chemistry Dashboard was used to look up CAS numbers by chemical name.

2.2.   Cross walks

To understand what information was available for each chemical, the aggregated list of chemicals derived from the literature review was compared by CAS number to a variety of databases. These databases included:

i.   Quantitative measures of dose-effect from databases identified by the U.S. EPA Office of Solid Waste and Emergency Response (OSWER) Directive 9285.7–53 (U.S. EPA, 2003);

ii.   Chemistry Dashboard database of toxicity values (ToxValDB) (Williams et al., 2017);

iii. Select chemical lists compiled by regulatory authorities, including the list of National Primary Drinking Water Standards (NPDWS) (U.S. EPA, 2009), the U.S. EPA Priority Pollutant List (PPL) compiled under the Clean Water Act (40 CFR §423, Appendix A, 2015), and hazardous waste defined under Appendix VIII of Part 261 of the Resource Conservation and Recovery Act (RCRA);

iv. Chemicals listed under the U.S. EPA-approved test methods for the analysis of environmental media (Hazardous Waste Test Methods/SW-846), those listed in Title 40 of the Code of Federal Regulations Part 136, which establishes the guidelines for test procedures for the analysis of pollutants (40 CFR §136, 2017), and those listed on the National Environmental Methods Index (NEMI), which is a database of analytical and field methods for environmental monitoring ("National Environmental Methods Index," 2017); and

v. Chemicals reported in FracFocus from January 2013 to January 24, 2018 (FracFocus, 2018).

2.3. Collection and analysis of toxicity data for produced water chemicals

Chemicals that were considered in the toxicity prioritization analyses were restricted to constituents that (1) had been detected in produced water more than once, (2) had reported concentration data, and (3) were known to have some toxicity data available, specifically those chemicals that were included in the Toxicity Value Database (ToxValDB) (Williams et al., 2017) and the Toxicity Forecaster (ToxCast) program ("Chemistry Dashboard," 2019; "Toxicity Forecasting," 2019). Chemicals that did not have these attributes were not included in further analyses. For the list of included chemicals, available toxicity data were collected, analyzed, and integrated for the purposes of prioritization of potential toxicity. These data were collected from three sources: the Ecotoxicology Knowledgebase (U.S. EPA, 2019b), the Chemistry Dashboard, and the Conditional Toxicity Value (CTV) Predictor (Wignall et al., 2018). These three sources reported hazards based on ecotoxicity testing, bioactivity from high throughput *in vitro* screening, and regulatory toxicity values for human health, respectively.

2.3.1. ECOTOX database—The ECOTOX database is maintained by the U.S. EPA's National Health and Environmental Effects Research Laboratory's Mid-Continent Ecology Division and includes toxicity data for aquatic organisms and terrestrial plants and wildlife (U.S. EPA, 2019b). These data are primarily from peer-reviewed literature. Data included in the ECOTOX database are for single chemicals with verifiable CAS numbers that demonstrate a biological effect on a live, whole organism. Concentration data are converted into units of parts per million where information reported in the literature make such conversions possible. All available data for the half-maximal effective concentration (EC50) were extracted and analyzed here; however, the original references for those data were not examined. EC50 for aquatic organisms was chosen as the representative endpoint for ecotoxicological hazard data as this was the richest dataset (i.e. most data available for each chemical) in the ECOTOX database over other effective concentrations or endpoints (e.g., EC15 or LD50) or for terrestrial organisms. While the database contains species-level information, data were aggregated for each chemical by species group. For

example, green algae (*Pseudokirchneriella subcapitata*) and diatoms (*Skeletonema costatum*) were considered collectively as the species group "algae." To establish a representative characteristic effective concentration for each chemical, the lower quartile (Q1) of EC50 was calculated across all effects for the most sensitive species group after removing outliers. Outliers were detected and removed using the ROUT (**R**obust regression and **Out**lier removal) method (Motulsky and Brown, 2006) in GraphPad Prism version 8.0.1 for Windows (GraphPad Prism 8, 2019). If less than four results were available for a chemical, then instead of the Q1, the lowest value was used. This approach represents a reasonably, but not overly, conservative estimate of the potential hazard that may not have considered all sensitive species or endpoints. The robustness of the results to a number of these assumptions was assessed using sensitivity analyses, described below in Section 2.5.

2.3.2.   EPA chemistry dashboard—The ToxCast screening library contained 4746 substances that had been tested in a panel of *in vitro* assays with a variety of cell types, design, or bioactivity type. For each chemical and assay type, whether the chemical were active and at what half-maximal concentration (AC50), was reported in the ToxCast database. These data were accessed through the Chemistry Dashboard and, similar to the EC50 data, AC50 for all available assays were analyzed collectively. Chemicals that did not show activity below the highest tested concentration were considered inactive and not included for that component in ToxPi. A representative concentration of AC50 for each chemical was determined in the same manner as described above for ecotoxicology data (Section 2.3.1). *In vitro* assay data for chemicals with a boiling point less than 200 degrees Fahrenheit were excluded, as the compound volatility may lead to artificially low concentrations in the assay or may create false negative results. In addition to the representative AC50 value, the percent of active assays was used as another indicator of potential hazard.

2.3.3.   CTV predictor—The Conditional Toxicity Value (CTV) predictor (Wignall et al., 2018) is a web-based quantitative structure activity relationship (QSAR) modelbased *in silico* approach for developing human health quantitative risk estimates for chemicals where there are none. The CTV predictor generates quantitative predictions for a variety of toxicity values (oral and inhalation exposures), with an uncertainty of an order of magnitude or less. Here, only chemicals within the applicability domain were considered. Additionally, if a chemical were part of the "training set" with existing published, peer-reviewed regulatory toxicity values, then the regulatory value was used instead of the QSAR-generated value. Furthermore, only toxicity values for oral exposure were included for this data set as the primary potential route of exposure to chemicals in produced water was assumed to be ingestion over inhalation.

2.4.   Data integration and prioritization using ToxPi

The Toxicological Prioritization Index (ToxPi) is an analytical framework that can be used to integrate multiple lines of evidence from a variety of sources into a dimensionless index score, as described elsewhere (Marvel et al., 2018; Reif et al., 2010). In this analysis, ToxPi was used to calculate a weighted combination of each data source to illustrate how each component (or slice) contributed to the overall toxicity profile for each chemical. The higher

the ToxPi score, the higher the potential toxicity of that chemical relative to the others in the study based on the components considered for that analysis. The visual representation of the index score allows for a rapid understanding of the underlying weight-of-evidence scheme employed in the analysis. To harmonize quantitative data, data are scaled or inverted, where appropriate, so that the compound with the lowest toxicity value is represented by the smallest, positive non-zero number in the set. To normalize ToxPi scores, each slice was scaled so that the data were distributed linearly. The scaling equations are presented in the Supplemental Information (SI Table 3). By default, component data that are missing for a particular chemical are considered to be zero by ToxPi. Therefore, where data were missing, the minimum values for that component data set (i.e. least toxic value) of all individual assays was imputed; this assumes chemicals with missing data caused a minimal effect or activity (To et al., 2018). In this analysis, toxicity data were used for each chemical to build component slices in three domains. Each slice was derived using characteristic concentrations for the toxic effect the chemical had on aquatic organisms (ECOTOX), cell-based high throughput screening assays (ToxCast), or by using known or calculated human toxicity values. The three domains were: *in vivo* endpoints, *in vitro* assays, and known or conditional toxicity values (CTV domain). The definition and notation of ToxPi is presented in Fig. 1.

2.5. Sensitivity analysis

To understand the impacts of data analysis or inclusion decisions for the three data sets, the following sensitivity analyses were performed: (1) for missing values, using no values (blanks) instead of the minimum in the analysis; (2) for both the ecotoxicity data and *in vitro* activity data, using median values (Q2) or geometric mean versus the lower quartile and minimum values in the main analysis for characteristic half-maximal concentrations; (3) including only data-rich chemicals defined as those having at least five out of nine component slices; and (4) using different weighting schemes in ToxPi that gave equal weight to each endpoint ("unweighted," Fig. 1A) versus giving equal weight to each domain in the main analysis ("weighted," Fig. 1B). Additionally, chemicals listed on the EPA's Priority Pollutants List (PPL), which are known to be hazardous, were included in the ToxPi analysis as reference chemicals to benchmark our methodology.

3. Results

3.1. Identification of chemicals in produced water through comprehensive literature review

Based on the literature search, a total of 15,661 records were returned (11,996 from Web of Science and 3,665 from PubMed), with an additional 33 studies identified through review articles (SI Table 2). After title/abstract screening, 302 articles were assessed for full-text eligibility and a final total of 129 articles were included for data extraction (Fig. 2, SI Table 4). These papers detailed chemical analysis for 173 sources of produced water (by state and hydrocarbon development type) collected from 27 locations in North America: 24 States in the USA, 2 Canadian Provinces, and 1 Mexican State (SI Fig. 1). The majority of produced water reported in these studies was sourced from unconventional developments (112). Produced water collected from conventional developments (35) and coalbed methane

(26) were represented more similarly. The source of produced water in these studies was predominately from the Marcellus basin in Pennsylvania, even though the majority of produced water volume is generated in areas such as Texas, California, and New Mexico (SI Fig. 1). From the comprehensive literature search, a total of 1,198 unique chemicals were identified as having been detected in produced water (SI Table 4). Of these, 296 chemicals satisfied the additional criteria of being detected more than once and having associated concentration data; 122 chemicals had the further criteria of having at least two types of toxicity data (Fig. 3, SI Table 5).

3.2. Comparison of produced water chemicals to other databases

The initial literature review identified 1,198 chemical constituents in produced water for which there were a CAS number identified (SI Table 4). The following sections describe their presence on selected databases; the results are summarized in Fig. 3.

3.2.1. Quantitative toxicity values—To perform a human health risk assessment for particular contaminants of concern in the environment, decision makers need appropriate chemical-specific toxicity values, which are a quantitative measure of dose-effect. These values can be reassessed or updated over time; therefore, the OSWER Directive 9285.7–53 (U.S. EPA, 2003) suggests a hierarchy of sources to identify appropriate toxicity values, defined as follows: Tier 1 toxicity information from the U.S. EPA's Integrated Risk Information System (IRIS), Tier 2 information from the U.S. EPA's Provisional Peer Reviewed Toxicity Values (PPRTV), and Tier 3 from "other sources." Priority is given to those sources that use similar methods to develop toxicity information as in Tiers 1 and 2. Examples of other sources include the California EPA's toxicity value database (CalEPA), the Agency for Toxic Substances and Disease Registry (ATSDR) Minimal Risk Levels (MRL), and the U.S. EPA Health Effects Assessment Summary Tables (HEAST). To identify which produced water chemicals found in the literature had chemical-specific toxicity values that may be considered appropriate for a risk assessment, the above databases, as well as the EPA's Office of Pesticide Programs (OPP), were cross-referenced. Of the 1,198 constituents identified in produced water, we found that 167 have existing toxicity values (14%) that can be used for a risk assessment in the United States (Fig. 4).

3.2.2. ToxVal database—Next, the list of produced water constituents was compared to the Chemistry Dashboard database of expanded toxicity Values (ToxValDB), which is a collection of databases that summarize *in vivo* data (Williams et al., 2017). Presence on the ToxValDB list indicates the availability of toxicity dose-effect values collected from multiple databases, including the databases listed above, but also European databases, including ECHA (European Chemicals Agency), and ecotoxicity data. When cross-referenced to this list, there were 527 chemicals identified in produced water with some associated toxicity data in one of these public databases, indicating that more than half (56%) of chemicals identified in produced water lack publically available toxicity dose-effect related values (Fig. 4).

3.2.3. Lists made by regulatory authorities—To further understand which produced water constituents may be regulated currently at the federal level, we compared constituents

to some key chemicals lists compiled by regulatory authorities, including NPDWS, the PPL, and hazardous waste defined under RCRA. The NPDWS are legally enforceable standards for public water systems that impose maximum concentrations or treatment techniques determined to be protective of public health. The PPL was compiled from a list of chemicals that were known to be toxic, however, these chemicals were highlighted for the PPL because they were measurable, had been detected and reported in water at a frequency of 2.5%, and are "produced in significant quantities." Chemicals that are listed under the NPDWS and PPL are all likely to cause adverse effects to human health at concentrations above standards. Finally, the Resource Conservation and Recovery Act is the law that governs the disposal of hazardous and non-hazardous solid waste. Hazardous waste is considered to be ignitable, corrosive, reactive, or toxic. Chemicals listed under Appendix VIII (40 CFR §261, Appendix VIII, 2010) may be considered hazardous to human health or the environment if improperly managed. Of the chemicals identified in produced water, there were 46 (3.8%) chemicals listed under the NPDWS, 76 (6.3%) on the PPL, and 81 (6.8%) that are listed under RCRA. However, it should be noted that produced water, along with other exploration and production (E&P) wastes have been exempt from Federal Hazardous Waste Regulations since 1980 (U.S. EPA, 2002).

3.2.4.   Standard analytical methods—Chemical analysis that is performed on waste or wastewater in a regulatory context, such as those described above, is limited to a proscribed set of methods. The produced water chemicals were cross-referenced by CAS number to chemicals that can be analyzed by U.S. EPA-approved test methods. For the 1,198 chemicals identified in this review, 290 (24%) can be identified through standard analytical methods.

3.2.5.   FracFocus—Only 91 (7.6%) chemicals that were reported in FracFocus were found in the constituents identified in produced water.

3.3. Collection and analysis of toxicity data for chemicals in produced water

Ecotoxicology data are summarized in Fig. 5. The Q1 of the EC50 was calculated for 37 out of the 92 chemicals that had dose-response data by species group. Because the remaining 55 chemicals had limited data—less than four studies per species group—quartiles were not calculated. Here, the lowest EC50 for the most sensitive species group of the remaining chemicals, which represents an entire study rather than a single assay, was included as the characteristic concentration in the analysis.

Data from *in vitro* assays in the ToxCast screening are summarized in Fig. 6. The Q1 of the AC50 was calculated for 73 chemicals. The minimum AC50 was selected for 30 chemicals that had less than four active assays. Thirteen chemicals were inactive at the maximum concentration assayed, and six chemicals had a boiling point below 200 °F.

Finally, "conditional" and existing toxicity values were predicted or identified using the CTV Predictor (toxvalue.org, Wignall et al., 2018). The availability of toxicity values for the chemicals in produced water are listed in Table 1.

3.4. Data integration and prioritization using ToxPi

Data from various toxicity data streams detailed above, ecotoxicity, *in vitro* methods, and predicted/existing toxicity values, were integrated using the ToxPi approach. The results of the ToxPi evaluation are summarized in Fig. 7 and SI Fig. 2. The toxicity prioritization schemes (weighted and unweighted), identified 57 chemicals as the top-third most hazardous compounds, relative to the restricted list of 122 chemicals (Table 2). There were twenty-three chemicals that ranked in the top-third of both weighting schemes (Table 3). The weighted scheme gives equal weight to ecotoxicology, high-throughput *in vitro* activity, and human health hazards. However, because there are more endpoints for human health hazards (the CTV domain), in the unweighted scheme, the ToxPi is dominated by human health hazards. While weighting data is somewhat subjective—which underscores the need for careful consideration of data inputted and assumptions of each ToxPi analysis—the facility to change weighting schemes allows the user to prioritize receptors of higher concern for the context of the analysis. When these two methodologies are considered concurrently, the 23 chemicals identified potentially represent the more data-rich chemicals with known toxicological hazard.

3.5. Sensitivity analysis

To assess how the ranking of chemicals based on ToxPi scores may be impacted by the types and amounts of data included in the assessment (data-rich versus data-poor chemicals) or how the characteristic concentrations were calculated (Q1 and minimum value vs Q2 and geometric mean), a sensitivity analysis was performed by comparing the ToxPi ranking and values for multiple analyses (Fig. 8). The sensitivity analysis showed that the overall ToxPi ranking was largely stable regardless of the choices made on input parameters of weighting. When comparing imputing of minimum (least toxic) values for missing values, the overall rank change was three or less rank-changes for 69% of the chemicals, and within ten rank places for 94% of the chemicals. Completing the ToxPi analysis with median values and geometric means gives an over-all rank change by three or less places for 51% of the chemicals and ten or less places for 84% of the chemicals. The analysis was found to be most sensitive to changes in how the domain is weighted or if data were removed by only considering chemicals that are data-rich. Here, the ToxPi rankings fluctuated more by chemical (16% and 17% moved three or less places, and 26% and 76% moved ten or less places for weighted and data-rich analyses, respectively).

4. Discussion

This study presents a method for identification and prioritization of constituents of concern in produced water from onshore oil and gas extraction. This unique approach identified the potential universe of chemicals associated with an understudied, industry-specific wastewater, and considered multiple lines of toxicity evidence for those constituents. The results of this analysis will be highly informative to measuring and mitigating exposure to constituents of concern in produced water. The method is particularly appropriate to survey toxicity of identified constituents, given the large number of compounds reported in produced water, as well as address a lack of data available on the chemical hazard of many of the chemicals detected. Given the variability in produced water composition,

the nearly 1,200 chemicals identified through our comprehensive literature review are not likely be present in each wastewater sample. However, this and other analyses indicate that there is not a comprehensive understanding of what constituents are of primary concern in regional-specific wastewaters (Luek and Gonsior, 2017; Oetjen et al., 2017; U.S. EPA, 2016). This lack of understanding represents a major hurdle for designing fit-for-purpose treatment and monitoring strategies for alternative management of specific produced water.

4.1. Chemical constituents in produced water

While this literature review identified 1,198 chemicals in produced water with CAS numbers, there are additional constituents of potential concern, such as isomers or homologous series of various compounds (Thurman et al., 2017), which cannot be represented by a single CAS number, and therefore were not included in this study. For the purposes of this study, requiring a CAS number allowed for rapid cross-reference to numerous other databases and facilitated identification of associated data. It is likely that substances without CAS numbers have even more limited toxicity information.

Our literature review found that only 14% of chemicals identified in produced water have existing toxicity values for risk assessment in the United States. While this percentage is much smaller than what has been reported by others (Elliott et al., 2017; Yost et al., 2016a), it is important to note that those studies considered only about 130 potential constituents in produced water, which had been primarily identified through targeted analytical methodologies because they were known to be of toxicological concern. Currently, only 24% (290) of the 1,198 chemicals identified in produced water can be detected through standard analytical methods. This limited chemical characterization prohibits thorough understanding of potential effects on human health, which is important given that some regions across the country are considering—or have implemented—alternative strategies to reuse the wastewater, including some that involve agricultural use. This bias toward "known knowns," as compared to the nearly 1,200 chemical constituents identified through the literature search, highlights the need to use a combination of targeted, suspect-screening, and non-targeted analytical methodologies to characterize the chemical composition of complex wastewaters and to identify a wider range of potential contaminants of concern.

Our analysis additionally indicated that many of the chemicals used in hydraulic fracturing are not present in produced water. However, this analysis does not account for likely transformation of compounds or degradants of hydraulic fracturing additives; meaning, it is difficult to estimate the percentage of geogenic versus anthropogenic compounds that are detected in produced water. Furthermore, while there have been very positive advancements in disclosure, a full understanding of chemicals used in operations remains limited due to trade secret provisions. Konschnik and Dayalu (2016) found that approximately 18.9% of CAS numbers were "intentionally withheld from public disclosure" in their analysis of FracFocus (Konschnik and Dayalu, 2016). Additionally, there are no requirements for the disclosure of chemicals used for oil and gas activities outside of well simulation (i.e. hydraulic fracturing) (Stringfellow et al., 2017).

4.2. Prioritization of produced water chemicals

This study devised a method to identify and integrate multiple sources of available toxicity hazard data in order to ascertain and prioritize constituents of concern. Known chemicals of concern were included in the ToxPi analysis to gain understanding as to how they may be distributed in the final ranking (Fig. 7). There were 36 chemicals listed on the EPA's Priority Pollutants List (PPL) in the subset of 122 chemicals used for the ToxPi analysis (SI Table 5). While these reference chemicals are classified as hazardous, they are being compared to chemicals that are being ranked based on a different set of considerations. Whereas PPL chemicals are selected based on toxicity to humans as well as production volume and detection frequency, the chemicals in this framework were ranked by toxicity to humans, ecotoxicity, and *in vitro* activity. Therefore, it is not unexpected that the reference chemicals are distributed throughout the ToxPi rankings. PPL chemicals, however, are more weighted at the top of the ToxPi ranking.

To address potential bias based on availability of data, two different weighting schemes— one weighted by endpoint and one weighted by domain—were considered concurrently. It was assumed that chemicals that are in the top-third of both are most likely to be a priority. The ToxPi evaluation, using these data and assumptions, yielded twenty-three chemicals of concern (Table 3). One finding of this work is the identification of many compounds with known profiles of toxicity, including 1,4-dioxane, nitrobenzene, 1,1,2,2-tetrachloroethane, and toluene. Implications of those compounds identified with high ToxPi scores indicate that some chemicals of known hazard are present in some produced water samples. Additionally, acetic acid was identified as a chemical of concern. While acetic acid is the primary component of vinegar, it is harmful to aquatic life at low concentrations. Acetic acid was found at elevated concentrations (1,600 mg/L) in produced water (Lester et al., 2015), a level which exceeds demonstrated acute and chronic ecotoxicity effects (Locke et al., 2009) and can further impact aquatic systems through eutrophication.

Apparent throughout this analysis is the identification of legacy pollutants that may be reintroduced to the environment through produced water discharges, including the presence of endosulfan, an organochlorine insecticide that was banned in the US in 2010, and tetrachloroethylene, which is a suspected human carcinogen. It is unclear how these legacy pollutants became associated with produced water. One possibility is that they are present in the make-up water used to create stimulation fluids (Yost et al., 2016b). Tetrachloroethylene, which is a commonly used industrial solvent (Cichocki et al., 2016), may be used to maintain drilling or other equipment. Regardless of the process, it is clear that some of these legacy contaminants are present in produced water and are another factor that should be considered in assessing potential risks associated with reuse applications that introduce this wastewater to the environment.

An important limitation of this ranking analysis is the restricted geographic area from which produced water samples were collected. As noted, the majority of samples came from the Marcellus Shale play. The implication is that the extent to which the type and quantity of compounds of concern are present in produced water across the country is not well understood. More representative sampling would have provided a broader perspective on chemicals that may be of concern nationally or regionally where treatment for discharge is

being considered. The framework described here can be applied to create a more targeted prioritization effort once regional constituents are identified. The flexibility, and potential concern, of the ToxPi approach in this application is that the selection of ToxPi evaluation criteria are inherently subjective. Different types of data—including physicochemical properties to consider fate and transport, as well as those chemicals that are persistent, bioaccumulative, and toxic—can be included and adjusted, leading to identification of different priority chemicals for various contexts and applications. This is helpful when a search for a particular set of chemicals, such as those that exhibit aquatic toxicity, is warranted, but may be a weakness if evaluation criteria are not considered carefully, potentially leading to misidentification of the highest priority chemicals of concern. Additionally, when considering particular chemicals of concern for a context-specific risk assessment, a review of all the original studies included in each of the databases should be completed to assess data quality, reliability, and applicability.

The bioactivity data underlying the ToxPi hazard evaluation also provides an initial step toward performing risk assessments for produced water. For context, we compared the maximum concentrations reported in produced water to the bioactivity data for ecotoxicity (Fig. 5) as well as from ToxCast (Fig. 6). However, we elected not to include a "margin of exposure" in the ToxPi evaluation, as produced water is a highly-variable, under-characterized waste stream, and there is no assurance that the reported concentrations are in any way representative of the typical or even maximal concentrations across all sources of produced water. Thus, in the context of site-specific risk evaluations, our ToxPi hazard evaluation methodology represents an example methodology that can also be applied in a fit-for-purpose, context-specific case. For instance, after the chemical constituents in produced water are characterized and potential receptors (ecological or human) identified, then the underlying bioactivity data can be used to provide a screening-level indication of what chemicals may be of greater or lesser concern from a risk-based perspective. However, even in this case, many chemicals potentially present in produced water do not have either bioactivity data or toxicity values, so whole effluent assessments could have an important role in establishing hazard and risk.

Thus, while providing useful insights, prioritizing chemicals and focusing on a subset of constituents does not eliminate potential concern from the broader list of chemicals. Moreover, a demonstrated need exists for a more robust analytical evaluation of wastewater and new methods to characterize the potential impacts of its release, which will also allow for smarter monitoring and better development of treatment regimes. This, and other knowledge gaps associated with human exposure to oil and gas development was highlighted in a recent literature review and research planning document developed by the Health Effects Institute (2019). In addition to targeted and non-targeted chemical characterization, whole-effluent toxicity evaluation techniques, such as novel high-throughput "toxicogenomics" assays (Gao et al., 2015; Lan et al., 2016) or effects-directed assessments combined with toxicity identification evaluation (Burgess et al., 2013) may be useful to characterize the potential toxicological hazard of "unknown unknown" chemicals. Evaluation of analytical and treatment options for compounds identified through this work is an obvious next step in characterizing mitigation strategies to reduce potential risks from compounds identified in produced water.

5. Conclusions

We have developed a methodology to address the critical issue of identifying and prioritizing constituents of concern in produced water by combining a comprehensive literature search process to rapidly and methodically collect useful literature on chemicals that have been identified in produced water, with a quantitative integration method to evaluate toxicity data for those constituents across multiple domains. One of the more salient findings of this work is the lack of sufficient toxicity data for 86% of compounds identified in the limited analysis of produced water. Our results also provide a systematic basis for prioritizing chemicals for further study. Furthermore, this method can be extended to allow for additional types of data to be considered, including physicochemical properties to consider chemical fate and transport, as well as what may be persistent in the environment, bioaccumulative, and/or toxic. This flexibility will facilitate context-specific application of the approach, investigating ways to reuse produced water while also being informed about risk and risk management needs. Overall, additional research investigating both "known unknowns" and "unknown unknowns" is vital in order to update existing, and potentially expanding, regulatory programs for produced water, particularly as interest grows in the intentional treatment and release of produced water for disposal or reuse outside of the oilfield.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

40 CFR §136, 2017. Guidelines Establishing Test Procedures for the Analysis of Pollutants.

40 CFR §261, Appendix VIII, 2010. Hazardous Waste Management System.

40 CFR §423, Appendix A, 2015. Toxic and Priority Pollutants under the Clean Water Act.

40 CFR §435.53, 1995. Oil and Gas Extraction Point Source Category, Subpart E—Agricultural and Wildlife Water Use Subcategory.

Akob DM, Cozzarelli IM, Dunlap DS, Rowan EL, Lorah MM, 2015. Organic and inorganic composition and microbiology of produced waters from Pennsylvania shale gas wells. Appl. Geochem 60, 116–125. 10.1016/j.apgeochem.2015.04.011.

Bell EA, Poynor TE, Newhart KB, Regnery J, Coday BD, Cath TY, 2017. Produced water treatment using forward osmosis membranes: Evaluation of extended-time performance and fouling. J. Membr. Sci 525, 77–88. 10.1016/j.memsci.2016.10.032.

Blewett TA, Delompré PLM, He Y, Folkerts EJ, Flynn SL, Alessi DS, Goss GG, 2017a. Sublethal and reproductive effects of acute and chronic exposure to flowback and produced water from hydraulic fracturing on the water flea *Daphnia magna*. Environ. Sci. Technol 51, 3032–3039. 10.1021/acs.est.6b05179. [PubMed: 28140571]

Blewett TA, Weinrauch AM, Delompré PLM, Goss GG, 2017b. The effect of hydraulic flowback and produced water on gill morphology, oxidative stress and antioxidant response in rainbow trout (Oncorhynchus mykiss). Sci. Rep 7, 46582. 10.1038/srep46582. [PubMed: 28425455]

Burgess RM, Ho KT, Brack W, Lamoree M, 2013. Effects-directed analysis (EDA) and toxicity identification evaluation (TIE): Complementary but different approaches for diagnosing causes of environmental toxicity: Environmental diagnostics: EDA and TIE. Environ. Toxicol. Chem 32, 1935–1945. 10.1002/etc.2299. [PubMed: 23893495]

Carey J, Zaidi A, Ribo J, 1992. Specific toxic organics in produced waters from in-situ heavy oil recovery operations in western Canada. In: Ray JP, Engelhardt FR (Eds.), Produced Water. Plenum Press, New York, NY, pp. 133–150. 10.1007/978-1-4615-2902-6_11.

CDPHE (Colorado Department of Health and Environment), 2012. Colorado Discharge Permit System (CDPS) Fact Sheet to Permit Number COG840011.

CDPHE (Colorado Department of Health and Environment), 2011. Colorado Discharge Permit System (CDPS) Fact Sheet to Permit Number COG840002.

Chemistry Dashboard [WWW Document], 2019. URL https://comptox.epa.gov/dashboard/ (accessed 3.3.19).

Chittick EA, Srebotnjak T, 2017. An analysis of chemicals and other constituents found in produced water from hydraulically fractured wells in California and the challenges for wastewater management. J. Environ. Manage 204, 502–509. 10.1016/j.jenvman.2017.09.002. [PubMed: 28934673]

Cichocki JA, Guyton KZ, Guha N, Chiu WA, Rusyn I, Lash LH, 2016. Target organ metabolism, toxicity, and mechanisms of trichloroethylene and perchloroethylene: key similarities, differences, and data gaps. J. Pharmacol. Exp. Ther 359, 110–123. 10.1124/jpet.116.232629. [PubMed: 27511820]

Cluff MA, Hartsock A, MacRae JD, Carter K, Mouser PJ, 2014. Temporal changes in microbial ecology and geochemistry in produced water from hydraulically fractured marcellus shale gas wells. Environ. Sci. Technol 48, 6508–6517. 10.1021/es501173p. [PubMed: 24803059]

Colborn T, Kwiatkowski C, Schultz K, Bachran M, 2011. Natural gas operations from a public health perspective. Hum. Ecol. Risk Assess 1039–1056. 10.1080/10807039.2011.605662.

Dahm KG, Guerra KL, Xu P, Drewes JE, 2011. Composite geochemical database for coalbed methane produced water quality in the rocky mountain region. Environ. Sci. Technol 45, 7655–7663. 10.1021/es201021n. [PubMed: 21790201]

DiGiulio DC, Jackson RB, 2016. Impact to underground sources of drinking water and domestic wells from production well stimulation and completion practices in the pavillion, wyoming, field. Environ. Sci. Technol 50, 4524–4536. 10.1021/acs.est.5b04970. [PubMed: 27022977]

Elliott EG, Ettinger AS, Leaderer BP, Bracken MB, Deziel NC, 2017. A systematic evaluation of chemicals in hydraulic-fracturing fluids and wastewater for reproductive and developmental toxicity. J. Eposure Sci. Environ. Epidemiol 27, 90–99. 10.1038/jes.2015.81.

Elsner M, Hoelzer K, 2016. Quantitative survey and structural classification of hydraulic fracturing chemicals reported in unconventional gas production. Environ. Sci. Technol 50, 3290–3314. 10.1021/acs.est.5b02818. [PubMed: 26902161]

Estrada JM, Bhamidimarri R, 2016. A review of the issues and treatment options for wastewater from shale gas extraction by hydraulic fracturing. Fuel 182, 292–303. 10.1016/j.fuel.2016.05.051.

Faber A-H, Annevelink M, Gilissen HK, Schot P, van Rijswick M, de Voogt P, van Wezel A, 2019. How to adapt chemical risk assessment for unconventional hydrocarbon extraction related to the water system. In: de Voogt P (Ed.), Reviews of Environmental Contamination and Toxicology Volume 246, Reviews of Environmental Contamination and Toxicology. Springer International Publishing, Cham, pp. 1–32. 10.1007/398_2017_10.

Fakhru'l-Razi A, Pendashteh A, Abdullah LC, Biak DRA, Madaeni SS, Abidin ZZ, 2009. Review of technologies for oil and gas produced water treatment. J. Hazard. Mater 170, 530–551. 10.1016/j.jhazmat.2009.05.044. [PubMed: 19505758]

Ferrer I, Thurman EM, 2015. Chemical constituents and analytical approaches for hydraulic fracturing waters. Trends Environ. Anal. Chem 5, 18–25. 10.1016/j.teac.2015.01.003.

FracFocus, 2018. FracFocus Chemical Disclosure Registry [WWW Document]. URL https://fracfocus.org/ (accessed 6.30.18).

Gao C, Weisman D, Lan J, Gou N, Gu AZ, 2015. Toxicity mechanisms identification via gene set enrichment analysis of time-series toxicogenomics data: impact of time and concentration. Environ. Sci. Technol 49, 4618–4626. 10.1021/es505199f. [PubMed: 25785649]

Gieg LM, Davidova IA, Duncan KE, Suflita JM, 2010. Methanogenesis, sulfate reduction and crude oil biodegradation in hot Alaskan oilfields: Methanogenic crude oil biodegradation in hot oilfields. Environ. Microbiol 12, 3074–3086. 10.1111/j.1462-2920.2010.02282.x. [PubMed: 20602630]

Gordalla BC, Ewers U, Frimmel FH, 2013. Hydraulic fracturing: a toxicological threat for groundwater and drinking-water? Environ. Earth Sci 70, 3875–3893. 10.1007/s12665-013-2672-9.

GraphPad Prism 8, 2019. . La Jolla, California.

Hayes T, 2009. Sampling and Analysis of Water Streams Associated with the Development of Marcellus Shale Gas. Gas Technology Institute.

Hayes T, Severin BF, 2012. Barnett and Appalachian Shale Water Management and Reuse Technologies (No. RPSEA 08122–05). Gas Technology Institute, Des Plaines, IL.

He Y, Flynn SL, Folkerts EJ, Zhang Y, Ruan D, Alessi DS, Martin JW, Goss GG, 2017. Chemical and toxicological characterizations of hydraulic fracturing flowback and produced water. Water Res. 114, 78–87. 10.1016/j.watres.2017.02.027. [PubMed: 28229951]

He Y, Sun C, Zhang Y, Folkerts EJ, Martin JW, Goss GG, 2018. Developmental toxicity of the organic fraction from hydraulic fracturing flowback and produced waters to early life stages of Zebrafish (*Danio rerio*). Environ. Sci. Technol 52, 3820–3830. 10.1021/acs.est.7b06557. [PubMed: 29376370]

HEI-Energy Research Committee, 2019. Human Exposure to Unconventional Oil and Gas Development: A literature Survey for Research Planning. Special Report 2. Boston, MA: Health Effects Institutes-Energy. https://hei-energy.org/publication/humanexposure-unconventional-oil-and-gas-development-literature-survey-research.

Hoelzer K, Sumner AJ, Karatum O, Nelson RK, Drollette BD, O'Connor MP, D'Ambro EL, Getzinger GJ, Ferguson PL, Reddy CM, Elsner M, Plata DL, 2016. Indications of transformation products from hydraulic fracturing additives in shale-gas wastewater. Environ. Sci. Technol 50, 8036–8048. 10.1021/acs.est.6b00430. [PubMed: 27419914]

Kassotis CD, Klemp KC, Vu DC, Lin C-H, Meng C-X, Besch-Williford CL, Pinatti L, Zoeller RT, Drobnis EZ, Balise VD, Isiguzo CJ, Williams MA, Tillitt DE, Nagel SC, 2015. Endocrine-disrupting activity of hydraulic fracturing chemicals and adverse health outcomes after prenatal exposure in male mice. Endocrinology 156, 4458–4473. 10.1210/en.2015-1375. [PubMed: 26465197]

Khan NA, Engle M, Dungan B, Holguin FO, Xu P, Carroll KC, 2016. Volatileorganic molecular characterization of shale-oil produced water from the Permian Basin. Chemosphere 148, 126–136. 10.1016/j.chemosphere.2015.12.116. [PubMed: 26802271]

Konschnik K, Dayalu A, 2016. Hydraulic fracturing chemicals reporting: Analysis of available data and recommendations for policymakers. Energy Policy 88, 504–514. 10.1016/j.enpol.2015.11.002.

Lan J, Gou N, Rahman SM, Gao C, He M, Gu AZ, 2016. A quantitative toxicogenomics assay for high-throughput and mechanistic genotoxicity assessment and screening of environmental pollutants. Environ. Sci. Technol 50, 3202–3214. 10.1021/acs.est.5b05097. [PubMed: 26855253]

Lester Y, Ferrer I, Thurman EM, Sitterley KA, Korak JA, Aiken G, Linden KG, 2015. Characterization of hydraulic fracturing flowback water in Colorado: Implications for water treatment. Sci. Total Environ 512–513, 637–644. 10.1016/j.scitotenv.2015.01.043.

Liden T, Santos IC, Hildenbrand ZL, Schug KA, 2018. Treatment modalities for the reuse of produced waste from oil and gas development. Sci. Total Environ 643, 107–118. 10.1016/j.scitotenv.2018.05.386. [PubMed: 29936154]

Locke A, Doe KG, Fairchild WL, Jackman PM, Reese EJ, Carman MR, 2009. Preliminary evaluation of effects of invasive tunicate management with acetic acid and calcium hydroxide on non-target marine organisms in Prince Edward Island, Canada. 10.3391/ai.2009.4.1.23.

Luek JL, Gonsior M, 2017. Organic compounds in hydraulic fracturing fluids and wastewaters: A review. Water Res. 123, 536–548. 10.1016/j.watres.2017.07.012. [PubMed: 28697484]

Lyman SN, Mansfield ML, Tran HNQ, Evans JD, Jones C, O'Neil T, Bowers R, Smith A, Keslar C, 2018. Emissions of organic compounds from produced water ponds I: Characteristics and speciation. Sci. Total Environ 619–620, 896–905. 10.1016/j.scitotenv.2017.11.161.

Maguire-Boyle SJ, Barron AR, 2014. Organic compounds in produced waters from shale gas wells. Environ. Sci.: Processes Impacts 16, 2237–2248. 10.1039/C4EM00376D.

Marvel SW, To K, Grimm FA, Wright FA, Rusyn I, Reif DM, 2018. ToxPi Graphical User Interface 2.0: Dynamic exploration, visualization, and sharing of integrated data models. BMC Bioinf. 19 (1). 10.1186/s12859-018-2089-2.

Motulsky HJ, Brown RE, 2006. Detecting outliers when fitting data with nonlinear regression – a new method based on robust nonlinear regression and the false discovery rate. BMC Bioinf. 7, 123. 10.1186/1471-2105-7-123.

National Environmental Methods Index [WWW Document], 2017. URL https://www.nemi.gov/home/ (accessed 3.13.18).

Nell M, Helbling DE, 2018. Exploring matrix effects and quantifying organic additives in hydraulic fracturing associated fluids using liquid chromatography electrospray ionization mass spectrometry. Environ. Sci.: Processes & Impacts 10.1039/C8EM00135A.

NYSDEC (New York State Department of Environmental Conservation), 2015. Natural Gas Development Activities & High-Volume Hydraulic Fracturing (Supplemental Generic Environmental Impact Statement (SEGIS)). New York State Department of Environmental Conservation, Albany, NY.

Oetjen K, Giddings CGS, McLaughlin M, Nell M, Blotevogel J, Helbling DE, Mueller D, Higgins CP, 2017. Emerging analytical methods for the characterization and quantification of organic contaminants in flowback and produced water. Trends Environ. Anal. Chem 15, 12–23. 10.1016/j.teac.2017.07.002.

Orem W, Tatu C, Varonka M, Lerch H, Bates A, Engle M, Crosby L, McIntosh J, 2014. Organic substances in produced and formation water from unconventional natural gas extraction in coal and shale. Int. J. Coal Geol 126, 20–31. 10.1016/j.coal.2014.01.003.

Orem WH, Tatu CA, Lerch HE, Rice CA, Bartos TT, Bates AL, Tewalt S, Corum MD, 2007. Organic compounds in produced waters from coalbed natural gas wells in the Powder River Basin, Wyoming, USA. Appl. Geochem 22, 2240–2256. 10.1016/j.apgeochem.2007.04.010.

Pashin JC, McIntyre-Redden MR, Mann SD, Kopaska-Merkel DC, 2014a. Water Management Strategies for Improved Coalbed Methane Production in the Black Warrior Basins (Geological Survey of Alabama No. DE-FE0000888). U.S. Department of Energy, National Energy Technology Library, Washington, DC.

Pashin JC, McIntyre-Redden MR, Mann SD, Kopaska-Merkel DC, Varonka M, Orem W, 2014b. Relationships between water and gas chemistry in mature coalbed methane reservoirs of the Black Warrior Basin. Int. J. Coal Geol 126, 92–105. 10.1016/j.coal.2013.10.002.

Regnery J, Coday BD, Riley SM, Cath TY, 2016. Solid-phase extraction followed by gas chromatography-mass spectrometry for the quantitative analysis of semi-volatile hydrocarbons in hydraulic fracturing wastewaters. Anal. Methods 8, 2058–2068. 10.1039/C6AY00169F.

Reif DM, Martin MT, Tan SW, Houck KA, Judson RS, Richard AM, Knudsen TB, Dix DJ, Kavlock RJ, 2010. Endocrine profiling and prioritization of environmental chemicals using ToxCast data. Environ. Health Perspect 118, 1714–1720. 10.1289/ehp.1002180. [PubMed: 20826373]

Rogers JD, Burke TL, Osborn SG, Ryan JN, 2015. A framework for identifying organic compounds of concern in hydraulic fracturing fluids based on their mobility and persistence in groundwater. Environ. Sci. Technol. Lett 2, 158–164. 10.1021/acs.estlett.5b00090.

Rogers JL, Hicks RT, Shaw B, Jensen J, 1992. Procedure for development of contingency plans to mitigate produced water releases on BLM Lands. In: Ray JP, Engelhardt FR (Eds.), Produced Water. Plenum Press, New York, NY, pp. 35–44. 10.1007/978-1-4615-2902-6_4.

Rosenblum J, Thurman EM, Ferrer I, Aiken G, Linden KG, 2017. Organic chemical characterization and mass balance of a hydraulically fractured well: from fracturing fluid to produced water over 405 days. Environ. Sci. Technol 51, 14006–14015. 10.1021/acs.est.7b03362. [PubMed: 29132208]

Santos IC, Hildenbrand ZL, Schug KA, 2019. A review of analytical methods for characterizing the potential environmental impacts of unconventional oil and gas development. Anal. Chem 91, 689–703. 10.1021/acs.analchem.8b04750. [PubMed: 30392348]

Shih J-S, Saiers JE, Anisfeld SC, Chu Z, Muehlenbachs LA, Olmstead SM, 2015. Characterization and analysis of liquid waste from Marcellus shale gas development. Environ. Sci. Technol 49, 9557–9565. 10.1021/acs.est.5b01780. [PubMed: 26140412]

Sirivedhin T, Dallbauman L, 2004. Organic matrix in produced water from the Osageskiatook petroleum environmental research site, Osage county, Oklahoma. Chemosphere 57, 463–469. 10.1016/j.chemosphere.2004.05.034. [PubMed: 15350408]

Sitterley KA, Linden KG, Ferrer I, Thurman EM, 2018. Identification of proprietary amino ethoxylates in hydraulic fracturing wastewater using liquid chromatography/time-of-flight mass spectrometry with solid-phase extraction. Anal. Chem 90, 10927–10934. 10.1021/acs.analchem.8b02439. [PubMed: 30139247]

Stringfellow WT, Camarillo MK, 2019. Flowback verses first-flush: new information on the geochemistry of produced water from mandatory reporting. Environ. Sci.: Processes Impacts 21, 370–383. 10.1039/C8EM00351C.

Stringfellow WT, Camarillo MK, Domen JK, Shonkoff SBC, 2017. Comparison of chemical-use between hydraulic fracturing, acidizing, and routine oil and gas development. e0175344. PLoS ONE 12. 10.1371/journal.pone.0175344.

Strong LC, Gould T, Kasinkas L, Sadowsky MJ, Aksan A, Wackett LP, 2014. Biodegradation in waters from hydraulic fracturing: chemistry, microbiology, and engineering. J. Environ. Eng 140, B4013001. 10.1061/(ASCE)EE.1943-7870.0000792.

Thacker J, Carlton D, Hildenbrand Z, Kadjo A, Schug K, 2015. Chemical analysis of wastewater from unconventional drilling operations. Water 7, 1568–1579. 10.3390/w7041568.

Thurman EM, Ferrer I, Blotevogel J, Borch T, 2014. Analysis of hydraulic fracturing flowback and produced waters using accurate mass: identification of ethoxylated surfactants. Anal. Chem 86, 9653–9661. 10.1021/ac502163k. [PubMed: 25164376]

Thurman EM, Ferrer I, Rosenblum J, Linden K, Ryan JN, 2017. Identification of polypropylene glycols and polyethylene glycol carboxylates in flowback and produced water from hydraulic fracturing. J. Hazard. Mater 323, 11–17. 10.1016/j.jhazmat.2016.02.041. [PubMed: 26947804]

To KT, Fry RC, Reif DM, 2018. Characterizing the effects of missing data and evaluating imputation methods for chemical prioritization applications using ToxPi. BioData Mining 11, 10. 10.1186/s13040-018-0169-5. [PubMed: 29942350]

Toxicity Forecasting [WWW Document], 2019. URL https://www.epa.gov/chemical-research/toxicity-forecasting (accessed 4.23.19).

U.S. EPA (U.S. Environmental Protection Agency), 2019a. Study of Oil and Gas Extraction Wastewater Management under the Clean Water Act (No. EPA-821-R19-001). U.S. Environmental Protection Agency, Office of Water, Washington, D.C.

U.S. EPA (U.S. Environmental Protection Agency), 2019b. ECOTOX User Guide: ECOTOXicology Knowledgebase System. Version 5.0. U.S. Environmental Protection Agency.

U.S. EPA (U.S. Environmental Protection Agency), 2018. Detailed Study of the Centralized Waste Treatment Point Source Category for Facilities Managing Oil and Gas Extraction Wastes (No. EPA-821-R-18–004). U.S. Environmental Protection Agency, Washington, D.C.

U.S. EPA (U.S. Environmental Protection Agency), 2016. Hydraulic Fracturing for Oil and Gas: Impacts from the Hydraulic Fracturing Water Cycle on Drinking Water Resources in the United States (No. EPA/600/R-16/236Fa). U.S. Environmental Protection Agency, Office of Research and Development, Washington DC.

U.S. EPA (U.S. Environmental Protection Agency), 2015. Compilation of Physicochemical and Toxicological Information about Hydraulic Fracturing-Related Chemicals (Draft Database) (No. EPA/600/R-15/134). U.S. Environmental Protection Agency, Washington, DC.

U.S. EPA (U.S. Environmental Protection Agency), 2009. National Primary Drinking Water Regulations (No. EPA 816-F-09–004). U.S. Environmental Protection Agency, Office of Groundwater and Drinking Water, Washington, DC.

U.S. EPA (U.S. Environmental Protection Agency), 2003. Human Health Toxicity Values in Superfund Risk Assessments (Memorandum No. OSWER Directive 9285.7–53). U.S. Environmental Protection Agency, Office of Solid Waste and Emergency Response, Washington, D.C.

U.S. EPA (U.S. Environmental Protection Agency), 2002. Exemption of Oil and Gas Exploration and Production Wastes from Federal Hazardous Waste Regulations (No. EPA530- K- 01–004). U.S. Environmental Protection Agency, Office of Solid Waste, Washington, D.C.

Veil JA, 2015. U.S. Produced Water Volumes and Management Practices in 2012. Ground Water Protection Council.

Wesolowski D, Broughton A, Hansotte CA, Koraido SM, Fillo JP, 1987. Characterization of Produced Waters from Natural Gas Production Operations (Topical Report No. GRI Report No. GRI-87/0335.1). Gas Research Institute, Chicago, Illinois.

Wignall JA, Muratov E, Sedykh A, Guyton KZ, Tropsha A, Rusyn I, Chiu WA, 2018. Conditional toxicity value (CTV) predictor: An in silico approach for generating quantitative risk estimates for chemicals. 057008. Environ. Health Perspect 126. 10.1289/EHP2998.

Williams AJ, Grulke CM, Edwards J, McEachran AD, Mansouri K, Baker NC, Patlewicz G, Shah I, Wambaugh JF, Judson RS, Richard AM, 2017. The CompTox chemistry dashboard: a community data resource for environmental chemistry. J. Cheminf 9. 10.1186/s13321-017-0247-6.

Yost EE, Stanek J, Burgoon LD, 2017. A decision analysis framework for estimating the potential hazards for drinking water resources of chemicals used in hydraulic fracturing fluids. Sci. Total Environ 574, 1544–1558. 10.1016/j.scitotenv.2016.08.167. [PubMed: 27666475]

Yost EE, Stanek J, DeWoskin RS, Burgoon LD, 2016a. Overview of chronic oral toxicity values for chemicals present in hydraulic fracturing fluids, flowback, and produced waters. Environ. Sci. Technol 50, 4788–4797. 10.1021/acs.est.5b04645. [PubMed: 27050380]

Yost EE, Stanek J, DeWoskin RS, Burgoon LD, 2016b. Estimating the potential toxicity of chemicals associated with hydraulic fracturing operations using quantitative structure-activity relationship modeling. Environ. Sci. Technol 50, 7732–7742. 10.1021/acs.est.5b05327. [PubMed: 27172125]

Ziemkiewicz PF, Quaranta JD, Darnell A, Wise R, 2014. Exposure pathways related to shale gas development and procedures for reducing environmental and public risk. J. Nat. Gas Sci. Eng 16, 77–84. 10.1016/j.jngse.2013.11.003.

Ziemkiewicz PF, Thomas He Y, 2015. Evolution of water chemistry during Marcellus shale gas development: a case study in West Virginia. Chemosphere 134, 224–231. 10.1016/j.chemosphere.2015.04.040. [PubMed: 25957035]
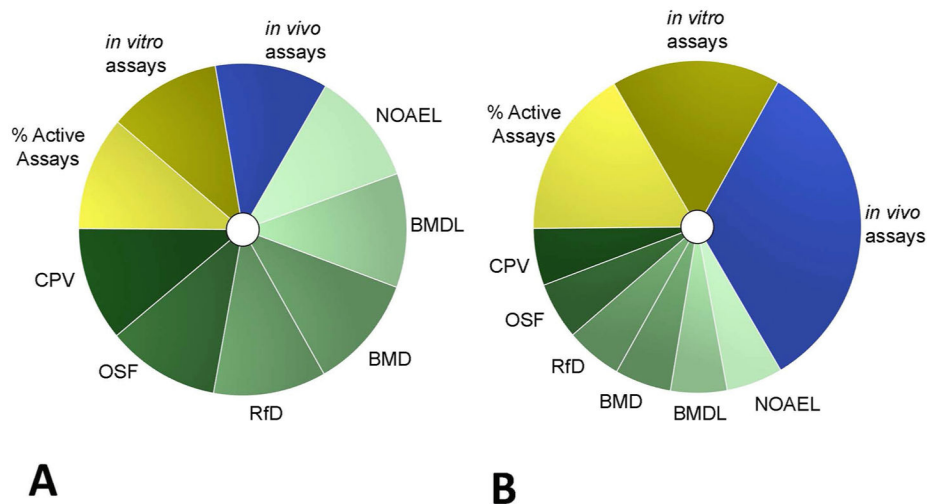
Fig. 1.
ToxPi definitions and notations. A: original analysis where equal weighting was given for each endpoint; B: alternative weighting, where each domain was given equal weight. Each chemical was analyzed using unweighted (A) and weighted (B) combinations of data from multiple domains, which are represented by slices of a similar color: *in vivo* ecotoxicology endpoints (blue), *in vitro* high throughput screening assays (yellow), and known or conditional human health toxicity values (green). Individual slices represent data from related assays, endpoints, or analyses. The distance of each slice from the center indicates the normalized value of the component. The angle of the slice represents how that component is weighted relative to the other components in the overall ToxPi calculation. Cancer potency value (CPV), Oral slope factor (OSF), Reference dose (RfD), Rfd Benchmark dose (BMD), RfD Benchmark dose lower limit (BMDL), RfD No observed adverse effect level (NOAEL).

Records identified through database search:
- Web of Science (n = 11996)
- PubMed (n = 3665)

Records screened (n = 15661)

Records excluded by title/abstract (n = 15333)

Full-text articles assessed for eligibility (n = 302)

Studies identified from hand-searching reviews (n = 33)

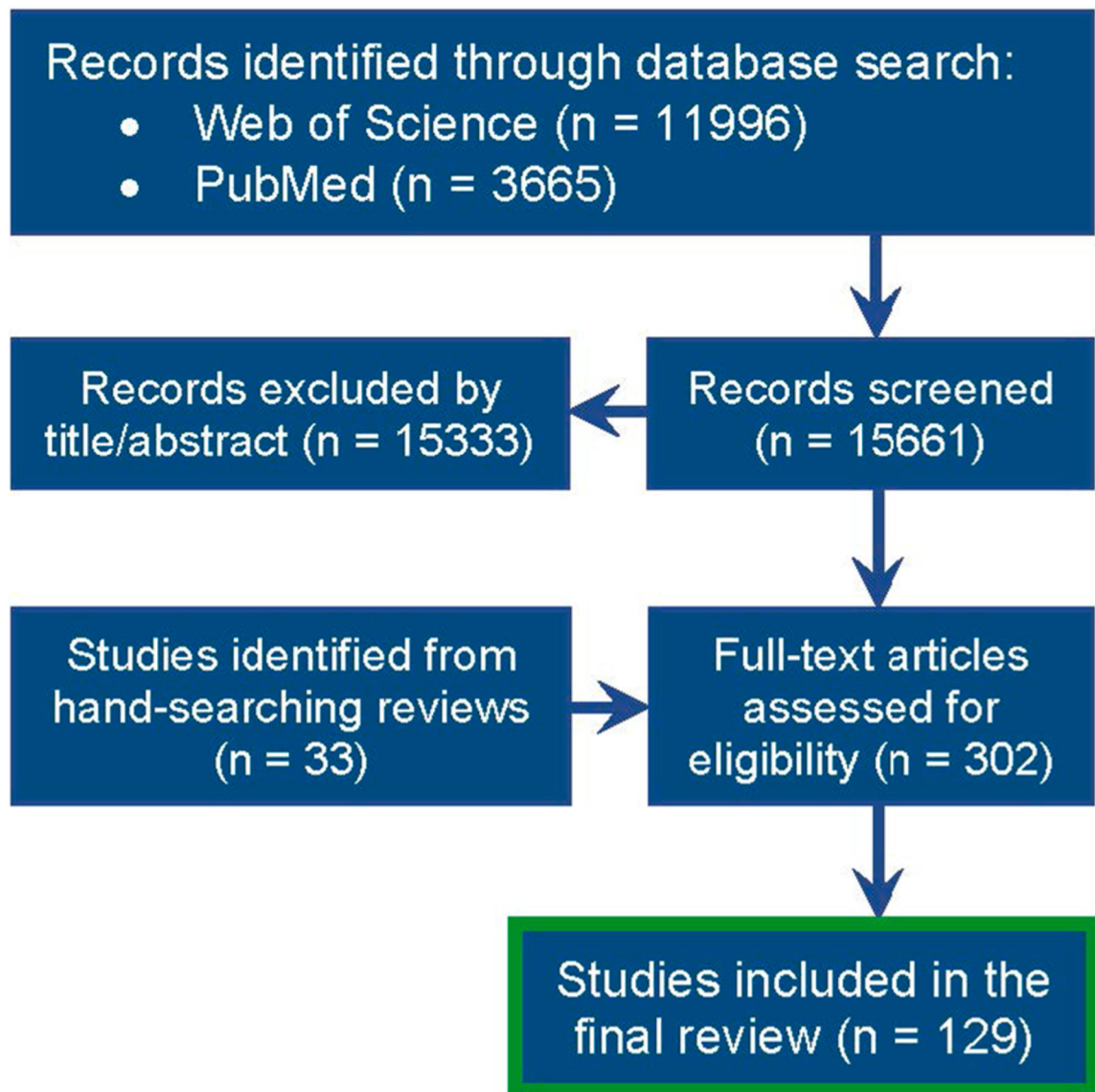Studies included in the final review (n = 129)

Fig. 2.
Flow diagram of the literature search strategy.

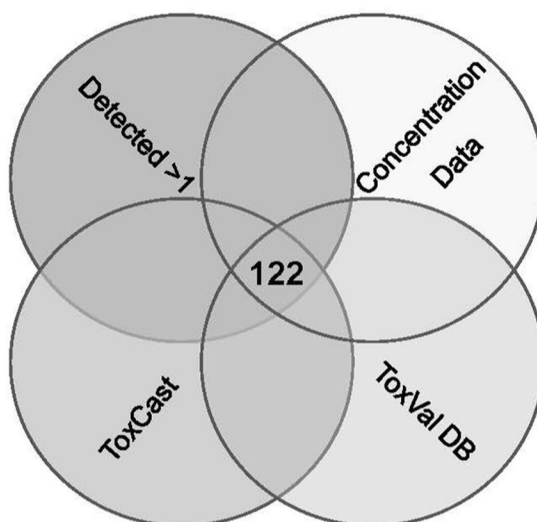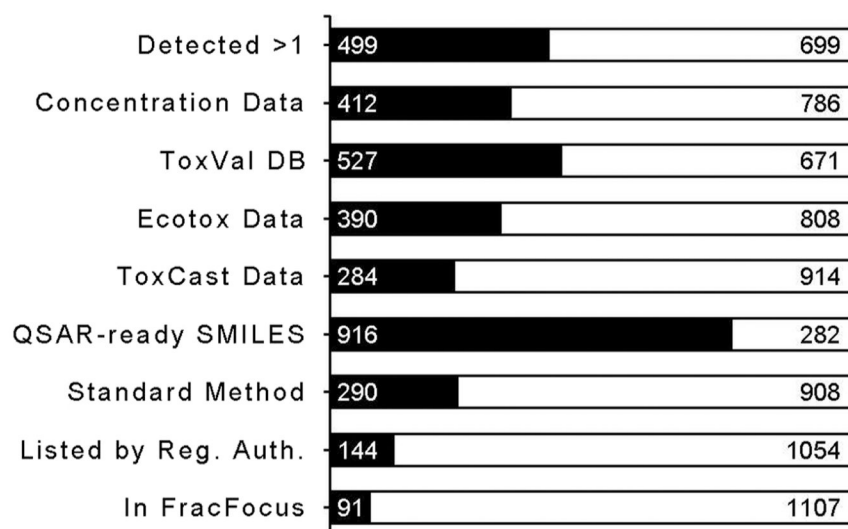## 1198 Compounds in Produced Water

Fig. 3.

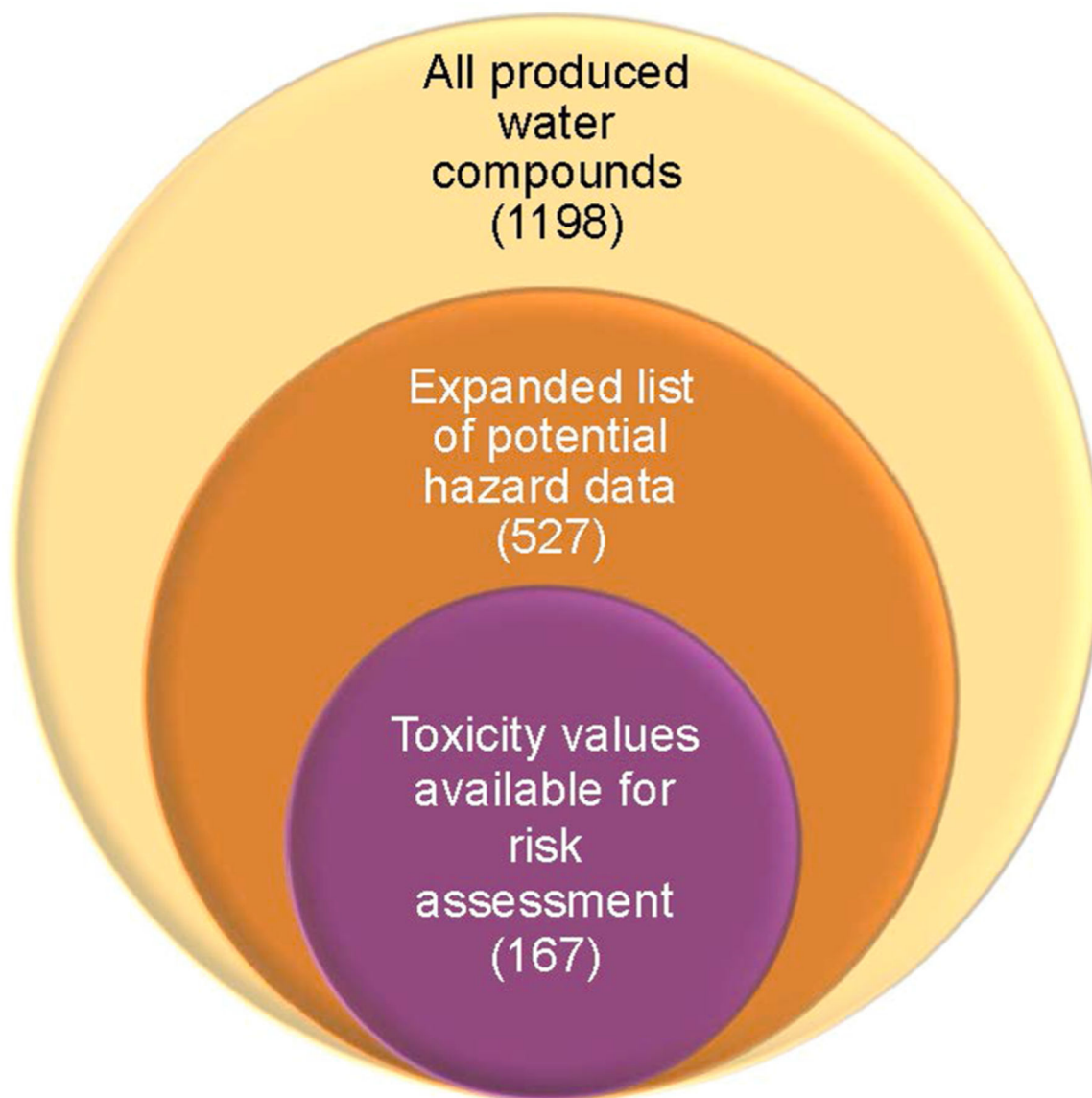Selection of data-rich chemicals for ToxPi rank analysis, based on available data.

Fig. 4.
Cross reference of produced water constituents to available toxicity data. The majority of constituents in produced water (56%) have no data available on the databases searched here to understand or indicate potential toxicological hazard.

Fig. 5.
Characteristic concentrations for EC50 (lower quartile or minimum concentration) for aquatic organisms by species group, compared to maximum concentration measured in produced water (mg/L). Inset displays an example box and whisker plot of EC50 that was used to determine quartiles across all effects by phenol on each species group. The lower quartile concentration for the most sensitive species group was chosen for each chemical (i.e. amphibians for phenol) as the characteristic concentration for ToxPi analysis. The pie chart indicates the available data for the chemicals included in this analysis.
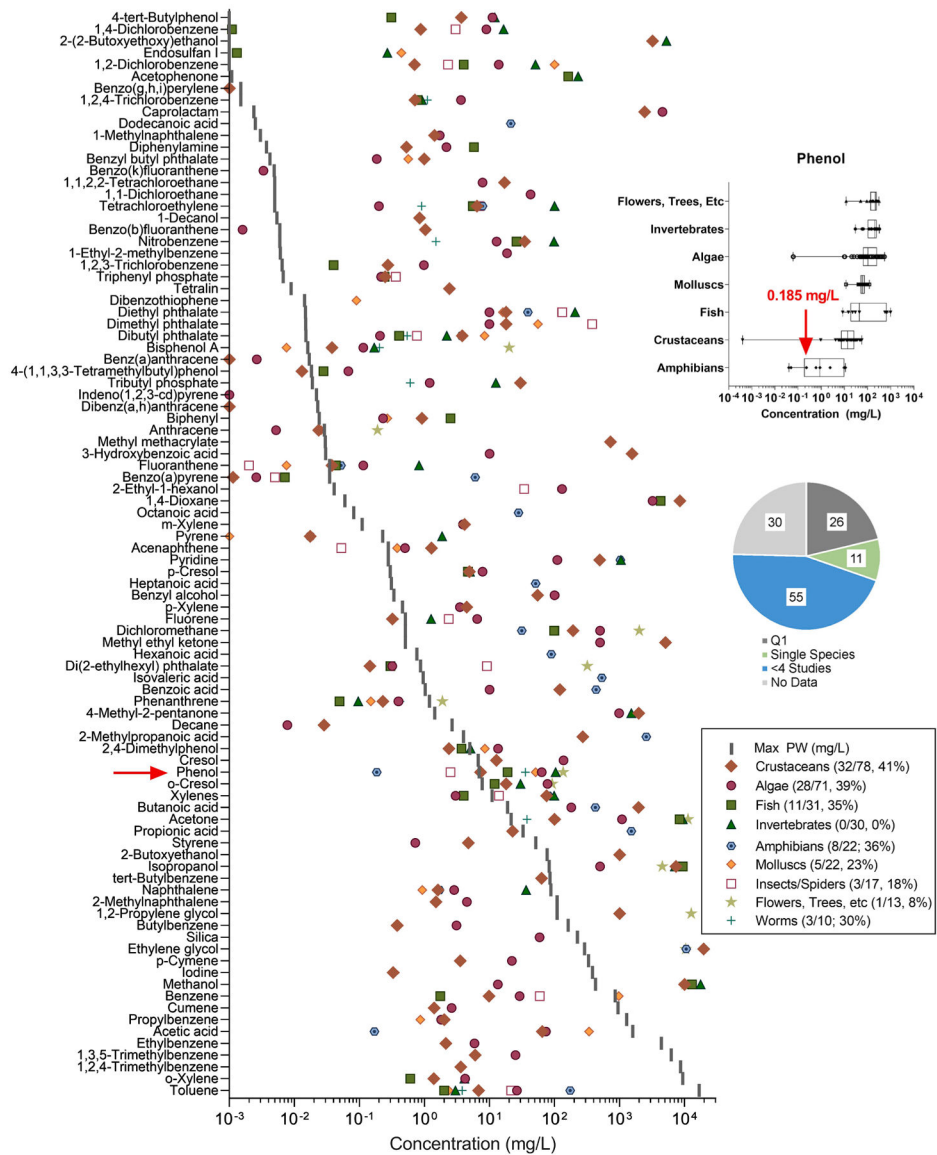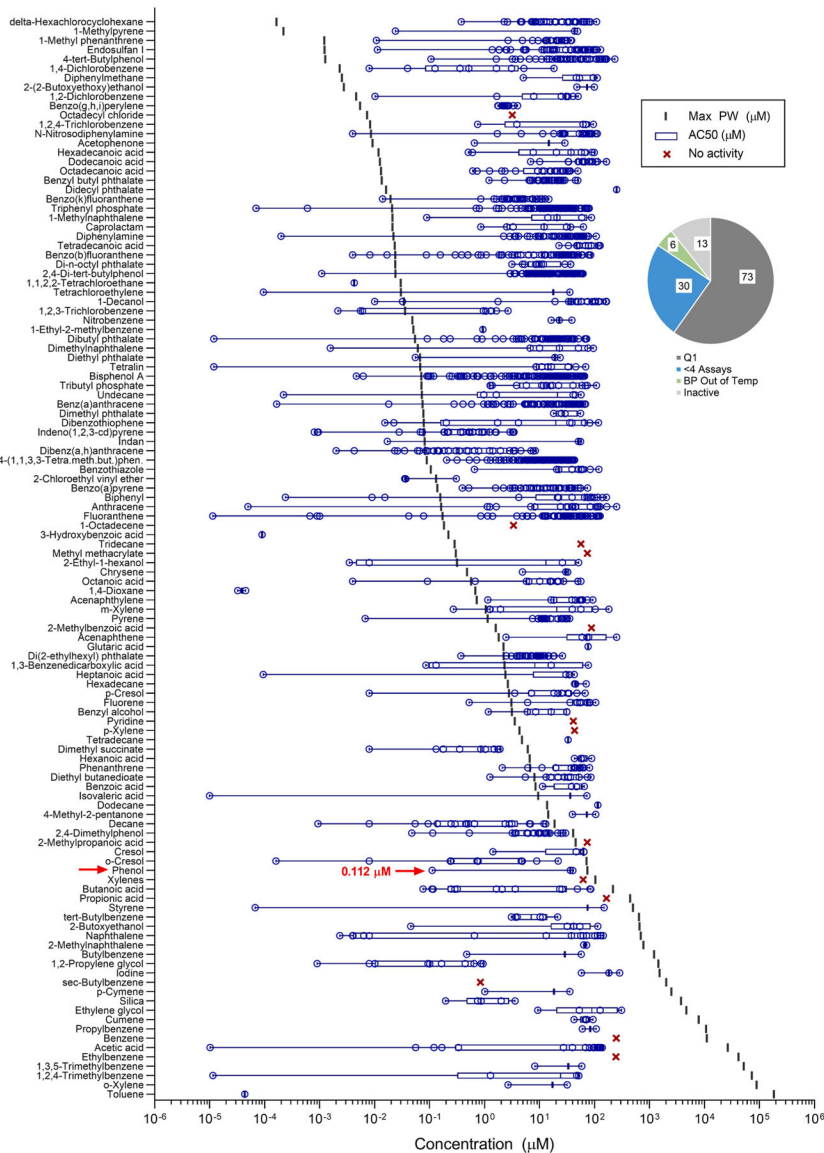
Fig. 6.
Characteristic concentrations for AC50 (lower quartile or minimum concentration) for ToxPi analysis, compared to maximum concentration measured in produced water (μM). For *in vitro* assays, the characteristic effective concentration was calculated as the Q1 of AC50 across all active assays, if greater than four assays. Else, the minimum value was chosen. An example compound (phenol) is called out. The pie chart indicates the available data for the chemicals included in this analysis.

Fig. 7.
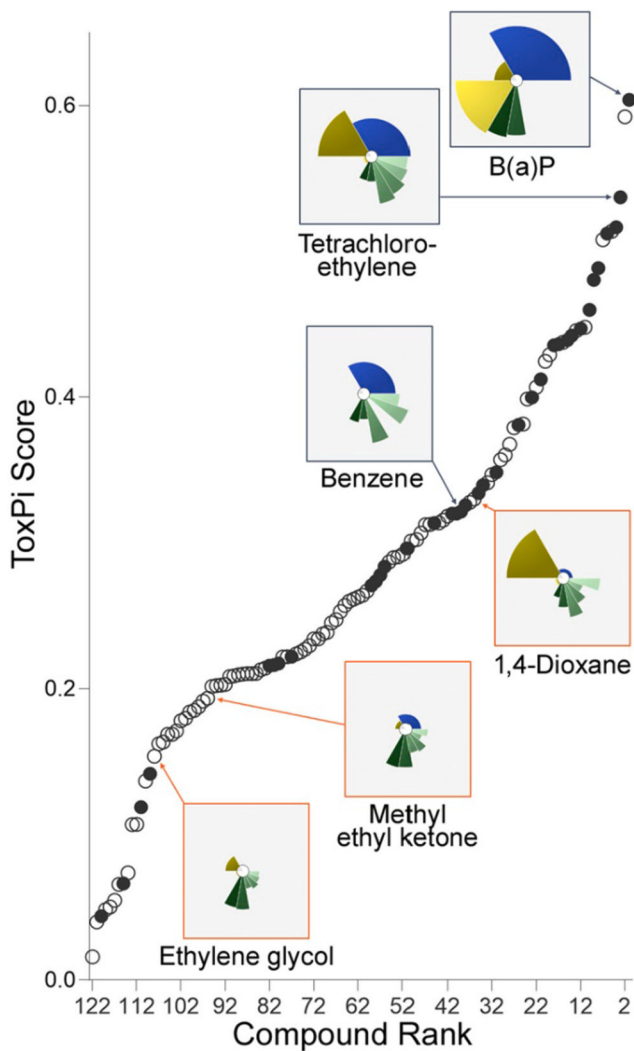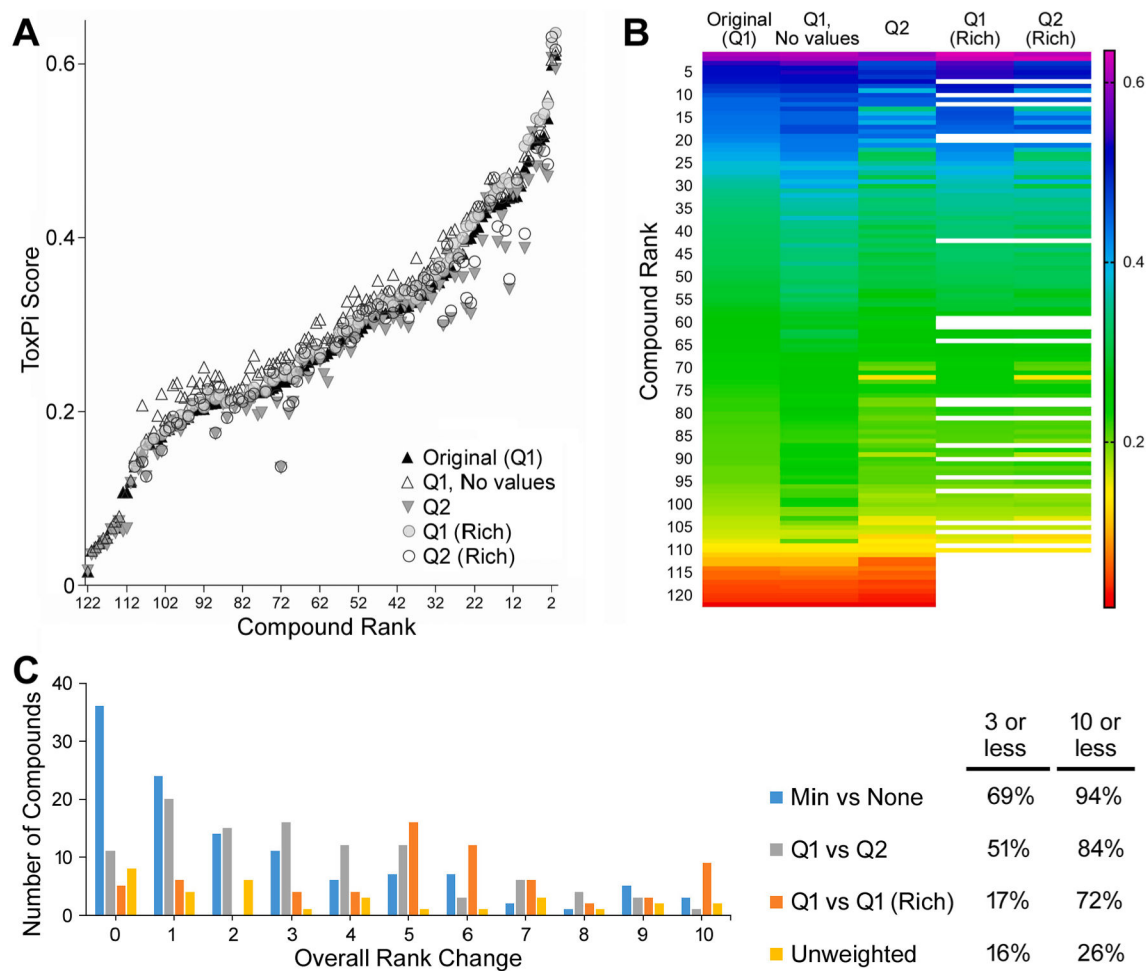Distribution dot plot of ToxPi Scores for all 122 chemicals using the weighted analysis. The dots represent an individual chemical, whereas the 36 chemicals also listed on the EPA's Priority Pollutant List (PPL) are denoted by a solid dot. Example ToxPi profiles are represented in the insets.

Fig. 8.

Sensitivity of ToxPi analysis. A: Sensitivity analysis comparing the rank versus ToxPi score of original weighted ToxPi analysis (Original, Q1) to imputing missing data using no values instead of the minimum in the main analysis (Q1, No values); using median values or geometric mean instead of the lower quartile and minimum values for both the ecotoxicity data and *in vitro* activity data for characteristic half-maximal concentrations (Q2); and including only data-rich chemicals defined as those having at least five out of nine component slices (Q1 and Q2 Rich). The rank derived in the original analysis was held constant to understand how the overall ToxPi score changed. B: A heat map of the same data derived in the sensitivity analysis to visualize change in rank order. C: Additionally, the sensitivity analysis examined how the overall rank changed using different weighting schemes in ToxPi that gave equal weight to each endpoint (unweighted) versus giving equal weight to each domain in the main analysis with using an unweighted analysis. The histogram indicates the total number of chemicals that moved rank (0–10) places.

Table 1

Availability of conditional (predicted) and known toxicity values for the restricted set of 122 chemicals. Existing toxicity values were collected by the CTV Predictor from the following sources: IRIS, Office of Pesticide Programs (OPP), Superfund Regional Screening Levels (CDC/ATSDR, PPRTV, HEAST), and California EPA Office of Environmental Health Hazard.

Toxicity value	"Conditional" predicted	Existing	No value
Cancer potency value (CPV)	46	16	60
Oral slope factor (OSF)	37	23	62
Reference dose (RfD)	29	62	31
Rfd Benchmark dose (RfD BMD)	39	24	59
RfD Benchmark dose lower limit (RfD BMDL)	44	10	68
RfD No observed adverse effect level (RfD NOAEL)	53	34	35

Table 2

Top third of chemicals as ranked by two ToxPi analyses. Bolded chemicals were found in both weighting schemes.

Rank	CAS	Domain weighted	CAS	Unweighted
1	**50-32-8**	**Benzo(a)pyrene**	**319-86-8**	**delta-Hexachlorocyclohexane**
2	**959-98-8**	**Endosulfan I**	**959-98-8**	**Endosulfan I**
3	**127-18-4**	**Tetrachloroethylene**	**127-18-4**	**Tetrachloroethylene**
4	**53-70-3**	**Dibenz(a,h)anthracene**	**126-73-8**	**Tributyl phosphate**
5	124-18-5	Decane	98-95-3	Nitrobenzene
6	**56-55-3**	**Benz(a)anthracene**	79-34-5	1,1,2,2-Tetrachloroethane
7	207-08-9	Benzo(k)fluoranthene	124-18-5	Decane
8	**205-99-2**	**Benzo(b)fluoranthene**	50-32-8	**Benzo(a)pyrene**
9	**106-46-7**	**1,4-Dichlorobenzene**	64-19-7	**Acetic acid**
10	193-39-5	Indeno(1,2,3-cd)pyrene	143-07-7	Dodecanoic acid
11	**126-73-8**	**Tributyl phosphate**	110-75-8	2-Chloroethyl vinyl ether
12	191-24-2	Benzo(g,h,i)perylene	104-76-7	**2-Ethyl-1-hexanol**
13	**64-19-7**	**Acetic acid**	**112-30-1**	**1-Decanol**
14	129-00-0	Pyrene	**123-91-1**	**1,4-Dioxane**
15	206-44-0	Fluoranthene	**108-88-3**	**Toluene**
16	**87-61-6**	**1,2,3-Trichlorobenzene**	111-76-2	2-Butoxyethanol
17	**108-88-3**	**Toluene**	**503-74-2**	**Isovaleric acid**
18	**79-34-5**	**1,1,2,2-Tetrachloroethane**	79-09-4	Propionic acid
19	140-66-9	4-(1,1,3,3-Tetramethylbutyl)phenol	544-63-8	Tetradecanoic acid
20	80-05-7	Bisphenol A	**56-55-3**	**Benz(a)anthracene**
21	**98-95-3**	**Nitrobenzene**	107-92-6	Butanoic acid
22	**108-95-2**	**Phenol**	544-76-3	Hexadecane
23	**91-20-3**	**Naphthalene**	1120-21-4	Undecane
24	100-42-5	Styrene	629-59-4	Tetradecane
25	132-65-0	Dibenzothiophene	**71-43-2**	**Benzene**
26	**319-86-8**	**delta-Hexachlorocyclohexane**	75-09-2	Dichloromethane
27	**112-30-1**	**1-Decanol**	106-65-0	Dimethyl succinate
28	**104-76-7**	**2-Ethyl-1-hexanol**	124-07-2	Octanoic acid

Rank	CAS	Domain weighted	CAS	Unweighted
29	99-06-9	3-Hydroxybenzoic acid	123-25-1	Diethyl butanedioate
30	**503-74-2**	**Isovaleric acid**	**108-95-2**	**Phenol**
31	120-12-7	Anthracene	**53-70-3**	**Dibenz(a,h)anthracene**
32	**143-07-7**	**Dodecanoic acid**	**205-99-2**	**Benzo(b)fluoranthene**
33	104-51-8	Butylbenzene	57-55-6	1,2-Propylene glycol
34	84-74-2	Dibutyl phthalate	112-40-3	Dodecane
35	85-68-7	Benzyl butyl phthalate	629-50-5	Tridecane
36	122-39-4	Diphenylamine	111-14-8	Heptanoic acid
37	**123-91-1**	**1,4-Dioxane**	**91-20-3**	**Naphthalene**
38	117-81-7	Di(2-ethylhexyl) phthalate	**106-46-7**	**1,4-Dichlorobenzene**
39	**71-43-2**	**Benzene**	**87-61-6**	**1,2,3-Trichlorobenzene**
40	83-32-9	Acenaphthene	57-10-3	Hexadecanoic acid

Table 3

Characteristics of the chemicals in top third for both ToxPi weighting schemes.

CASRN	Compound	# of samples	Age of PW (range)	Type	Reported to FracFocus	Toxicity Value?§	Citation
50-32-8	Benzo(a)pyrene	32	1–90 d	U		Yes	(Bell et al., 2017; Chittick and Srebotnjak, 2017; Hayes, 2009; Hayes and Severin, 2012; He et al., 2017; Regnery et al., 2016; Shih et al., 2015)
959-98-8	Endosulfan I	4	0–90 d	U			(Hayes, 2009; Shih et al., 2015)
127-18-4	Tetrachloroethylene	5	1 d - 38 mo	U		Yes	(Akob et al., 2015; Hayes, 2009; Hayes and Severin, 2012; NYSDEC, 2015; Shih et al., 2015)
53-70-3	Dibenz(a,h)anthracene	13	1–90 d	U		Yes	(Bell et al., 2017; Chittick and Srebotnjak, 2017; Hayes, 2009; Hayes and Severin, 2012; Shih et al., 2015)
124-18-5	Decane	18	1 wk - 18 mo	U	Breaker		(Hoelzer et al., 2016; Maguire-Boyle and Barron, 2014; Strong et al., 2014; Wesolowski et al., 1987)
56-55-3	Benz(a)anthracene	76	1–90 d	U		Yes	(Bell et al., 2017; Chittick and Srebotnjak, 2017; Hayes and Severin, 2012; He et al., 2017; Regnery et al., 2016; Shih et al., 2015)
205-99-2	Benzo(b)fluoranthene	40	1–90 d	U		Yes	(Bell et al., 2017; Chittick and Srebotnjak, 2017; Hayes, 2009; Hayes and Severin, 2012; He et al., 2017; Shih et al., 2015)
106-46-7	1,4-Dichlorobenzene	2	ns	U		Yes	(Shih et al., 2015)
126-73-8	Tributyl phosphate†	50	PW	CBM	Defoamer	Yes	(Orem et al., 2014, 2007; Pashin et al., 2014a)
64-19-7	Acetic acid	11	14–328 d	U, C, CBM	Corrosion inhibitor/iron control		(Cluff et al., 2014; Hayes, 2009; Lester et al., 2015; Shih et al., 2015; Sirivedhin and Dallbauman, 2004; U.S. EPA, 2016)
87-61-6	1,2,3-Trichlorobenzene	4	14–90 d	U		Yes	(Hayes, 2009; Shih et al., 2015)
108-88-3	Toluene	734	1 d to > 2 yrs	U, C, CBM, EOR	Paraffin Inhibitor	Yes	(Akob et al., 2015; Carey et al., 1992; Chittick and Srebotnjak, 2017; Cluff et al., 2014; Dahm et al., 2011; DiGiulio and Jackson, 2016; Hayes, 2009; Hayes and Severin, 2012; Hoelzer et al., 2016; Kassotis et al., 2015; Khan et al., 2016; Maguire-Boyle and Barron, 2014; NYSDEC, 2015; Rogers et al., 1992; Rosenblum et al., 2017; Shih et al., 2015; Sirivedhin and Dallbauman, 2004; Thacker et al., 2015; U.S. EPA, 2016; Wesolowski et al., 1987; Ziemkiewicz et al., 2014; Ziemkiewicz and Thomas He, 2015)
79-34-5	1,1,2,2-Tetrachloroethane	2	PW	C		Yes	(Sirivedhin and Dallbauman, 2004)
98-95-3	Nitrobenzene	3	PW	C, CBM		Yes	(Sirivedhin and Dallbauman, 2004)
108-95-2	Phenol	49	1 d to >10 yrs	U, C, EOR	Proppant, Other	Yes	(Carey et al., 1992; DiGiulio and Jackson, 2016; Gieg et al., 2010; Hayes, 2009; Hayes and Severin, 2012; Hoelzer et al., 2016; Lester et al., 2015; NYSDEC, 2015; Shih et al., 2015; U.S. EPA, 2016; Wesolowski et al., 1987)
91-20-3	Naphthalene	712	1 d to >10 yrs	U, C, CBM	Other*	Yes	(Bell et al., 2017; Chittick and Srebotnjak, 2017; Cluff et al., 2014; DiGiulio and Jackson, 2016; Hayes, 2009; Hayes and Severin, 2012; He et al., 2017; Hoelzer et al., 2016; Kassotis et al., 2015; Lyman et al., 2018; NYSDEC, 2015; Orem et al., 2014, 2007; Pashin et al., 2014b, 2014a; Regnery et al.,

CASRN	Compound	# of samples	Age of PW (range)	Type	Reported to FracFocus	Toxicity Value?§	Citation
							2016; Shih et al., 2015; Sirivedhin and Dallbauman, 2004; U.S. EPA, 2016; Wesolowski et al., 1987)
319-86-8	delta-Hexachloro-cyclohexane	3	0–14d	U			(Hayes, 2009; Shih et al., 2015)
112-30-1	1-Decanol†	4	1–3 wks	U	Acid Corrosion Inhibitors		(Hoelzer et al., 2016; Shih et al., 2015)
104-76-7	2-Ethyl-1-hexanol	4	1–3 wks	U	Corrosion Inhibitor		(Hoelzer et al., 2016; Shih et al., 2015)
503-74-2	Isovaleric acid†	6	1–3 wks	U, C, CBM			(Hoelzer et al., 2016; Sirivedhin and Dallbauman, 2004; U.S. EPA, 2016)
143-07-7	Dodecanoic acid†	67	1–3 wks	U, CBM			(Hoelzer et al., 2016; Orem et al., 2007; Pashin et al., 2014b, 2014a)
123-91-1	1,4-Dioxane	41	1–90 d	U	Other*	Yes	(Hayes and Severin, 2012; Lester et al., 2015; Shih et al., 2015; U.S. EPA, 2016)
71-43-2	Benzene	647	1 d to > 2yrs	U, C, CBM	Carrier Fluid, Surfactant	Yes	(Akob et al., 2015; Chittick and Srebotnjak, 2017; Cluff et al., 2014; Dahm et al., 2011; DiGiulio and Jackson, 2016; Hayes, 2009; Kassotis et al., 2015; Khan et al., 2016; NYSDEC, 2015; Rogers et al., 1992; Rosenblum et al., 2017; Shih et al., 2015; Sirivedhin and Dallbauman, 2004; U.S. EPA, 2016; Wesolowski et al., 1987; Ziemkiewicz et al., 2014; Ziemkiewicz and Thomas He, 2015)

PW – Produced Water; U – Unconventional; C – Conventional; CBM – Coalbed Methane; EOR – Enhanced Oil Recovery.

*
Other – multiple usages listed.

†No U.S. EPA-approved analytical test method exists for this chemical.

§Quantitative toxicity values available from the databases described in the U.S. EPA Office of Solid Waste and Emergency Response (OSWER) Guidance document under Directive 9285.7–53 (U.S. EPA, 2003).